

GAO

Fact Sheet for the Chairman, Committee
on Governmental Affairs, U.S. Senate

AD-A256 736



July 1992

INFORMATION DISSEMINATION

Case Studies on Electronic Dissemination at Four Agencies



DISTRIBUTION STATEMENT
Approved for public release;
Distribution Unlimited.

92-28793



92 11 03 016

United States
General Accounting Office
Washington, D.C. 20548

DTIC QUALITY INSPECTED 4

Information Management and
Technology Division

B-248094

July 22, 1992

The Honorable John Glenn
Chairman, Committee on
Governmental Affairs
United States Senate

Dear Mr. Chairman:

Accession For	
NTIS	CRA&I ☒
DTIC	TAB ☐
Unannounced ☐	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Availability or Special
A-1	

This report responds to your request that we examine electronic information dissemination processes at four government agencies. Our review focused on the Department of Agriculture's Computerized Information Delivery Service, the National Technical Information Service's Bibliographic Database, the Bureau of the Census' general information dissemination programs, and the National Library of Medicine's (NLM) Medical Literature Analysis and Retrieval System (MEDLARS). Each of the agencies' dissemination practices is highlighted in appendixes I through IV, respectively. Appendix V provides additional detail on the individual databases comprising the MEDLARS system.

For each agency, we describe users and uses of the data, agency planning activities, data integrity practices and controls, agency costs for information dissemination, and prices paid by information users. However, we did not attempt to evaluate the effectiveness or efficiency of agency dissemination practices or whether agency dissemination activities were performed in accordance with existing laws and regulations.

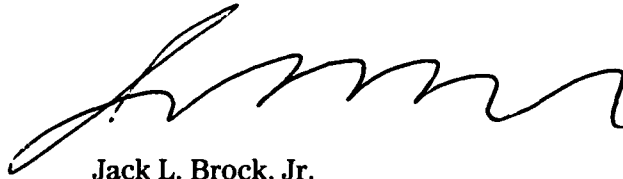
As arranged with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution until 30 days from the date of this letter. At that time, we will send copies to the Directors of NLM, NTIS, and Bureau of the Census; the Secretaries of the Departments of Health and Human Services, Agriculture, and Commerce; and the Chairmen, House Committee on Government Operations and Senate and House Committees on Appropriations. We also will send copies to other interested parties and will make copies available upon request.

DISSEMINATION STATEMENT
Approved for public release
Distribution Unlimited

1992

Should you have any questions about this report, please contact me at (202) 512-6406. Details of our objectives, scope, and methodology are in appendix VI and major contributors to this report are listed in appendix VII.

Sincerely yours,

A handwritten signature in black ink, appearing to read "J. Brock, Jr.", with a stylized, cursive flourish extending to the right.

Jack L. Brock, Jr.
Director, Government Information
and Financial Management

Contents

Letter		1
--------	--	---

Appendix I		8
U.S. Department of	Users of Computerized Information Delivery Service	8
Agriculture's	How CIDS Information Is Used	9
Computerized	Information Dissemination Planning Activities	10
Information Delivery	Maintaining Data Integrity	10
Service	Prices Charged to CIDS Users	11

Appendix II		13
The National Technical	Users of the Bibliographic Database	14
Information Service's	How the NTIS Bibliographic Database Is Used	14
Bibliographic Database	Information Dissemination Planning Activities	14
	Controls to Maintain Data Integrity	15
	Cost and Price Recovery Policy	16

Appendix III		19
Bureau of the Census'	Users of Electronic Information Dissemination	20
Information	How the Census Data Are Used	21
Dissemination Practices	Planning for Electronic Information Dissemination	21
	Controls to Maintain Data Integrity	22
	Costs for Electronic Information Dissemination	23
	Prices Charged for Electronic Dissemination	24

Appendix IV		26
The National Library of	MEDLARS Users	27
Medicine's Information	How MEDLARS Is Used	28
Dissemination Practices	Information Dissemination Planning	29
	Controls to Maintain Data Integrity	30
	Cost Recovery Process	32
	Charges for On-line Services	34
	Charges to Licensees	35
	Billing and Collection Services	37

Contents

Appendix V		39
Description of	Information Available on the NLM Computer	39
MEDLARS Databases	Information Available on TOXNET	45

Appendix VI		46
Objectives, Scope, and		
Methodology		

Appendix VII		47
Major Contributors to		
This Report		

Tables	Table I.1: CIDS Charges by Function	12
	Table II.1: Fee Structure for Lessees	17
	Table II.2: Costs and Revenue for NTIS' Bibliographic Database	18
	Table III.1: Electronic Data Dissemination Revenue	25
	Table IV.1: MEDLARS Database Costs and Billings, Fiscal Year 1989	37
	Table IV.2: MEDLARS Collections, NTIS Fees, and Excess Returned to the Treasury	38

Figures	Figure I.1: CIDS User Composition As of March 1992	9
	Figure III.1: Users of Census Products From November 1990 Through October 1991	20
	Figure IV.1: Composition of MEDLARS Domestic On-line Users As of April 1991	27
	Figure IV.2: Fiscal Year 1989 MEDLARS Costs	33

Abbreviations

AIDS	acquired immunodeficiency syndrome
AIDSLINE	AIDS Information On-line
AIDSTRIALS	AIDS Clinical Trials
AVLINE	Audio Visuals On-line
BIOETHICSLINE	Bioethics On-line

Contents

CANCERLIT	Cancer Literature
CATLINE	Catalog On-line
CCRIS	Chemical Carcinogenesis Research Information System
CHEMLINE	Chemical Dictionary On-line
ChemID	Chemical Identification
CLINPROT	Clinical Cancer Protocols
CD-ROM	compact disc read only memory
CIDS	Computer Information Delivery Service
COTR	contracting officer's technical representative
DART	Developmental and Reproductive Toxicology
DBIR	Directory of Biotechnology Information Resources
DIRLINE	Directory of Information Resources On-line
DOCUSER	Document Delivery User
EMICBACK	Environmental Mutagen Information Center Backfile
ETICBACK	Environmental Teratology Information Center Backfile
GAO	General Accounting Office
HEALTH	Health Planning and Administration
HISTLINE	History of Medicine On-line
HSDB	Hazardous Substances Data Bank
IMTEC	Information Management and Technology Division
IRIS	Integrated Risk Information System
MEDLARS	Medical Literature Analysis and Retrieval System
MeSH	Medical Subject Headings
MEDLINE	MEDLARS On-line
NLM	National Library of Medicine
NTIS	National Technical Information Service
OMB	Office of Management and Budget
PDQ	Physician Data Query
POPLINE	Population Information On-line
RTECS	Registry of Toxic Effects of Chemical Substances
SDILINE	Selective Dissemination of Information On-line
SERLINE	Serials On-line
TRI	Toxic Chemical Release Inventory
TOXNET	Toxicology Data Network
TOXLINE	Toxicology Information On-line
TOXLIT	Toxicology Literature from special sources
USDA	U.S. Department of Agriculture

U.S. Department of Agriculture's Computerized Information Delivery Service

The U.S. Department of Agriculture's (USDA) Computerized Information Delivery Service (CIDS) is a system to electronically disseminate a wide range of perishable and time-sensitive agricultural data to news and information services for distribution to their subscribers or to others interested in large amounts of USDA information. Perishable and time-sensitive data are defined as data with a limited useful life; data that lose their significance if they do not reach the proper users in a timely manner; and data that, when replaced, are completely replaced. Some of the information provided by CIDS includes crop, livestock, and price statistics; export sales reports; and foreign agriculture trade leads.

Although CIDS is a departmental approach to electronic data dissemination, each agency within USDA is responsible for loading, deleting, and specifying data release times and the length of time data will remain in the system. Since CIDS was created to augment rather than replace the published reports of USDA agencies, the agencies still make their data available in printed form as well as through electronic media such as magnetic tapes and floppy diskettes. As of March 1992, there were 16 USDA agencies participating in this service.

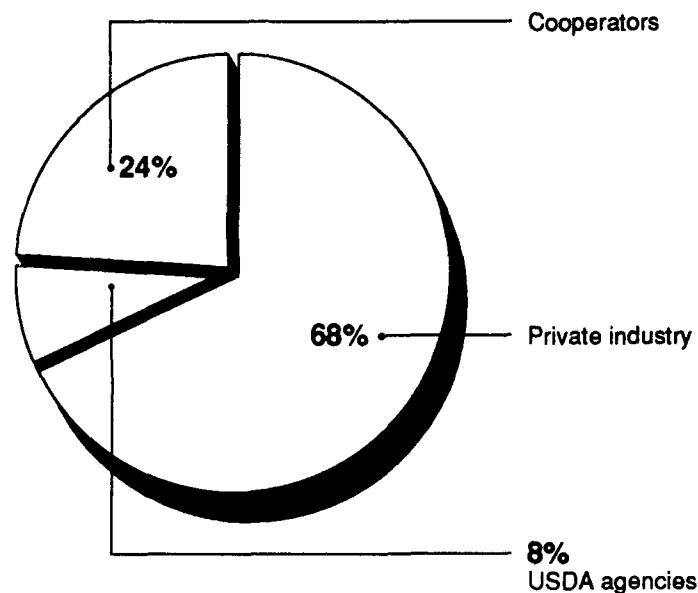
Martin Marietta Data Systems, under contract to USDA, designed the system and has operated it since 1985. According to USDA, the use of a contractor allows information to be available from a single source, avoids security risks associated with allowing the public to access the Department's computer, and provides more responsive service to the public. CIDS data are available to users from a central database on mainframe computers at the Martin Marietta data center in Orlando, Florida. Access to the data is provided by Martin Marietta's telecommunications network, or by common carriers such as Telenet. Network access is available 24 hours a day, 7 days a week from anywhere in the U.S., Canada, the Virgin Islands, and any other location with access to the telecommunications network.

Users of Computerized Information Delivery Service

USDA estimates that there were over 200 on-line subscribers, many with multiple users, as of March 1992. These users fall into three categories: private industry (publishers, news media, agribusiness, information services, and commodity and market investment brokers); USDA agencies; and other government agencies that USDA refers to as cooperators. Cooperators currently include non-USDA federal agencies such as Bureau of Labor Statistics, Tennessee Valley Authority, and Central Intelligence Agency; state departments of agriculture; state extension directors; state universities; and libraries in land-grant colleges and in the Library

Depository Program. Figure I.1 shows the composition of users as of March 1992.

Figure I.1: CIDS User Composition as of March 1992



How CIDS Information Is Used

According to USDA's CIDS coordinator, federal government users access CIDS to improve their efficiency and effectiveness in managing their programs, while nongovernment users use the information to enhance their customer services or make farm management decisions. For example, USDA's National Agricultural Statistics Service uses CIDS data to prepare their publications in a more timely and accurate manner. The Service's CIDS coordinator told us that prior to using CIDS, field staff had to wait 2 weeks or more for the information to be prepared and delivered by mail.

Nongovernment users use the information to enhance customer services in the sale of their products and services. For example, one nongovernment user we surveyed used information such as market place analysis, progress reports on acreage, and agriculture industry updates to estimate market share.

Information Dissemination Planning Activities

USDA relies primarily on the CIDS Policy Board and user input to plan CIDS activities. In 1984, the Department established the Board consisting of senior USDA officials. The Board is responsible for policy, overall direction, and management of CIDS.

The CIDS users group, consisting of data administrators and nongovernment users, meets monthly to discuss user service problems and provide feedback to the contractor or USDA's technical representative. The group also receives from the contractor a monthly report on the types and frequency of agriculture data that have been accessed. The report helps agencies decide on information changes to the database. For example, one USDA agency eliminated 13 reports from the system because the report indicated that no one had accessed those reports for several weeks. The user group's monthly meetings also provide an opportunity for nongovernment users to request changes or additions to the system to better meet their needs.

Maintaining Data Integrity

USDA employs various techniques to maintain the integrity of the information in the system. Data integrity controls include electronically reading the information that was loaded into CIDS and using high-speed error-correcting modems and dedicated telecommunications lines. USDA also manually verifies the accuracy of the data loaded into the system. For example, USDA's World Agricultural Outlook Board has USDA representatives review printouts of data to ensure that the data are correct before entering the data into CIDS. In addition, the contractor provides routine and real-time backup of all necessary files to ensure there is no loss of data in the event of hardware or software failures.

As stipulated in the CIDS contract, the system is accessible only to those who have been authorized to load, administer, and retrieve data, and they each are assigned specific identification numbers and passwords. User identification numbers and passwords are maintained by the contractor, and the users have the flexibility to modify their passwords at any time. The system also allows for password changes at fixed intervals and masks users' passwords on the users' terminals.

The Department monitors and enforces the terms and conditions of the contract primarily through a contracting officer's technical representative (COTR), who is responsible for ensuring that the contractor meets the technical requirements of the contract. For example, the contract provides for CIDS to be available at least 98 percent each operational day and 99

percent for each operational month. The contractor is required to notify the COTR as soon as the system is not operational and again when the system is available. The COTR independently monitors the availability of the system by accessing the system every morning and extracting an agriculture news report developed by the Office of Public Affairs. The COTR puts the report on the Department's internal electronic bulletin board, which can then be accessed by agriculture officials overseas, in the field, and at headquarters. According to the COTR, the Department is satisfied with the system's performance and there have been no major technical problems with the service.

Prices Charged to CIDS Users

Martin Marietta's charges for CIDS vary depending on the type of user. The contract between USDA and Martin Marietta specifies that federal agencies (including USDA), and cooperators are charged the same unit price for services; all other users pay a rate established by the contractor based on standard commercial rates. The prices for federal agencies and cooperators are subject to volume discounts, and cover services such as computer utilization, hourly connect time, file loading and storing, peripheral device utilization, software surcharges, and microfiche. In addition, the contract provides for a rebate to USDA based on overall CIDS usage. The unit prices, discount rates, and rebate provisions are proprietary and thus are not included in this report. Table I.1 shows USDA and cooperator charges for fiscal years 1988 through 1991. For example, in 1991 the total charges to USDA and cooperators were \$131,225. After subtracting the rebate and collections from cooperators, the net charges to USDA were \$80,335—\$27,540 for loading and storing and \$52,795 for accessing the system.

Appendix I
U.S. Department of Agriculture's
Computerized Information Delivery Service

Table I.1: CIDS Charges by Function

	Fiscal Year ^a			
	1988	1989	1990	1991
Usage Charges:				
Total	\$121,892	\$163,409	\$113,988	\$131,225
Rebate	10,470	12,466	17,212	28,667
Collections from Cooperators	6,509	5,935	8,228	22,223
Net Charge to USDA	104,913	145,008	88,548	80,335
Charges by Function:				
Loading and Storing	66,339	83,826	35,315	27,540
User Access	38,574	61,182	53,233	52,795
Net Total	104,913	145,008	88,548	80,335

^aMonetary amounts are in current dollars.

The pricing structure for information services provided to private industry is established by the contractor and is based on the contractor's standard commercial rates for the use of timesharing services. The charges, which are set forth in separate contracts with each user, include a \$75 minimum monthly usage fee plus other fees based on type of service. For example, users are charged \$.95 per 100 lines of text retrieved and \$.04 per menu entry for each selection made from a menu screen. In addition, rates per connect hour vary from \$12 to \$18 depending on the speed of data transmission and \$500 for a high-speed, dedicated line.

The National Technical Information Service's Bibliographic Database

The National Technical Information Service (NTIS), an agency of the Department of Commerce, sells scientific and technical reports, computer software, and data files; and leases intellectual properties generated by and for the government. Traditionally, federal agencies voluntarily submitted their scientific, technical, and engineering information for dissemination through NTIS. However, the American Technology Preeminence Act of 1992 now requires federal agencies to transfer this information to NTIS to be disseminated to the private sector, academia, state and local governments, and federal agencies. By law, NTIS is authorized to be self-supporting; and thus, its costs, including salaries, marketing, printing, product distribution, and all associated operating costs, are recovered primarily through sales income. NTIS acquires reports, data files, bibliographic products, and software from source agencies and sells this material in various media, including paper, microfiche, audiovisual tape, floppy disk, and compact disc read only memory (CD-ROM). NTIS also provides bibliographic control, document distribution, patent licensing, and accounts receivable management services to other agencies.

NTIS' largest and most widely used database, the Bibliographic Database, contains about 1.6 million citations covering documents issued since 1964. The database includes research summaries of U.S. and foreign government-sponsored research, applied engineering studies, software, conference presentations and proceedings, and technical manuals. The database is updated twice a month with about 70,000 citations being added annually. This database is available on-line through vendors that lease it from NTIS and from libraries that subscribe to it through vendors. Lessees may also make the database available on CD-ROM by entering into a separate lease agreement with NTIS. The CD-ROM version of the database, which is available from two vendors, covers the previous 12 years and is updated quarterly.

NTIS maintains a help desk to answer questions on the database and provides a free search guide to assist users in searching the database using NTIS subject categories. NTIS also offers on-line training and group presentations on the Bibliographic Database. Once users identify documents through the database, they can contact NTIS or the source agency to order copies. Complete reports are available from NTIS for about 88 percent of the citations in the database. The remaining documents are available from the originating agency, journal publisher, or conference publisher. Users can place orders by telephone, mail, facsimile machine, or through on-line ordering services made available by NTIS and the database vendors.

Users of the Bibliographic Database

The database is leased to about 24 customers including DIALOG Information Services; Georgia Institute of Technology; Canada Institute of Scientific and Technical Information; DATA-STAR; European Space Agency; STN International; Orbit Search Service and BRS Information Technologies, divisions of Maxwell Online; and SilverPlatter Information, Incorporated. These lessees make the database available internationally either by on-line access or on CD-ROM. Users, such as libraries or the public, pay for on-line use and receive a password to the on-line database or receive an annual subscription of the database on CD-ROM.

How the NTIS Bibliographic Database Is Used

According to an NTIS product management official, private businesses lease the database to further their own research and product development efforts. Vendors make the database accessible to their customers who may use it to identify past research to avoid duplication of effort. Academic institutions make the database available to faculty, students, and staff for educational and research purposes.

Information Dissemination Planning Activities

NTIS' planning activities for the electronic dissemination of information cover all of NTIS' products, including the Bibliographic Database. Planning activities include preparing product overview documents; preparing business, strategic, and information technology plans; and conducting user conferences.

Each year, NTIS product managers prepare product overview documents that examine product sales, product markets, product revenue contributions, and long-term product outlook. In addition, product overview documents include recommendations to ensure each product line contributes, to the extent feasible, to the agency's total cost recovery. For example, the NTIS Bibliographic Database product overview contained information on multiyear sales trends, market opportunities, database usage, vendor prices, industry trends, planned promotional activities, cost and revenue projections, and recommended actions.

NTIS also prepares an annual business plan that outlines areas for management consideration such as problem identification, strengths and weaknesses in problem resolution, actions to be taken in solving problems, and actions needed to take advantage of technological advances to improve service. In addition, NTIS annually updates a strategic plan that describes NTIS' vision for improving and modernizing information products, services, and facilities. In addition, NTIS annually prepares a 5-year information

technology plan that describes how NTIS will use information resources to support operations, including information dissemination.

Other planning mechanisms include conducting annual user conferences attended by users from government and the private sector. In addition, NTIS holds formal meetings with its information providers, such as the National Aeronautics and Space Administration and the Department of Energy, to identify information those agencies will provide NTIS to make available for dissemination.

Controls to Maintain Data Integrity

NTIS employs various techniques to maintain the integrity of the citations in the database including key verification, machine edits that check consistency, and alphanumeric field verification. In addition, NTIS uses edits that check source codes in the citations against a pre-existing authority file that includes all approved source codes. NTIS annually prepares a correction tape that contains corrected citations based on feedback from the originating agencies and sends it to the lessees so they can update the NTIS Bibliographic Database.

NTIS' Bibliographic Database lease agreement describes the terms, conditions, and restrictions for using the database. Generally, the lease agreement prohibits copying or reproducing any portion of the database in machine-readable form without prior written permission. Copies of citations may be made but they are subject to usage fees. The lease agreement also warns users that NTIS does not guarantee that the contents of the database are accurate or complete or that the results obtained will be error free.

The lease agreement allows lessees to make computerized access to the database available to others by requiring the users to enter into a subscription agreement. Like the lease agreement, the subscription agreement describes the conditions and restrictions for using the database. The subscription agreement also notifies users that neither the lessee nor NTIS guarantees that the contents of the database are accurate or complete, or that the results obtained will be error free. Although most of the information in the database is in the public domain, lessees warn users every time they access the database that it contains copyrighted information. Each copyrighted citation contains an identifier that indicates that it is copyrighted.

NTIS does not formally monitor the provisions of the lease agreements between lessees and users, or the integrity of the NTIS Bibliographic Database once it is loaded on the lessees' systems. However, NTIS requires lessees to report monthly on database usage including connect hours used, number of records printed or displayed, number of search terms entered, and number of current awareness profiles maintained. CD-ROM vendors report the total number of subscriptions. This information is used to collect the appropriate use fees from the lessees. NTIS does not independently verify lessees' use statistics.

Cost and Price Recovery Policy

Title 15 Section 1153 of the U.S. Code authorizes the Director of NTIS to recover costs from fees paid by users of information products and services. NTIS has established a pricing committee whose members have 2-year terms and include NTIS finance, marketing, and production managers and others appointed by the Director. The Committee reviews cost and revenue plans and evaluates alternatives for reaching a financial break-even point considering product characteristics, marketing conditions, and industry trends. Annually, the Committee recommends a new pricing schedule for products and services for the next calendar year, subject to the Director's approval.

NTIS' goal is to establish prices that will recover the variable and fixed costs of producing its products and services. The prices charged for a product, such as the Bibliographic Database, may not recover all costs because NTIS considers other factors in setting prices. For example, NTIS considers what the vendor is willing to pay for the database, how private sector databases are priced, whether the database will be used commercially or in-house, and how sales will be affected at various price levels; as well as the value of the database as an announcement that generates sales of information items in the NTIS collection. In general, NTIS attempts to recover shortfalls from the sale of one product through revenue generated from the sale of other products.

NTIS' fee structure includes a yearly flat fee of \$9,000 for the database tape, plus fees based on usage. To determine usage, lessees are required by the lease agreement to maintain and report to NTIS database use statistics including connect hours, and number of search terms entered, records displayed, and records printed. The lessees are required to pay NTIS fees for this usage and for using the database on CD-ROM. Table II.1 shows NTIS' fees for obtaining and using the database.

Appendix II
The National Technical Information Service's
Bibliographic Database

Table II.1: Fee Structure for Lessees

Category	Fee^a
Tape	
Data from 1992	\$9,000 per year
Data from 1991	\$8,000 per year
Data from 1990	\$7,000 per year
Backfile of the database (1964-1989)	\$1,150 per year
On-line use	
Connect time	\$30.00 per hour
On-line display of data	
Full record	\$0.28 per record
Partial record	\$0.25 per record
Off-line print of data	
Full record	\$0.28 per record
Partial record	\$0.25 per record
Search term per term	\$0.03 per term
Profile search charges	
First 500 profiles	\$1.50 per profile
Next 999 profiles	\$1.25 per profile
For each additional profile	\$1.00 per profile
"Hit" Charges ^b	
Partial record	\$0.28 per record
Full record	\$0.25 per record
Off-line use	
Retrospective search charges	
Current year data	\$1.50 per issue
Prior years data	\$3.00 per year
"Hit" Charges	
Partial record	\$0.28 per record
Full record	\$0.25 per record
CD-ROM	
CD-ROM leased or sold by vendor	
Current year data	\$500 per year
Data from 1991	\$300 per year
Data from 1990	\$250 per year
Data from 1985-1989	\$50.00 per year

^aForeign lessees pay somewhat higher fees in order to help U.S. businesses be more competitive with their foreign counterparts.

^bFor each bibliographic record displayed, typed, printed or otherwise derived or extracted from each execution of a search profile, whether on-line or off-line, the customer pays for each partial or full record.

NTIS breaks down the cost of the Bibliographic Database into two types, variable and fixed. Variable costs include direct costs and a portion of processing costs, whereas fixed costs include overhead and a portion of processing costs. Processing costs include staff time spent for subscription order processing, computer operations, scheduling and development, promotions, and registration services. Table II.2 shows the total costs and revenues for fiscal years 1990 and 1991 for the NTIS Bibliographic Database.

Table II.2: Costs and Revenue For NTIS' Bibliographic Database

	Fiscal Years	
	1990	1991
Costs^a		
Variable:		
Direct	\$139,570	\$98,899
Processing	8,353	14,900
Fixed:		
Overhead	1,745,447	1,738,034
Processing	1,312,764	1,454,862
Total Costs^b	\$3,206,134	\$3,306,695
Total Revenues	\$2,154,317	\$1,821,954

^aMonetary amounts are in current dollars.

^bNTIS attempts to recover shortfalls from the sale of one product through revenue generated from the sale of other products.

To determine the adequacy, reliability, and accuracy of NTIS' financial statements and to comply with the National Technical Information Act of 1988 (Public Law 100-519), an annual independent audit of NTIS' financial statements is performed. In 1990, the Department of Commerce's Office of Inspector General issued the results of the first financial audit in NTIS' 45-year history.¹ The Inspector General found that NTIS' financial operations were not profitable with most product lines and programs losing money; the accounting system was seriously deficient, producing inaccurate and unreliable information; and internal controls were virtually nonexistent. NTIS' management has agreed with both the findings and recommendations and has agreed to take action to correct the inadequacies.

¹Review of National Technical Information Services' Financial Statement and Accounting System, Report Number ATD-0024-0-0004, September 1990.

Bureau of the Census' Information Dissemination Practices

The Bureau of the Census collects, compiles, and publishes general-purpose statistics. It is responsible for taking all federal censuses, including those of population, housing, agriculture, retail and wholesale trade, service industries, manufacturers, mineral industries, transportation, construction industries, and governments. Census conducts national economic, agriculture, and government censuses every 5 years; and population and housing censuses every 10 years. In addition, Census administers about 250 sample surveys each year to obtain current information on social and economic conditions, which it uses to prepare estimates and projections. Census publishes its results in more than 2,000 reports in a non-census year and about 5,000 reports in a census year for the general public, the Congress, states, and local governments. Census also conducts international statistical studies of foreign social and economic systems, maintains socioeconomic databases for all nations, and provides users with analyses of statistical data.

Census disseminates its statistical data in various electronic formats, including tapes, CD-ROM, floppy disks, and microfiche. These may be purchased directly from Census' Data User Services Division. The data are also available via CENDATA, a database that is made available by Census through two vendors. Users access CENDATA by subscribing to either of the two vendors' systems. CENDATA includes 1990 census data; sub-state data, such as per capita income estimates for states, counties, cities, and towns; economic indicators; population projections; and demographic profiles on 208 foreign countries.

Census' Data User Services Division markets Census' electronic media. Organizationally, this division is divided into several components including data access and use, systems and programming, customer services, user training, state and regional programs, and statistical abstracts. The Data User Services Division produces machine readable data products, assists data users in accessing and using Census products and services, sells computer products, and develops training guidelines and standards. In addition, it conducts briefings, workshops, and presentations to assist organizations in obtaining and using Census data.

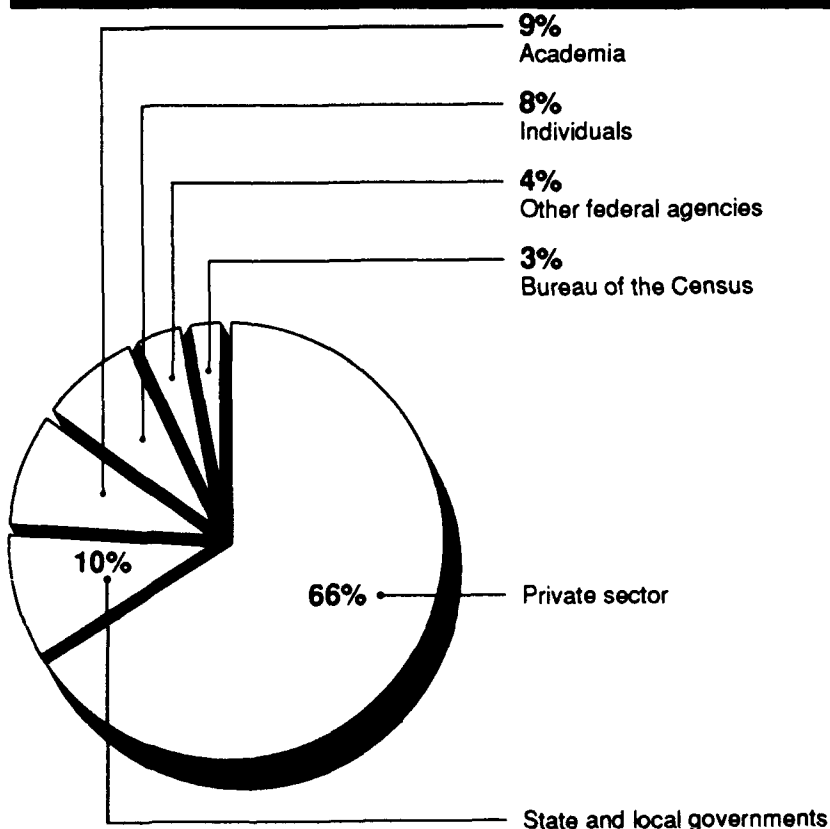
Census also has programs that allow users to buy or access Census products at various locations throughout the country. The State Data Center and Business and Industry Data Center Programs enable Census to furnish data products, training, and technical assistance in 50 states, the District of Columbia, Puerto Rico, Guam, and the Virgin Islands. The Business and Industry Data Center Program was created to allow

participants to receive economic data that will assist businesses and to further state economic development. Over 1,500 State Data Center and Business and Industry Data Center organizations and affiliates offer products and assistance to community leaders, planners, business people, and other users. Generally, both programs charge users for information disseminated; however, some centers disseminate Census data at no cost to promote economic development.

Users of Electronic Information Dissemination

Users obtaining electronic information products directly from Census can be grouped into six categories: private sector, state and local governments, academia, individuals, federal agencies, and the Bureau of the Census. Figure III.1 shows the users of electronic products from November 1990 through October 1991.

Figure III.1: Users of Census Products
from November 1990 Through October
1991^a



^aThis figure does not include CENDATA users because the vendors do not identify their users by type

How the Census Data Are Used

Census data are used in a variety of ways. For example:

- The private sector uses census data for a number of business reasons such as forecasting general economic conditions; measuring potential markets in terms of size, geographic areas, and kinds of businesses; calculating market share; determining markets; deciding on new plant locations; laying out sales territories; analyzing sales performance; and allocating funds for expenses such as advertising. They also sell Census data to their customers.
- Many federal, state, and local government plans, individual grants, grants-in-aid, and revenue sharing programs have been based on factors calculated from census statistics. State governments also use the decennial census as the basis for reapportioning their congressional seats in the U.S. House of Representatives. In addition, state and local governments use census data to redraw their election districts.
- Census uses statistical data to answer specific questions from the public and state and local government officials. It also uses the information to structure future samples, prepare studies, and reconcile import and export data. Analysts use census data for such tasks as assigning geographic classification codes to addresses, preparing maps, delineating high-crime areas, and plotting projected population growth.

Planning for Electronic Information Dissemination

Census uses several mechanisms to help plan its electronic information activities. Census meets with and obtains input from advisory committees and local, regional, and national officials; polls users with mail surveys; evaluates previous census data and new dissemination media; and prepares marketing plans in some subject areas.

Census prepared for the 21st decennial census, which began on April 1, 1990, with several planning activities at various levels. For instance, Census reviewed recommendations from data users at 10 regional census product meetings and subsequently held a national conference on data products to gather additional recommendations for refining the product proposals discussed at the regional level. In addition, Census held 65 local public meetings in major cities to seek advice from the public on planning the 1990 census. Further, Census met with and considered recommendations from data users at conferences and meetings of professional organizations including the Population Association of America, the Association of Public Data Users, and the American Statistical Association.

Census has advisory committees on subject areas such as agriculture and population. The advisory committees, which are composed of people outside Census, provide input and recommendations to Census on what information to collect and publish. For example, the advisory committee on agriculture statistics expressed concern over Census' plan to no longer collect data on irrigation organizations during the next agriculture census. Because of the committee's view that this is important data, Census indicated that it hoped to reinstate collection of data for irrigation organizations depending on the availability of funding.

To improve the dissemination of information, Census conducted an evaluation to determine how to enhance data products and make census information more accessible. Based on the evaluation, Census decided to use CD-ROM to disseminate census information. Census also develops marketing plans to promote the general awareness of the value of the censuses and target markets of data users.

Controls to Maintain Data Integrity

Census performs field edits to monitor data for consistency, completeness, and acceptability. Data are also reviewed for omissions, inconsistencies, and population coverage. For example, during the 1990 decennial census, Census had 344 census district offices and 6 processing offices check and edit questionnaires for completeness and consistency of responses. The questionnaires were then microfilmed and processed by optical sensing devices for input to computers. The information was placed on tape and subsequently transferred to computer disks.

In addition, Census provides other mechanisms for maintaining data integrity, including computer system controls, procedures for notifying users of errors, and replacing defective electronic media products. For example, Census uses software that has embedded routines that ensure data are copied correctly from the original computer media to CD-ROMs, tapes, and floppy disks.

Further, Census has controls in place to monitor data disseminated by the two CENDATA vendors. Census determines when subject area information may be released to the public by accessing and previewing the data on the system. According to the Chief of the Technical Information Section, Census provides the CENDATA vendors with census data tapes, which the vendors process. After processing, Census checks the CENDATA records for accuracy by sampling the data and comparing the samples to the original data. For example, a disk containing economic indicators is compared

against the CENDATA record. Also, a hard copy of each sample is forwarded to the offices in charge of the relevant subject areas where it is reviewed for correctness. After these checks are completed, the vendors are notified that the information can be released to the public.

Costs for Electronic Information Dissemination

Census can recover costs for dissemination under authority contained in Title 13 Section 8 of the U.S. Code. Costs incurred for information dissemination vary by media and include factors such as order processing, packaging, computer and personnel support, and documentation. For example, the total cost for disseminating magnetic tape would include the cost to copy the tape, staff costs to reproduce and package the tape, the cost of the tape itself, and postage.

All overhead costs are tracked using an internal fund, called the interfund, authorized by Congress in 1962. The objective of the interfund is to determine the portion of overhead costs for each program or project and then establish a rate based on that analysis. This is determined by factors such as hours of operation and direct labor costs. According to an agency official in the budget division, the revenue that is collected for information activities helps pay for overhead costs such as personnel, budget, finance, facilities, computers, and depreciation. The revenues collected from information dissemination activities are returned to Census' various organizations depending on their level of participation.

Special services and special censuses, including any resulting information products, are completely paid for by the requesting organization. Special services include designing and conducting sample surveys, providing population estimates and projections, making special tabulations, and giving other technical assistance. These services are provided for private groups, individuals, and other government agencies. Special censuses are population censuses requested by counties, cities, villages, townships, and school districts. After completion, these special services and censuses are part of the public domain.

The charge for special services and censuses is designed to recover Census' costs and is determined by factors such as the quantity of data requested, the amount of work required to prepare the data, the complexity of data specifications requested, costs for Census personnel and equipment, the medium in which the data are available, planning costs, and overhead expenses. However, the charges do not include the cost of collecting the data unless special questions have been added to a

questionnaire or an entire survey is undertaken for the sponsoring organization.

Prices Charged for Electronic Dissemination

The price for each electronic data product is based on actual dissemination costs and is determined by the costs incurred to complete a typical order plus associated indirect costs. The prices established for electronic census products are based on associated costs in creating the data product or service, including all direct costs such as salaries and travel, as well as applicable indirect costs for such items as space, employee benefits, and administrative expenses. In addition, the prices for these products include the associated costs incurred in disseminating the product or service, including such items as processing the order, accounting for product sales, copying the product (including the cost of the media and the staff time to make the copy), operating the computer, packaging the product, preparing technical documentation, and providing customer assistance to the user.

Census charges \$175 per reel for computer tape and most CD-ROM prices range from \$150 to \$250, depending on the amount of information. The charge for each floppy disk is \$6 plus a \$60 processing fee per order. Microfiche prices range from \$5 to \$40 for up to 100 copies, depending on the number ordered, and \$.40 each for 100 copies or more. The pricing structure for accessing on-line CENDATA is established by the vendor.

The pricing structure for special services, special censuses, and resulting electronic data products is based on the actual cost of creation and dissemination activities. Census collects, in advance, estimated costs on orders placed for special services and censuses and on orders placed for the resulting electronic data products.

Census' revenue reported from the sales of tapes, disks, and CD-ROM's for fiscal years 1988-1990 is shown in table III.1.

Appendix III
Bureau of the Census' Information
Dissemination Practices

Table III.1: Electronic Data
Dissemination Revenue^a

Media	Fiscal Year					
	1988		1989		1990	
	Number of Units	Revenue	Number of Units	Revenue	Number of Units	Revenue
Tapes	2,749	\$683,965	3,125	\$749,879	3,833	\$1,168,705
Floppy Disks	300	49,083	633	97,835	422	80,818
CD-ROM	199	15,745	133	12,677	657	153,196
Total		\$748,793		\$860,391		\$1,402,719

^aMonetary amounts are in current dollars.

The National Library of Medicine's Information Dissemination Practices

The National Library of Medicine (NLM), located in Bethesda, Maryland, provides medical information to the health community. Its holdings include more than 4.5 million books, journals, technical reports, theses, pamphlets, photographs, and audiovisual materials covering more than 40 biomedical areas and related subjects, such as chemistry, botany, psychology, and zoology. Material may be consulted at NLM, borrowed through an interlibrary loan, or searched by means of NLM's computerized on-line databases.

NLM electronically disseminates information through its Medical Literature Analysis and Retrieval System (MEDLARS). MEDLARS, which originated in 1964, consists of 42 individual databases and data banks most of which contain references to and abstracts of articles published in biomedical journals.¹ Thirty-one of the databases reside on a computer system located at NLM, referred to as the NLM computer, and 11 databases reside on the Toxicology Data Network (TOXNET), located at a Kensington, Maryland, contractor.

The NLM computer runs software called Elhill, created in 1972, which handles the information retrieval function for 31 of the MEDLARS databases. It was designed to perform bibliographic retrieval of records and handle simultaneous users. The largest and most extensively used database on the NLM computer is MEDLINE. It includes citations for about 6 million articles from approximately 3,700 biomedical journals. MEDLINE can be accessed directly with the Elhill information retrieval software, or with GRATEFUL MED—information retrieval software used on personal computers and designed for inexperienced users.

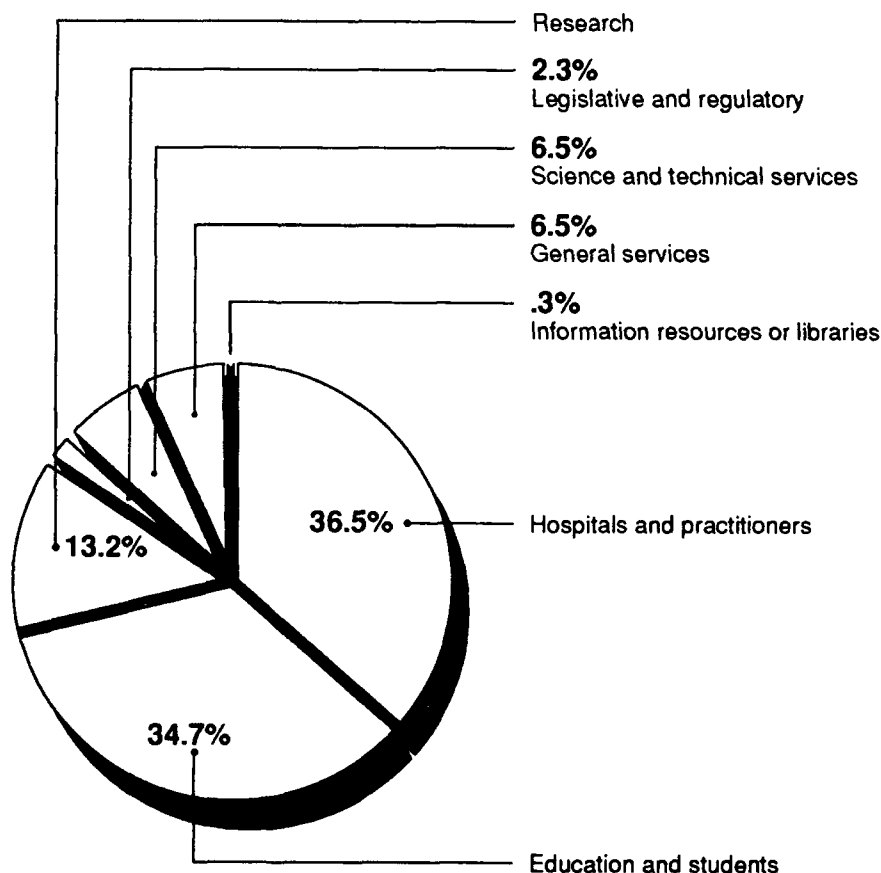
TOXNET is NLM's on-line network of 11 databases on toxicology, environmental health, hazardous substances, and other related subjects. TOXNET was developed separately when it became clear that Elhill would not be able to handle the length of data records that were in the Hazardous Substances Data Bank and other toxicology databases without a major rewrite of the Elhill program.

¹NLM defines a database as a collection of records that are bibliographic in nature, and a data bank as a collection of records containing factual data that may or may not refer back to primary or secondary sources.

MEDLARS Users

NLM users include on-line users who access MEDLARS at NLM and licensees who lease MEDLARS data on magnetic tape. NLM groups on-line users into different categories to assist in answering questions about the composition and activities of their user base. NLM determines what group a new user is in using information from the user's application for services. From the application, on-line users are grouped into seven categories: hospitals and practitioners; education—including medical, dental, and other allied health professional schools—and students; legislative and regulatory; information resources or libraries; research; science and technical services; and general services. NLM estimates that there were 38,756 accounts billed for domestic on-line users and 4,581 accounts for foreign on-line users as of April 1991. Figure IV.1 shows the composition of domestic on-line users as of April 1991.²

Figure IV.1: Composition of MEDLARS
Domestic On-line Users as of April 1991



²NLM does not maintain data on the composition of foreign on-line users.

Over 80 licensees—identified as universities, governments, and vendors—lease one or more of the MEDLARS databases on magnetic tape. Some licensees also use NLM's on-line services. In addition, licensees may obtain a one-time, personal use subset of MEDLINE data on magnetic tape. These subsets are usually for a one-time research effort and are not updated. There are also licensees who produce and distribute compact discs with MEDLARS data.

All licensees sign a standard license agreement with NLM that specifies terms and conditions under which the licensee is authorized to use NLM's databases. The vendors and universities that we contacted said that their users included health professionals, students, university faculty, research scientists, librarians, hospitals, pharmaceutical institutions, and the general public.

How MEDLARS Is Used

According to the Associate Director for Specialized Information Services, NLM does not routinely maintain data on the topics searched because it has assured users that their specific use of MEDLARS databases will not be divulged to third parties. However, in 1982, the Office of Technology Assessment issued a report that stated that hospitals and medical schools use MEDLARS to conduct searches for physicians, nurses, lab technicians, and administrators, often for the same reasons—patient care, preparation of articles and lectures, and planning.³ Hospitals receive about one-half of their search requests from physicians wanting information directly applicable to patient care. In medical schools, the majority of searches are related to research.

Through our contacts with licensees, we found that licensees vary in the products and services offered. For example, some university licensees indicated that they offer NLM's databases without modification. Other licensees said that they make available to their users software that can be used to search the databases; features that eliminate variations in the spelling of medical terms; links to full texts of referenced journal articles; access to databases from other sources, both medical and nonmedical; training; toll-free telephone assistance; copies and deliveries of requested journal articles; and compact discs containing information.

³MEDLARS and Health Information Policy: A Technical Memorandum, Office of Technology Assessment, September 1982.

Information Dissemination Planning

In January 1985, NLM's Board of Regents began to develop a long-range plan to guide NLM's use of human, physical, and financial resources for the next 20 years.⁴ To develop the plan, the Board convened five panels covering various aspects of NLM's work; these planning panels were composed of librarians, physicians, nurses, and other health professionals.

The long-range plan, completed in January 1987, consists of seven reports—an executive summary, a report containing detailed recommendations for accomplishing NLM's goals, and reports from each of the five planning panels. These goals include making information more accessible to health professionals, reviewing the public's need for and access to health information, and strengthening competence in medical informatics in the health professions.

NLM uses the annual budgeting process to implement the strategic plan. Each NLM operating division prepares budget estimates and supporting justifications for projects that support strategic objectives within the division. The divisions semiannually assess their projects to ensure that funds are being spent as intended.

NLM supports program planning by obtaining input from users and others. For instance, NLM has established a Literature Selection Technical Review Committee consisting of physicians, researchers, educators, editors, librarians, and historians to evaluate literature included at NLM. In addition, NLM asks users for feedback on program development by polling various users to obtain reactions about new or revised features. Finally, NLM contracted with the American Institutes for Research to study the impact of MEDLINE use by health professionals.⁵ The study uncovered different reasons for searching MEDLINE and how retrieved information is used for medical decision-making. NLM has used this study to plan improvements to its services such as its plan to improve its GRATEFUL MED software to include an expanded medical subject heading vocabulary.

⁴The Board of Regents is an advisory board operating under Public Law 99-158 to advise, consult with, and make recommendations to the Secretary of Health and Human Services on matters of NLM policy, including the rules under which its materials, publications, facilities, and services shall be made available to users.

⁵Use of the Critical Incident Technique to Evaluate the Impact of MEDLINE, American Institutes for Research, September 30, 1989.

Controls to Maintain Data Integrity

To maintain data integrity, NLM uses various techniques such as data entry controls; links between MEDLARS citations and subsequent errata, retractions, and comments; and requirements contained in the license agreement.

NLM sends journals to a contractor who enters descriptive data for each article, including author, title, and abstract. The contractor types the entries twice to eliminate errors, and returns the data to NLM for machine validation of fields and format. The contractor is responsible for correcting identified errors. On a random basis NLM compares on-line records to journal issues and inspects records for completeness and accuracy. In-house and contract indexers review data to describe each article's content using medical subject headings. The work of indexers who are less experienced or are contract staff is reviewed and revised if necessary. In addition, specialists in languages, science, history, or other areas may review articles for accuracy in their fields of expertise. The system will not allow a record to be added to the database until all identified errors are corrected.

When anyone notifies NLM of an error in the database, indexers retrieve and review the original article and make any necessary changes in the bibliographic citation or index. When NLM has been alerted to a possible error in an original journal article, NLM notifies the journal's editor to verify the correction. Corrected records are loaded into MEDLARS on a weekly basis. If the error may cause bodily harm—which occurs several times a year, it is immediately corrected in the database and all licensees are notified by facsimile. In addition, when errata or retractions related to existing citations are published, NLM labels the affected citations so that users retrieving them are notified that they have been corrected or retracted. Also, citations are linked with published comments challenging, supporting, or expanding upon the citation.

NLM's license agreement specifies that all licensees must receive a corrected update of the database annually. A licensee can request weekly or monthly updates of the current year's data in addition to the annual update. As of November 1991, most licensees requested and received monthly updates; a smaller number receive weekly updates. Compact discs produced and distributed by the licensees must be updated at least annually according to the license agreement. If the agreement is terminated or revoked, the licensee must destroy or erase all data in machine-readable form obtained under the agreement.

NLM's license agreement specifies that licensees must update, maintain, correct, and modify the database in such a way that no elements in the database records become incorrect. To help ensure data integrity, licensees who provide remote access to their users receive test codes that can be used without charge to verify that records are correct in the NLM on-line system. NLM can tell when a test code is being used; however, no records are kept of how often test codes are run or why they were used.

Similarly, licensees must provide a test code that will allow NLM access to the licensee's database. NLM uses these test codes to ensure data quality in two ways. Initially, when the licensee and NLM have first entered into an agreement and the licensee is ready to publish the database, NLM uses the test code to make sure the database is accurate and complete. In addition, if NLM is informed that a licensee's database is inaccurate, incomplete, or not up-to-date, NLM may use the test code to determine whether this is true. According to the Chief of the Bibliographic Services Division, NLM uses the test codes for this purpose a few times a year. NLM does not keep records about the number of times this occurs or with what licensees.

Occasionally, a licensee will request approval to merge MEDLARS data with data from other sources. When this occurs, licensees must obtain written approval from NLM. Before approval is obtained, NLM examines physical samples from the licensee's database to ensure that MEDLINE records are accurate and clearly differentiated from other records.

In May 1990, NLM requested that MEDLARS licensees voluntarily complete a self-administered evaluation covering the quality assurance clauses of the license agreement. Evaluations were sent to 23 licensees that provide access to MEDLINE data outside of their own organizations, but not to licensees who limit access to users within their own organizations. Seventeen out of 23 licensees responded to the request. The licensees were requested to answer questions about their databases that included a sample search to verify that their database included the same data as NLM's computer. NLM plans to do this on a bi-annual basis and has recently sent out the questions for this year. NLM believes this is the only way to evaluate the licensees in a consistent and comparable way.

By analyzing the responses, NLM identified the following five problem areas that need to be addressed by the licensees: (1) providing a timely announcement mechanism for significant errata or retractions identified by NLM; (2) correcting records at the time new records are added, or at least annually; (3) performing searches the results of which are comparable to

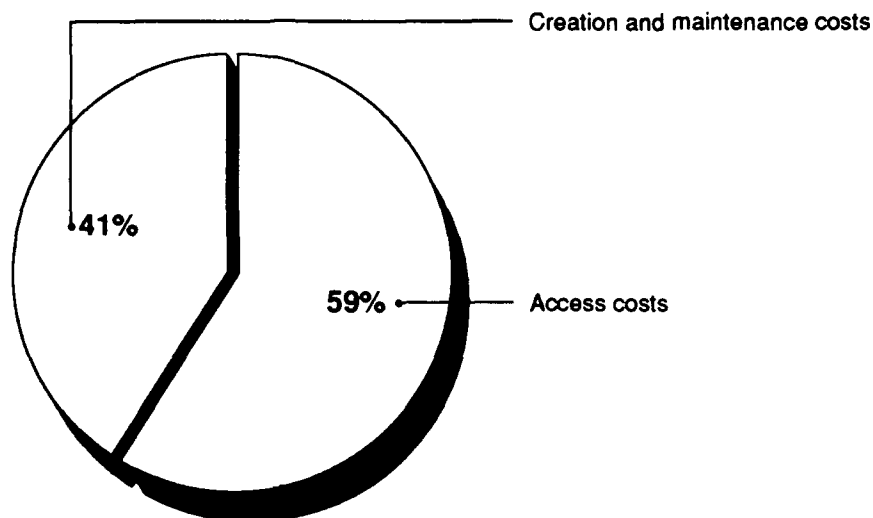
results of searches performed on MEDLINE; (4) retaining a minimal set of data elements in each record as specified by NLM for each database; and (5) retaining a comments field in a data record. NLM discussed with each responding licensee the problems they had found and possible solutions. Because this evaluation was outside the requirements of the license agreement, NLM limited its follow-up action with the nonrespondents to a letter. No other follow-up action took place.

Cost Recovery Process

The Health Research Extension Act of 1985 authorizes the Secretary, Department of Health and Human Services, with the advice of the Board of Regents, to set NLM's pricing policies. The Board delegated to NLM's director the authority to set prices, but annually reviews pricing policies and submits any necessary changes to the Secretary. The Board uses four principles to guide its decisions: (1) medical information is a public good and should be readily accessible; (2) data quality and integrity must be protected at all times; (3) users, regardless of location or situation, should have equal access to data at equal prices; and (4) the access costs should be shared by users at the lowest feasible price.

Prices set by NLM are related to costs incurred by NLM. To determine the costs that should be recovered, NLM first separates MEDLARS-related costs from costs not related to MEDLARS. NLM further separates MEDLARS-related costs into two categories—data creation and data access costs. Data creation costs are covered by appropriated funds because that activity is considered a basic library function. Data creation costs include computer equipment costs, personnel costs, and 50 percent of indexing costs. Costs incurred to provide access to the data via on-line services and magnetic tape are also considered recoverable. Access costs for the on-line database include network management, personnel, user support, shipping, equipment, agreements, contracts, and some overhead costs; access costs for leasing of data tapes include blank tapes, postage, computer equipment, and staff costs related to tape productions and distribution. While NLM incurs costs for magnetic tape and on-line services, it incurs no direct cost for compact disc production because licensees produce and distribute the discs. Figure IV.2 shows the fiscal year 1989 MEDLARS costs.

Figure IV.2: Fiscal Year 1989 MEDLARS
Costs



NLM states that it sets and collects user fees to cover only the full cost of accessing its databases. NLM's intention is to set prices that do not differentiate on the basis of database or method of distribution. NLM's rationale for this approach is based primarily on the desire to ensure that access to data files that contain infrequently used or highly specialized data is not inhibited by high user charges. Consequently, NLM does not recover access costs by individual MEDLARS database but rather recovers total access costs for all databases. In addition, foreign users, as non-U.S. taxpayers, pay a surcharge that represents their share of database creation costs.

Charges may be reduced or waived in special cases. For instance, technical services databases—NLM's on-line public access catalogs—are provided by NLM without usage charges.⁶ These databases are used by librarians and other users to locate items in other databases; they are not considered end products. In addition, NLM decided not to assess usage charges for products related to acquired immunodeficiency syndrome (AIDS) because of special government initiatives in this field. Finally, while all government agencies have to pay to use MEDLARS, including the National Institutes of Health, the cost of NLM employees using MEDLARS is covered by appropriations and is not passed on to other MEDLARS users.

⁶These databases include AVLINE, MeSH VOCABULARY FILE, CATLINE, SERLINE, and NAME AUTHORITY FILE, which are described in appendix V.

NLM's pricing policy covers two different users—on-line users and licensees. MEDLARS charges vary by user type. About two-thirds of the revenues collected by NLM are derived from on-line charges and about one-third are obtained from licensee fees.

Charges for On-line Services

Prior to 1983, NLM computed charges for on-line use based solely on the amount of time the user was connected to the system. In 1983, because the speed of access varied considerably among users, NLM began computing charges based on the length of connect time and measures of resources used, such as the number of searches attempted, citations printed, interactions processed, characters printed, and times the storage disks are accessed. Also, charges are less for services provided during nonprime-time hours than during prime-time hours.⁷ NLM estimates that users pay approximately \$2.00 to perform an average MEDLARS search using GRATEFUL MED.

Different rates are charged depending on whether the user is accessing the NLM system or TOXNET because TOXNET and Elhill perform differently. This differential rate is intended to overcome the differences in the two computers' operating systems. Pricing algorithms were developed so that if the same file were retrieved from the two computers, the cost to the user would be similar.

On the NLM system, hourly charges for time the domestic user is connected to the system are \$6.10 during nonprime time and \$8.85 during prime time. Foreign users are charged an additional \$3.00 per connect hour. Charges for computer resources used vary depending on the time of day. For example, during prime time, users are charged \$.02 for every search, \$.01 for every interaction, \$.14 for every 100 disk accesses, \$.02 for every citation printed, and \$.12 for every 1,000 characters printed.

On the TOXNET computer, hourly charges for time the user is connected to the system are \$8.92 during nonprime time and \$11.28 during prime time. As on the NLM computer, foreign users are charged an additional \$3.00 per connect hour. The charges for computer resources during prime time are \$.05 for every search statement, \$.03 for every interaction, and \$.13 for every 1,000 characters printed.

⁷Prime time is defined as 10:00 a.m. to 5:00 p.m. Eastern Time, Monday through Friday. All other times are nonprime.

Three MEDLARS databases—TOXLIT, TOXLIT65, and CHEMLINE—incur an additional charge to recover royalties for the private organizations that compile them. For TOXLIT and TOXLIT65, the additional charges are \$27.00 per connect hour, \$.73 per search and \$.53 per citation. For CHEMLINE, the additional charges are \$27.00 per connect hour, \$.60 per search and \$.53 per citation. For users who can't afford the higher cost, NLM has parallel files—TOXLINE, TOXLINE65, and ChemID—that do not have information that would require the additional royalty charge.

Charges to Licensees

In the early 1970s, NLM began leasing MEDLARS data on magnetic tape in response to a Board of Regents recommendation. NLM requires individuals or institutions that want to obtain MEDLARS on magnetic tape to sign a license agreement that authorizes the licensees to provide information services and products related to MEDLARS data. The license agreement contains several provisions, such as those covering the duration of the agreement, copyright constraints on certain databases, NLM's and licensees' roles related to quality assurance, restrictions on redistribution of data, requirements for licensees to provide usage reports, and prices for data obtained. The license agreement does not require licensees to provide a report of the revenue they receive from their users.

Prices charged to licensees for MEDLARS vary depending on the databases obtained, the number of records requested, the frequency of updates required, and the intended use of the data. NLM's latest pricing policy, which became effective October 1, 1991, considers a licensee to be either a vendor who makes MEDLARS data available to users outside its own organization or an institution that limits access to MEDLARS data to its own employees, faculty, and students. Since 1983, NLM has determined the annual minimum charges based on the average cost to produce and distribute a tape, the number of tapes required to distribute the databases annually, shipping costs, and the average cost for license management and quality assurance. For example, distribution of MEDLINE with weekly updates requires 89 tapes annually per licensee at a cost of \$75 per tape. Therefore, costs associated with tape production and distribution for MEDLINE are \$6,675. The license management cost per licensee is \$1,200, bringing the total annual cost to \$7,875. As a result, the annual minimum fee for MEDLINE is \$8,000.

Licensees, both vendors and institutions, pay an annual minimum fee. For instance, licensees providing MEDLINE via on-line service pay an annual minimum fee of \$8,000 to receive weekly updates of the current year's

data, or \$4,000 to receive monthly updates. Annual minimums for older data range from \$1,250 to \$5,000 depending on the number of records requested.⁸

In addition to the annual minimum fee, licensees pay different usage charges depending on whether they are a vendor or an institution. The usage charge for a vendor is \$1.90 per connect hour, \$.013 per 1,000 characters transmitted, and \$.01 per off-line citation. The usage charge offsets the annual fee, so that the vendor is charged either the annual fee or the total of all usage charges, whichever is larger. The usage fee for an institution is \$100 per year for a single-user system or \$200 per concurrent user per year for a multiuser system. Special arrangements can be made for statewide educational systems under one administrative body.

Subscription Fees Related to Compact Discs

In 1988, NLM began assessing use charges to licensees that produced and distributed compact discs containing MEDLARS data. The current pricing policy sets use fees for a domestic institution at \$100 per year (\$125 foreign) and \$200 per concurrent domestic user for a multiuser system (\$250 foreign).

NLM established a reduced price schedule for individual health professionals of \$40 for domestic and \$50 for international users. This pricing policy was instituted to encourage the use of CD-ROM technology by NLM's largest audience and to assist its licensees in creating an individual compact disc market.

Fiscal Year 1989 Costs and Amounts Billed

As shown in table IV.1, during fiscal year 1989 the amount billed to MEDLARS users exceeded costs incurred by NLM to provide access to MEDLARS. The difference is attributed to the surcharge collected from foreign users to cover the cost of creating MEDLARS, as well as the time lag between costs incurred and billings. Specifically, billings include some services provided during the latter part of fiscal year 1988 and exclude some services provided near the end of fiscal year 1989. Also, NLM's pricing policy is designed to recover the total cost of providing access to all MEDLARS databases, not just individual databases. Therefore, NLM recognizes that billings may not equal operating costs for each database.

⁸Foreign annual minimum rates are the same as U.S. annual minimum rates.

Appendix IV
The National Library of Medicine's
Information Dissemination Practices

Table IV.1: MEDLARS Database Costs and Billings, Fiscal Year 1989

Database^a	Costs Incurred^b	Amount billed
MEDLINE/SDILINE	\$7,224,253	\$9,975,757
HEALTH	220,438	289,030
TOXLINE/TOXLIT	1,220,836	798,185
AIDSLINE/AIDSDRUGS/ AIDSTRIALS	64,230	52,258
CHEMLINE	367,571	145,136
TOXNET	1,834,536	265,061
CANCERLIT/CLINPROT	301,622	256,283
PDQ	523,874	262,293
AVLINE/CATLINE/ MeSH VOCABULARY FILE/ SERLINE/NAME AUTHORITY FILE	397,843	243,740
BIOETHICSLINE/ POPLINE/DENTALPROJ	38,547	45,893
DIRLINE	70,635	10,651
HISTLINE	6,842	6,064
INFORM	13,760	6,464
DOCUSER	5,261	3,555
REFLINE/YEAR86 ^c	39,410	0
Other ^d	0	1,049,357
Total	\$12,329,658	\$13,409,727

^aDescriptions of these databases can be found in appendix V. ChemID, DART, and IRIS appear in appendix V but are not included in this table because they were added to MEDLARS after fiscal year 1989. CLINPROT is not included in appendix V because it has been deleted from MEDLARS.

^bCost incurred and amount billed are in current dollars.

^cA portion of MEDLINE onsite at NLM.

^dOther includes primarily foreign surcharges.

Billing and Collection Services

According to an interagency agreement, NTIS handles all of NLM's billing and collection responsibilities for information dissemination services. NLM regularly reports to NTIS the amount of usage for on-line and tape dissemination services, and NTIS subsequently sends bills to and collects payments from the system users. Until recently, NTIS charged NLM 9 percent of total annual collections. In October 1990, NLM agreed that NTIS would apply the 9 percent charge to billings instead of collections. NTIS had several overdue NLM accounts and NTIS felt it was losing money when payment was based on collections. NTIS could not account for the costs that contribute to the 9 percent charge; an NTIS accounting official also did not know whether charges exceeded the costs of administering the NLM account. Table IV.2 shows NLM's collections, NTIS fees, other expenses, and excess funds for fiscal years 1985 through 1989.

Appendix IV
The National Library of Medicine's
Information Dissemination Practices

Table IV.2: MEDLARS Collections, NTIS Fees, and Excess Returned to the Treasury^a

Fiscal year	Amount collected^b	Fee paid to NTIS	Other Expenses^c	Excess
1989	\$13,053,000	\$1,174,770	\$7,040,000	\$5,373,000
1988	12,180,000	1,096,200	5,389,000	5,335,000
1987	10,955,000	985,950	5,053,000	4,600,000
1986	8,966,000	806,940	4,155,000	3,800,000
1985	8,179,000	736,110	3,406,000	3,800,000

^aAll monetary amounts are in current dollars.

^bCollections for a fiscal year differ from the amount billed because of the time lag between the two. Specifically, fiscal year 1989 collections would include payments of bills distributed near the end of fiscal year 1988, and would exclude bills distributed late in fiscal year 1989 and collected in fiscal year 1990.

^cAccess expenses paid directly by NTIS.

As can be seen from table IV.2, NLM has excess funds each year. These are returned to the Department of the Treasury and include surcharges to foreign users, and management and overhead costs that already are covered by appropriated funds. In 1984, the Office of Management and Budget (OMB) proposed that NLM keep any excess funds, and OMB would reduce funding the following year. The appropriations committee objected, and the proposal was dropped. NLM's Deputy Director believes that NLM should recover all costs attributed to distribution and does not want to exclude those covered by appropriated funds.

Description of MEDLARS Databases

MEDLARS consists of 42 individual databases and data banks accessible via two computer systems, which are referred to as the NLM computer and TOXNET. The following database descriptions were obtained from NLM documents.

Information Available on the NLM Computer

The following 31 databases and backfiles are available on the NLM computer:

AIDSDRUGS is a dictionary of chemical and biological agents currently being evaluated in acquired immunodeficiency syndrome (AIDS) clinical trials that are in the companion AIDSTRIALS database. Each record represents a single substance and provides information such as standard chemical names, synonyms and trade names, Chemical Abstracts Service registry numbers, pharmacological action, adverse reactions, physical and chemical properties, and manufacturers' names.

AIDSLINE contains a collection of bibliographic citations on aspects of AIDS. The file contains citations from other files, such as MEDLINE and CANCERLIT, and covers the years 1980 forward. The structure of the file is similar to MEDLINE and the database is searched in the same manner.

AIDSTRIALS contains records on clinical trials of substances being tested for use against AIDS, AIDS-related complex, AIDS-related opportunistic diseases, and human immunodeficiency virus infection. Each record covers a single trial and provides information such as title and purpose of the trial, diseases studied, patient eligibility criteria, contact persons, and trial locations. Information about the trials is supplied by the National Institute of Allergy and Infectious Diseases and the Food and Drug Administration.

AVLINE contains bibliographic citations to audiovisual material and computer software in clinical medicine. Citations cover audiovisuals and software cataloged by NLM since 1975.

BIOETHICSLINE contains citations to documents on such topics as euthanasia, organ donation and transplantation, the allocation of health care resources, patients' rights, codes of professional ethics, and in vitro fertilization. This file is produced by the Bioethics Information Retrieval Project of the Kennedy Institute of Ethics at Georgetown University. The file contains citations covering the years 1973 forward.

CANCERLIT contains bibliographic citations covering major cancer topics. Citations are derived from MEDLINE, foreign language journals, and published literature references. Records added since January 1980 have been indexed using NLM's medical subject headings and are retrievable using this controlled vocabulary. All records are retrievable by free-text searching.

CATLINE provides users with access to virtually all of the cataloged titles in NLM's collection, ranging from fifteenth-century imprints to the present, as well as manuscripts. The file is also used by NLM to produce printed materials, verify interlibrary loan requests, and provide reference services.

ChemID is a dictionary of chemicals for over 182,000 compounds of regulatory and biomedical interest. Records include Chemical Abstracts Service registry numbers and other identifying numbers, molecular formulae, generic and trivial names, medical subject headings, and file locators that lead users to other MEDLARS files.

CHEMLINE is an interactive dictionary of over 1,000,000 chemical substances. It provides a mechanism whereby information on these substances can be searched and retrieved on-line. The file contains a locator designation that points to other MEDLARS databases. This file can also be used to locate classes of chemical substances. File content is obtained from the Chemical Abstracts Service under a usage royalties arrangement.

DENTALPROJ is a joint effort between NLM and the National Institute of Dental Research to provide information on active and recently active dental research projects. The file covers those research projects being supported by the Department of Health and Human Services, Department of Defense, Department of Veterans Affairs, and other sources.

DIRLINE is an on-line directory primarily of health and biomedical organizations providing information in specific subject areas. The organizations covered by this file fall into many categories, including all levels of government, information and referral centers, professional societies, voluntary associations, academic and research institutions, libraries, and museums. Records may contain information on the services provided, publications, and holdings. DIRLINE may be searched by organization name, text words, or subject-related keywords and medical subject headings in order to identify organizations specializing in a particular subject area.

DOCUSER contains descriptive and administrative information about libraries and other organizations that use NLM's interlibrary loan services or are a part of the Regional Medical Library Network. Each record contains the institution's name, address, and telephone number when available. In addition, some records have been supplemented with interlibrary loan policy data provided to NLM by the Regional Medical Libraries. DOCUSER contains records for over 10,000 health-related libraries located within the United States and approximately 1,100 foreign libraries.

HEALTH contains bibliographic citations covering nonclinical aspects of health care delivery. Subject areas emphasized include the administration and planning of all types of health facilities, services and manpower, health insurance, health policy, financial management, regulation, personnel administration, quality assurance, and licensure and accreditation that apply to health care delivery. Citations are prepared by NLM and the American Hospital Association. This file contains citations covering the years 1975 forward.

HISTLINE contains citations covering the history of medicine and related sciences. The scope of the records include specialties, professions, individuals, institutions, drugs, and diseases of given chronological periods and geographic areas. Data in this file come from MEDLINE, CATLINE, journals on the history of medicine and the sciences, and recent publications in fields such as general history and classics. HISTLINE contains citations covering 1964 forward.

MEDLINE, the largest and most extensively used database in MEDLARS, contains information on the science and practice of medicine and public health, bioengineering, bioethics, and other health-related disciplines. This database includes citations for about 6 million articles from approximately 3,700 biomedical journals published in the United States and abroad. Citations are indexed using NLM's controlled medical subject heading vocabulary. MEDLINE contains all citations published in Index Medicus and corresponds in part to the International Nursing Index and the Index to Dental Literature. This file contains citations covering the years January 1989 to the present.

MED86, MED83, MED80, MED77, MED72, and MED66 are backfiles that contain the same information as MEDLINE but for different time frames. These databases contain citations covering January 1966 through December 1988.

MeSH VOCABULARY FILE is an on-line dictionary of medical subject headings, subheadings, and supplementary chemical terms used in indexing and searching several MEDLARS databases. Records for MeSH headings include synonyms, the scope of the heading, its position in the MeSH hierarchical structure, the date it became a heading, previous indexing notes, and additional information useful in indexing and searching. The file also includes the generic name, systematic name, pharmacologic action, Chemical Abstracts Service registry numbers, and enzyme codes. This file contains headings for the current year.

NAME AUTHORITY FILE contains records that may be used in the cataloging process to identify personal, conference, or corporate names; uniform titles; and series to be used to access other bibliographic records resident in AVLINE and CATLINE. Name records provide detailed information on a name, and the name's relationship to other names, as well as cross-references.

PDQ contains information on advances in cancer treatment and clinical trials. Designed by the National Cancer Institute for use by physicians, PDQ uses a series of menus to provide access to (1) detailed summaries of all major tumor types; (2) the names, addresses, telephone numbers, medical specialties, and affiliations of physicians who devote a major portion of their clinical practice to the treatment of cancer patients; (3) similar information on organizations affiliated with societies that are associated with cancer treatment; and (4) detailed information on more than 1,000 active treatment protocols and standard therapy protocols approved by the National Cancer Institute.

POPLINE contains citations covering population and family planning. This database includes citations related to human fertility; contraceptive methods; community-based services; program evaluation; demography; censuses; vital statistics; and related health, law, and policy issues. POPLINE contains citations covering the years 1970 forward with selected citations dating back to 1886.

SDILINE contains bibliographic citations to journal articles from biomedical journals in the United States and abroad. This file contains the current month's input to the MEDLINE database, including all citations to the forthcoming printed edition of the monthly Index Medicus.

SERLINE contains bibliographic records for all biomedical serials and numbered congresses that are currently in NLM's collection, in process, or

on order. This file also contains a limited number of serials not in NLM's collection.

TOXLINE and TOXLINE65 contain citations covering the toxicological, pharmacological, biochemical, and physiological effects of drugs and other chemicals. The citations are derived from secondary sources that do not require royalty charges based on usage. Citations on a given subject may be retrieved by entering text terms as they appear in titles, keywords, and abstracts of articles. In addition, certain portions of the database may be searched using MeSH vocabulary. Chemical substances can be searched by entering their corresponding Chemical Abstracts Service registry numbers or synonyms. TOXLINE contains citations for 1981 forward, and TOXLINE65 covers 1965 through 1980.

TOXLIT and TOXLIT65 contain citations covering the toxicological, pharmacological, biochemical, and physiological effects of drugs and other chemicals. The citations are derived from a secondary source, Chemical Abstracts Service, that requires usage royalty charges. Citations on a given subject may be retrieved by entering text terms as they appear in titles, keywords, and abstracts of articles. Chemical substances can be searched by entering their corresponding Chemical Abstracts Service registry numbers and/or synonyms. TOXLIT contains citations for 1981 forward and TOXLIT65 covers 1965 through 1980.

Information Available On TOXNET

TOXNET is a computerized system of databases related to toxicology. The following 11 databases and data banks are accessible on TOXNET.

CCRIS is a scientifically evaluated and fully referenced data bank developed and maintained by the National Cancer Institute. It contains carcinogenicity, tumor promotion, and mutagenicity test results. Data are derived from primary journals, current awareness tools, and a specific core set of sources, including National Cancer Institute reports. This data bank is organized by chemical record and contains approximately 2,000 records.

DART is a bibliographic database covering teratology and other aspects of developmental and reproductive toxicology. It continues the Environmental Teratology Information Center Backfile (ETICBACK) database, also on TOXNET, which covers primarily teratology information published from 1950-1988. DART contains citations to literature published since 1989. The records include bibliographic citations, abstracts, medical subject headings, chemical names, and Chemical Abstracts Service registry

numbers. DART and ETICBACK are funded by the Environmental Protection Agency and the National Institute of Environmental Health Sciences. NLM manages and provides access to the databases.

DBIR is a multicomponent data bank containing information on a wide range of resources related to biotechnology. These resources may be on-line databases and networks, publications such as books and compendiums, organizations, collections, and repositories of cells and subcellular elements. This data bank also identifies groups and agencies working on issues of nomenclature in biotechnology and molecular biology. Records contain information on the name of the resource, a contact person's name and address, a general description, and special services provided by the resource. This database is also available as part of NLM's DIRLINE database.

EMICBACK is a bibliographic database that contains citations to publications concerning chemical, biological, and physical agents that have been tested for genotoxic activity. These records include full bibliographic references, keywords, chemical names, and Chemical Abstracts Service registry numbers. This file is produced and funded by the Environmental Protection Agency and the National Institute of Environmental Health Sciences. EMICBACK contains over 67,000 citations to articles published from 1950 to 1988. A small number of older citations are also included. The citations in EMICBACK are also available as part of NLM's TOXLINE database.

ETICBACK is a bibliographic database that contains citations to literature on agents that may cause birth defects. These records contain bibliographic citations, Environmental Teratology Information Center keywords, chemical names, and Chemical Abstracts Service registry numbers. This file is funded by other agencies. ETICBACK contains over 49,000 citations to literature published from 1950 to 1989. A small number of older citations are also included. The citations in ETICBACK are also available as part of NLM's TOXLINE database.

HSDB is a factual, nonbibliographic data bank focusing on the toxicology of potentially hazardous chemicals. It also contains data from such related areas as emergency handling procedures, environmental fate, disposal methods, detection methods, and regulatory requirements. Data are derived from a core set of standard texts and monographs, government documents, technical reports, and journal literature. This database is organized by chemical record, and now contains approximately 4,200 chemical substance records.

IRIS is an on-line database of chemical-specific risk information prepared by the Environmental Protection Agency to serve as guidance for Environmental Protection Agency risk assessments. The TOXNET version of the file includes reference doses and slope factors, which are determined following an agencywide consensus review.

RTECS contains toxic effects data on some 96,000 chemicals. Both acute and chronic effects are covered, including data on skin and eye irritation, carcinogenicity, mutagenicity, and reproductive consequences. Selected federal regulatory requirements and exposure levels are also presented. This database is built and maintained by the National Institute for Occupational Safety and Health, which also periodically publishes paper and microfiche versions of the file.

TRI87, TRI88, and TRI89 databases contain information on the annual estimated releases of toxic chemicals to the environment. The file is based upon data collected by the Environmental Protection Agency for the years 1987, 1988, and 1989, under the Emergency Planning and Community Right-To-Know Act. Records include such data as the names and addresses of industrial facilities releasing the chemicals; the amounts released to air, water, land, or by underground injection; and the amounts transferred to waste sites.

Objectives, Scope, and Methodology

In February 1991, the Senate Committee on Governmental Affairs asked us to examine federal agency information dissemination practices, particularly those involving information maintained in electronic format. Specifically, we were asked to identify, at four agencies, information uses and users, information dissemination planning, controls over data integrity, costs incurred by the agency, and prices paid by users to access the information. We did not attempt to form any conclusions on the adequacy or legality of the agencies' dissemination activities.

We determined our audit scope at each agency by identifying the primary approach each agency takes to disseminate information electronically. At USDA, we focused on the Computerized Information Delivery Service, an on-line system that provides time-sensitive and perishable data to the agricultural community. Our scope at Census was broader because Census disseminates information using various media almost equally. For example, Census disseminates information on tape and CD-ROM, as well as via CENDATA, an on-line system. At NTIS, we concentrated on the Bibliographic Database, NTIS' largest and most widely used database. Similarly, our focus at NLM was the Medical Literature Analysis and Retrieval System (MEDLARS), which electronically disseminates NLM's large store of biomedical data.

The information in this report was obtained from agency studies and documents, as well as through interviews with senior agency officials, program managers, budget officers, and other agency staff. In addition, we reviewed related Office of Technology Assessment and GAO reports. We obtained and reviewed relevant licenses, leases, and contracts that are used by agencies. We also contacted a random sample of vendors, libraries, and other information users and obtained descriptions of their users and the types of services they offer.

Our audit work was performed in accordance with generally accepted government auditing standards, between February 1991 and January 1992, in Washington, D.C.; Suitland and Bethesda, Maryland; and Springfield, Virginia. We obtained oral comments on this report from each agency. They generally agreed with the facts as presented. We incorporated their comments where appropriate.

Major Contributors to This Report

Information Management and Technology Division, Washington, D.C.

Nancy A. Simmons, Assistant Director
James Rinaldi, Evaluator-in-Charge
William Dunahay, Senior Evaluator
Eugene Kudla, Staff Evaluator
Keith Landrum, Staff Evaluator
Alicia Wright, Staff Evaluator
Anita L. Cook, Secretary

Office of the General Counsel

Richard Seldin, Senior Attorney