

Session 1

ELECTRONIC DISSEMINATION OF FEDERAL STATISTICS

Electronic Dissemination of Energy Data and Analysis

Mark Rodekoeh, and Henry Weigel

Energy Information Administration

Abstract

This paper provides a discussion of the Energy Information Administration's (EIA) Electronic Dissemination program. This program is designed to deliver EIA energy information, statistics and analysis in the most timely fashion possible using the latest technologies to provide economies to both the provider (the U.S. government) and users of EIA information products. The EIA is responsible for producing the nation's statistics and analysis on energy production, consumption, imports and prices. These statistics are often available by month, year, region or other disaggregation. Just one of our databases (the Oil and Gas Resource Information Database) contains 50 MB of energy data. The delivery of these data in an efficient and timely manner is critical to EIA.

The paper is organized around several sections which describe:

- Electronic Dissemination Goals and Strategies
- Dissemination Techniques
- EIA's Experience with Electronic Dissemination Methods
- Possibilities for the Future
- Electronic Dissemination Policy Issues

Strategy, techniques usage statistics and other policy related factors are discussed in some detail.

Goals and Strategy

The goal of electronic dissemination is to maximize the use and ease of use of EIA's energy data and analysis products. Minimizing the cost to both users and producers of these products goes a long way to achieving this objective. The rapid improvement in electronic publishing technologies is lowering costs to both users and producers but increases the costs and frustration of technology choice.

Electronic dissemination strategy involves matching appropriate technologies with users. For example some users do not have World Wide Web access but do have access to

fax machines. This implies that products should be made available in fax formats where appropriate. In addition there are classes of users who again do not have Internet access but who do have modems, this suggests that some type of bulletin board system may be a useful addition to the electronic dissemination technologies in use at EIA.

The rapid improvement in electronic dissemination techniques has clearly changed information dissemination techniques. The old paradigm (excluding data collection and processing) involved:

- Writing reports in long hand,
- Having a secretary type the report,
- Using a desktop publishing system to format and produce print quality masters,
- Print reports,
- Mail reports,
- Answer questions on the phone and by hand fax out selected tables.

The new paradigm involves:

- Analysts type and largely format reports,
- Publish electronically,
- Publish traditional paper products where needed,
- Customize information reports using a variety of technologies to meet user needs.

Clearly the time involved with the old and new paradigms in meeting customer needs is significantly different. From user surveys conducted at the Energy Information Administration **the number one user concern continues to be timeliness.** Electronic dissemination strategies can and are being used to address this concern. Simply put, customers want more timely data more than ever and are consistently demanding that these needs be met.

Dissemination Techniques

In this section the advantages, disadvantages and primary customer for each electronic dissemination technique is discussed in turn.

Internet Servers	
Advantages	Disadvantages
Distribution is instantaneous	Users must be fairly high tech.
10 - 30 million users	Moving large amounts of data is time consuming
Customer base growing rapidly	Can be slow during peak hours
Fairly low marginal cost	

This description of the Internet clearly shows some of the reasons for its rapidly increasing popularity, the point and click ease of use and almost instantaneous distribution characteristics make it one of the more popular technologies.

Bulletin Board Server	
Advantages	Disadvantages
One of the first electronic dissemination technologies	Still requires a modem
Has an existing customer base of 1,800 users who have been using this system for years	Slower than the Internet for many users
Does not require an Internet connection	Cannot use some of the more user friendly Internet applications
Can be replaced with a dialup ftp server to reduce costs	

This system has been in use for several years and many customers have constructed custom scripts to retrieve and format data to meet their special needs. It is likely that the use of this technology will decline over time and be replaced with dialup access to the Internet ftp server.

Applications Server	
Advantages	Disadvantages
Meets specific customer needs by providing only those data needed	Development of new applications can be very expensive
Does not require the downloading and installation of large software packages	Somewhat higher marginal cost than some of the other Internet technologies
User costs low	

The Applications Server (or the Interactive Query System) allows users to specify by clicking on a menu exactly what data elements they desire. Once the menu has been specified the request is transmitted to the applications server which then, using a query system constructs tables of the requested data and then returns the data to the user. This technology allows users to view only those data that they are interested in. Most other database technologies require users to download entire databases and then using either downloadable run time engines, or client based software construct queries to produce the needed data aggregations. Clearly the applications server technology is less demanding from a users point of view and therefore superior to providing specific data needs. The downside to use of this technology is the cost to the producer organization. It is somewhat more costly but widely used by some Federal data providers.

Listserv	
Advantages	Disadvantages
Very low marginal cost	Can only be used for short products
Ability to provide interested readers with short report very quickly	
Allows interested customers to subscribe	

The listserv technology has turned out to be a very popular electronic feature. At the Energy Information Administration's Web site users are offered a menu of listserv products, such as press releases, data summaries etc., to which they can subscribe. Once they have subscribed the products are sent automatically to their mailboxes. They can subscribe to as many or as few products as they wish. At present there are about 18 separate listserv products with a total distribution of about 3000 mailings per month.

Fax on Demand/Broadcast Fax	
Advantage	Disadvantages
Only requires that users have a fax machine	User pays phone charges
Menu systems allows users to get a list of all products	Only a limited amount of information can be transmitted in paper formats
Good for the occasional user	Transmission time is slow

This technology is especially suited to users with more limited access to technology. Since fax machines are very commonly used throughout the world they enable users to receive small amounts of information in a fairly timely fashion. It should be noted that most broadcast fax systems work best in the late evenings when phone lines are not usually tied up however this produces greater delays in receiving information.

CD-ROM	
Advantages	Disadvantages
Reaches a broad customer group since over 10 million cd readers are in use	Relies on the mail system
Libraries are very attracted to this medium since it saves shelf space	Will not reach low tech users
Low production cost	Users still must print to get hard copy
Holds up to 650 MB of information	
Easy to search	

The search ability and high volume make this medium ideal for a variety of users. Libraries find that this is ideal medium since rows and rows of shelf space can be saved. At the same time it is ideal for archival purposes since it does not degrade over time. The ease of conducting searches makes it ideal for users who are interested in very specific subjects. It is also the only medium that can deal well with large databases since space is not a limiting factor. As stated earlier one EIA database contains about 50 MB of data which is too much to download over most web and ftp sites.

The Electronic Dissemination Experience at the Energy Information Administration

The experience of using electronic dissemination techniques at the Energy Information Administration has been very instructive. While many things have gone as expected there have been several surprises. The table shown below presents some aggregate statistics on the use of electronic dissemination products on a monthly basis starting in January 1996. While some of the products have been available before this date this is when the EIA product line was largely established.

Electronic Product Statistics			
Month	Web ¹	Ftp ²	Listserv ³
Jan	NA	NA	NA
Feb	NA	NA	NA
March	12,723	81,026	NA
April	15,732	101,440	2,500
May	19,711	113,987	3,050
June	14,280	93,118	3,750
July	12,161	96,385	NA
August	12,498	100,898	NA
September	16,657	122,981	6,100
October	22,892	175,399	NA

These data show that starting in January traffic on the Web and Ftp sites typical grew at a rate of 20 to 25 percent per month except during the summer. Several factors caused the decline in June, these included a large increase in May figures due to the rapid increase in gasoline prices which by June had largely abated. This increase in interest increased the figures for May dramatically but was a one time event. Other significant factors included the end of the school year and the summer vacation season which tends to reduce commercial

¹Number of unique daily users per month.

²Number of file requests from Web, Ftp and EPUB per month.

³Number of individuals subscribing to E-mail products, not necessarily unique.

use of the Internet site. Starting in September the Internet traffic continued at a higher rate. In October it jumped to about 40 percent. It is interesting to note that the listserv traffic continues to increase by 15 to 25 percent per month over the period following the establishment of this product.

An examination of the detailed logs produced by the Internet server turns up a number of other interesting observations including:

- Over 60 percent of our users are from U.S. commercial firms and universities.
- About 15 percent of users are from other Federal agencies who tend to look at a larger number of files than commercial users.
- About 15 percent of users generate about 70 percent of the hits in other words the repeat customers are the heaviest users.
- About 10 percent of our users are from foreign sources.

Not shown in these statistics are the number of customers for the CD-ROM. It turns out that these users are much more limited in number than other users. By June 1996 about 1,000 CD's were being distributed every quarter. A large number of these users were represented by libraries who have expressed a great interest in this medium. Since library distributions can be looked at by a large number of users it is difficult to get an accurate measure of the number of total number of users of this technology.

The Future of Electronic Dissemination Techniques

Rapid changes in technology make the prediction of future dissemination techniques very difficult. However, there are some promising technologies that are being developed. One of the first that comes to mind is the use of a corporate database, structured using a database management system such as Oracle, which could permit a common user interface that would generate databases, tables and even permit the use of query systems that would operate across all data series maintained by a statistical agency. Another concerns client-server software that allows users to seamlessly get database updates and additions given the existence of a predefined database structure. The advantage of this approach would be that users would no longer have to download entire databases but only the database outline. Given this outline they could fill the data they need over the Internet, as desired. This could greatly reduce download time. In addition CD technologies already exist which allow users with Internet connections to get updates of CD databases over the Internet. Technologies such as Amber could allow users to download pdf files a page

at a time thereby making the pdf file a much more usable technology and at the same time greatly reduce the need for the construction of time consuming html file formats.

Policy Issues

There are many policy issues that remain to be resolved with regard to electronic dissemination. A few include:

- Choice of file formats. Private sector companies are very sensitive to the choice of file formats because these tend to be proprietary in nature and can have a impact on future sales. Agencies must be aware of these impacts on production and usage costs. For example the choice of one spreadsheet format over another can greatly affect the kind of software users are required to purchase. Another example relates to the choice of document formats such as WordPerfect, pdf etc. which are tied to specific company products. Since no generic format exists these policy decisions could be difficult.

- Client-Server software. Again, this can impact company sales and user requirements and become a sensitive issue.

- Browser support. Most web sites will support many browsers but are developed to be easiest to use with one specific browser. Companies are already complaining of the choices of Federal agencies in this area and could be more vocal in the future. From a resource point of view developing web sites that work well across a wide range of browsers can be very costly.

- Fees. Some Federal Agencies are already charging users for access to specialized statistics. The pressure to expand or start this practice in the future could become greater. Given the long history of providing low cost data to users this transition could pose some very difficult policy choices.

COMMENTS

Robyn Levine
Congressional Research Service
Library of Congress

Note: The views contained herein are those of the author and not necessarily those of Congressional Research Service or the Library of Congress.

Session 1: Electronic Dissemination of Federal Statistics-- Census and Energy Data

Note: My remarks were directed to two papers I had received beforehand, although the Census Bureau did not present the paper it provided me and gave a hands-on demonstration of DADS (Data Access and Dissemination System) instead. Thus some of my remarks were not necessarily addressed by the Census speakers/presenters.

Obviously, there are myriad benefits a reference librarian gains from the electronic dissemination of agency reports and data, particularly if the data tables can be user-defined. We can get the information when and where we want and be assured that what we see on the screen is the latest, most up-to-date version of the report or data, reflecting the most recent revision. Both the Energy Information Data System and the Census DADS system take advantage of these inherent benefits of the Web environment.

My particular areas of specialization are transportation and the census, although at CRS librarians are expected to be generalists when we work in our reference centers in the House and Senate office buildings and in our reading rooms. Each is a mini-library with a basic core of reference sources and finding aids. Given the relatively small size of the facilities and the ever-present budget crunch, the Internet environment expands our collections to reflect the expertise and imagination of the librarian. The librarian, thus, has tools heretofore provided only for the specialist.

Because we have various possible places that can supply statistical answers to questions, depending on the question (e.g., via our Inquiry Section, to a member of a particular team or to an analyst in our research divisions, or in a reference center or reading room) by having the data online we can ensure that the same question will be answered with the same data.

While CRS functions much like a special library, we are not in a position to consider payment mechanisms such as user fees. We certainly can appreciate the tough budgetary choices Congress has imposed on the Census Bureau, but Congress has placed similar restraints on us. Thus, we would hope that the policy of waiving fees to Congress will extend to us, as Congress is our sole client.

Another area of concern for government librarians in general, and us at CRS in particular, is the depository library network and its Census collections. The mandate CRS received from Congress is to support Congress not only in its legislative needs, but also in its representational duties. Some of our work is in response to constituent inquiries, supplying information for Members to pass on to their constituents in their districts. In this respect, we view our work as much educational as informational. We in the Congressional Reference Division (CRD) are 43 professional librarians/researchers out of 750 people in the entire Service. CRD answers about 65 per cent of the 340,000 congressional requests that come into the Service annually. We rely on the depository network as a referral; we expect it to be able to provide certain government and congressional publications.

Following the 1990 Census, depositories received CD-ROMs whether or not they had the necessary equipment to make them available to the public. There were still paper products available. The Web environment opens up exciting possibilities for obtaining Census information, but it also opens a Pandora's box as to how depositories will handle the potential costs for any special products created by using the DADS system. The Library of Congress is wrestling with a similar question in enabling patrons to use fee-based electronic systems in the reading rooms. If this is a problem here, it is very likely to be a problem elsewhere. For some depository libraries, their ability to provide service could be limited by the electronic dissemination of publications by government agencies. Thus far, EIA is attempting to offer paper products as well as the electronic ones, including Faxed documents, and I urge them to remember the depository network when making future decisions concerning data dissemination.

I can see very positive benefits from the DADS system. After the 1990 CD-ROMs became available, it was obvious that the GO software that Census provided had some significant drawbacks, namely in the ability to manipulate the data. We spent considerable in-house resources to convert the CD-ROMs, with SAS-Assist, into a system where we would be able to rank and sort the data to our clients' specifications. Our system is very time-consuming to use and not very user-friendly. Thus, how easy DADS will be in producing custom-tailored tables is of great interest to us. The more user-friendly the better, since we have had numerous occasions to assist Congressional staff directly in their own use of the Internet.

That CRS has a vested interest in the Internet as a warehouse of government information can be seen in our Home Page, which is available to our own staff and to all Congressional offices. We make available our own databases and products, of course, and we enable our patrons to reach agency home pages with direct links. Staffers approach CRS staff in our reference centers for help in using the variety of Internet sites. In our prepared information, we include the URLs to government agencies and associations when relevant, and if we include any products obtained from the Internet, we provide the URL for further information.

In terms of government information, CRS may potentially be the primary intermediary outside the agency congressional liaison offices to assist

congressional users of government statistical databases. In terms of Census data, Congress's need for small-area data is going to be more demanding than in the past. Hopefully, DADS will be able to integrate various data sources by comparable geographic levels, especially smaller geographic entities like congressional districts.

With a devolution of authority for many government programs, federal involvement in data production may be more significant in statistical areas than ever before. Now that Congress has created devolution, it will have an oversight role with the shift in program responsibility to the states. We cannot rely solely on the states and local areas to produce data, as they are hard-pressed financially, and there is no overall standard for comparable data. Thus, for oversight, it will be imperative that Congress has comparable data to assess the impact of programs across state and local areas.

CRS has begun to examine resources and consider options for the information that will be required due to the impact of devolution. We have established a Federal-State Task Force since we believe devolution will be of significant congressional interest in the 105th Congress. We hope that other agencies are looking into how to present their data for very small geographic entities and that there will be interagency cooperation to enable this to happen in as consistent and unified a manner as possible. This likely will be a challenging and difficult situation for all of us, and the cost of our information gathering and production potentially will increase as Congress assumes this oversight role. In the end, the use of certain data may be very specialized and for Congress' eyes only (given inherent confidentiality problems).

While I used the EIA system in anticipation of this meeting, my familiarity with energy-related questions is not as keen as with the census. Thus I tested the EIA system against some typical questions one of my colleagues receives in terms of energy production, consumption, imports and prices, and found it to be easy to use. At the same time it allowed us to customize certain reports. Most of the data, however, is national in scope, with limited state information.

The rest of these comments will be addressed to questions for the Census Bureau that are active concerns of data users at CRS, although EIA may want to consider them also.

1. As a historical tool, how long will data be available electronically? Will this data be archived?
2. How accessible will raw data be on DADS or from EIA, and will it be available across state lines or just on a state-by-state basis?
3. How does the Census Bureau decide which of the printed reports are being discontinued or changed, and are users involved in the decision-making process?
4. Concerning congressional districts, as they are redrawn as a result of court decisions, will new data be issued for the reconfigured districts?
5. Congress wants more and more detailed data for smaller areas such as congressional districts, something the data user community at large is not

particularly interested in. Will the Bureau, and other agencies as well, be prepared to face the potential burden of this need?

6. One suggestion made in the paper I read concerned partnerships with non-governmental parties to conduct joint research with the Census Bureau. This has advantages in being able to trust data supplied by non-governmental sources if in sync with the Census Bureau (e.g., private marketing demographic systems and the formulas employed to update decennial census data). Where this partnership will be used to find ways to combine various federal agency data sets has far-reaching applications, especially if small-area geography will be available across the board.

But if the information will not be made available for up to 5 years due to the contractual nature of the joint partnership, then the benefits discussed in the outset of these comments, namely the supply of current and timely data, has been negated.

Thank you for allowing me the opportunity to be here this morning.