

Proceedings of the 7th Annual Federal Depository Library Conference

April 20 - 23, 1998

Public Access to Government Information

Deborah Klein, Bureau of Labor Statistics
Washington, DC

Rather than talk about specific new products from the Bureau of Labor Statistics (BLS), I would like to use my time today to discuss my views on my role in the process of disseminating BLS information and how it interfaces with the role of the FDLP. Let me stress that I am not speaking for the administration, or even for the statistical agencies as a whole. I am planning to tell you how I view these issues, and how the situation looks from my perspective. I think the value of these meetings is the opportunity they provide to broaden our understanding. I am speaking as someone with the responsibility for the publishing and information dissemination at one statistical agency and as a person who firmly believes in the importance of disseminating Government information widely to users along with assistance in understanding and using the information that their Government collected and developed.

BLS is an agency with a long history of reporting on the studies it conducted; the first annual reports of the Commissioner of Labor were published in the 1880's.

In the course of carrying out its mission, BLS--like other Federal agencies--develops, collects, and distributes information. To whom does this information belong? In my view, the public paid for it with their tax dollars and the producing agency developed and understands its properties. Attempts to improve access to Government information and to ensure accountability need to deal with this reality as well as the current and future technological possibilities. This point is particularly important to me, and much that I have read on this subject seems to ignore the role of the producing organization.

Criteria for Information Access

Many groups have enumerated basic principles of information access. One set of principles was developed by the National Commission on Libraries and Information Science (NCLIS):

- The public has the right of access to Government information.
- The Government has an obligation to disseminate and provide broad public access to its information.

- The Government has an obligation to guarantee the authenticity and integrity of its information.
- The Government has an obligation to preserve its information.
- Government information created or compiled by Government employees or at Government expense should remain in the public domain.

[http://library.berkeley.edu/GODORT/gpo_prin.html]

Current Challenges

Among the current challenges to the existing system are:

- the vast increase in Government information;
- the emergence of new forms of information, particularly electronic;
- the impact of tightened budgetary constraints;
- the emergence of private sector information providers; and
- the changing views about the proper roles of the legislative and executive branches in this arena, as well as changing views on the public's right to Government information.

The emergence of low-cost printing technology coupled with the availability of information in electronic form has given Government agencies improved ability to reach the public without using either GPO or the FDLP. It has also expanded the explosive volume of Government information. At the same time the public has become both more demanding in its quest for Government information, and more sophisticated in its use.

Moreover, Government concern over cost control has shifted from paper production to applications of technology for electronic production and distribution. At the same time, I understand that participating FDLP libraries have serious and legitimate concerns with costs for cataloguing, provision of public access, and storage.

The 1995 Paperwork Reduction Act "reaffirms that dissemination applies to all publicly-available agency information, regardless of format." It also mandates that "agencies develop Government information locator services (GILS), and requires the integration of access and dissemination planning into life-cycle management."

The February 1996 revision to OMB Circular A-130, which implemented the 1995 PRA, requires that agencies should provide information about the nature of their information resources and how the public may gain access to them. Further, agencies are required to take advantage of as many dissemination channels as possible in recognition that information reaches the public through many secondary sources. The role of newspapers and other private publication vehicles is recognized, as well as information specialists, and database providers. This approach reflects positions detailed in the recommendations of the

National Performance Review, which supported executive branch responsibility for printing and the development of agency locators for public access to Federal information.

Statistical data

Electronic access is particularly important in the statistical area because users are becoming more sophisticated and want more than just published tables. They want the data in a form that they can put in spreadsheets and statistical formulas. This suggests that the solution is not to restrict the Federal agencies from making their data available electronically, but rather to improve the ability of the depository libraries to handle material in this format.

This will become even more important in the future because the statistical agencies are engaged in activities to improve the overall information infrastructure. Some of the directions in which we are heading include the development of better tools for information finding, extraction, and reuse. In these activities, we are partnering with the library and information service community that has long been a leader in this field.

Our goal is to increase statistical literacy and the ability of the public to transform information into knowledge. To accomplish this we are working to link metadata – by which we mean information about a survey and the information it provides – to the actual data from a survey. This would permit users to find definitions, reference periods, survey size, and other pertinent facts that can help them evaluate the relevance of the data to their objective. Again, these are all areas in which the library community should be a valuable resource.

Should Dissemination be Centralized?

The GPO-FDLP model envisions a system in which all Government information goes through GPO to Federal depository libraries. Viewed from the user perspective, those who want information go to their local public or university library, if it happens to be a depository library it would have the information needed, if not it would contact the regional depository library for the appropriate material.

A more decentralized approach, such as that underlying the procedures described in OMB Circular A-130, relies on information producing agencies to select appropriate means of access, balancing the usefulness of the information to the public against the cost of distribution. This approach recognizes that information reaches the public in many ways: a news release may only be sent to a mailing list of 1,000, but if the information is picked up on the network evening news, it becomes available to millions. This is the rationale behind locator systems, a recognition that all information does not have to go directly between the developer of the information and the end user.

To some extent the current FDLP centralizes access to Federal information in distributed locations. Centralized access is the cost determinant in today's information-demanding society, whereas centralized publication and distribution was the cost determinant a century ago. Accountability for access and preservation of information must also be assigned.

Locator Models

In theory, the major advantage of a centralized model of information dissemination is ease for the end user. The user could find all Government information at one location. A major advantage of a decentralized model is that information would be provided in the context of related information from providers who best understood the properties of the material.

In reality, neither of the extreme versions of the two models would be realistic. There is no way that every piece of Government information could be collected in one place, nor is it realistic to expect the public to search every possible location for a specific piece of information.

One strategy that, at least, some producers of Government information are pursuing is the development of locator systems. A prime example is in the area of Federal statistics. The United States has a decentralized statistical system in which statistical agencies are generally located in a parent agency responsible for the overall topic (e.g., agricultural statistics in the Department of Agriculture, education statistics in the Department of Education, labor statistics in the Department of Labor).

Because of this decentralization, there are more than 70 agencies that produce statistics that are of interest to the public. Moreover, many of the questions that the public wants answered require information from multiple sources.

In order to improve public access, the statistical agencies have developed a single Web site on the Internet (FedStats) that organizes and provides access to the entire range of Government statistics while leaving the individual agencies responsible for maintaining the data and ensuring their accuracy
[<http://www.fedstats.gov/>].

The user can obtain Government information without having to know which agency produces it by using a table of contents, a keyword search, a listing of subject matter contacts, or a directory of statistical programs. The Web site also includes links to statistical policy working papers, statistical news releases, and a clickable map to identify regional information.

This approach is consistent with the broader GILS (Government Information Locator Service), which has been defined as a decentralized collection of locators and associated information services used by the public either directly or indirectly or through intermediaries to find information. The developers of GILS envision the utilization of a standard syntax for records describing Government documents that can be organized for access.

The FedStats home page has proven to be a popular site, and it has gotten favorable write-ups in the press. (Paul Bugg of OMB will be discussing the FedStats site at the conference tomorrow.)

Because many users of statistics do know which agency has the data they want, more of them go directly to the agency sites than to FedStats. Thus, in February (a short month) there were 300,000 visits to the BLS Web site, and nearly 200,000 distinct hosts serviced. A

visit consists of a series of consecutive views of a Web site by the same user. You need to multiply that figure by about 20 to get the approximate number of hits (6.5 million for the BLS site alone in February)-- the statistic most often quoted to judge a site's popularity. People stayed at the BLS site for an average of more than 7 minutes, they looked at about 8 pages per visit, and they downloaded many many megabytes of data. My understanding is that the Census site is getting around three times that amount of traffic and other Government sites also are recording large and increasing numbers of visits.

I cite these numbers to show that electronic dissemination is an important and useful way for Federal agencies to distribute their information. It is also cost effective, because the marginal cost of each additional user is virtually zero. Without minimizing the importance of depository libraries, the most recent statistics I've seen show that there are less than a million customer assistance events per month on all subjects in all depository libraries combined.

While I am comfortable with the progress we are making on disseminating statistical data, I am less sanguine about the situation with respect to preservation and archiving. BLS has not yet completely replaced any print publications with electronic materials, and I hope we never will. But new information, information that would never before have made publicly available, is now being put up on the Internet. This is to the public good in the short run, but we have not fully resolved all of the issues involved in ensuring that this information is archived. Our goal should be to harness the technology that gave us the ability to disseminate data more broadly to insure that the functions of preserving and archiving information also are achieved.

Also, in the broad area of archiving is the question about how to preserve all the print material. BLS has paper documents that are more than 100 years old, and we need to preserve them for future generations. As resources permit, our plan is to scan them into both ASCII and PDF files, so that they can be searched and loaded into databases, while bibliographic control can be maintained and we will be able to preserve the authenticity of the specific data.

While there are many views on these issues, it isn't a question of the good guys versus the bad guys. I believe that all of the players wish to improve public access to information. There are serious issues about ensuring that all users have access to information, about ensuring that users can find out that the information exists so that they can access it, and about ensuring that there be permanent access to this information. The disagreements largely center on determining the best methods to achieve these goals, and open discussion is the best way to reach a consensus that will work. I look forward to continuing this dialog with you.