

Proceedings of the 6th Annual Federal Depository Library Conference

April 14-17, 1997

USDA Publications: Creating a Preservation Action Plan: Planning Summary

Gregory W. Lawrence, Cornell University
Ithaca, NY

Introduction

As you all know, the U.S. Government collects an enormous amount of public information and is in turn a huge publisher. The U.S. Department of Agriculture (USDA) economic agencies, the Economic Research Service, the National Agricultural Statistics Service, and the World Agricultural Outlook Board, are no exception to this phenomenon. Where they are different is how early each of these agencies adopted digital technology in order to disseminate their information. In the mid-1980's, these agencies disseminated time sensitive commodity reports via the Martin Marietta CIDS system, and a decade later, they are working with the Albert R. Mann Library at Cornell to disseminate this information on the Internet. Their partnership with Mann Library has created the USDA Economics and Statistics System, a network accessible, research-level core collection in agricultural economics. The collection now includes 79 full text serials, over 200 historic data series, and grows by some 400 Mb per year.

When this partnership began three years ago, USDA digital publications duplicated print documents. Issues concerning the authority and preservation of the digital files were minimized by the broad availability of the print copy. Unfortunately, the economic agencies are slowly phasing out print publications. For instance, ERS (Economic Research Service) has converted the Situation and Outlook commodities series to a digital only format. Or, the agencies are creating new publications which do not have a print counterpart, such as the NASS (National Agricultural Statistics Service) Crops County Data annual. As the USDA Economics and Statistics System grows it is gradually becoming a unique, definitive collection of digital USDA information.

The Problem

Maintaining this collection for future data users presents serious management problems.

First, we are confronted with the need for data migration. Obsolescence of the original data format is a barrier to the long term use of the data. Data-sets and reports are prepared with a wide variety of formats, including plain ASCII, formatted text such as PDF, spreadsheet files and binary files used by proprietary software. It is unclear how, and by whom, these publications with different file structures will be converted to new file formats as computing standards change. Clearly, a non-Federal organization such as Mann Library should not modify USDA publications without official oversight, since this runs the risk of violating the authority of the document.

The second problem is the need to move the data to an appropriate storage medium. We perform a routine backup of the USDA system, but the procedure saves only a copy of the information as organized in a hierarchical file system. To retire aged data in an appropriate manner requires a full range of processes, such as file collation and verification, and the use of archiving and compression software. Ultimately, the data must be copied to the proper medium such as magnetic tape cartridge or CD-ROM.

Unfortunately, the data management procedures and the storage medium for these documents have not been specified, either by the three agencies, or the Department. This inhibits the library from taking proactive steps to preserve this information, since without official guidance the effort may be wasted if it doesn't ultimately conform to a Departmental standard.

If the problem were local to the USDA Economics and Statistics System there might not be much alarm. But we have been contacted several times to verify if we will preserve the data we hold, and it has become clear the problem is generating considerable concern. In the meantime, agency home pages are filling up with digital publications, and other academic/USDA partnerships, similar to the one at Mann Library, are developing large data collections elsewhere on the Internet. The result is a rapidly expanding body of digital information which potentially is at risk because there isn't a general policy to coordinate the safe and systematic transfer of digital documents to a final repository.

This issue was raised at the USAIN (United States Agricultural Information Network) Preservation Steering Committee meeting held last June at the Library of Congress. Committee members felt this topic was important enough to recommend a national meeting to discuss extending the national preservation plan for agriculture to include USDA digital information. The National Agricultural Library was the first organization to commit strong financial support for this meeting, and was later followed by the Economic Research Service, the Government Printing Office, and the Farm Foundation.

The Stakeholders

Any coordinated preservation effort will have a large number of interest groups, or stakeholders, who need representation. Stakeholders for USDA information sort themselves into three groups:

1. Those directly involved or affected by the preservation of USDA information or those who coordinate or support similar activities in other Federal agencies.

2. The dedicated information user involved in business, education or policy development.

3. The less demanding, yet very important user found in the general public.

We felt a planning committee for the conference would represent primary members of these three groups, and broader representation could be established by selectively inviting the other 60 attendees.

Putting together a planning committee to represent the broad interests of agriculture is at best a series of trade-offs. It must be large enough to represent and deliberate, but also to make decisions. It must have members with sufficient expertise to advise and contribute to problem solving. And it must have members who can perform a variety of roles during the planning stage and at the conference. In the end we constituted a planning committee with the following members:

- USDA NAL - Pamela André, who is the Director at NAL. NAL is one of four national libraries, and serves as the library for the Department of Agriculture.
- USDA ERS - Jim Horsfield, an administrator for the Economic Research Service who helped establish the USDA site at Mann Library.
- USDA CIO - Bob Whiting, sitting in for Anne Thomson Reed, the newly appointed Chief Information Officer.
- GPO - Duncan Aldrich, a depository librarian, with a one-year appointment to the Electronic Transition Staff at GPO.
- USAIN - Antoinette Powell, Director of the Agricultural library at the University of Kentucky, and President of USAIN, the United States Agricultural Information Network, an organization which represents the agricultural information interests of land grant universities and colleges, agribusinesses, and private organizations.
- Mann Library - Sam Demas, Head of Collection Development and Preservation. Sam has coordinated several cooperative preservation programs, including one which will have 16 participating institutions.
- Farm Foundation - Steve Halbrook, Assistant Director. The Farm Foundation is an unconventional organization with an interest in the social and economic implications of new technologies applied in agriculture and rural life.

Our committee selections were excellent. Each member contributed considerable time, energy, and many constructive suggestions.

The Resources

The major preconference Planning Committee meeting was in October of last year. This meeting had two purposes. One was to identify and discuss the pressing issues that would

be at the focus of a national meeting. The second was to establish the personal relationships that would help facilitate the Committee process through the many months of planning.

Helping us set an agenda for a preservation conference were two documents: one a report on digital preservation, the other an action plan for preserving print information.

The first item is the report *Preserving Digital Information*, prepared by a task force appointed by the Commission on Preservation and Access and the Research Libraries Group (RLG). This report, written by Co-Chairs Donald Waters, Associate University Librarian at Yale, and John Garrett, CyberVillages Corporation, with input from task force members, framed the key problems that need to be resolved to ensure continuing access to digital documents. The Planning Committee drew four major points from that report:

1. That digital migration is more than "refreshing" or rewriting digital information. Data migration is a specific set of tasks which must be followed when transferring information from one hardware/software configuration to another, or when moving from one generation of computing to another.
2. Digital archives are distinct from digital libraries in the sense that digital libraries are repositories that collect or provide access to digital information but may or may not provide the long term storage for that information.
3. Digital archives must be certified by an independent program.
4. The preservation responsibility initially rests with the creator or owner of the information. This last point is critical. A digital dissemination system is fluid and dynamic, with a single master generating multiple copies. Responsibility for the preservation of that master copy must reside with the creator of that document. This responsibility may be delegated to others, but when this is done, the delegation must be visible, clear and lawful.

The report also calls for discipline-specific preservation plans and pilot programs.

The second document is a preservation action plan adopted in 1993 by USAIN. This document, *A National Preservation Program for Agricultural Literature*, was prepared by a consultant, Nancy Gwinn, of the Smithsonian Institution, with assistance from the USAIN Advisory Panel on Preservation. The plan envisioned a systematic, organized, and coordinated national program combined with local initiatives. The plan also anticipated a need for sustained digital preservation efforts by recommending the implementation of scanning technologies to lay the groundwork for electronic access to preserved agricultural literature.

This plan has received national recognition as a sound model for a discipline-specific preservation program. This approach fits well with the objectives of the conference, since we were attempting to preserve not all government information, just the narrow slice from the USDA.

Key points of the plan include:

1. Preservation efforts must employ all relevant archival preservation standards and guidelines.
2. Preservation techniques used will not require uncommon or proprietary equipment.
3. All preserved materials will have bibliographic access.
4. All preserved materials are to be made easily available at low or no cost to the user.

These points either matched or fit closely those of the RLG report. The two documents provided a strong policy and organizational base detailed enough to facilitate action and flexible enough to accommodate unusual situations.

The national plan provides a framework for action by dividing the broad universe of agricultural literature into smaller, more manageable pieces. A key player is NAL, which has assumed responsibility for the preservation of Federal documents in print. It seemed worthwhile for the conference participants, including NAL, to consider if it were practical to extend the responsibility of preservation of digital information to one central organization, or if the responsibility needed to be distributed over numerous institutions, possibly making the process more manageable.

The Sharpshooter

Experience with the USAIN national preservation plan also led the Planning Committee to agree to hire a consultant to prepare a digital action plan. Beyond the fact that preparing the plan would be extremely complex, the credibility of the plan author would be a crucial factor when the document was distributed for peer review within the USDA, USAIN, professional societies and other bodies.

We were pleased to have Paul F. Uhlir agree to work with us. Paul, a lawyer and experienced staff member of the National Research Council of the National Academy of Sciences, recently coordinated a national panel of experts associated with the earth sciences data and edited the final report. Paul advised the Planning Committee on many topics as the planning process developed, and later served as Conference Facilitator.

Conclusion

As you can see, planning for the Conference required coordinating many issues. In the end we had a successful meeting and established the foundation for a digital preservation plan. The following two speakers, Evelyn Frangakis and Paul Uhlir, will describe these to you in more detail.

Resources

Preserving Digital Information: Report of the Task Force on Archiving of Digital Information

Commission on Preservation and Access, Research Library Group

Principal authors: Donald Waters, Yale University, John Garrett, CyberVillages Corp.

<http://lyra.rig.org/ArchTF/TFADI/index.htm>

National Preservation Plan for Agricultural Libraries

USAIN Advisory Panel on Preservation

Principal author: Nancy Gwinn, Smithsonian Institution

(Copies available from Gregory Lawrence)