

Digitizing the EPA Libraries: What they thought we didn't know

By Virginia Luehrsen

Presented 03/29/2008, Ann Arbor, MI, at "Questioning Authority" conference.

First, thank you to my colleagues who have discussed very well how the digitization of documents may lead to challenges to information's authority and integrity. Also thank you to the Ann Arbor SLA student group, who not only gave me an audience today, but a reason to finally complete this paper that has been writing itself in the back of my brain for some time now.

I am sure most people in this room are familiar with the recent closing of several EPA libraries (including the Regional Library in Chicago). However, data related to these library closings, and the implicated consequences, are often scattered throughout newspaper articles, congressional reports, and letters circulated via the many listservs devoted to library and environmental issues.

This paper is an attempt to draw together the facts of the EPA library closings, the consequences and implications of these closings, and recent developments as to how the EPA can move forward after more than 2 years of controversy.

HISTORY of the ISSUE

When the FY 2007 federal budget was passed in early 2006, there was some concern within small circles about a 77% cut to the EPA budget that targeted the agency's physical library system. Within one month of the budget being approved, administrators at the Chicago library sent an internal email claiming they would be the first to close.

[SLIDE 2] By October 1, 2006, regional libraries in Dallas, Kansas City, and Chicago were closed, as well as the Headquarters library and chemical library in D.C.

Shortly after, an anonymous EPA librarian posted on YubaNet that librarians were not assured the digitization process would completely replicate the physical collection, and there were already instances where librarians were "unable to obtain a document they need that was previously available" (Sternstein, 2006).

EPA Deputy Administrator Marcus Peacock responded on November 2nd with a claim that all the "unique EPA" documents from Chicago were digitized already, and other documents in the EPA system would be digitized in 2 to 3 years. He also denounced the claim that documents were not available, even to librarians, and that the EPA had the money to undertake a massive digitization effort, despite the \$2 million budget cut that led to the shutting down and reduced hours of libraries in the first place.

However, by the early part of 2007, complaints were already surfacing from scientists and librarians, including the EPA's own employees, who were finding the incomplete digitized records of the closed libraries nearly inaccessible and difficult to use. The newly digitized documents were made available through the National Environmental Publications Internet Site (NEPIS), but the interface was difficult, and corresponding metadata with which to search inventory and documents was incomplete or missing entirely. This fact was confirmed by EPA employees at the ALA meeting in Seattle in January, 2007. It was become unclear what will happen to the thousands of other documents the EPA was responsible for preserving and making accessible to the public that were authored privately or by other institutions.

[SLIDE 3] This next slide shows how many requests were received at libraries from the EPA's own scientific and enforcement staff in 2005 – more than half of these were extended research questions.

Internal information requests handled at EPA libraries in 2005
All 26 libraries: more than 134,000

Region 5, Chicago: more than 32,000
(McClure, 2007).

The first congressional hearing related to the closings of the EPA libraries was not held until early February, 2007. EPA Administrator Stephen Johnson testified that only five of the 26 EPA libraries had been shut; however, the truth was, more libraries were being in the process of being shut down at that time, including the Atlanta library who's materials were being sent to Cincinnati as Mr. Johnson was testifying (Goldberg, 2007b).

[SLIDE 4] Also at this hearing, ALA President Leslie Burger testified:

"The EPA's policy with regard to its libraries is a very important issue, not just for the library, science, and environmental communities, but also for the general public. This information is vital to public health and needs to remain accessible. Further, while having the information digitized is indeed important, it should not be the driving force behind shuttering the libraries and, thus, taking from the public one of the libraries' most important assets: the librarians." (Leslie Burger, President of ALA: Feb. 6, 2007)

Following congressional hearings, a moratorium was placed on further closings until the Government Accountability Office and Congress itself had a chance to review the impact and consequences of the library closings, and what actions should be taken next – including reopening the libraries.

IMPLICATIONS

So, what is wrong with digitization? Three key elements present a challenge to the authority and integrity of the documents the EPA proposes to digitize as part of their library reorganization:

1. the digitization process employed by the agency;
2. lacking policy for dealing with copyrighted materials that cannot be replicated;
3. and lacking research and knowledge of preserving such a large digital collection in perpetuity.

The EPA libraries are used by scientists, legal aides, and regulators to perform environmental assessments, develop environmental policy, and enforce environmental laws; in addition to requests of the general public and private companies.

There is indeed a growing interest in accessing government information online, and a modernization effort of government documents is certainly needed – over time.

[SLIDE 5; 6] In a recent study by University of Illinois Urbana-Champaign and the PEW

Internet Project:

- 66% of Americans prefer electronic access to government documents and information for school or work research (viii). Additionally, 58% of Americans contacted a government agency in the year prior the study's publication
- For information about community issues, such as crime, traffic, and environmental concerns, 35% of Americans prefer the internet, 32% prefer face-to-face interaction, 26% prefer use of the telephone, and 7% prefer to contact via mail (ix). [Make a pie chart]

This seems inline with EPA's modernization efforts (though it should be noted that this report was released in December, 2007, more than a year and half after the budget cuts spurred the modernization efforts). [SLIDE 7] However, it must also be noted that this survey revealed:

- 56% of Americans are not entirely trusting of the government;
- 65% of low or no-use internet Americans do not entirely trust the government (28).

The digitization project of the EPA is likely to reveal higher numbers, especially amongst those who may distrust government motives for eliminating the paper copies they prefer and trust. The EPA's own employees have raised questions as to the motive behind the closings and the lack of apparent coordination in efforts to modernize the library, as well as the method of digitizing the collections,.

The motive for closing the libraries was supposedly in response to the FY2007 budget cut, along with a growing demand for digital access to government information and materials. [SLIDE 8]

"I support the [Bush] Administration's efforts because the proposed cut in the EPA's budget is an attempt to eliminate inefficiency by encouraging the EPA to digitize its library holdings and move toward electronic record keeping...Some major concerns have been raised regarding potential negative ramifications that could result if EPA libraries were to shut down. However, I understand that these concerns have been adequately addressed."
(Senator Wayne Allard, R-Col., Senate Budget Committee; personal correspondence to Carol Edwards, dated May 29, 2007; sent through listserv.)

Clearly, there are some errors in this statement, but what else did Senator Allard, and indeed other politicians and EPA agency officials, think we didn't know?

First, the \$2 million cut in the FY2007 budget was done by the agency itself –they had selected to include the cut in their budget proposal (Pearson, 2008). Second, the closing of libraries to achieve budgetary limits comes into question when considering the very costly business of digitizing tens of thousands of documents, and the lack of cost-benefit analysis from the EPA has certainly created speculation about the cost-effectiveness for the agency if regulators and scientists are left to find information on their own. [SLIDE 9]

But, besides money, there are clearly other issues at hand here that EPA officials thought we didn't know - or at least care – about. [SLIDE 10]

The first consideration, "Who will digitize the documents?" A white paper circulated in the agency in late 2007, but as yesterday, the agency was still in the review process for finalizing their digitization plans, and reviewing current digitization practices for cost-effectiveness and accuracy (EPA website). Because the plans are not finalized, it is difficult to know if the agency is merely putting up pictures of documents, or actually

going to lengths to conduct Optical CR scanning to enable in-text searching. Already, it is clear that metadata related to the already digitized documents is lacking or insufficient as librarians are having difficulty accessing information. If readable text is being created, the high error rate of 1-5% will create a costly correction process and/or room for data modification (intentional or unintentional).

The second consideration of how users will access digitized materials is already presenting problems. Digitized documents are to be housed in the NEPIS (GET NAME) network, but evidence is surfacing that access to the digitized data is severely impeded, even for information professionals. While users should be able to access documents from home, time will tell how effective that search method will ultimately be.

The third consideration: regarding where the digital items will be housed, organized, and managed. As already mentioned, digitized documents are to be held in the NEPIS network, but their organization and search criteria standards are left to be determined. Already digitized documents are available, with some difficulty, but there seems to be no clearly set location specified for those documents awaiting digitizing, or how such documents will ultimately be organized.

The fourth consideration is probably the most widely discussed in newspaper articles, congressional hearings, and library conferences such as this one. How long the digitization process will take is anyone's guess, though the EPA claims 2 to 3 years; and the affect on public and professional access is unquestionable. Priority for digitization of documents also does not seem to be set, and it can be anyone's guess when documents will be made available electronically, or whether or not they will be made available at all. The recently released GAO report found that only 15,260 titles have so far been digitized.

Long-term interruption of service for many patrons is undeniable. [SLIDE 11]
Only 10% of the EPA collections are "unique" materials –over 500,000 copyrighted materials will not be digitized (Lee, 2008). The lack of overlap for the physical and digital collections has drawn sharp criticism from members of congress.

In March, 2007, the first claims that the library closings were threatening Community Right to Know rules for toxic chemicals were made by Senator Barbara Boxer (D-Calif.) (K.D., 2007). Not only are some collections not being digitized, they are reportedly being "dispersed" to locations where accessibility is difficult if not impossible (Goldberg, 2007) – in contrast to the Agency's own policy [SLIDE 12].

"Although it may be tempting to dispose of library materials quickly, the loss of important and unique materials could have serious future consequences if the Agency cannot document scientific findings or enforcement actions" (EPA Library Network Procedure: Library Materials Dispersion." Issued by the EPA Chief Information Officer pursuant to delegation 1-19 dated 07/07/05).

However, internal e-mails from the EPA apparently give instructions to dispose of library materials, including potentially "unique" documents (K.D., 2007). The GAO report from February, 2008 found that the Washington, D.C Headquarters Library and Chemical Library had disposed of 800 and 3,000 journals and books respectively (9).
[SLIDE 13]

The last consideration of preservation of digital surrogates can certainly not be overlooked. While digitization, or the preservation of analog content via digital means, brings greater access to information, the digital medium is by no means permanent. Of course, many of the documents printed prior to 1990 (PL 101-423) were printed on acid paper – or brittle paper – and preservation of the content of these materials may best be done digitally. Therefore, digital has its advantages, but it also has some consequences.

While digitization preserves content of documents, it disregards the physical form. Why does this matter for government documents? The authority of these documents rests in their basis in trusted research and modes of production. When dismantled from their bindings or physical status, these documents are easily altered or mismanaged within the digital space.

Some may argue that those documents born digital will pave the way for ongoing preservation of those that are digitized; however, as Galloway (2004) states, most born-digital documents are not considered for preservation until after creation, when formats, visual content, and organization so long-term preservation becomes more complicated and costly (552). The EPA is once again reviewing their digitizing procedures, leaving to question the authority and integrity of what has already been done. The process of digitizing, then, is too costly to be temporary – the life of the digital surrogate must be at least as long as the life it had in print. [SLIDE 14]

CONCLUSSIONS

How can it be fixed? Well, to be honest, I'm not sure that everything can be fixed to restore the complete authority of the complete EPA records. Perhaps I am as skeptical as the 56% of Americans who do not completely trust the government – perhaps I have good reason based on the reality that there are limitations to digitization processes, preservation of digital documents, and the ability to account for all the documents two years after the closing of several libraries and the over-night dismantling of their collections. There is room to question what black hole has been created, and into which much information is lost – or may become lost.

That said, there is some hope that the situation will improve, at least in the short term, as the closed libraries have been ordered re-opened, and a line item in the 2008 congressional budget allocated an additional \$1,000,000 for the explicit purpose of restoring the libraries and library services. In the EPA's report to congress on Wednesday, March 26, 2008, they pledged to re-open the closed libraries and work with the Federal Library and Information Center Committee of the Library of Congress to plan future services and delivery of information. However, the report is short and vague about how they actually plan to move forward. Certainly, it does not address many of the issues raised by the GAO. [SLIDE 15]

Overall, reading the GAO report is not likely to make anyone optimistic about the future of these library materials. While digitization will certainly fulfill a need, and even an expectation of government information users, it is the process of how the digitization is being carried out that is in question here.

Authority of documents no longer readily available when modified from their original format

- what documents are altered, intentionally or unintentionally?
- What documents are missing as the collection is re-arranged digitally (again, intentionally or unintentionally.)

Indeed, the integrity and authority of the EPA documents may be in jeopardy, not to mention the documents themselves. [SLIDE 14]

Bibliography

Dames, K. Matthew and Jill Hurst-Wahl. 2007. Digitizing 101. *Library Journal*, xx(x): xx.

Estabrook, Leigh, Witt Evans, and Lee Rainie. 2007. Information searches that solve problems: How people use the internet, libraries, and government agencies when the need help. PEW Internet & American Life Project; University of Illinois at Urbana-Champaign.

Galloway, Patricia. 2004. Preservation of digital objects. *Annual Review of Information Science and Technology*, xx(xx): 549-xx.

Goldberg, Carol. January 31, 2007a. EPA library system in shambles. Public Employees for Environmental Responsibility News Release.

Goldberg, Carol. February 8, 2007b. EPA library system continues to implode: Union charges EPA with unfair labor practice for refusing to consult on closures. PEER Press Release.

K.D. 2007. Senate committee questions EPA actions. *American Libraries*, 38(3): 15.

Lee, Christopher. March 14, 2008. EPA closure of libraries faulted for curbing access to key data.

McClure, Robert. January 22, 2007. EPA gets an earful on library closures: Decision derided as harmful to agency's own employees. *Seattle Post-Intelligencer*.

Moraff, Christopher. February 1, 2007. Why are EPA libraries closing? *In These Times*.

Pearson, Dan. March 10, 2008. Hearing on the closure of EPA libraries. Memo to Member and Legislative Assistants, Subcommittee on Investigations, and Oversight of the House Science and Technology Committee.

Sternstein, Aliya. November 3, 2006. EPA official tries to allay library fears. FCW.com.