

Appendix H

Assessment of Electronic Government Information Products List of Expert Interviews and Interview Questions

**Assessment of Electronic Government Information Products
List of Expert Interviews and Discussion Questions**

Interviewees	Date of Telephone Interview
Jerry Malitz, Webmaster National Center for Education Statistics U.S. Department of Education Washington, D.C.	October 27, 1998
Linda Wallace, Chief Electronic Information Services Internal Revenue Service Washington, D.C.	October 27, 1998
Evelyn Frangakis, Preservation Officer National Agricultural Library U.S. Department of Agriculture Beltsville, Maryland	November 10, 1998
Abby Smith, Director of Programs Council on Library and Information Resources Washington, D.C.	November 10, 1998
John Bertot, Associate Professor State University of New York at Albany Albany, New York	November 18, 1998
Charles McClure, Distinguished Professor, School of Information Studies Syracuse University Syracuse, New York	November 24, 1998

**Interview Questions for Webmasters:
Jerry Malitz and Linda Wallace
(October 27, 1998)**

Role of Webmaster

1. How long have you been in your current position as webmaster? When and how was the position created? Were you the first webmaster in your agency? How does the position reside administratively in the structure of your agency? What office or unit do you report to?
2. Please describe your current job responsibilities and duties. What portion of the following skills, experience, and knowledge do you use to perform your job: technical, administrative, analytical, program, other?
3. Is there a formal or informal structure for working with staff and administrators in other departments or units (e.g., program staff, IT, publications, public relations, records managers, librarians, etc.)? If yes, please describe how you interact with them?
4. What do you envision as the future role of the webmaster in Federal Government agencies? Do you see your role as being very different in 5 years than it is now? How?

Format Standards and Public Accessibility

5. Please describe the website development process in your agency from the time you receive or generate requests through design, development, evaluation, testing, and implementation, etc.
6. Has your agency developed policies or guidelines including format standards to ensure technical consistency in the development of web products that are intended for public dissemination? What are the most frequently used file formats and why? Can you identify any formats you plan to use in the future?
7. Are there limitations or specific designations of software tools that may be used to develop and implement web pages or sites? What standards are applied to configuration control and arrangement of web-based applications? Do you have a direct role in determining these standards, or are they developed at an agency or departmental level?
8. Are there general security standards applied to the availability or distribution of web-based information? What is your role in determining or implementing these standards? Which software products do you use to implement these standards?

9. How do you evaluate the effectiveness of your websites? Who is involved in the process? What methodology and criteria have you or others used to evaluate websites?

10. Has your agency discussed the concept of permanent public accessibility as it relates to Government electronic information products intended for public dissemination? How is your agency addressing the concerns of librarians, GPO, and others for ensuring permanent public accessibility for electronic Government information products?

11. What consideration are you giving to creating a metadata record for your information resources or services on the web (e.g., GILS, MARC, or specific agency locators)?

Cost Analysis

12. Linda, we know you have collected data on the comparative costs of delivering services to customers via different delivery mechanisms such as mail, e-mail, Fax on Demand, kiosks, Internet, telephone, walk-in, CD-ROM, etc. What have you learned about the costs of delivering services to customers using these different systems? Which delivery mechanisms are the most cost-effective for what types of services?

NCLIS Assessment of Electronic Government Information Products

Summary Notes from Interview with Linda Wallace (IRS) and Jerry Malitz (NCES)

1. Please describe your current job responsibilities and duties.

WALLACE (IRS):

(Wallace was a telecommunications expert and technical advisor to CIO when IRS asked her to be webmaster. She also holds the title of Chief, Electronic Information Services. She is responsible for all electronic information products including Fax on Demand, Internet, e-mail, etc.)

Wallace's three major areas are content, applications, and development. Her office:

- **Generates new services and authorware, including creating automated filters and templates through core knowledge repository.**
- **Participated in the development of SGML format (standard format for IRS since 1970s).**
- **Interacts with customers to automate a standard way to build a core knowledge repository. Repository contains automated templates and filters to generate media output to serve customers via Internet, Fax on Demand, CD-ROM, bulletin board system, telephone, or mail requests.**
 - a. **Core repository can satisfy 95 percent of requests using 86 different variables or attributes that are indexed so everything is searchable. All documents include individual catalog and document numbers.**
 - b. **Documents are always authored in SGML and can automatically be converted into a different format or posted on the web, BBS, etc., in 10-12 hours to fill customer requests.**
 - c. **Filters and templates are solely by this group. They also track history of a document.**
 - d. **A knowledge base is being built by developing a database of frequently asked questions.**
 - e. **They use ICON tagging to provide accessibility to the visually impaired. All IRS documents are ADA-compliant, online searchable, and downloadable.**

MALITZ (NCES):

As Technology Outreach Officer for NCES, Malitz services state education agencies, school districts, etc.

- a. **Each one of NCES's 30 programs has a web publisher and a web liaison.**
- b. **Contractors actually prepare materials for the web once program officer has approved content.**
- c. **Malitz sets standards, guidelines, and procedures for web publisher to follow.**

- d. **Web publisher in program area develops website on a separate development server; Malitz reviews and makes technical changes to ensure that the site meets minimum standards and guidelines.**
- e. **Sometimes, he develops a new application for others to use (e.g., NEWS FLASH subscription service featuring daily breaking news from the Department of Education).**

2. Do you have a formal or informal structure for working with staff and administrators in other departments or units (e.g., program staff, IT, publications, public relations, records managers, librarians, etc.)? If yes, please describe how you interact with them.

WALLACE:

Wallace deals with high-level senior executives, reviewing their business plans, problems, and goals, then recommending solutions that include productivity measures, production rates, and cost per person. In one case, she recommended a business CD-ROM.

- **Establishes strong liaisons with industry (has marketing person on staff). They receive one-half of funding from industry to support business projects that benefit industry and IRS customers.**
- **Established various delivery service programs: Internet in '96, Fax on Demand in '96, CD-ROM in '95, BBS a while ago.**

MALITZ

- **Each one of NCES's 30 programs has a web publisher and a web liaison.**
- **Contractors prepare materials for the web once content has been approved by the program officer.**
- **Web publisher in program area develops web site on a separate development server; when finished, Malitz reviews and makes technical changes to ensure that site meets minimum standards and guidelines.**
- **Sometimes, Malitz develops a new application for others to use such as the NEWS FLASH subscription service that features daily breaking news from the Department of Education.**
- **Malitz's work is divided fairly evenly between technical, administrative, analytical, and program areas.**

3. What do you envision as the future role of the webmaster in Federal Government agencies? How you see your role as in 5 years as compared to now?

MALITZ:

The role of webmaster will be completely different. In the future, he/she will have more of a coordinating function and will set policies and procedures. The program staff will be forced to do their own work on the web, just as they now do their own word processing and e-mail.

WALLACE:

The role of the webmaster will be that of an enabler for business units with everyone involved. There will be more of a focus on multimedia (e.g., BBS, CD-ROM, Fax on Demand) and not just the Internet.

4. Please describe the website development process in your agency from the time you receive or generate requests through design, development, evaluation, testing, and implementation, etc.

WALLACE:

Her unit receives and generates requests. Requests from the core repository can fit into an existing template filter application.

- a. **Staff and contractors conduct testing and implementation.**
- b. **Second year after web in operation, requests for paper copies of forms dropped by 50 percent.**
- c. **Provide hidden codes to track where returns come from: fax, Internet, libraries, phone requests, etc.**
- d. **Evaluation includes a simple three-question customer service survey on content: did you get what you needed, where would you have gone if not here? They build evaluation into every step of the process.**
- e. **A panel of experts measures the effectiveness of their websites. Also, they have partnered with schools to recruit instructors and students to review site before they go "live."**
- f. **One person reviews all e-mail messages that contain feedback on website; use automated sorters by key words to batch the type of feedback received.**

MALITZ:

Before website, customers were very specialized. Most were data file users. After they created their website, their customer based increased tremendously. Now the culture is different and NCES is dealing with questions from the general public.

- a. NCES uses a developmental server but is planning to implement a Point-to-Point Tunnel Protocol (PPTP) so the developmental server is behind the firewall and no longer open to everyone. Only web publishers and contractors will have access to server.
- b. Malitz never reviews content; that is done by individual program staff.
- c. NCES conducts customer surveys of users to develop and refine sites.
- d. Malitz does database development and tests multiple browsers. NCES is UNIX-based; rest of ED is Windows-based.

5. Has your agency developed policies or guidelines including format standards to ensure technical consistency in the development of web products that are intended for public dissemination? What are the most frequently used file formats and why? Can you identify any formats you plan to use in the future?

WALLACE:

Formats most frequently used are SGML, PDF, HTML, and Postscript, respectively. They will add XML soon. They train authors to use SGML. SGML is “intelligent” data that can automatically generate other formats. Most agencies do not use SGML because it is harder to author in. Wallace’s agency uses it because it is much more robust, and it is easy to change a document format to match customer needs (e.g., tax law information for consumer and for lawyers).

MALITZ:

NCES uses PDF, then HTML (optional). They rarely put entire publication in HTML.

6. Are there limitations or specific designations of software tools that may be used to develop and implement web pages or sites? What standards are applied to configuration control and arrangement of web-based applications? Do you have a direct role in determining these standards, or are they developed at an agency or departmental level?

WALLACE:

Their focus is knowledge-based, not web application. Her department sets the standards. They use C++ and Perl.

MALITZ:

All publications are in PDF; all else in HTML. They use SQL databases to support the web. Malitz has a direct role in determining standards.

7. Are there general security standards applied to the availability or distribution of web-based information? What is your role in determining or implementing these standards? Which software products do you use to implement these standards?

WALLACE:

The IRS uses an automated redacting scheme; with one keystroke, they can create a public and specialized version of the same document. They apply all security standards from the Government, including SSA, Treasury, etc. They cannot reveal security software.

MALITZ:

Only a few people who use developmental server have access; all must be registered users. NCES uses PPTP encryption for the developmental server.

8. Has your agency discussed the concept of permanent public accessibility to electronic Government information products intended for public dissemination? How is your agency addressing the concerns of librarians, GPO, and others for ensuring permanent public accessibility for electronic Government information products?

WALLACE:

All tax forms, instructions, publications etc., are available for 5-6 years online. The core knowledge repository maintains material for 14 years, but they do not keep every application back that far. IRS can fill e-mail requests for information or forms from earlier years. In addition, they provide GOLD CARD SERVICES for librarians. Librarians have their own page, track orders, and talk "live" with one another. IRS gives their orders priority.

MALITZ:

The issue of permanent public accessibility is currently under discussion.

9. What consideration are you giving to creating a metadata record for your information resources or services on the web? (e.g., GILS, MARC or specific agency locators)?

WALLACE:

GILS records are a subset of the 86 variables that go into the core knowledge repository.

MALITZ:

The Dept. of Education has an agency locator with total search capability. They also participate in FedStats, White House Briefing Room, etc.

10. Linda, we know you have collected data on the comparative costs of delivering services to customers via different delivery mechanisms such as mail, e-mail, Fax on Demand, kiosks, Internet, telephone, walk-in, CD-ROM, etc. What have you learned about the costs of delivering services to customers using these different systems?

WALLACE:

Breakdown of comparative costs follows:

- a. **It costs IRS \$3 per call for the public to call into their toll-free number and for IRS to fill the request. The cost to IRS for the public to use the Internet to access and use forms is 1 cent, a difference of 300 to 1.**
- b. **The costs to create forms on Internet have gone down, but the cost to fill phone requests remains the same.**
- c. **It costs IRS \$2.50 to make and distribute to public libraries each CD-ROM containing 5 years of tax forms, instructions, and publications. This is less than it takes for IRS to respond to one telephone call. The IRS also sends tax CD-ROMs to the depository libraries. They can mount them on their PCs or allow customers to check them out.**
- d. **They found that kiosks are very expensive; ATMs are cost-effective**

Final Comments from Wallace

The answer to public accessibility is *not* the Internet; it is multimedia. Delivery mechanisms must meet the individual needs of the customers; no one size fits all.

**Interview Questions for Preservation Specialists:
Evelyn Frangakis and Abby Smith
(November 10, 1998)**

1. How long have you been in your current positions? Please describe your current job responsibilities and duties. What portion of the following skills, experience, and knowledge do you use to perform your job: technical, administrative, analytical, and other?
2. What are the key problems associated with digital preservation?
3. What key policy, organizational, economic and other non-technical issues need to be addressed or solved to facilitate digital preservation?
4. What technological strategies or models have various organizations such as the Association of Research Libraries, the Digital Library Federation, National Archives, etc., identified to address these problems? Evelyn, one of the NCLIS staff mentioned that NAL has established a structure or framework that addresses this problem. Could you please talk more about that? If you have any handouts you can fax to us, that would also be helpful. Abby, can you describe some of CILR's recent efforts to address the issue of digital preservation, including the survey by Jeff Rothenberg of the RAND Corporation?
5. What do we know about specific file formats or mediums that might facilitate digital preservation such as SGML, CD-ROM, etc.?
6. Are there any important preservation issues that we have not addressed in the above-listed questions? If so, please discuss them.
7. Could you please refer us to any important articles on this topic that have been published in the last year?

**Summary of Notes from Conference Call with Two Preservation Specialists:
Evelyn Frangakis (NAL) and
Abby Smith (Council on Library and Information Resources)**

<p>1. How long have you been in your current position? Please describe your current job responsibilities and duties. What portion of the following skills, experience, and knowledge do you use to perform your job: technical, administrative, analytical, and other?</p>
--

ABBY SMITH (CLIR)

- **Been with CLIR since Sept. 1997 as director of programs.**
- **Provide program coordination among the four areas: economics of information, leadership in libraries and archives, digital libraries, preservation and access.**
- **Her primary program responsibility is in preservation and access in libraries, traditional and digital.**
- **10 staff members; 6 professionals, 4 admin. support staff.**
- **Spends 75 percent of time on policy-related issues and the remaining 25 percent of time spent on administrative functions (i.e., coordinating publications program)**

EVELYN FRANGAKIS (NAL)

- **Been in current position since January 1997. She is NAL's first preservation officer.**
- **Duties: plan, direct, and implement agency-wide programs for ensuring permanent and future accessibility of the foremost national collection of materials in agriculture.**
- **Coordinates activities with other national efforts, such as the U.S. Agricultural Information Network (USAIN). Established in 1988, USAIN provides a forum for discussion of agricultural issues, takes a leadership role in the formation of a national information policy as related to agriculture, makes recommendations to the National Agricultural Library on agricultural information matters, and promotes cooperation and communication among its members. NAL participates in implementing USAIN's preservation plan for print materials, *A National Preservation Program for Agricultural Literature*. The USAIN Preservation Steering Committee, on which Frangakis serves, oversees this national cooperative plan. Under the auspices of Cornell University, the USAIN plan has received two NEH grants to microfilm core national and state agricultural literature. To date, 15 states are participating in these grants. Other components of the national plan and program include determining what are the important archival and manuscript collections of agricultural materials and what approaches can be used for their preservation.**
- **NAL efforts include developing their own preservation program that includes a traditional preservation program and digital efforts.**
- **Digital efforts are two-pronged: conversion of brittle paper materials into digital products by working with best available guidelines to implement good preservation practices (this digital**

material will be available on the web); develop a program to preserve USDA digital materials (i.e., materials that are born digitally).

- **Helps develop preservation policies and analyze other policies that come to NAL or USDA that affect preservation of the collection.**
- **Time spent on different types of work at different times. Duties fairly split among policy, technical, administrative, analytical areas.**
- **Staff consists of two assistants at present. However, NAL leverages its preservation resources by establishing cooperative inter-institutional agreements and contributing funds to sister institutions in order to further develop the USAIN preservation program (e.g., cooperative agreements with Cornell University to establish copyright clearance for core historical literature, developing NEH grant proposals).**

2. What is the distinction between digital preservation and permanent public accessibility of electronic records (as it relates to, for example, the Federal Depository Library Program)? How long is "long-term" preservation vs. "permanent public accessibility?"
--

FRANGAKIS

- **Some background from the USDA perspective: The USDA Digital Publications Preservation Steering Committee was established this past summer to oversee the implementation of the plan, *A Framework for the Preservation of and Permanent Public Access to USDA Digital Publications*. This group met for the first time in October 1998. There was a discussion of definitions in order to place into context the universe of material covered by the *Framework*. Publication was defined as “a data or information product prepared by the USDA in digital form intended to be disseminated to the public.” The *Framework* defines preservation as “the act of permanently maintaining and making available data or information, with all original content intact.”**
- **Other experts, such as Don Waters of the Digital Library Federation, talk about preserving integrity and ensuring persistence of digital information. The Commission’s SGML report talks about preservation goals such as enhancing the long-term preservation of and access to information of enduring value for as long into the future as possible.**
- **THE CPA Digital Archiving Task Force was charged to investigate the means of ensuring “continued access indefinitely into the future of records stored in digital electronic form.”**
- **Concept of preservation in traditional preservation world examines the concept of permanence, but in the print world the concept of permanence relates to chemical inertness and mechanical durability. These concepts do not translate easily into a digital world. In the digital world, we are no longer dependent on original copies (i.e., original copies do not have the same meaning.).**
- **Within NAL, they use digital preservation fairly loosely to speak about both digital efforts: conversion of brittle materials and USDA digital publication preservation efforts. Not sure professionals in the library and preservation community have a common understanding of what it means, even though it is important to come to a common understanding. GPO defines permanent access as “Government information products within the scope of the Federal**

Depository Library Program that remain available for continuous no-fee public access through the program.” The 1996 GPO report to Congress, *Study to Identify Measures Necessary for a Successful Transition to a More Electronic Federal Depository Library Program*, states that “‘preservation’ means that official records of the Federal Government, including Government information products made available through the FDLP, which have been determined to have sufficient historical or other value to warrant being held and maintained in trust for future generations of Americans, are retained by the National Archives and Records Administration (NARA).”

- At NAL, the mission of its preservation program is "to preserve and ensure access to the intellectual content and physical composition of agricultural works of national and international importance indefinitely into the future." No timeframe is mentioned because no one at this time can say how long into the future information will be needed.

SMITH

- There is no standard accepted method of ensuring long-term access to digital information. She described preservation goals as permanent or persistent or perhaps more accurate to say that one of the primary goals of preservation is to set up systems that "sustain predictable levels of loss."
- Difference between preservation in a digital world and in an analog world is that in a digital world, information is completely independent from the medium on which it is carried. In an analog world, people try to preserve the media in which information is recorded. No analogy in the digital world. No concept of preserving the artifact as an artifact that has its own level of information.
- Problem in digital preservation is that there is no way to ensure that digitally stored information can move from one software-hardware configuration onto another through generations. Two problems:
 - (1) Problem of instability of media in which information is stored (don't know how long CDs or other media will last)
 - (2) More serious issue from CLIR point of view is that software/hardware configurations on which information is stored become obsolete so quickly that even when you migrate information from one system to another, much of the information is lost (data and functionality).
- CLIR tries not to talk about "digital preservation" but they cannot avoid it.
- Other countries may view preservation differently. England interested in American concept of digital archiving -preserving the integrity of data; that is, information is original and authentic and it can be proved that data have not been changed. According to Smith, scientists say that we will solve the problem of authentication, but hasn't been solved thus far.

Permanent Public Accessibility Issues (FRANGAKIS)

- Federal agencies relying on FDLP to serve an "archiving" function for retrospective materials.
- Question of how to preserve digital information indefinitely into the future has not been answered. CLIR and NAL are discussing strategies by opening up dialogue and promoting research in this area.
- Digital Archiving Task Force Report discussed ensuring the integrity and long-term availability of digital information through migration. Information on CD-ROM and other media is in a format that may or may not be readable into the future due to hardware/software obsolescence. Even if media could be preserved, no guarantee that it would be accessible and functional indefinitely into the future. No answers anytime soon.

3. What are the key problems associated with digital preservation?

SMITH

- Additional problems: fragility of media and platform dependence issue
 - (1) Two additional issues: difficulty of understanding what we can and cannot do under current copyright law. Latest iteration of copyright law clarifies copyrighting for preservation purposes, but still unclear for access purposes. Library of Congress is currently studying this. Copyright law may not have any implications for Government information but many vendors create derivatives of Government information and copyright it. Government should never be in the position of depending upon the private sector to preserve some of this information.
 - (2) Any transmission link is as strong as the weakest link. The weak link in the transmission of electronic information is not technology; it's human beings. Human infrastructure is not in place yet that would ensure permanent access.

FRANGAKIS

- Agrees that human error is far more prevalent than technology error. Key problems in digital preservation are infrastructure, technology, and media. Humans need to learn to live, exist, and operate in a digital world. It's still very new to us as compared to the print world.

4. What key policy, organizational, economic, and other non-technical issues need to be addressed or solved to facilitate digital preservation?

SMITH

- Digital Archiving Task Force Report met with consensus in the community. CLIR has been lobbying people to pay attention to these issues. No single community has stepped forward and said this is our problem and we are going to work on solving the problem. Therefore, one of difficulties is that organizations that collect, preserve, and disseminate information, as opposed to create information, find themselves in this digital world in which the preservation of that

information must be thought about at the creation stage, not after the fact. Need to forge partnerships with the computer science industry, publications industry, scholarly and scientific publishing communities to address some of these issues.

- One of perhaps intractable core infrastructure problems is the issue of creating a failsafe archives mechanism for materials that disappear from the web. What happens when information is created and the people who created it do not have responsibility for preserving it? Who is going to authorize a failsafe archive that is going to take and preserve that information for the public good? This would be the equivalent of libraries, but so far, it doesn't exist and no one has expressed interest in creating it at the Governmental level.
- CLIR's role in above: Not in a position to do much more than alert people about the problems. NAL and literature that agriculture creates are one of few examples where this failsafe archive might work because NAL is a national library dedicated to one type of literature. Not the case with other literatures except for medicine (NLM). CLIR is looking for partners like ARL to address this issue, but has not made much progress. CLIR has been fairly effective in talking to National Science Foundation (NSF) in getting their second round of digital library initiatives grants to address the issue of preservation as a distinct issue. No luck with archiving part.

FRANGAKIS

Refer to the report Framework for the Preservation of and Permanent Public Access to USDA Digital Publications by Paul Uhler, November 1997 (listed in bibliography).

- Three areas of issues:
 - (1) management structure and organizational relationships within and outside USDA,
 - (2) funding of program on a permanent basis (keeping in mind the need to minimize costs of access and retrieval to information users), and
 - (3) identification of legislative or administrative actions or policies required to implement a digital publications preservation program.

Sub-Issues: Needs and Considerations for USDA

- **Inventory and life cycle management:** a comprehensive inventory of all departmental digital information products and how they are being managed needs to be conducted. A system for tracking the creation of each new USDA digital information product that is intended for public distribution needs to be recommended.
- **Technical requirements:** identification of acceptable document formats and media, and related standards, for long-term retention; development of processes for transferring all digital publications from old storage media to new media; establishment of one or more separate back-up facilities for all digital publications; review and establishment of system security protocols; review and establishment of system interoperability requirements; and identification and review of other permanent digital preservation and access initiatives.
- **User access and retrieval:** provide equitable access and retrieval services to all potential users; minimize technical, regulatory, and cost barriers to access and retrieval; assure the integrity of

the information that is made publicly available; make the information as easy to find and use as possible, with directories and documentation (metadata), consistent with the Government Information Locator System, while protecting confidential or proprietary information; and establish a means for users to provide feedback and a mechanism for responding to user feedback.

- **Status: Moving ahead with implementation. USDA CIO accepted the report, and under her guidance NAL established a national steering committee made up of representatives from USDA and from agribusiness, research library community, USAIN, Federal partners, etc.**
 - **Group will meet on a quarterly basis for first 2 years.**
 - **Will establish test groups to explore issues such as inventory and life cycle management, technical requirements, and user access and retrieval, as well as funding issues.**
 - **Hoping to get funding for a pilot project and then take entire framework and test it on an agency within USDA to see how manageable Framework will be for full-scale implementation.**

5. What technological strategies or models have various organizations such as the Association of Research Libraries, the Digital Library Federation, National Archives, etc., identified to address these problems? Evelyn, one of the NCLIS staff mentioned that NAL has established a structure or framework that addresses this problem. Could you please talk more about that? If you have any handouts you can fax to us, that would also be helpful. Abby, can you describe some of CILR's recent efforts to address the issue of digital preservation, including the survey by Jeff Rothenberg of the RAND Corporation?

SMITH

Three CLIR initiatives:

- **Commissioned report by Jeff Rothenberg from RAND Corporation on emulation. [Emulation is the process of imitating one system with another so both accept the same data, execute the same programs, and achieve the same results.] Report complete and may be published by January 1999. Since report is highly controversial, CILR will partner with National Research Council to convene a group of computer scientists to engage Rothenberg on issues of emulation to stimulate research. Report describes the weaknesses of migration and the strengths of emulation and sets up a research agenda to develop emulation.**
- **Commissioned an analysis of migrating file formats to do a risk assessment associated with those file formats during migration. Study commissioned from Cornell using data from the Mann Library (agricultural library) and will use numeric file formats and databases and text formats. Report will be finished by September 1999 and will include analysis and a template that others can use for doing a risk assessment of migration of those file formats. Purpose: to stimulate further research.**

- **Identified a computer scientist at Carnegie-Mellon University (CMU), John Ockerbloom, who has developed a system of file conversion; type of migration that converts web-based materials to different file formats, called TOM (Typed Object Model).**
(www.cs.cmu.edu/afs/cs.cmu.edu/user/spok/www/defense/index.html). **He developed this as part of his thesis. Working with CMU to see if they can bring his concepts into fuller application to do an assessment about its scalability.**
- **Log on to publications on CLIR site, which summarize Rothenberg report. Water's report addresses definition of digital preservation.**

6. What do we know about specific file formats or mediums that might facilitate digital preservation such as SGML, CD-ROM, etc.?

SMITH

Nothing to say about this

FRANGAKIS

- **For conversion efforts from paper to digital images, SGML serves as an important descriptive markup tool. Thinks it will be valuable to them. CD-ROM serves specific functions in NAL's Preservation Program but right now has a limited life expectancy. NAL is looking for things that are non-proprietary, platform independent, things that will allow user full access to the content of digital products. They know that media will continue to change.**

7. Are there any important preservation issues that we have not addressed in the above-listed questions? If so, please discuss them.

No.

8. Could you please refer us to any important articles on this topic that have been published in the last year?

1. **Margaret Hedstrom at the University of Michigan School of Information believes that there is a reliable way of preserving, with predictable levels of loss, migration of digital information, through ASCII. It happens now. She is a leading authority in the field.**
2. **Check ARL's website.**
3. **Coalition of Networked Information - Dedicated to computer use in education.** www.cni.org
4. **White paper on access authorization. Developing infrastructure for digital libraries.**
5. **www.RLG.org/preserv - (includes information on Hedstrom's research).**
6. **Reference Model for Open Archival Information Systems:**
http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html
White Book, issue #4, Sept. 1998- preservation of digital information; technical recommendation for use in developing consensus on what's required of any archive to provide permanent preservation. Hoping to turn this into an ISO standard, but now just in draft form.

**Interview Questions for Information Resources Specialist:
John Bertot
(November 18, 1998)**

1. What are the primary obstacles to successful information resources management (IRM) practices in the Federal Government, in priority order? What changes should occur to eliminate or alleviate the barriers?
2. In your article on the impact of Federal IRM on agency missions, you mention the reinvention of IRM to be the key link between agency information and agency performance. Can you describe a small and a large Federal agency that currently meet this goal in spite of the lack of a concentrated, coordinated Federal IRM policy? Why are they more effective than other agencies?
3. We have found in our interviews with Federal agency personnel that many agencies have not come to terms with two important issues: information life-cycle management, and the concept of permanent public availability of electronic Government information. What are some of the larger policy issues that have prevented agencies from addressing these important issues?
4. In our site visits to Federal depository libraries in the D.C. metropolitan area, we are keenly aware that the problems and issues faced by FDLs here are different than they might be for FDLs located in more isolated, rural areas. Based on your survey of public libraries connected to the Internet, what are some of the key concerns or problems faced by users who want to access electronic Government information who live in small, isolated communities with limited resources?
5. Based on your experience in working with Federal Government agencies that are analyzing their web usage, what are the key questions they want to answer and how are they using the data? Are they analyzing the websites for technical or content-related purposes? What techniques, other than log file analysis, are being used? What agency units or departments (e.g., IT, program areas, CIO) are involved in the process?

**Interview Questions for Information Resources Specialist:
Charles McClure
(November 24, 1998)**

1. What is the current status of IRM policy since your 1994 article, "Federal Information Resources Management: New Challenges for the Nineties?" Specifically, has OMB or another appropriate agency begun addressing the issue of developing a broad vision that reflects the evolving role of IRM within the Government with general guidelines and standards for all Federal agencies to better manage the life cycles of information? If so, how?
2. We have found in our interviews with Federal agency personnel that many agencies have not dealt with two important issues: information life-cycle management, and permanent public accessibility of electronic Government information. What are some of the larger planning, policy, and organizational issues that are preventing agencies from addressing these important problems?
3. Could you talk a little more about the design-based assessment for evaluating Government websites that you described in the 1997 Proceedings of the 60th ASIS annual meeting? What were the technical and policy problems in the design-based assessment? What specific policy issues did you assess?
4. What is the status of electronic record management (ERM) guidance for Federal agencies since the 1998 conference? Is there another conference planned next year? What specific guidance are the NARA Working Group and other agencies planning?
5. What would you say are the top three Federal IRM challenges in the next decade?

Summary of Notes
Interview with
John Bertot, Associate Professor, SUNY/Albany

1. What are the primary obstacles to successful IRM practices in the Federal Government, in priority order? What changes should occur to eliminate or alleviate the barriers?

- **IRM is not on the radar for top-level agency managers and it will never be raised up to the point of where it matters.**
- **IRM has been lost in the transition to the CIO. CIO is the next iteration of IRM. There has been 20 years worth of talking and it has never seemed to make it out of the administrative trenches of the agencies. Typically, IRM has been a low-level position located within the printing, reprographic, or records management units of agencies. IRM is not viewed as strategic or long-range function.**

2. In your article on the impact of Federal IRM on agency missions, you mention the reinvention of IRM to be the key link between agency information and agency performance. Can you describe a small and a large Federal agency that currently meets this goal in spite of the lack of a concentrated, coordinated Federal IRM policy? Why are they more effective than other agencies?

The article is based on Bertot's dissertation.

The purpose of the survey was to get an internal assessment on what IRM is trying to do, and to get an external assessment of what IRM is doing. Bertot tried to compare the two in relation to strategic planning.

Generally, those agencies that understood IRM tended to be the smaller agencies. There is a scale factor, and much more attention was given to IRM in the smaller to mid-sized agencies. FDIC and the Peace Corps were doing some interesting things.

Other factors relating to size:

- **Small to medium-sized agencies have fewer programs and staff; with fewer administrative layers, there are fewer communication and organizational barriers.**
- **One can more easily work collaboratively in a small organization.**
- **Top administrators are not as removed from day-to-day operations and can ideally participate more in implementing new initiatives because they have a vested interest in the projects' working.**
- **There is less oversight from OMB and Congress for smaller agencies. The smaller agencies tend to have less mandated legislation that can interrupt work, so they tend to have higher motivation to finish projects.**

- **Larger agencies tend to have the greater expertise. Smaller agencies may have better levels of management, but they do not always have the experts.**

Models

- **As far as larger agencies were concerned, Treasury was moving along. However, many have a central agency component that is not very powerful, although subagencies might be very powerful. For example, Treasury has IRS and the FBI, and they are pretty powerful players.**
- **Another agency that has done a great deal of work with IRM is EPA. EPA, along with AID, are strange models. They have large data shops, but they are all contractors. The model that is adopted for information and information technology management makes a difference. Whether IRM is in-house or out-sourced has a real impact on how it is implemented inside the agency, and the choice of contractor really matters. Another aspect that makes EPA unique is that a large portion of their system management function occurs in North Carolina.**

3. We have found in our interviews with Federal agency personnel that many agencies have not come to terms with two important issues: information life cycle management, and the concept of permanent public availability of electronic Government information. What are some of the larger policy issues that have prevented agencies from addressing these important issues?

- **The biggest barrier to successful implementation of IRM is that agencies do not view information as a resource. There is little or no understanding of the concept of information as a life cycle; it's not linear.**
- **IRM policy initiatives and legislation do not fully address the life cycle of information. It is mentioned in some of the policies developed within the last 20 years. but not adequately addressed. Most policy initiatives focus on the technology side of the issue, probably because it is tangible.**
- **The web has created problems that have not been handled. Many agencies believe that if it is up on the web, it has been published. Along with the pressure over Title 44 Reform, there is no discussion of preservation and public accessibility. Should we move to an electronic FDL program? What does that mean and how will that work?**
- **GPO is under attack for being deficient in distributing Government information to the public. One reason is because GPO (centralized print environment) is so slow, and the technology allows distribution to be handled more efficiently (decentralized, electronic environment). Agencies are under the gun to cut costs, so by putting information on their websites and contracting printing jobs with outside sources, they don't have to go through GPO.**
- **Going electronic does offer potential. The back end means, however, that anyone with access is a vehicle for getting Government information. Putting information up on the web does cut costs, but we have not figured out a systematic way of distributing Government information to the public, making sure it is preserved for posterity and provided to the public on a long-term basis.**

4. In our site visits to Federal Depository Libraries in the D.C. metropolitan area, we are keenly aware that the problems and issues faced by FDLs here are different than they might be for FDLs located in more isolated, rural areas. Based on your survey of public libraries connected to the Internet, what are some of the key concerns or problems faced by users who want to access electronic Government information who live in small, isolated communities with limited resources?

Recently Bertot conducted some research in rural Pennsylvania to study public libraries. Public libraries in rural communities face large problems with access and technology.

- **These areas are composed of populations that tend not to have computers in the home, so they rely totally on the library for Internet access.**
- **The public libraries tend to only have one station in these rural areas.**
- **Computers are slow; libraries have 56 K modems, but they do not necessarily have access to a 56 K Internet provider.**
- **Patrons can only reserve computers in half-hour time slots.**
- **Libraries may or may not have access to print equipment.**
- **Staff training is minimal due to cost and little access to computers. (Models like GPO Access have been useful, but patrons and librarians still need one site for access to all Government agencies rather than many individual websites. Most agencies have more than one site.)**
- **Staff are competing with patrons for access because there are only one or two computers.**
- **Libraries cannot always pay for printing services. Some are passing the cost to the user, but they are trying to avoid that approach (e.g., first 5 pages are free and then it is 10 cents per page).**

5. Based on your experience in working with Federal Government agencies that are analyzing their web usage, what are the key questions they want to answer and how are they using the data? Are they analyzing the websites for technical or content-related purposes? What techniques, other than log file analysis, are being used? What agency units or departments (e.g., IT, program areas, CIO) are involved in the process?

- **Bertot has not seen agencies doing very much with their web statistics. Part of the reason is that they do not want to make the information public. For example, one agency had a request for all of their agency log file records. They panicked and rejected the request on the condition of privacy. If one can access the log, one can get IP addresses, and they were afraid that someone would use this information as a reverse directory mailer.**
- **A second reason is that sometimes it is difficult to get the statistics if a different administrative unit within the agency is managing the website. They won't necessarily turn them over to the unit that needs the statistics because it crosses administrative barriers.**

- **A third reason relates to records management of log files. Should we "schedule" log files for NARA? NARA doesn't want this to happen either. They would then have to schedule the information for retention. However, this raises the question of whether the logs are public information.**
- **It was not until recently that there was a demand to look at web statistics as a management and strategic decision-making tool. Managers are still learning how to use them. Right now, these statistics are used primarily by network and system administrators.**
- **Bertot wonders how many agencies are doing web analysis and evaluation given Circular 130-A, which cautions agencies not to do so if it creates a paperwork burden for them.**

Summary of Notes
Interview with Charles McClure,
Distinguished Professor, School of Information Studies, Syracuse University

- | |
|---|
| <p>1. What is the current status of IRM policy since your 1994 article, “Federal Information Resources Management: New Challenges for the Nineties”? Specifically, has OMB or another appropriate agency begun addressing the issue of developing a broad vision that reflects the evolving role of IRM within the Government with general guidelines and standards for all Federal agencies to better manage the life cycles of information? If so, how?</p> |
|---|

IRM policy came and went and no one noticed:

- **GSA is mounting its CIO university effort to provide education and training to CIOs.**
- **IRM in Government policy is now whole world to CIO.**
- **Many agencies do not now know what to do with IRM.**
- **ITMRA (Information Technology Management Reform Act of 1996)—McClure thought this policy would strengthen IRM, but in reality, it took responsibility away from existing IRM people and gave it to CIO; it gave more attention to technology management.**
- **A few agencies don't know what to do with IRM staff since CIO is on board; in other agencies there is conflict between IRM and CIO functions.**
- **Eighty-two percent of technology efforts in agencies are currently focused on Y2K efforts.**

- | |
|---|
| <p>2. We have found in our interviews with Federal agency personnel that many agencies have not dealt with two important issues: information life cycle management, and permanent public accessibility of electronic Government information. What are some of the larger planning, policy, and organizational issues that are preventing agencies from addressing these important problems?</p> |
|---|

- **There is no staff or time to devote to standards and interoperability.**
- **Even if agencies had staff and time, staff need to upgrade skills and knowledge.**
- **Information life cycle and permanent public accessibility are not priorities for agencies; they don't seem to understand the issues.**
- **For example, GPO Reform Bill is dead in the water; no one in Congress cared about it.**
- **No one is concerned about long-term accessibility.**

3. Could you talk a little more about the design-based assessment for evaluating Government websites that you described in the 1997 Proceedings of the 60th ASIS annual meeting? What were the technical and policy problems in the design-based assessment? What specific policy issues did you assess?

They are now using more advanced website methodology (i.e., a 4-legged approach):

- **User-based: Usability testing; simulates user searching that is videotaped. With fairly sophisticated graduate students, they use scripted search analysis with a range of criteria. System and design staff are showing videotapes to so they can see the problems with searching information on specific sites. Agencies have used different audiences to do testing based on objectives and purpose of sites.**
- **Log analysis: Using in-house scripts beyond WebTrends and Log Tracker that allow them to do cross-file analysis with access vs. error and browser files. Perl scripts allow them to dump selected variables in log files into SASS or SPSS. Commercial products do not do cross-log analysis well.**
- **Policy analysis: Internal policies (who's in charge) and external policies (e.g., Freedom of Information Act, public access, privacy issues, etc.).**
- **Management assessment: How is agency department organized for web maintenance and evaluation? Information is gathered through interviews and focus groups with managers.**

4. What is the status of electronic records management (ERM) guidance for Federal agencies since the 1998 conference? Is there another conference planned next year? What specific guidance are the NARA Working Group and other agencies planning?

- **Update: McClure and Tim Spreche are working on a new project, PARS- Public Access Rating System. The purpose of this project is to create a core set of performance measures and indicators (now being developed for four Government agencies) with public access criteria to help agencies rate their websites. Agencies will be able to determine the degree to which the site is publicly accessible. It's difficult to sell ERM by itself, so they are taking a public access approach.**
- **National Archives hasn't done much. Court ruling delayed 6-8 months.**

5. What would you say are the top three Federal IRM challenges in the next decade?

- **IT management policy development is on hold due to Y2K. No one currently knows how good or bad preparation is for this.**
- **How best to integrate and coordinate IT and IT management. Agencies do not have a good handle on this yet.**

- **Issues of interoperability and standards that cut across all agencies. Need to be able to access Government information horizontally rather than vertically. (For example, for public access sites; GILS, gov.doc 1, and one more; no way public can access specific information from one point of entry. GILS does not work the way it was originally conceived.)**
- **Lack of money for training and education. IRM graduate students' degrees are useful for about 1-2 years. After that, their skills are 50 percent out of date. Government agencies have well meaning people who don't have the knowledge and skills to implement policies. For example, agencies say they don't need to send copies of all products to GPO because they are on their website. Then you ask them will it be there in 6 months and they have given no thought to this issue.**