

3:45 - 5:15 pm Thursday, February 4

Roundtable 4 : : Technology Room

- Gone today, Here tomorrow: assuring access to government information in the digital age--ShinJoung Yeo, University of Illinois; and James R. Jacobs, Stanford University

Panelists:

James Jacobs

Thomas Bruce (Legal Information Institute, Cornell University)

Daniel Schuman (Sunlight Foundation policy director)

Cindy Etkin (Govt Printing Office)

[SLIDE 1: govt documents]

Right up front, I'm a librarian and a collaborator in the LOCKSS distributed digital preservation project (Lots of Copies Keep Stuff Safe). I've been in academia/education my whole life as a student, teacher, librarian and technologist. I've been a government information/FDLP librarian since 2002 and currently am serving a 3 year term on the Depository Library Council, the body which informs and advises the Govt Printing Office regarding issues of the Federal Depository Library Program (which Cindy talked about). So my mindset/perspective/bias is from one who assists in the scholarly communication process, one who believes that libraries have a place in the digital information landscape, and one who believes strongly in the idea that access to govt information is a fundamental right. As Ralph Nader has said, "There can be no daily democracy without daily citizenship." And there can be no citizenship without access to government information.

[SLIDE 2: mmm documents]

With that in mind, I'd like to talk about the underlying historical ideals of the FDLP, discuss how those ideals have been under fire from both within and without the library community and argue that those ideals applied to today's information landscape give us the best chance at access to and long-term preservation and assurance of govt information.

[SLIDE 3: FDLP logo]

The federal depository library program (FDLP) has been around since 1813 in one form or another. The basis underlying the need for an FDLP is to give the public free access to government information. Depository libraries have long safeguarded the public's right to know by cooperating with and receiving for free the govt publications published by the Govt Printing Office (GPO), organizing, maintaining, and preserving those publications, assisting users in accessing said information in a geographically dispersed system and most importantly, assured that govt information is freely available and tamper-proof -- think Napster for govt information. Taken together, the collections of the 1238 depository libraries make up the historic corpus of govt information available for free to every citizen. Jessamyn West of librarian.net, recently called the FDLP the longest running open source project. I would add that it's the longest government-run public-centric open-source project to support the democratic ideal.

[SLIDE CHUCK QUOTE]

Over the last 20-30 years, developments in publishing and Internet technologies have affected the way government information is produced, disseminated, controlled, and preserved. These changes have affected the policies and procedures of the GPO and, in turn, have affected the depository library program. Despite the often-heard promises that Web technologies will bring more information to more people more quickly and easily, the actual effects have been decidedly mixed. The highly visible, short-term successes of rapid dissemination of single titles directly to citizens (e.g., the large number of downloads of the 9/11 report) mask the loss of a secure infrastructure (GPO's Federal Digital System (FDsys) notwithstanding) for long-term preservation of and access to government information as more and more agencies publish content on their own Web sites rather than using the GPO conduit (which librarians call "fugitive documents") and very few agencies publish to any standards or have policies in place that deal with archiving and preservation. As Chuck Humphrey, a data librarian friend of mine, once said, "there seems to be an inverse relationship between convenience of dissemination and preservation standards."

In addition to this lack of a secure infrastructure, the growing din of the call for digitization of historic govt publications (most recently the Ithaka/ARL report "Documents for a Digital Democracy: A Model for the Federal Depository Library Program in the 21st Century"), while no doubt a boon for access today, is somewhat of a red herring that makes library administrators believe that they will soon be able to dispose of their physical collections and use that space for today or tomorrow's buzz word. This call for digitization may instead have the deleterious affect of damaging the long-term preservation of govt publications.

Lastly, the growing trend toward privatization of govt information has actually caused a decrease in public access despite it's digital nature. This is not a new trend. Herbert Schiller noted this in 1986 in his book "Information and the Crisis Economy." Speaking of machine readable formats, he wrote that, "Library information capability is greatly enhanced. Yet this benefit is accompanied by the abandonment of libraries' historical free access policy. User charges are introduced. The public character of the library is weakening as its commercial connection deepens. No less important, the composition and character of its holdings change as the clientele shifts from general public to the ability-to-pay user."

[SLIDE: GAO contract]

We've seen over the last 30 years a disturbing rise in Federal Agencies entering into contracts with private companies whereby public domain govt documents are digitized and then taken out of the commons via licensing agreements. See for example, the Government Accountability Office (GAO)'s deal with Thomson-West whereby Thomson-West digitized the GAO's 20,597 legislative histories of most public laws from 1915-1995 and in return received exclusive license to sell access to the content. GAO received nothing in return but an account on Thomson's service while the public received nothing at all.

Rapid technological change and the misplaced assumption that "it's all in google" have caused some in the FDLP community to question the need for the FDLP and some others to drop out of the program

altogether. I believe that the inherent nature of digital information actually increases the need for a distributed network of dedicated, legislatively authorized libraries. It would be prudent to draw upon the existing infrastructure of FDLP libraries and the almost 200 years of cumulative experience of these institutions in assuring preservation of and access to government information. We must reinforce FDLP's traditional mission of selection, collection, free access, and preservation of government information in the digital era in order to assure free access to this information into the foreseeable future. Some in the depository community, like my library, are doing just that by participating in the LOCKSS-USDOCS network, harvesting digital govt information -- for example, harvesting openCRS that Daniel mentioned along with other sites that post CRS reports -- and yes digitizing parts of their collections. But we need more libraries not less.

[SLIDE: FDLP ecosystem]

Nobody knows for sure how to preserve digital content for the long-term. This means to me that a loosely coupled, independently administered, distributed ecosystem is the best way to assure long-term preservation -- many organizations with many funding models and a distributed technical infrastructure(s) have a better shot at preservation than 1 or 2 organizations -- especially if one of those organizations has a tenuous budget, or is a private corporation etc.

Imagine if you will 2 future govt information systems: on the one hand, the system where there are one or two digital collections (say for example GPO's Federal Digital System (fdsys) and Portico, the dark archive currently housing digital journals); and on the other hand, one with many digital collections in fdlp libraries. How would each of these deal with or react to different stress situations or threat models (e.g., reduced budgets, increased demand for privatization, increased demand for censorship or control or removal of information, media/hardware/software/network failure, natural disaster, organizational failure etc.)? It's easy to see that a highly replicated, distributed FDLP model of preservation would deal with these situations much better than a centralized model. A web is much stronger than a silo.

[SLIDE: Federal Register XML]

law.gov, Carl Malamud's proposal for a registry and repository of all legal information -- from what I've seen and heard and read, is a compelling proposal for a significant piece of the federal (and state) legal information ecosystem. What we ought to be doing is a) figuring out how to make law.gov a reality; b) figuring out how to expand it beyond legal materials to include ALL federal information -- information from all 3 branches of government, federal agencies as well as the regional and local offices of those agencies, data and statistics, the entire Congressional/legislative process including the funding that goes into that process to grease the skids so to speak, and making sure public information stays in public control; and c) MOST IMPORTANTLY from my perspective as a librarian, figure out how to preserve that ecosystem for the long term so that the public can inform itself not just today or tomorrow but 100 years from now. Now the 4 of us on this panel are just 4 players with dogs in this fight. But if we agree on the goals, then we ought to work together to proceed toward them and mobilize our communities and the public to support this endeavor.

It's going to take the government (and not just GPO) being serious about transparency and funding the necessary changes in its own federal information distribution system to include open format standards with no DRM, bulk data channels, indexing, description, collection and authentication of information resources, multiple digital preservation strategies to not only assure preservation but also to insure against tampering and deletion of vital information (which, as I've stated earlier, the FDLP historically has done very well!). It's also going to take libraries being serious about and applying the ideals of the FDLP to build a distributed digital infrastructure that takes into account access to as well as preservation of digital govt information.

I agree with Tom and am absolutely convinced that the changes in the information ecosystem that are needed should not be left to the market because the information market leans heavily toward monopoly, proprietary standards, licensing restrictions, lack of access, "rights management" and the like.

If an evolving ecosystem that is free, open, standards-based, authenticated, and privacy-protecting is built and sustained correctly then citizens, libraries, non-profit watchdogs, hackers, activists, AND government will thrive.

[SLIDE 7: THANKS! lockss, archive-it]

digital changes a lot of things about information, but it doesn't change the need to fund it, collect it, share it, preserve it, and give access to it. As my friend and colleague Jim Jacobs recently stated, "lots of collections keep stuff safe!" (yes there are 2 of us working on FGI!)