Appendix A

Technological Opportunities for Managing and Disseminating Federal Scientific and Technical Information

Introduction to the Electronic STI Revolution

Dissemination of Federal STI is being transformed by the ongoing revolution in electronic information and telecommunication technologies. The scientific and technical community is one of the heaviest and most advanced users of computers. The vast majority of U.S. scientists and engineers have a microcomputer at work and/or at home, and some have access to mainframe and high-performance computer resources either onsite or through telecommunication networks. The microcomputer or workstation provides the scientist or engineer with a versatile tool. Continuous, steady improvement in the price/performance of microcomputers has resulted in the power of a 1970s-vintage mainframe computer now being on the desktop of the typical scientist. The microcomputer can be used to search, recover, and store STI on magnetic or optical media, manipulate and analyze STI using a variety of software, and access STI remotely via online bulletin boards, computer conferences, and database networks.[1]

Online information networks serve at least three important needs of the scientific and technical community. First, they are used for the transfer of very large streams of STI, for example, from a central repository of data collected by Earth-observing satellites to regional data repositories and to individual research institutions or user groups. Second, online networks are used to search STI bibliographic databases and to remotely access large-scale high-performance computers. Third, online networks are used for informal exchange of STI among researchers, for example, an electronic bulletin board on research in progress or upcoming key events, a computer conference for exchanging working notes and ideas among scientists conducting related research, and electronic mail for submission of manuscripts and review comments to scientific and technical journals and to funding agencies.[2] Online STI dissemination benefits from both a proliferation of online gateways that provide channels for electronic information exchange (offered by telecommunication common carriers, value-added carriers, and not-for-profit and governmental systems), and a growing variety of STI services (especially bibliographic and reference services offered by commercial and not-for-profit organizations as well as some government agencies). Advances in online STI gateways and information services are made possible in part by progress in underlying digital telecommunication technologies (e.g., packet switching, fiber optics, and satellite networking). The net result is that online is feasible over a broader range of STI dissemination applications than ever before.

Dissemination of Federal STI is being transformed by the ongoing revolution in electronic information and telecommunication technologies.

The package of online and optical disk technologies offers a powerful combination. Online can be effectively used when time or geographic factors are most important (e.g., bibliographic updates on just-published research, access to remote computing resources or to international STI databases) and offline optical disks can be used for large data sets and/or extensive data manipulation and analysis requirements that are not time-sensitive and would be much more expensive online (even at off-peak rates).

The future of STI dissemination will be dominated by electronic formats. Some major types of STI--e.g., satellite remote sensing data or the results of large-scale computer models-are created, stored, transmitted, and used in electronic form. These data are rarely, if ever, converted to paper or microfiche, except when summarized and analyzed in technical reports and scientific papers. By comparison, STI bibliographic and reference materials are currently offered and used in paper, microfiche, and electronic formats (principally online and Compact Disk-Read Only Memory (CD-ROM)). Full-length reports and documents are still largely distributed on paper or microfiche. However, electronic publishing is rapidly taking over the document preparation and production process. Most STI documents are created electronically with word processing systems or software, even though the output is still on paper or microfiche.

[1] U.S. Congress, Office of Technology Assessment, *Informing the Nation:* Federal Information Dissemination in an Electronic Age, OTA-CIT-396 (Washington, DC: U.S. Government Printing Office, October 1988).

[2] See National Academy of Sciences, Committee on Science, Engineering, and Public Policy, *Information Technology and the Conduct of Research: The User's View* (Washington DC: National Academy Press, 1989); U.S. Congress, Office of Technology Assessment, *High Performance Computing and Networking for Science*, OTA-BP-CIT-59 (Washington, DC: U.S. Government Printing Office, September 1989).

> The package of online and optical disk technologies offers a powerful combination.

Electronic publishing makes it possible to carry the advantages of electronic word processing through all stages of document preparation and information dissemination. Electronic publishing creates an electronic document database that can be accessed online, stored on magnetic or optical media, or printed out in whole or in part on paper or microfiche. The major barrier to realization of the ''intelligent database" is standardization of data structures and file formats for graphics and datasets as well as text.

The price/performance of all electronic publishing components continues to improve. This is resulting in a continued narrowing of the gap between relatively inexpensive desktop systems and expensive, high-end electronic publishing and phototypesetting systems. Desktop systems can be linked to very fast, very high-quality phototypesetters and printers.

Desktop publishing and dissemination functions benefit from steady progress in development of expert systems. The expert systems applicable to STI dissemination are no different in principle from the systems that have been successfully applied to other scientific, industrial, and educational areas. Expert systems with sophisticated search strategies can be used to retrieve and deliver bibliographic or full-text STI from offline (e.g., CD-ROM) or online information systems. Expert systems can improve the dissemination process by accounting for such factors as: the profile of the information product (number of pages, layout, type style, use of graphics, etc.), anticipated user needs (e.g., size of demand by format), and the modes of dissemination (press run, provisions for demand printing in paper or microform, online database access, optical disk distribution, etc.). Expert systems can also assist SDI (selective dissemination of information) by matching user interest profiles with available databases, and, potentially, in translation of STI from foreign languages to English (and vice versa).

Over the next 3 to 5 years, use of printed Federal STI is likely to decline modestly, while the use of electronic formats will likely increase dramatically. Some transitional effects are already evident. For example, the National Technical Information Service (NTIS) experienced a roughly 50 percent reduction in sales of paper and microfiche copies of reports between 1980 and 1989. The reduction is attributed in part to the effectiveness of online searching of the NTIS bibliographic database (offered via private vendors).[3] The fastest growing NTIS product line now is computer products. The Office of Scientific and Technical Information at the Department of Energy has noted a similar declining demand for paper and microfiche copies over the past decade as reliance on computerized bibliographic databases increases.[4]

Surveys conducted by the General Accounting Office have documented the plans of Federal agencies to increase their use of electronic formats for STI, and the growing demand of STI users for electronic formats. The survey results indicated a 50 percent or greater anticipated increase over a 3-year period in the number of civilian agencies using electronic mail, electronic bulletin boards, floppy disks, and compact optical disks for STI dissemination. The results showed a doubling over the next 3 years in the number of scientific and technical associations desiring Federal STI in electronic formats. For Federal depository libraries, the results indicated, for example, about an eight-fold increase over the next 3 years in demand for Federal STI on compact optical disks. In contrast, the results showed a projected decline in demand for paper and microfiche formats of about 15 to 20 percent.[5]

A key to realizing the potential for technology-enhanced dissemination is the "information life cycle, " where STI dissemination is part of the larger process of collection/creation, storage, processing, and archiving. The stages in the STI process need to be integrated with interconnected technologies to be cost-effective. Thus the cost and delays associated with rekeyboarding, incompatible equipment, and the like can be reduced.

Another key is to substantially upgrade technology training for the scientific and technical community. Recent surveys have concluded that many scientists and engineers are still not comfortable with online and ondisk systems.[6] At present, user education and training receive

[3]U.S. Congress, Office of Technology Assessment *Informing the* Nation, op. cit., footnote 1, pp. 112-114; U.S. National Technical Information Service, "Annual Report to the Congress on NTIS: Operations, Audit, and Modernization" January 1989.

[4]Bonnie C. Carroll, Office of Scientific and Technical Information, U.S. Department of Energy, ''DOE Reports Distribution Program: Current System and Why Change Is Needed, " Apr. 30, 1986.

[5]U.S. General Accounting Office, *Federal Information:* Agency Needs *and Practices*, Fact Sheet for the Chairman, Joint Committee on Printing, U.S. Congress, GAO/GGD-88-115FS, September 1988; and U.S. General Accounting Office, Federal Information: Users' Current and Future Technology Needs, Fact Sheet for the Chairman, Joint Committee on Printing, U.S. Congress, GAO/GGD-89-20FS, November 1988.

[6]See statement of Charles R. McClure, Professor of Information Studies, Syracuse University, before a hearing of the House Committee On Science, Space, and Technology, Subcommittee on Science, Research, and Technology, Oct. 12, 1989.

Over the next 3 to 5 years, use of printed Federal STI is likely to decline modestly, while the use of electronic formats will likely increase dramatically.

minimal attention. In order to fully realize the potential of electronic technologies for STI, user training needs to be viewed and funded as an integral part of STI system development. Once exposed to electronic STI and trained in its use (whether formally or through self-or collegial-learning), the users frequently become technology enthusiasts.[7]

The convergence of trends in technology and in user preference for electronic data, combined with the emergence of systems integration and standards for the STI life cycle, offer an almost limitless array of possibilities for STI dissemination. Several of these are highlighted below in the context of Federal science agency applications.

Cartographic/Geographic Information

Many aspects of science and technology depend on geographic information-frequently in the form of maps that show the location of transportation networks, natural resources, climate regimes, environmental sources, and the like. In the past, these maps were preparedly hand and printed on paper. Over the last 15 years or so, mapmaking has been computerized, and satellite imagery has been incorporated along with data from field surveys and aerial photogr ammetry. But the final product was and still is largely printed on paper. Over the last 5 years advances in computer technologies have culminated in ‘‘nothing less than a cartographic revolution. ”

This revolution is being driven by digital cartography combined with powerful hardware and software that can access and manipulate geographic data from multiple sources. By collecting information in digital (as opposed to analog form), or by converting analog data (e.g., aerial photographs) to digital form, the data can be readily processed by computers to produce a vast array of computer products. Digitized maps can be displayed on computer screens and recorded on magnetic and optical media, for example, as well as used to produce traditional printed maps.[8]

The U.S. Geological Survey (USGS) expects that many of these digitized maps will be produced in CD-ROM format at a fraction of the cost of the equivalent magnetic tapes or printed paper documents. USGS pilot tests of CD-ROM indicate that it is likely to be an order of magnitude less expensive than computer tapes, and will require only a microcomputer and CD-ROM reader rather than a more expensive mini- or main-frame computer needed for tapes.

Optical disks will revolutionize STI storage and dissemination. Optical disk technology uses a laser beam to record data on plastic disks by engraving pits in the surface. Encoded disks can be read by a low-power laser beam to retrieve the data. Other members of the optical disk family include: WORM (Write Once Ready Many-times); Erasable disks; Videodisk (for storing film or still photos); and CD-I (Compact Disk-Interactive) that combines text, data, video, audio, and software capabilities on one disk.

The CD-ROM is rapidly gaining acceptance, and the basic technical standards are already in place. The marginal cost of producing CD-ROMs is very low—currently about $2 per disk at volumes of several hundred or more. The full cost can be as much as $50 to $500 per disk for several hundred, if the costs of data preparation, premastering, and mastering are included. But even this compares favorably with other storage media. Each CD-ROM can store up to about 600 megabytes (millions of bytes) of data. This is equivalent to about 300,000 text pages (assuming 250 words or about 2,000 bytes per page). One CD-ROM can store the equivalent of about 1,650 floppy diskettes, 30 of the 20-megabyte hard disks, 15 of the 1,600 bits-per-inch 9-track magnetic computer tapes, or 4 of the 6,250 bits-per-inch computer tapes. Thus a $2 CD-ROM can store as much as several hundreds of dollars’ worth of magnetic media. Microcomputer-based CD-ROM authoring software now costs less than $1,000. Agencies with heavy CD-ROM activity might be able to justify purchase of a premastering system ($50,000 to $100,000), but will need to contract out for mastering and duplication.

USGS issuing CD-ROMs with cartographic and geographic information on a variety of topics, such as:

- Ž Gloria Sidescan Sonar Data--contains data for the Gulf of Mexico and parts of the eastern Pacific Ocean, produced by USGS, NOAA, and NASA, and available from USGS;

[7]See John R.B. Clement, “Increasing Research Productivity Through Information Technology: A User-Centered Viewpoint,'' manuscript submitted to “Research Reviews in Information and Documentation,'' October 1989; also see National Academy of Sciences, *Information* Technology and the Conduct of Research: The User’s View, op. cit., footnote 2.

[8]U.S. Federal Interagency Coordinating Committee on Digital Cartography, “Coordinating of Digital Cartographic Activities in the Federal Government,” Sixth Annual Report to the OMB Director, 1988.

The U.S. Geological Survey expects that many of these digitized maps will be produced in CD-ROM format at a fraction of the cost of the equivalent magnetic tapes or printed paper documents.

- • Aerial Photography Records-contains aerial photographs from the USGS National Cartographic Information Center (recently renamed the Earth Science Information Center);
- • Joint Earth Sciences--contains sidelooking airborne radar data, prototype produced by and available from USGS, Bureau of Land Management, and Soil Conservation Service;
- • Hydrodata--contains daily measurement data for USGS water gage stations, produced and sold by Earth Info., Inc. (for profit, formerly U.S. West Optical Publishing); and
- • USGS Reference Materials-contains GEO Index (a database of geologic maps) and Earth Science Data Directory, produced and sold by OCLC, Inc. (not-for-profit, Online Computer Library Center).

Space Science Data

The collection of scientific data by satellites and rockets-already very extensive-will increase further over the next few years, as a new generation of Earth- and space-observing satellites, manned space missions, and interplanetary and deep space probes is launched. The storage and dissemination of these data pose a major challenge to the Federal science agencies—and especially to the National Aeronautics and Space Administration (NASA). Several new electronic technologies have the potential to avoid total systems overload from the expected avalanche of space data.

NASA's primary institution for space data management and dissemination is the National Space Science Data Center (NSSDC) located at the Goddard Space Flight Center in Greenbelt, Maryland. NSSDC is the largest space data-archive in the world, with about 85,000 magnetic tapes of digital data currently on file (along with another 35,000 backup magnetic tapes). The NSSDC archives only processed data, not the raw telemetry data received directly from space. The center also archives a large volume of photographs and film taken by satellites and space missions. Some data are maintained on microform or hard copy. At present, the center archives about 4,000 different data sets, mostly from NASA missions but with a few from Department of Defense or foreign missions. The center retains no classified data, and the primary users are researchers from the disciplines of astronomy, astrophysics, lunar and planetary science, solar terrestrial physics, space plasma physics, and earth sciences.[9]

The opportunities are substantial for use of optical disks to store and disseminate space science data. NASA is beginning to experiment with both 12-inch WORM and 4.75-inch CD-ROM. One WORM product is currently available for dissemination (the data from 20 magnetic tapes were transferred to one WORM disk). And four prototype CD-ROM products are available: 1) a CD-ROM space science sampler that includes a cross-section of planetary, land, oceans, astronomy, and solar-terrestrial data ($50 for the CD-ROM, software on floppy disk, and documentation); 2) a 3-disk CD-ROM set of Voyager/ Uranus images ($100 for the disks, software, and documentation); 3) a 5-disk CD-ROM set of Voyager/Jupiter and Saturn images ($175 for the disks, software, and documentation); and 4) a CD-ROM produced by the NOAA National Geophysical Data Center that includes solar wind and magnetic field data from NASA and various geomagnetic and solar data from NOAA (disk and basic software free while they last; $100 for advanced software and updates).

An understanding of the potential of optical disks can be gained from the following hypothetical examples. The Apollo 17 lunar mission generated about 240 magnetic computer tapes of digital data, 32,000 feet of 16 mm color photographs, and 39,000 feet of 16mm black-and-white photographs.[10] These digital data could be stored on about 4 double-sided 12-inch WORM disks. (One 12-inch WORM disk can store 1.2 gigabytes per side--equivalent to 30 of the 1,600 bits-per-inch magnetic tapes. A two-sided WORM disk can store 2.4 gigabytes or 60 tapes of data.) With 4:1 data compression, it would be possible to store the Apollo 17 data on one WORM disk. The 16mm photographic data, which in this example are equivalent to roughly 850,000 individual photographs, could be stored on about 17 analog videodisks (at the standard 54,000 images per videodisk).

For some of the earlier missions, data for entire series of mission activities could be consolidated. For example, the Mariner interplanetary mission series generated the following volumes of digital data in number of magnetic tapes: Mariner 2 (5 tapes); Mariner 4 (10 tapes); Mariner 9 (42 tapes); and Mariner 10 (184 tapes).[11] The total of

[9] See U.S. National Aeronautics and Space Administration, Goddard Space Flight Center, The National Space Science Data Center, NSSDC-88-26, January 1989.

[10] U.S. National Aeronautics and Space Administration, Goddard Space Flight Center, *NSSDC* Data Listings, NSSDC-88-01, January 1988.

[11] Ibid.

286 magnetic tapes could be stored on about 5 double-sided 12-inch WORM disks (without data compression). The NSSDC archive provides clear evidence of the proliferation of space data over time, as the number and sophistication of space missions increased.

New optical and magnetic storage technologies make it possible for NSSDC to carry out a gradual transition from magnetic tapes and photographic film to higher density storage media such as optical disks or digital tape cartridges (not tape reels, see later discussion on earth sciences data) for digital data and videodisks for analog data. This transition will be quickest for newly acquired data, and for historical data that needs to be re-recorded on new media (i.e., due to deterioration of magnetic tapes, many of which are more than 10 years old and written on obsolete technology).

At the same time, demand for online data dissemination is also increasing. NSSDC is making more data sets available online either over networks or on a dial-up basis. Network options currently include: SPAN (the Space Physics Analysis Network) that links DECnet-based computers in the United States, Canada, Europe, Japan, Australia, New Zealand, and South America; NSN (NASA Science Network) that links with NSFnet and the ARPANET-based Internet; BITNET that links various universities and research organizations; and Telenet, a public packet-switching data network.[12]

Second, technical evaluations and guidelines will need to be developed on when and how to use these media for storing and disseminating data. How fast should high-density storage media be phased in, and what kinds of data sets are best suited for WORM, CD-ROM, videodisk, digital tape cartridge, and other storage technologies? These guidelines will need to take into account the ability of users to accommodate high-density storage media, in terms of training, equipment, and cost. What are the highly leveraged data sets that are both best suited for the new media and matched to user capabilities to handle high-density storage? And the guidelines will need to consider the appropriate balance between offline high-density storage media and online dissemination.

At present, NSSDC includes only a small number of data sets in the online program, and generally limits online time to one-half hour or less. This restriction is based in part on the limited transmission speeds (e.g., still 9.6 kilobits or occasionally 56 kilobits per second, for many universities) such that longer transmissions cost more than offline dissemination. However, online will become more cost-effective as transmission speeds increase. NASA itself already has a l-megabit/second transmission network for use by NASA laboratories and centers. And the proposed multi-agency national research and education network (NREN) anticipates transmission speeds of l-gigabit/second or more in the future.[13] Some current online space science data sets include:

- International Ultraviolet Explorer Satellite, contains ultraviolet spectral data, sponsored by NASA, European Space Agency, and British Science and Engineering Council;
- Total Ozone Mapping Spectrometer, contains 120 days of ozone data from the Nimbus 7 satellite, sponsored by NASA;
- Space Telescope Archive and Catalog, contains catalogs of astronomical data and various observing logs from spaceborne astronomy missions, sponsored by European Space Telescope and Southern Observatory; and
- Crustal Dynamics Data Information System, contains catalog of data from Satellite Laser Ranging, Lunar Laser Ranging, Very Long Baseline Interferometry, and Global Positioning System experiments, sponsored by NASA, National Geodetic Service, and various universities.

Earth Sciences Data

Over the last several years, the Federal science agencies, and the scientific community generally, have made a significant effort to improve the collection, management, and dissemination of earth sciences data. This effort is driven by the widespread concern over problems of global change—ranging from climate change and deforestation to air and water pollution to soil erosion and demineralization to drought-and the recognition that better understanding of these global problems requires much better information. The concept of the Earth system has emerged as an important organizing principle, since global change involves all major earth subsystems—the atmosphere, oceans, snow and ice, lakes and rivers, land formations, and the biosphere (e.g., trees, plants, and animals) and can be affected by forces from deep within

[12] U.S. NASA, *Data* Center, op. cit., footnote 9, pp.15-16; also see U.S. Congress, Office of Technology Assessment, High *Performance Computing*, op. cit., footnote 2.

[13] See U.S. Office of Science and Technology Policy, Executive Office of the President, The Federal High *Performance Computing* program, Sept. 8, 1989; and U.S. Congress, Office of Technology Assessment, High *Performance* Computing, op. cit., footnote 2.

the Earth (e.g., volcanoes and earthquakes) and from far in space (e.g., changes in solar radiation) .[14]

The earth system concept is being used to organize the vast array of data relevant to the disciplines that comprise the earth sciences-climatology, oceanography, glaciology, hydrology, biology, biogeochemistry, geology, etc. In the U.S. Government, the long-term objective is to develop a “virtual” interagency information system for global change data. ‘‘Virtual’ means that the information system will be a family of decentralized data centers, most of which already exist in some form, linked together by common directories, standards, and policies on access, user charges, quality control, and the like. The goal is to have the system fully implemented by the time that NASA’s planned Earth-observing system is operational in the late 1990s (and thus generating a large additional volume of earth sciences data) .[15]

> In the U.S. Government, the long-term objective is to develop a "virtual” interagency information system for global change data. 44 Virtual” means that the information system will be a family of decentralized data centers.

As is the case for space science data, the most effective technology for managing this massive volume of data is high-density storage. Some of the smaller data centers could be converted entirely to a combination of WORM and CD-ROM. For example, the National Oceanographic Data Center, operated by NOAA, maintains about 12 gigabytes of processed data in the following categories: chemical data (marine chemistry), pollutants/toxic substances); biological data (e.g., fish/shellfish, marine birds, plankton); and physical data (e.g., wind/waves, current, subsurface temperature). NODC also maintains about 12 gigabytes of raw, unprocessed data. The entire NODC database of 24 gigabytes would fit on about two to twelve double-sided 12-inch WORM optical disks, depending on the data compression ratio. As new data accumulate, the WORM disks could be updated. Those portions of the database in high demand could be extracted, mastered, and duplicated at very low cost on CD-ROM, and updated CD-ROMs could be issued periodically.

Several Federal earth science data centers are experimenting with CD-ROM. One is the National Snow and Ice Data Center, operated by the University of Colorado for NOAA’s National Geophysical Data Center (NGDC). The Snow and Ice Data Center has issued a prototype CD-ROM with data on Northern Hemisphere ‘brightness temperature grids,” which are collected by a NASA satellite and used to estimate the polar sea ice parameters. The CD-ROM disk comes with a software diskette and a user’s guide, and is available free while supplies last. This is the first in what is planned as a series of CD-ROMs, and reflects a shift in data dissemination philosophy to offline low-cost optical disks for many research purposes.[16] In general, the NGDC believes that CD-ROM will greatly improve the accessibility and usability of STI by the research community, as well as by governmental and private-sector organizations that depend on geophysical data.

The larger data centers are also considering high-density magnetic as well as optical storage. For example, the EROS (Earth Resources Observation Systems) Data Center, operated by USGS, archives about 6 million frames of aerial photographs and over 1 million Landsat and other remotely sensed satellite images. The Landsat imagery alone is roughly equivalent to 75 terabytes (or 75,000 gigabytes) of digital data. Because of this large volume, the EROS Data Center is considering the digital tape cassette as the next generation high-density storage medium. Each cassette can store up to 50 gigabytes of data, much more than either CD-ROM (about 0.6 gigabyte per disk, or 4 gigabytes with 6:1 data compression) or WORM (1.2 gigabytes per disk up to about 12 gigabytes for a two-sided disk with 6:1 data compression). Digital cassettes have a faster data transfer rate than optical disks. On the other hand, the digital tape cassette is a magnetic medium that, like magnetic computer tape reels, deteriorates over time and needs a tape refresh every 7 to 15 years. This compares with a projected lifetime of 20 to 30 years or more for optical disks (the longevity of optical media is still uncertain). The cassettes and equipment cost considerably more than comparable optical disk systems. Optical disks also have the advantage of random (as opposed to sequential) access and

[14] See J.A. Eddy, “The Earth As A System, ” *Earth Quest*, 1987, vol. 1, No. 1, pp. 1-2, available from the Office of Interdisciplinary Earth Studies, University Corporation for Atmospheric Research, Boulder, CO; U.S. National Aeronautics and Space Administration, Earth Systems Science Committee, Earth Systems *Science: A* Closer View (Washington, DC: NASA, January 1988); F.B. Wood, Jr., ‘‘The Need for Systems Research on Global Climate Change, ’ SystemsResearch, 1988, vol. 5, No. 3, pp. 225-240; U.S. National Oceanic and Atmospheric Administration Panel on Global Climate Change, The Vision: A Rededication of *NOAA*, January 1989; and R. Corell, “A Paradigm Emerging,” Earth Quest, 1990, vol. 4, No. 1, pp. 1-4.

[15] See, for example, U.S. Interagency Working Group for Data Management of Global Change, “Interagency Session on Data Management for Global Change, ” minutes of meetings dated Sept. 18, 1987, Nov. 24, 1987, and Mar. 18, 1988.

[16] U.S. National Geophysical Da@ Center, National Snow and I@ Data Center, Data Announcement, ‘ ‘Scanning Multichannel Microwave Radiometer (SMMR) Brightness Temperatures for the Northern Hemisphere,’ June 1, 1989; also see R. Weaver, C. Morns, and R.G. Barry, “Passive Microwave Data for Snow and Ice Research: Planned Products From the DMSP SSM/I System, ” EOS, Sept. 29, 1987, pp. 776-777.

microcomputer compatibility (with inexpensive, user-friendly software). Optical tape is another storage technology that warrants consideration. One 12-inch reel of optical tape can store up to a terabyte of data. Preparation and duplication cost, expected level of use, storage capacity, data transfer rate, data access time, longevity, and equipment and training requirements are among the factors that need to be considered in evaluating alternative storage media.

Drought Monitoring Information

Electronic technologies open up new alternatives for dissemination of time-sensitive Federal STI, such as drought information, that is widely used (contrasted with the very large space and earth sciences data sets that are less time-sensitive and have fewer users). Drought information is collected and disseminated by the U.S. Department of Agriculture and NOAA. The NOAA/USDA Joint Agricultural Weather Facility and NOAA Climate Analysis Center produce several drought-related reports and bulletins, such as the Weekly Weather and Crop Bulletin.

Should the government decide to prepare and distribute a weekly or monthly electronic drought bulletin, it might include: temperature and precipitation trends and forecasts; streamflow, lake (and reservoir) level, and snow pack trends and forecasts; soil and plant (including forest) moisture conditions; soil quality conditions (e.g., mineral content, depth of topsoil); crop conditions; and overall drought indices (e.g., the Palmer drought severity index). The information could be presented on a county, State, regional, and national (and international) level, and would be ideally suited for use with analytical and presentation software (e.g., using spreadsheet or graphics techniques).

Agency pilot tests and experience in other areas suggest several prototypes for electronic dissemination that could be applied to drought (or other) time-sensitive Federal STI. The "best" approach depends on the type of information, number and types of users, importance to the agency mission and statutory guidance, agency precedents, budgetary constraints, related private sector alternatives, and the historical and political context.

A weekly or monthly drought bulletin could be made available on an agency electronic bulletin board for dial-up access by users over commercial telecommunication lines. This approach is used, for example, by the

> Electronic technologies open up new alternatives for dissemination of time-sensitive Federal STI, such as drought information.

National Science Foundation's "science indicators" bulletin board and the Department of Commerce's ‘‘economic bulletin board. Or the drought bulletin board could be disseminated online via the computer center of a single agency contractor, an approach used by the Securities and Exchange Commission (for corporate financial information) and USDA (for various agricultural reports and bulletins). Alternatively, an agency computer center could be employed. For instance, the Environmental Protection Agency is making its "toxic release inventory" available online via the National Library of Medicine computer center. And the U.S. Geological Survey provides data on earthquake epicenters online from the National Earthquake Information Center in Golden, Colorado. Finally, the drought bulletin board could be offered as a service of private sector commercial or not-for-profit value-added information gateways and vendors.[17]

Forest Monitoring Information

Concern about forest ecology is growing. The need for monitoring forest health reflects: 1) mounting evidence of forest decline due to changing environmental, climatic, and soil conditions, among other factors; and 2) growing appreciation of the key role of forests in stabilizing ecosystems and climate on local to global scales.[18]

Because of the rural, remote location of most forests, monitoring in these areas has, until recently, been logistically difficult and expensive. Advances in information and telecommunication technologies now make it possible to provide for efficient and cost-effective monitoring of remote locations. For example, before the microcomputer revolution, temperature and precipitation typically were measured at remote locations with a hydrothermograph that recorded the data on chart paper. In order to retrieve that data, someone had to periodically visit the monitoring site, tear off the chart paper, and carry the chart paper back to a central location where the data

[17] See statements of Edward J. Hanley, Director, Office of Information Resources Management, U.S. Environmental Protection Agency, John Penhollow, Director, Office of EDGAR Management, U.S. Securities and Exchange Commission, and John J. FranIce, Assistant Secretary of Administration, U.S. Department of Agriculture, before a hearing of the Subcommittee on Government Information, Justice, and Agriculture, House Committee on Government Operations, Apr. 18, 1989. Also see U.S. Department of Commerce, Under Secretary for Economic Affairs, "Request for Comments on the Preliminary Implementation Plan of Subtitle E, Part I of the Omnibus Trade and Competitiveness Act of 1988, the National Trade Data Bank," Apr. 21, 1989.

[18] See U.S. Department of Agriculture, Forest Service, "Forest Productivity and Health in a Changing Atmospheric Environment," Conceptual Plan for the Forest/Atmosphere Interaction Priority Research Program, 1988.

Advances in information and telecommunication technologies now make it possible to provide for efficient and cost-effective monitoring of remote locations.

was extracted and logged in. The data had to be typed in if computer processing (e.g., calculation of means and standard deviations) was desired. This largely mechanical and manual process was (and is, where still used) labor-intensive, expensive, and prone to data quality problems. Frequently, the data were collected only sporadically (e.g., by forest rangers on trail maintenance duty), and was logged in and compiled on an erratic basis. As a result, temperature and precipitation records for many rural monitoring stations were incomplete or nonexistent.

Microcomputers and satellites are transforming the nature of remote monitoring and providing a valuable information resource for both research and management purposes. Digital data recorders (at under $1,000 each) are replacing the hydrography. These simple devices record environmental and climatic data on a removable digital chip that stores 45 to 60 days of data. The chips still have to be picked up by someone and carried to an office, but once there, the data can be entered directly from the chip into a microcomputer for data storage and manipulation. Daily, monthly, seasonal, and annual means, among other statistics, can be readily calculated. Remote automated monitoring stations (costing $10,000 to $15,000 each, with a satellite dish) can be used to electronically collect and transmit the data via satellite to a central location. For example, the U.S. Interagency Fire Center in Boise, Idaho, collects climate data via a NOAA satellite from about 500 remote weather stations located on forested lands (260 stations operated by the U.S. Forest Service, 200 by the Bureau of Land Management, and 40 by the National Park Service and Fish and Wildlife Service).

Both the Forest Service and National Science Foundation are funding rural monitoring networks to meet forest ecology and climate research needs. The Forest Service has a ''long-term forest ecosystem monitoring program" to measure selected physical and biological parameters, establish baseline conditions, and detect changes over time. The program builds on existing stations located in experimental forests, experimental rangelands, and research natural areas. Most stations require equipment upgrades and more consistent data collection and analysis. For example, only 16 of 83 experimental forests have long-term data sets (i.e., more than 15 to 20 years), and some of these sets are incomplete or currently inactive.

The monitoring information will help researchers and resource managers determine changes in forest composition and species distribution resulting from air pollution, climate, and soil changes. This in turn should be important input to ecological models used to predict forest growth, commercial tree yield, forest and rangeland habitat and carrying capacity, fire frequency and severity, recreational opportunities, and the like.

To the extent possible, experimental forest monitoring stations are being colocated with the "long-term ecological research" field sites funded by NSF. These sites are designed to study the ecology of a diversity of natural landscapes-including forest, prairie, desert, and aquatic environments. The intent is to better understand the patterns of organic decomposition, primary photosynthetic production, food webs, biogeochemical cycling, nutrient movement through soils and groundwater, atmospheric deposition, and climatic change.[19] Standardized meteorological data collection is being implemented at the various ecological sites, so that baseline conditions are adequate for detection (and documentation) of both cyclic and long-term climatic change.[20] The ecological stations together with the more well-established experimental forest stations provide reasonably balanced geographic coverage of the United States, with one or more stations in the following States: Alaska, Washington, Oregon, California, Idaho, Arizona, New Mexico, Kansas, Minnesota, Wisconsin, Michigan, Illinois, Mississippi, Georgia, South Carolina, North Carolina, Virginia, West Virginia, New York, Massachusetts, and New Hampshire.

The Forest Service and NSF include the dissemination of monitoring information to users as an important objective. Advances in information technology can make this a cost-effective reality. For example, the monthly data for each monitoring station could be sent via inexpensive floppy disk or electronic mail or bulletin board to a designated central location. This could be a Forest Service office, or a research university, or the National Climatic Data Center (operated by NOAA) or National Technical Information Service. The central office or center would quality control and consolidate the data on one disk or magnetic tape. Depending on the data volume and demand, the consolidated data could be issued periodically in floppy disk, magnetic tape, and/or CD-ROM

[19] See James C. Halfpenny and Kathryn P. Ingraham (eds.), Long-Term Ecological Research *in* the United States: A Network *of Research Sites*, Forest Sciences Laboratory, Corvallis, OR, 1984; James T, Callahan, "Long-TermEcological Research,'' Bioscience, vol. 34, No. 6, June 1984, pp. 363-367.

[20] David Greenland (cd.), ''The Climate of the Long-Term Ecological Research Sites, " Occasional Paper No. 44, Institute of Arctic and Alpine Research, University of Colorado, 1987.

formats. An electronic bulletin board could also be cost-effective for disseminating monthly data sets.

In sum, electronic information technologies both: 1) help make the forest and ecological monitoring system a reality, and 2) help make it possible for the monitoring results to be shared among the research community and other public- and private-sector users at little marginal cost to either the government or the users.

Energy Research Documents

Electronic information technologies also open up new possibilities for the dissemination of Federal scientific and technical documents that traditionally have been maintained in paper and microfiche formats. An estimated 200,000 such documents are generated annually, with more than half of the total originating from the Department of Energy, Department of Defense, or NASA.

Advancing technologies create new alternatives for electronic dissemination of both Federal STI bibliographic databases and the STI documents themselves. The activities of the DOE Office of Scientific and Technical Information are illustrative. DOE/OSTI currently distributes about 14,000 documents per year in paper or microfiche format to NTIS and in microfiche to the Depository Library Program (DLP). Abstracts of the documents are included in both the DOE bibliographic database called "Energy Data Base" and the NTIS bibliographic database. While the depository libraries receive paper copies of Energy Research Abstracts, which contain abstracts of DOE-funded research, the libraries have online access to the DOE and NTIS bibliographic databases only through private vendors at commercial rates.

To meet its own internal needs, DOE has implemented an Integrated Technical Information System (ITIS), which provides DOE employees and contractors with online access to the most recent 14 months of the Energy Data Base. DOE has proposed a pilot test to offer depository libraries similar online access. Besides timely access to the Energy Data Base (compared with the paper format Energy Research Abstracts), the pilot would provide an electronic "gateway" to archival energy research summaries (maintained on a database by a commercial vendor), and "electronic cataloging" of DOE documents in a format compatible with that used by depository libraries (and the Library of Congress) .21

Another aspect of the DOE pilot test is a study of alternative formats for document distribution. Over the next few years, DOE, like other Federal science agencies, has the opportunity to convert from paper and microfiche

> Advancing technologies create new alternatives for electronic dissemination of both Federal STI bibliographic databases and the STI documents themselves.

to optical disk as the primary document format. One possibility is to require DOE research offices, laboratories, and contractors to submit all documents in an electronic form (e.g., magnetic tape, online, diskette) that can easily be converted to high-density optical disks (e.g., WORM or CD-ROM). Since the demand for STI documents is generally small, any desired paper copies could be printed on demand. (The more popular documents could be printed in larger volumes with traditional printing processes.)

The study may show, as a hypothetical example, that DOE could distribute copies of the documents via a bimonthly CD-ROM, rather than on microfiche. A standard double-sided CD-ROM can store about 300,000 pages of material (double-spaced, typewritten) or about 1,500 documents at 200 pages per document. Thus the 14,000 documents could fit on about 10 CD-ROMs. The CD-ROM cost probably would be significantly lower than microfiche (and much lower than paper). At present, DOE pays about $350,000 per year for microfiche production of depository library materials, compared to an estimated $210,000 for mastering CD-ROMs and duplicating 1,400 copies of each (one per depository library). If DOE was able to piggyback depository CD-ROM duplication onto mastering and production for internal and possibly NTIS needs, the cost could be even lower (and savings greater). Compared to microfiche, CD-ROM should be easier to use, permit full-text searching, and provide higher quality document resolution (on the screen or when printed out on demand).

One disadvantage of using a bimonthly CD-ROM is the up to 2-month delay in getting some energy research documents to the depository libraries (and other users). This delay could be alleviated by maintaining the most recent 2 (or perhaps 4) months of documents online in full-text format, for retrieval and printing on-demand. Many private vendors are adopting a similar approach, which combines the strengths of online with CD-ROM formats. Another possible disadvantage is that all participating depository libraries (and other users) would need to have adequate CD-ROM facilities (one or several

[21] U.S. Department of Energy, Office of Scientific and Technical Information, ''DOE/Depository Library Gateway: Access to DOE R&D Results in Electronic Form, A Pilot Project Proposal,' August 1986; U.S. Congress, Joint Committee on Printing, "Dissemination of Information in Electronic Format to Federal Depository Libraries: Proposed Project Descriptions," June 1988.

microcomputer, CD-ROM drive, and local printer set-ups, depending on the level of use). As CD-ROM readers continue to drop in price and become standard equipment on microcomputers, the availability of CD-ROM equipment will improve, at least in the larger research libraries. Special provisions may be needed-whether through the DLP or otherwise-to ensure that smaller, rural, or economically disadvantaged libraries have CD-ROM equipment.

An inherent advantage of electronic formats such as CD-ROM is that powerful bibliographic, retrieval, and even expert search system software can be included directly on the optical disk or loaded into the microcomputer via diskette. CD-ROM or online versions of the "Grateful Meal" user-friendly software developed by the National Library of Medicine (NLM) will be commonplace, whether developed by the government and/or private vendors. NLM developed "Grateful Meal" to facilitate user access to MEDLINE and other databases on the NLM MEDLARS (MEDical Literature And Retrieval System). Tens of thousands of copies at $29.95 each have been sold through NTIS. The package includes 2 floppy disks, a user's guide, and an application for a MEDLARS access code. The capabilities of user-friendly software such as "Grateful Meal" or numerous commercial software packages can be easily replicated on CD-ROM.

In considering the appropriate role for Federal agencies in online dissemination of STI bibliographic databases, three aspects warrant particular attention. First, most of the Federal scientific and technical agencies have a statutory charter and/or mission objective to promote the wide distribution of information on the results of Federal research and development. Even agencies that operate under restrictions (e.g., NASA) have a strong dissemination mandate. Bibliographic databases are key tools in facilitating access to information on R&D results, and online databases (or for some purposes CD-ROM) offer significant advantages in terms of timeliness and ease of search and retrieval. Thus agencies need to be sensitive to equity of access to Federal STI, and ensure that, whatever means of online dissemination may be employed, certain user groups are not disadvantaged. Students, teachers, retired scientists, small business persons, and the like may need special consideration.

Second, development and dissemination of online bibliographic databases (and now CD-ROM versions of same) are strengths of the private commercial and not-for-profit information industry. A wide range of excellent STI bibliographic databases has been developed by private vendors that offer a portfolio of STI databases (including some from Federal agencies) over information gateways and value-added networks. Again, equity of access is a concern since full commercial online rates can range from $75 to $150 per hour or higher for privately developed databases, and commercial rates range from about $40 to $80 per hour for government databases (two to four times the comparable government rate). On the other hand, commercial vendors increasingly are proposing or offering a variety of discounts for off-peak or bulk volume use, that are more affordable for students, teachers, and the general public. Private sector not-for-profit vendors are providing some databases at rates between full commercial and governmental levels.

Third, a Federal STI bibliographic database may or may not be less expensive if offered online by the government. There is no clear-cut answer. Each situation requires individual analysis. For example, adding an online database to an already existing online computer capability (e.g., at NLM) or providing expanded access to an existing online system (e.g., depository library access to the DOE system) may have minimal marginal costs, if the existing computer center could handle the additional file and/or users without costly upgrades or expansion. In these situations, the incremental or marginal cost of additional computer use may be minimal, and competitive with comparable private-sector costs. On the other hand, if this required an upgrade of agency computer capability, the cost could be higher. For setting up a small electronic bulletin board, the cost of a new system is likely to be modest, but for a large, heavily used bibliographic database, the cost could be substantial. In making decisions on online bibliographic (or other online) systems, agencies will need to consider the quality of service, agency mission, equity of access, and related private-sector activities, in addition to cost-effectiveness.

With respect to CD-ROM (and other optical storage media), the situation is clearer. It seems likely that for some types of Federal information, and especially various STI documents, high-density optical storage will largely supplant paper and microfiche. It is not a question of whether this will happen, but when. Federal agencies will, in all probability, make this transition themselves in order to meet their statutory mission and records management responsibilities. The agencies may employ any of several means to make this transition, including private contractors, NTIS, and/or GPO. But the end result is likely to be the availability of many or most Federal STI documents on optical disk, at affordable prices, with powerful built-in search and retrieval capabilities, that will be cost-effective compared to paper or microfiche. This upgrade may also offer many new opportunities for the private sector to develop more value-added applications and products.