The State of US Government Information: Toward a Sustainable Ecosystem
Canadian Govinfo Day 2018
Keynote presentation
Presenter notes
James R. Jacobs
http://bit.ly/govinfoday18

SLIDE 1:

Thank you all so much for your very kind invitation to attend and speak with you at the 20th anniversary of Govinfo Day. I'm truly honored to be part of this celebration. I recognize that our situations in Canada and the US are very different political and depository contexts and historical moments, but I'm sure through my talk today that we will find and share commonalities and hopefully ways forward toward solving -- IMHO -- one of the most critical library issues of the day, preserving and giving long-term access to government information.

SLIDE 2:

Soon after Donald Trump was elected as the 45th president of the US, environmentalists, academics, journalists, archivists, librarians and others began the DataRescue, a series of events across the US and Canada focused on downloading, saving, and nominating seeds to the End of term crawl as much government information and data as they could before Mr Trump and his administration of infamous climate-change skeptics could delete the data critical to continued research. While the wholesale erasure of the environmental record thankfully hasn't materialized as many feared, the public energy behind the DataRescue movement has energized a focus and willingness by the public, librarians and library administrators to work on the issues surrounding this swath of important public information. In some ways, it's never been a better time to be a govt information librarian!

SLIDE 3:

But our issues run deeper and longer than just this administration, and thats what I'd like to talk with you about today. I'll talk about some access breakdowns in the US govt information realm that no doubt have parallels in the Canadian context as well as some exciting projects and initiatives before concluding with some ideas about building and sustaining a govt information ecosystem.

I'd like to start with a more prosaic anecdote which I hope will help to contextualize the ideas I want to get across.

anecdotal story

A legal researcher at Stanford Law School recently wanted to do a meta-analysis of the reports of inspectors general in several federal agencies. IG's are the agency watchdogs insuring against mismanagement and budgetary waste. So she requested the "Inspector General Semiannual Reports to Congress" from 1995 to 2005 for six departments. What seemed to me on the surface to be a simple matter of document access became much more complex, and highlights several issues that government information librarians deal with regularly: online access and usability, findability, collection management, and agency compliance with depository law and regulations.

Because we are members of the Federal Depository Library Program (FDLP), Stanford has these semiannual reports to Congress in its collection either in paper or on microfiche, but due to some depository snafus, we had some gaps in our physical collection. Additionally, most of these departments began putting their reports online around 2000, but all of the agencies had online gaps in coverage, and included digital files in a variety of file formats and usability, including Microsoft Word, PDF with digital text obtained by optical character recognition (OCR), and PDF without OCR'd text. One agency had posted two files on its website that had become corrupted. (To its credit, the agency replaced those files quickly with uncorrupted files after I contacted them.)

In the end though, I found that the best way to meet the researcher's needs and fulfill this seemingly simple access query was to request the bound, complete runs of the reports of all of the agencies and years via interlibrary loan (ILL) from the Southwestern Law School library and then work with our Digital Library Systems and Services (DLSS) staff to scan, perform OCR, catalog, and save all of the reports to Stanford's digital repository (2 birds with 1 stone! We also filled our gaps and made all the reports accessible in usable digital formats). What seemed to be a simple request for specific known publications took approximately 6 weeks to complete and involved staff from across the library. It was made possible, not by the work of the issuing agencies or the seeming ubiquity of the Internet, but by the diligence of the depository staff at the Southwestern Law School.

This simple, everyday andecdote illustrates the issues that we as librarians are facing on a daily basis.

Our processes and systems are supposed to collect (or receive on deposit), describe, preserve and give access to the published record of our government. But the system is far from perfect and is set up in such a way that issues of fugitive documents, commercialization, technical issues and lack of control, are also an everyday nuisance.

The FDLP has been around in one form or another since 1813, and today there are approximately 1150 FDLP libraries in virtually every congressional district assisting in the efforts to build, maintain and give access to the -- admittedly aspirational! -- "national collection." This distributed network of libraries remains critical to maintaining that shared collection and providing information services to their communities.

SLIDE 4:

In order to more fully understand the complexities inherent in access to US govt documents, I'd like to talk about some Access Breakdown points as I see them in order to see access more critically and comprehensively.

I'll broadly organize my talk about these breakdown points into libraries as the traditional centers of information access, technical infrastructure, and political/economic issues. Please realize that these 3 are not mutually exclusive but are intertwined, inseparable, and go across country borders.

1. Libraries
2. Internet infrastructure
3. Political/economic issues

SLIDE 5:

There has been massive growth in the amount of government information on the Web over the last 25 years with a concomitant shrinking of paper documents distributed to US libraries. Today, the Govt Publishing Office (GPO) quotes 97% of FDLP materials as being online only, and virtually every agency has an online presence of some sort.

But what does that 97% figure really mean? This is a chart taken from a report my doppleganger JIM Jacobs wrote in March 2014 for the Center for Research Libraries (CRL) entitled "Born-Digital US Federal Govt Information: Preservation and Access." This chart more clearly shows what this 97% figure actually means. It compares the amount of documents distributed by GPO in 2011 (@10,000 items) to the estimated 2.3-3 million total items distributed throughout the FDLP since 1813 (nobody knows the exact number or the exact dimensions of the "national collection" which is why I earlier called it aspirational!), to the number of URLs harvested by the 2008 End of Term .gov Web Crawl (160 million which has nearly doubled to 310,000,000 for the 2016 crawl!).

even if only 1/1000 of those 160 million URLs are actual documents, that's still 160,000 documents, data sets and other published content across the .gov domain, 16 times more than the number of paper docs distributed to libraries in 2011. Only a very small portion of these ever make their way into GPO's catalog of govt publications, the supposed "national bibliography", or are preserved in any substantative way. That's a whole lot of content flowing around the net and not being curated in any way.

Now that we have an idea of the scope of born-digital govt information in the US, let's delve more closely into the access breakdown points. We're often blinded by the abundance of online information, but it's important to look at these access breakdown points in order to understand and forward the cause of access in a wholistic way.

In 2008, the Govt Accountability Office (GAO) sold exclusive rights to its 20,597 legislative histories of most public laws from 1915-1995 to Thomson West in order to have them digitized. In august, 2009, GPO's PURL server crashed and was not available for several weeks, rendering links to thousands of Federal publications in thousands of library catalogs 404 file not found. In 2012, NARA made a deal with Ancestry.com to have them serve out the 1940 US Census schedules, released and made public after a 72 year embargo. In march 2013, NASA took its technical reports server (http://ntrs.nasa.gov/) offline -- and cut access to hundreds of thousands of technical reports -- based on the off-hand comment of a Congressman. In august, 2014, FDLP.gov was hacked (http://freegovinfo.info/node/9014). And more recently, websites from the EPA, USDA and other agencies have been changed, skewed or had information deleted or obfuscated and the GAO has added the 2020 US census to its high risk list because of cancelled field tests, critical IT uncertainties, information security risks, and "unreliable" cost estimates which do not "conform to best practices."

Each of these incidents have implications for ongoing and longterm access of US government publications.

SLIDE 6:

* Ignoring Born digital collections

* In the US, many FDLP libraries are no longer building documents collections, but merely pointing to GPO and executive agency content via purls in catalog records ("virtual depositories" in GPO parlance), and licensing -- mostly historic -- .gov content from information publishers/vendors like Proquest and LexisNexis. Executive agency libraries are good but obscure sources of information, and many of them have been severely hampered in their collections duties by budget and staff cuts and outsourcing -- if not completely shut down -- over the last 10-15 years. This is doubly problematic because FDLP materials are not getting into the national bibliography, and library users are less likely to find FDLP materials. The fugitive document problem whereby documents in scope of the FDLP do not make their way into the program is accelerating in the born-digital era.

* Fugitives: When we depend on pointing instead of collecting
http://freegovinfo.info/node/3900
"'Issued for Gratuitous Distribution' The History of Fugitive Documents and the FDLP"
https://freegovinfo.info/node/12735

At the same time, a disconcerting amount of libraries are dismantling their historic documents collections. Many FDLP libraries are heavily weeding their low-use but poorly cataloged collections to re-purpose their spaces for "information commons" or cafes under the misplaced assumption (IMHO) that most historic documents are or soon will be digitized -- and more importantly under the false assumption that users will only want online access to digitized publications. We know this is a false dichotomy because there is much anecdotal evidence of users requesting paper documents after finding digitized versions online -- remember my earlier anecdote about Inspector General reports. Reasons for this range from missing and/or poorly scanned pages (lots of them! See Conway's research on HT), large documents not easily read online, redigitization for special purposes like scanning and keying of tabular data, and corpus analysis of documents scanned but not OCR'd etc.

SLIDE 7:

next let's talk about

II. INTERNET INFRASTRUCTURE

When we search Google, we always get results. But people generally don't think about what's missing from their Internet search results. There's a dark side of the Web called "link rot" and it's parallel but lesser known twin "content drift." Access today does not equal permanent public access or the building and curating of the national collection. There is largely an absence of a long-term curation and preservation system for large swaths of born-digital US govt information.

SLIDE 8:

While the average lifespan of a physical government document is 50 or more years, according to the Internet Archive, the estimated lifespan of a URL is 44 - 75 days. While .gov Web sites are slightly more stable than some Web sites, “link rot,” the process by which Internet hyperlinks disappear, is a real and growing concern. According to data collected by the Chesapeake Digital Preservation Group, which has been studying link rot of the .gov domain since 2008, 51% of the urls from their original 2008 data set were 404! There is a growing concern about this issue within certain areas of academia -- especially in the legal realm -- and a couple of great projects -- PERMA.CC (https://perma.cc) and memento project (http://timetravel.mementoweb.org) --

that have been working over the last few years to build a system that addresses the growing link rot issue and links up archived content across Web archives.

SLIDE 9:

"content drift" is a term used in the Web archiving community. It is the process whereby an archived file has changed since being archived. Andy Jackson from the British Library's UK Web Archive gave a presentation a couple of years ago at the International Internet Preservation Consortium (IIPC). In trying to answer the question “How much of the content of the UK Web Archive collection is still on the live web?” his research showed that 50 percent of content had gone, moved, or changed so as to be unrecognizable in only one year. After three years the figure rose to 65 percent.

In short order, born-digital content on the Web either changes or disappears in days or months, not years or decades.

SLIDE 10:

Another infrastructure access breakdown point is centered around GPO and executive agencies themselves as the producers of information.

* GPO is currently working on an internal TDR audit for its govinfo.gov content management system (nee FDsys), but GPO does not have an adequate preservation program in place for GOVINFO which includes a succession plan -- a critical requirement of the OAIS standard. ALA's Government Documents Round Table (GODORT) and the FDLP community have been asking since at least the early 2000s for GPO to at least create a mirror of their content, but nothing official is in place currently. The LOCKSS-USDOCS program -- of which I'll talk about more in a moment -- has stepped into that breach and is harvesting and collaboratively preserving all GOVINFO content, but there is no official MOU in place between GPO and LOCKSS -- that's on my todo list since I'm the program lead for LOCKSS-USDOCS.

* As for Executive agencies, though some are building databases of born-digital and digitized content, they are by and large not doing anything in the digital preservation space that I'm aware of (though I'd be happy to be proven otherwise!). Most are going it alone on the Web, creating 100s of single points of failure, and only a few are working w GPO to share their metadata and host their documents on GOVINFO.

The information at the beginning of the info lifecycle is sorely in need of large amounts of born-digital curation and preservation love.

SLIDE 11:

III. POLITICAL/ECONOMIC ISSUES

there are also political and economic issues at play. Government information is political by nature. At the same time, given the current political/economic context, govt information has turned increasingly into a valuable commodity. I don't have to tell you that FDLP libraries are under enormous budgetary contraints, while commercialization and privatization of govt information has accelerated.

SLIDE 12:

Commercialization is a long-term problem. Instead of promoting and supporting free access, libraries are increasingly relying on commercial vendors -- at least the ones that can afford to pay for access to commercial databases like Lexis Nexis, West Law, Proquest and neophites like Voxgov. Further, many private companies are offering agencies "no cost" contracts to digitize and commoditize historic govt information under the guise of "efficiency." The Thomson-West GAO project is but one particularly egregious example of this pernicious problem. But anywhere there are troves of historically important documents and "big data" like Weather, climate, satellite data, you'll find private companies looking to cash in on free public domain information.

SLIDE 13:

Fast forward to the current president. Along with the historically significant issues of fugitives and commercialization, the current administration has raised the specter of information obfuscation. Perhaps taking a page from your former prime minister Stephen Harper, the current administration has sought to shift and constrict the government's Web presence and published materials for crass political gain, especially in, but not limited to the area of climate change.

The Environmental Data and Governance Initiative (EDGI) (https://envirodatagov.org) a non-profit membership organization, has written several reports centered around federal agency website monitoring, documenting and analyzing data that disappears from public view, and also monitoring and analyzing how data, information, and their presentation have been changed, sometimes in subtle but significant ways. EDGI does great work and I'm happy to report that they just received a Packard Foundation grant of $500,000 to support their work for the next two years. So please check out their work.

Buried, altered, silenced: 4 ways government climate information has changed since Trump took office.

https://theconversation.com/buried-altered-silenced-4-ways-government-climate-information-has-changed-since-trump-took-office-92323

Unexplained censorship of women’s health website renews questions about Trump administration commitment to public health

https://sunlightfoundation.com/2018/04/02/unexplained-censorship-of-womens-health-website-renews-questions-about-trump-administration-commitment-to-public-health/

SLIDE 14:

So it seems on the surface that the public has amazing and unprecedented access to FDLP materials at their fingertips and keyboards. But I hope my brief examination of some access breakdown points has convinced you that that's not necessarily true and that access requires:

Thoughtful collection and curation
Systematic preservation
Human expertise
Understanding political economic context

SLIDE 15:

II. Second section: examples of what's currently happening:

But it's not all doom and gloom! Now that I've laid out the US govt information landscape, I'd like to shift gears a bit and highlight some exciting projects. There are a bunch that I could mention, but I'll limit myself to 5 because these projects represent possible key components to the building of a govt information ecosystem with which I'll conclude.

1. PEGI
2. Hathitrust govt documents registry
3. End of term crawl
4. LOCKSS-USDOCS
5. Policy efforts

SLIDE 16:

1. Preservation of Electronic Government Information (PEGI)

In 2016, a digital preservation of federal information summit was held. This summit brought together stakeholders from a variety of public and private organizations, including archivists, librarians, technologists, library directors and others. The aim of the meeting was to 1) engage in a structured and facilitated dialogue with national leaders specifically focusing on preservation of born-digital govt information, and 2) to begin the development of a national agenda to address preservation for the most pressing categories of at-risk digital government information.

A Reflections Report was published (bit.ly/dig-preservation-govinfo-summit) which outlined the findings of the summit and kickstarted the PEGI project into existence.

In 2017, PEGI received an $85,000 IMLS grant to host a series of forums to engage stakeholder communities in conversation about the importance of preserving born-digital govt information, perceived future needs of those communities, and potential barriers as they see them. The grant will end with a final report in early 2019 which we hope to be a blueprint for future preservation efforts. The project is also engaged in an environmental scan to identify aligned stakeholders and existing digital repositories within and outside the .gov domain.

The combination of our final report and environmental scan will give the documents community a solid basis to bring people together and figure out and target infrastructure and policy gaps in the ecosystem.

SLIDE 17:

The Next project of note is the Hathitrust US Federal Government Documents Registry.

Many of you will be familiar with -- if not members of -- the hathitrust digital library, a large digital archive originally built off of University of Michigan's google book project scans which continues digitization efforts to this day.

In 2011, Hathitrust members approved a ballot initiative to begin work on the HathiTrust US Federal Government Documents Registry. The registry is meant to define the field of govt documents held by hathitrust but also more widely. They have created a publicly available

database of metadata garnered from Hathitrust and from libraries across the US, representing the comprehensive corpus of historic U.S. federal documents produced from 1789 to the present.

By building this metadata inventory, HT hopes to homogenize disparate metadata and fill gaps in its archive. But other libraries, such as Stanford, are also starting to use the registry to find gaps in their own collections and help HT with theirs.

The registry provides the library community a reliable inventory of items in the US documents corpus, an amazing first step and building block to scoping out the dimensions of the historic national collection, long a goal of the FDLP.

SLIDE 18:

Next up is the End of term crawl.

In the late summer of 2016 a group of institutions -- Internet Archive, Library of Congress, CA Digital Library, GPO and libraries from the University of North Texas, Stanford University, and George Washington University -- organized to preserve a snapshot of the federal government .gov/.mil web domain. This is the third time this End of Term (EOT) group has gotten together to identify, harvest, preserve and provide access to the federal government web presence both as a way of documenting the changes caused by the transition of elected officials in the government and to provide a broad snapshot of the federal domain once every four years replicated among a number of organizations for long-term preservation.

Defining exactly what IS the .gov/.mil web presence has been one of the biggest challenges for the project. The .gov domain ebbs and flows, contracts and expands (mostly expands :-)), and goes beyond .gov/.mil. So we can't just go WGET *.gov and get it all.

At first, even the US govt didn't have a comprehensive list of top level domains and subdomains, but over the course of the crawl cycles, we've created and maintained a bulk list of seeds. This year, we've also received for the first time a bulk seed list from google, perhaps the largest Web crawler (I was told it's @3.3billion seeds!).

SLIDE 19:

Along with these bulk lists, we've relied on volunteer seed nominators to make sure we get all of the sites that people use and care about.

The number of volunteers has grown over the 3 cycles. This year with the perceived notion of the loss of environmental information and data and the extreme public interest in preserving and assuring access to govt information, we've had many more nominations and nominators. We received over 100,000 seeds from DataRescue and EDGI events!

2016: 15,000+ from 400+ nominators (via UNT form)
Plus!: Over 100,000 from events and tools hosted and created by DataRescue/Environmental Data & Governance Initiative (EDGI) events!

SLIDE 20:

This time around, the project will have archived somewhere in the neighborhood of 300 terabytes of content. We've also expanded our efforts to include FTP data and social media accounts for the first time. Along with 100TB of public websites and over 150TB of public data from federal FTP file servers, the crawl this time will total all together over 310 million URLs/files, over 70 million html pages, over 40 million PDFs and, towards the other end of the spectrum and for semantic web geeks in the room, 8 files of the text/turtle mime type.

End of Term crawl has become a critical stopgap measure of preservation of born-digital govt information.

SLIDE 21:

Let's now turn our attention to LOCKSS-USDOCS.

Lots of Copies Keep Stuff Safe or LOCKSS is software developed at Stanford in 1999 for the distributed and secure preservation of digital information. Many of you are already aware of and are participating in CGI-PLN. There's another LOCKSS network called LOCKSS-USDOCS. This network, began in 2008 with 15 libraries participating. Today, there are 36 libraries -- including Simon Fraser University, University of Alberta thank you very much! -- plus the Govt Publishing Office (GPO) involved in the project. The USDOCS archive harvests and preserves all of the collections of documents in GPO's govinfo system -- and as a biproduct, helps to inform the discussion about "digital deposit" a process for which I've been advocating for 15 years! Govinfo is only 1 database, but it includes all Congressional documents from 1995 - present (and some going further back historically) as well as some important executive branch and Judicial branch publications. In other words, it's not the entire universe of born-digital US govt information, but it's a major star in the galaxy and we're thankful that we've got that information into a distributed preservation network outside of the .gov domain. I'm also working to get a USDOCS cache into the Internet Archive so if GOVINFO goes down or GPO gets defunded and shuttered, we can point the DNS servers to the Internet Archive cache and maintain access.

SLIDE 22:

Lastly I'd like to highlight a couple of policy efforts.

While technical solutions and infrastructure are critical going forward, nothing will be sustainable without policy solutions, the "source code of govt" to paraphrase Carl Malamud. The policy efforts I'd like to highlight do much to broaden the scope of the fdlp, mandate agency buy-in, and define and expand on Govt and library roles. Public policy is perhaps the most important aspect of the efforts to build a long-term sustainable US govinfo ecosystem, so I'm excited to highlight a couple of policy initiatives that are coming to a head at the moment. These bills represent many years of grassroots efforts and advocacy by librarians, library- and open government organizations.

The first one is "H.R. 4631: Access to Congressionally Mandated Reports Act"
https://www.govtrack.us/congress/bills/115/hr4631

This bill will require that the 4000+ Congressionally mandated annual reports from federal agencies be made publicly available on GPO's GOVINFO system. This seems like such a no-brainer, but many of these mandated reports are never released or made publicly available. This would instantly expand and enhance the national collection. This bill has passed out of the Committee on House Administration and will soon come up for a vote in the House.

Another effort revolves around CRS reports.

Public access to CRS reports:

The Congressional Research Service (CRS) is Congress' think tank. But CRS has long resisted making their reports publicly available, considering them "privileged communication" with members of Congress.

Since 1916, these reports could only be accessed if a person knew of and requested a report from their Congress person, purchased one from a private publisher like Penny Hill Press, or if their library subscribed to a private service (CIS, then LexisNexis, and now Proquest had an agreement w CRS to microfiche (now digitize) and sell access to them). But a 20-year(!) grassroots campaign that included draft legislation not passed in the last few Congresses has *finally* borne fruit and public access to CRS reports was included in the most recent omnibus appropriations bill which passed the House and Senate just a few weeks ago! The Library of Congress will, within 90 days, begin publishing CRS reports on their site and GPO has promised that the reports will be within scope of the FDLP. This is HUGE! If you're unfamiliar with CRS reports, check out https://www.everycrsreport.com.

HR 5305 the FDLP Modernization Act
https://www.govtrack.us/congress/bills/115/hr5305

Lastly, another bill coming to a head is HR 5305 the FDLP Modernization Act. This one's particularly near and dear to my heart since Title 44 or the US Code, which defines the FDLP, hasn't been substantially updated since 1962. This bill also just passed out of committee and should be up for a vote in the House in the very near future.

While it doesn't update everything to my satisfaction (I'm an idealist after all :-)), the bill requires for the first time that GPO provide free access to digital content and to have a program of digital preservation; it changes and expands the scope of GPO and FDLP and for the first time mentions and defines the "national collection." It also strengthens privacy protections for users of govt information and even has "digital deposit" codified into the law (but not in a strong enough way for my liking). And most importantly, because of a strong and sustained effort by librarians and library associations, part of the bill was struck which purported to "reform" the Govt Publishing Office but would have slashed GPO's budget and made the positive FDLP gains of the bill unsupportable by GPO.

If you really want to get into the weeds on the FDLP Modernization Act, Jim and I have been closely covering and advocating for stronger changes to the bill over at freegovinfo.

SLIDE 23:

The grassroots and an admittedly small but growing number of libraries have clearly rallied around efforts at expanding public access and born-digital preservation. And we're beginning to see some movement within the government in that direction as well.

In the last few years, there are efforts within federal agencies which have started to take shape. For example, GPO is actively collaborating with library efforts and working toward a trusted digital repository audit for govinfo.

GPO, NARA and LC and a couple of other agencies have come together to form the Federal Web archiving working group to share information and best practices.

In 2013, the Office of Science Technology Policy (OSTP) published a memo entitled "Expanding Public Access to the Results of Federally Funded Research." This memo has resulted in the inter-agency Commerce, Energy, NASA, Defense Information Managers Group (CENDI) federal Sci and Technical Information managers group publishing public access plans for 12 scientific agencies including the Depts of Defense, Agriculture, Education, Energy, Homeland Security, NSF, and NASA. Some but not all of these public access plans mention preservation as part of their efforts toward information access.

And there are a number of Federal agency portals coming online like science.gov and scientific data sharing repositories driven by the NSF's data management requirements for federally funded research starting in 2011. And Institute of Museum and Library Services (IMLS) is beginning to focus on national digital infrastructure development as an area of priority in their Grants to States program.

But until there is a comprehensive strategy or strategies among and between the government and libraries to actively collect and preserve born-digital govt information, there will continue to be loss and erosion of the national collection.
SLIDE 24:

That brings me to my last segment. Please indulge me and share in some pie-in-the-sky for a few minutes. To my thinking, we need to put in place a govt information ecosystem which assures long-term control of and free public access to information outside of the .gov domain. I'd just like to highlight some things that have been knocking around my brain for a while, as a way for us to explore and discuss the possibilities. I use the metaphor of an "ecosystem" because I think it insinuates a community of interconnecting and interacting parts or entities that work together toward some common greater goal, and that's what I think we need.

SLIDE 25:

As Dilbert so aptly points out, we shouldn't let the perfect be the enemy of the good, but I've always felt that attempting to map out the ideal sytem is a good thought exercise and gives one something with which to evaluate the current situation and work towards.

SLIDE 26:

In order to scope out our ecosystem, we need some guiding principles. Luckily, there is an international preservation standard, "The Reference Model For An Open Archival Information System" or OAIS, that fits the bill and which gives us some great guiding principles for our hopefully destined ecosystem.

By the way, OAIS is already the basis for certifying Trusted Digital Repositories. OAIS tells us that just storing files or making them "accessible" is not the same thing as preserving them.

To preserve information you have to ensure that the information is all these things...

GUIDING PRINCIPLES

Information must be:

Not just preserved, but discoverable.
Not just discoverable, but deliverable.
Not just deliverable as bits, but readable.
Not just readable, but understandable.
Not just understandable, but usable.

SLIDE 27:

I was recently brainstorming what the ideal ecosystem would look like with 2 of my PEGI colleagues, Shari Laster and Deborah Caldwell. So a shoutout to them for helping me talk through the design of this ecosystem flower! Our ecosystem needs to have the following petals:

Publishing output
collections/curation
preservation
metadata/description
access

It needs to be publicly controlled and funded, collaborative, interoperable, and sustainable, be built on open standards like OAIS, with version control and links resolving. Perhaps most importantly, it must be based on public policy which requires .gov entities to produce open, findable, collectible, re-usable information.

The first step to achieving this system is that agencies produce structured information and well-designed Websites with site maps and sub-directory structures for ../data ../reports ../publications ../video ../audio etc. so that libraries, GPO and other interested groups and projects can easily collect and archive agency information. We also need to include in this petal the collection of .gov social media presences.

We need well-curated archives of interconnected, well-described, preservable govt content in re-usable formats. The big thing in libraries these days is "service," but a library can't build services without collections, and these well-curated archives form the basis of library services going forward, not just for .gov content but across the library.

We need a ubiquitous metadata layer that can be shared among and between archives, search engines, and the public, and allow libraries to build discovery layers that contain .gov and non-governmental materials (ie books etc) for their designated communities, either ongoing or on the fly.

And lastly, we need access and user-interface tools that go beyond just finding one "document" at a time or one archive at a time, that allow users to re-use content and metadata for their own purposes.

It's important in this ecosystem that each of the flower petals connect and facilitate the processes of collection, description, access and preservation. The projects that I mentioned earlier are proto-flower petals that fit into one or more of these ecosystem categories, but you can see that there are gaps and missing pieces that need to be grown and nurtured by the .gov community.

SLIDE 28:

I know that building and maintaining a functional and sustainable ecosystem sounds like a big and daunting task, and our day-to-day responsibilities probably do not allow much time to think of big issues and long-term strategies.

But that doesn't mean we are powerless.

So what can we do individually, every day?

First, we can use existing tools to collect and preserve the content of importance to our designated communities (and some were presented here today in the previous Cool Tools segment!). Many of these tools are easy to use and free or inexpensive.

This is necessary in order to complement and extend what our govt agencies can do on their own.

Second, we can take the lead in building a movement for a long-term, comprehensive plan for the life-cycle of government information. There are lots of small steps we can take every day to do that.

In the US context, we can't assume that government agencies will do this on their own. Though I'd love it to be otherwise, they do not have the legal mandate or the resources to do so. And they lack the knowledge that we as librarians have of the needs of our different communities of users. The silver lining of the current political climate in the US is that Library administrators and organizations are *finally* realizing the need for strategic and ongoing action by libraries.

The harder the struggle, the sweeter the eventual victory! One of my heroes, journalist IF Stone has this quote to which I often return, especially at those times when things look their darkest:

"The only kinds of fights worth fighting are those you are going to lose because somebody has to fight them and lose and lose and lose until someday, somebody who believes as you do wins. In order for somebody to win an important, major fight 100 years hence, a lot of other people have got to be willing -- for the sheer fun and joy of it -- to go right ahead and fight, knowing you're going to lose. You mustn't feel like a martyr. You've got to enjoy it."

Librarians can and must drive this change. Thank you all for the work that you do for Canadian documents. Together I feel certain that we will ring in the govinfo ecosystem!

SLIDE 29:

Thank you very much for your time today!