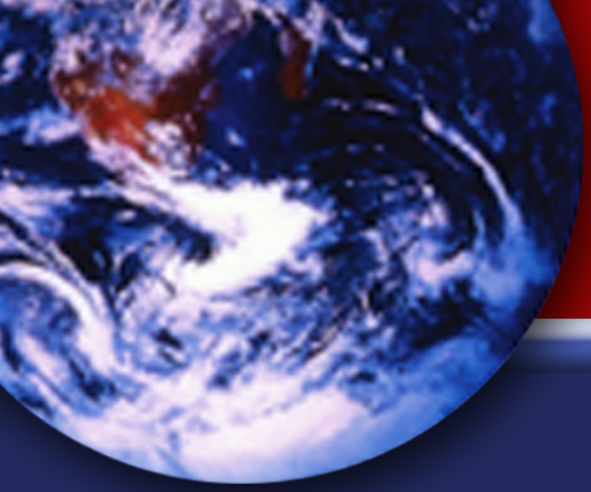
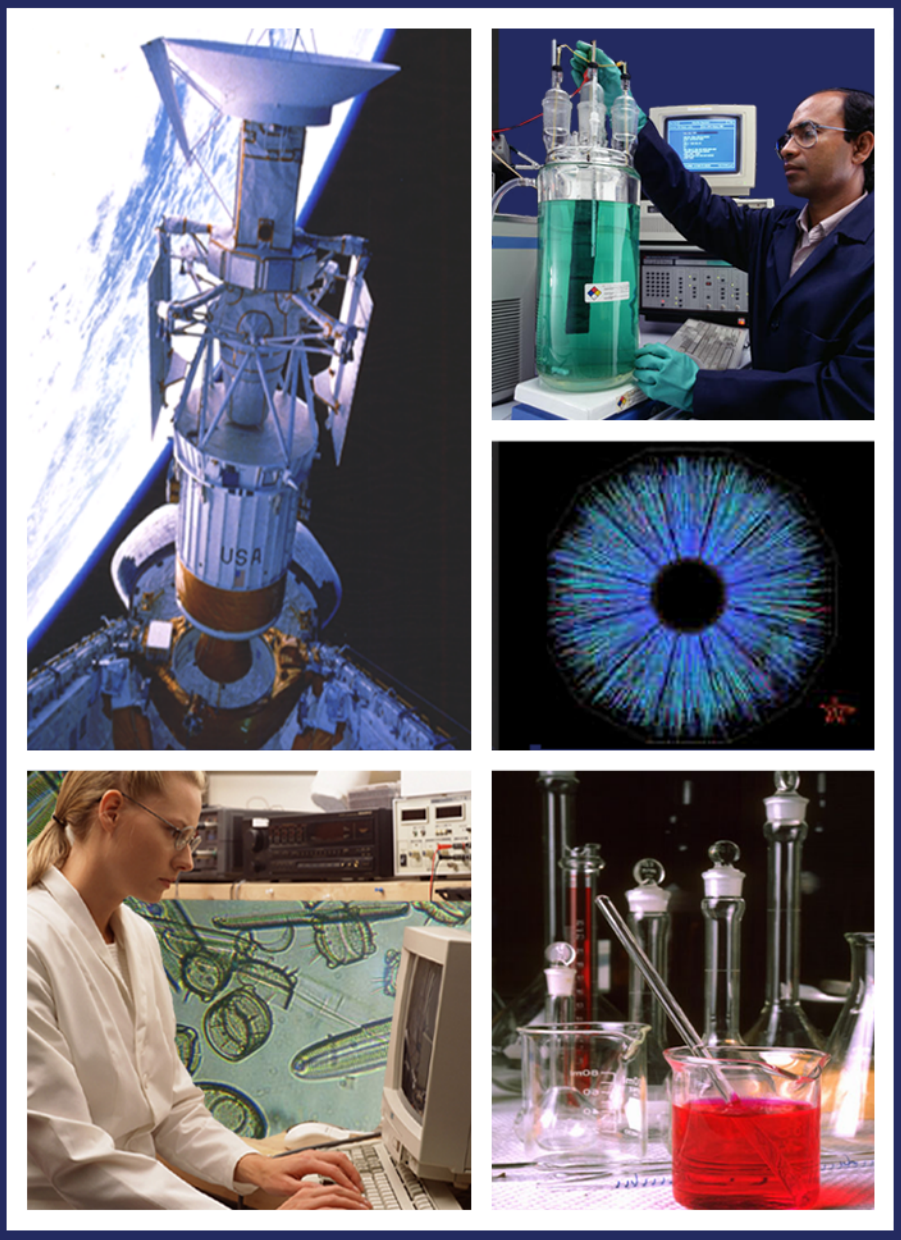


Strengthening the Public Information Infrastructure for Science

Report of the Workshop

Held April 18 - 19, 2001

At the National Institute of Standards and Technology



Issued by the Science.gov Alliance

Report of the Workshop

Strengthening the Public Information Infrastructure for Science

April 18-19, 2001

Federal agencies, academic experts, and other information professionals gathered for a Workshop April 18 and 19, 2001, at the National Institute of Standards and Technology (NIST) to examine how the public infrastructure for science information could be improved and how public access to science information of the Federal agencies could be enhanced. Over 60 participants from 35 organizations participated in the Workshop organized by the CENDI Information Managers Group, the University of Maryland Center for Information Policy, the Department of Energy (DOE), NIST, and the National Science Foundation and sponsored by DOE. This is a report of the Workshop.

Acknowledgment

The Science.gov Alliance gratefully acknowledges the foundation provided by this workshop through the enthusiastic participation and contributions of the Workshop organizers. The Alliance and the workshop organizers extend their appreciation to the speakers, panelists, and other participants who made this workshop an interesting and productive experience.

Workshop Organizers:
- CENDI* Information Managers Group
- Center for Information Policy, University of Maryland
- Department of Energy
- National Institute of Standards and Technology
- National Science Foundation

Science.gov Alliance:
- Department of Agriculture
  - National Agricultural Library
- Department of Commerce
  - National Institute of Standards and Technology
  - National Technical Information Service
- Department of Defense
  - Defense Technical Information Center
  - National Air Intelligence Center
- Department of Education
  - National Library of Education
- Department of Energy
  - Energy Library
  - Office of Science
  - Office of Scientific and Technical Information
- Department of Health and Human Services
  - National Institutes of Health
  - National Library of Medicine
- Department of Interior
  - United States Geological Survey
  - National Biological Information Infrastructure
- Environmental Protection Agency
  - Office of Research and Development
- National Aeronautics and Space Administration
  - Scientific and Technical Information Program
- National Science Foundation

*CENDI is an interagency working group of senior Scientific and Technical Information Managers from ten major programs in nine U.S. Federal Agencies: Department of Commerce, Department of Energy, Environmental Protection Agency, National Aeronautics and Space Administration, National Libraries of Agriculture, Education and Medicine, Department of Defense, and Department of the Interior.

Executive Summary

The Internet and World Wide Web have had a profound effect on the conduct and communication of science and on expectations of the science-attentive citizen. The role of Federal agencies is extremely important because they are significant sources of science information and are important sources of science information for education and research. Thus they have real resources of interest to the scientific community and useful to research universities, for example, as well as other audiences of government science information.

The Federal agencies, academic experts, and other information professionals who gathered April 18-19, 2001, at the National Institute of Standards and Technology examined the current environment and explored the means for improving public access to science information of the Federal agencies. Pursuant to the visions expressed at the May 2000 workshop on "Future Information Infrastructure for the Physical Sciences" and numerous recent initiatives (such as the February 2001 PITAC report, "Digital Libraries: Universal Access to Human Knowledge"), this Workshop was convened to examine the emerging new concepts for providing access to science information.

The context for science information is changing. Developments include the National Science, Mathematics, Engineering, and Technology Education (SMETE) Digital Library (NSDL) Program at the National Science Foundation, the organization of private-sector resources around SMETE.ORG, the new administration's policies on e-government, the success of FirstGov, the growth of open archives within the academic research community, and the emergence of technologies for improving access to materials on the Web. The transformation of the science information infrastructure is still in an early stage, but a rich, distributed network of science resources is already coalescing around the World Wide Web.

Federal agencies at the workshop recognized that the building blocks are available, and now the time is right for an interagency science information infrastructure. Agencies can make digital information accessible no matter where the information resides physically, so that no classroom, group or person is ever isolated from the world's greatest knowledge resources and researchers will have enhanced access to data and scientific information. Also agencies can take full advantage of the Internet and other technologies to overcome arbitrary boundaries between agencies, so that the government can provide the public with seamless, dependable online services.

Agency representatives at the Workshop agreed to work together on an interagency science information initiative, called "Science.gov." This science portal would strengthen the science information infrastructure and integrate various agencies' information content, tools, and technologies into a borderless digital resource available to researchers, teachers, and learners wherever they are located.

To that end, twelve agencies formed a "Science.gov Alliance" to work on the interagency science portal as a unified navigation path to science conducted by the government.

As a result of this workshop, the formation of the Science.gov Alliance is the first step toward achieving these goals for science information of government agencies.

An organized, comprehensive gateway to science information would provide a coherent government

R&D presence on the web and streamline the process of identifying and accessing government science information. In particular, a science gateway would provide these additional benefits:

- Foster greater and more accurate public understanding of the government's science and technology contribution to the U.S. taxpayer.
- Help users navigate through Federal science collections and resources, quickly and accurately.
- Provide the public with well-organized information, from practical information for the consumer to highly technical scientific data for the bench scientist.
- Support future scientists and engineers with information and science education resources.
- Raise scientific and technical literacy.
- Ensure use of scientific research as a foundation for future discoveries.

THE VISION

The improvement of understanding is for two ends: first, our own increase of knowledge; secondly, to enable us to deliver that knowledge to others.

-- John Locke

Over the years, the United States has developed a tradition of leadership in science and technology that has made it a model for nations throughout the world. Much of the fundamental scientific research in the U.S. is funded by the U.S. government, which is an investment of about $80 billion annually. The Federal government's research has made and continues to make a significant difference in the lives of all Americans. In recognition of this tremendous investment and resource, Federal agencies, academic experts, and other information professionals gathered April 18 and 19, 2001, at the National Institute of Standards and Technology, for a workshop to examine how the public infrastructure for science information could be improved and enhanced, the Workshop on "Strengthening the Public Information Infrastructure for Science."

Clearly, the Internet, developed by U.S. research agencies, has had a profound effect on the conduct and communication of science, and on the expectations of the science-attentive citizen. The growing wealth of information and related technology is impacting lives and industry as never before, at a time when half of American adults use the Internet every month, and a tremendous 73% of U.S. children under the age of 18 are now online.[1] Given these statistics, at any given minute in a day, the need for science information is tremendous.

- Mark is a ninth-grade student and, like many young students, he is fascinated with space, the known and the unknown. When he was recently assigned to research Mars, many of the resources he was able to obtain in his school or local library were several years old and he knew there had to be information on current research. But how could he find it?

- Dr. Hunter is a researcher at a National Laboratory. Recent research findings at her laboratory have given her new insight into a key problem in her research area. How can she find out about recent advances at other laboratories, and what opportunities for additional funding exist?

- Joe, an Idaho farmer, is asked to reduce the amount of land he farms in order to cut back on electricity consumption because Idaho's electrical generation using hydropower has been impacted by reduced snow pack. Where can he find information about renewable energy sources, such as wind power, to produce the power needed on his farmland?

- Jeff's life has been turned upside-down in recent years as he has faced numerous medical decisions regarding treatment and options. One thing he and his family have learned is the need to stay current on the latest research and the importance of being informed when decisions must be made and questions asked. But is he aware of the various resources?

What is the common thread that is at the foundation of these needs? **Science information**

Who is sponsoring scientific research which impacts these areas? **The Federal government**

What is the source for ready access to information in these areas? **The Internet**

In fact, when you multiply these few examples by just a fraction of the millions of people that turn to the Internet each day for science information, information that is available through government-sponsored web sites, the potential of an enhanced public information infrastructure for science comes into focus.

Impact of the Internet

Children today are growing up in the Internet age - soon many will never know what it is like not to have the Internet, just as most of us have never lived without electricity and the associated inconveniences. While a growing number of individuals turn to the Internet each and every day as a way of life, few know the chronology of events which birthed the Internet as it is known today.

The Internet was initially conceived over 30 years ago to support research. The Internet was first conceptualized in the early 1960s by the Department of Defense Advanced Research Project Agency (ARPA). The result was a small network called ARPANET, named after its Pentagon sponsor, intended to allow scientists and researchers to share data and access remote computers in the United States.

In the early 1970s, the ARPANET became a "high-speed digital post office"[2] as electronic mail was the most popular application for users who collaborated on research projects and discussed topics of various interests. In the late 1970s and early 1980s, the commercial version of the ARPANET went online and the general public got its first insight into the potential of networked computers.

Following the creation in the mid- to late-80s of TCP/IP, the common language of all Internet computers, the collection of networks which made up the ARPANET was seen as an "internet." This came around the same time as the boom in the personal computer and super-minicomputer industries, allowing many companies to join the Internet for the first time.

In 1984 the National Science Foundation played a key role in the Internet through its Office of Advanced Scientific Computing. The new NSFNET set a new, faster pace for technical advancement, linking newer and faster supercomputers, through faster links, links which were continually upgraded and expanded. Other agencies followed suit and jumped in: the National Aeronautics and Space Administration, the National Institutes of Health, and the Department of Energy.

By 1987, the number of Internet hosts exceeded 10,000; a number which would grow beyond 300,000 in three short years by the time the ARPANET was decommissioned, leaving the growing Internet - and the World Wide Web is born.

The nineties continued to bring major advancements, including the creation of Web browsers, Internet programming languages, electronic commerce, search engines, and phenomenal growth of traffic on the Internet. Bruce Sterling wrote in his "Short History of the Internet," published in 1993, "The Internet is especially popular among scientists, and is probably the most important scientific instrument of the late twentieth century. The powerful, sophisticated access that it provides to specialized data and personal communication has sped up the pace of scientific research enormously."[3]

Move ahead almost ten years later and this statement still rings true now. What is still needed is to open up the wealth of science information available on the Internet not just to the important research community, but to the "science-attentive citizen."

The Workshop on "Strengthening the Public Information Infrastructure for Science" represents a major step in achieving this goal.

PURPOSE OF THE WORKSHOP

Fulfilling the Call

Last year, the need for a comprehensive collection of science information easily available to researchers and students was recognized at the Workshop on a Future Information Infrastructure for the Physical Sciences[4] in May 2000, chaired by Dr. Alvin Trivelpiece, where views on the feasibility were heard from experts in the physical sciences and in science communication. In fact, the need has been well documented in a range of studies since the 1940s.

< 1945	Vannevar Bush Report to Roosevelt on science accessibility
< 1958	Humphrey recognizes Information Age; Eisenhower issues plan
< 1960	COSATI established
< 1963	Weinberg Report: "Science, Gov't, & Information"
< 1965	Licklider forecasts electronic publishing
< 1976	NSF suggests Federal government ensure scientific communication
< 1983	John Creps, Jr., describes vision for library of the future
< 1989	NAS recommends an interconnected national information technology network
< 1991	Loken Report calls for development of a National Physics Database
< 1994	AAU task force examines new options for collection and dissemination of STI
< 1999	PITAC issues information technology report on future directions
< 2000	Trivelpiece Report endorses Physical Sciences Information Infrastructure (PSII)

One of the earliest studies was the 1945 report[5] of Dr. Vannevar Bush, Director of the Office of Scientific Research and Development, to President Roosevelt, which called for scientists to make the vast store of knowledge more accessible and thus extend man's physical and mental power. Bush saw great potential for focusing scientific knowledge in a new direction and developing instruments to give command over information. He also noted that scientific progress was essential for the good of the country and that science was a proper concern of government.

Today, information technology has raised the expectations of researchers for immediate, online access to information in the physical sciences. Panelists and participants of the May 2000 workshop believed that the researchers' expectations could be met by deploying current technology to provide an integrated network of dispersed resources. The report from the workshop, referred to as the "Trivelpiece Report," gives a high-level vision for a comprehensive "Physical Sciences Information Infrastructure." The findings from that workshop set the foundation for "Strengthening the Public Information Infrastructure for Science."

May 2000 Workshop Findings Support Need for:

- **A common knowledge base** - to provide integrated, comprehensive access and to facilitate reuse of resources, regardless of where they reside.
- **A point of convergence** - to ensure awareness, availability, use, and development of information technologies and tools.
- **An openly available source** - to serve all users, from students to scientists to concerned citizens, in a highly efficient electronic environment.

Laying the Groundwork

The purpose of the "Strengthening the Public Information Infrastructure for Science" Workshop was to examine how the public infrastructure for science information could be improved and enhanced. Those gathered at the Workshop sought specifically to identify ways to improve public access to science information of the Federal agencies.

The first day of the workshop was an open meeting that reviewed recent developments in the changing context for science information. Developments include the National Science, Mathematics, Engineering, and Technology Education (SMETE) Digital Library (NSDL) Program at the National Science Foundation, the organization of private-sector resources around SMETE.ORG, the new Administration's policies on e-government, the success of FirstGov, the growth of open archives within the academic research community, and the emergence of technologies for improving access to materials on the Web.

On the second day of the workshop, federal agencies convened to identify and evaluate strategies for strengthening the information infrastructure within and across existing science agency programs. A significant gathering of over 20 government organizations from 13 different agencies, all of which had a stake in the public information infrastructure for science, participated in the forum. Individual agencies have worked to respond to new opportunities and constituency needs.

The agencies with science information are uniquely positioned to strengthen the public information infrastructure for science. The time is right, the momentum is growing, and an adequate number of studies have been done. This realization along with the convergence of the digital landscape, e-government initiatives, citizen interests, and technological capabilities moved the agencies to work together on an action plan that would bring the vision to reality.

Defining the Science-Attentive Citizen

The prolific use of the Internet and current e-government initiatives have raised expectations by the public to have immediate access to full and open information. Web search services are the most widely used information discovery tool in universities today. Users are becoming more familiar with search terminology and more "savvy" on how to use the Internet.

The public is also becoming more attuned to the impact of science on their everyday lives. Thus, citizens are more likely to look to the government for answers to their science and technology questions and to seek information from the government related to issues such as health, the human genome, space, defense, energy, food, the environment, and other science-related topics.

In addition to the researcher who needs access to scientific information and to the public seeker of science information described above, a range of other people require easy access too – including students, teachers, engineers, entrepreneurs and product developers, policy makers, and others. For the purposes of science.gov, these all are included in the definition of "science-attentive citizen."

THE WORKSHOP

DAY 1: COLLABORATIVE FORUM -- Laying the Groundwork and Exploring the Changing Context for Science Information

The workshop was opened by Dr. John Rumble, who welcomed the participants and highlighted the 100-year anniversary of NIST. He noted that, as agencies collectively look forward, there are many opportunities to make a difference because the information revolution has far reaching impacts. He pointed out NIST's role in the critical evaluation of data, much of which is found in gray literature and abstracts, and there remains a key need to access U.S. as well as worldwide information. A scientific and technical information (STI) infrastructure is important not only for NIST, but could open up a new era on how NIST approaches information in its second 100 years.

Following Dr. Rumble's welcome, Brian Kahin of the University of Maryland's Center for Information Policy provided an overview of the workshop and the planned strategy. He highlighted three areas for the participants' consideration:

- The non-commercial infrastructure which is in place and evolving.
- The user orientation approach, which implies that it is less about the collections but more about providing services to a community of users.
- The opportunities for leverage through an open vernacular infrastructure.

He noted that the role of Federal agencies is extremely important and they represent real resources, resources that are critical for universities. Agencies have a key role in announcing university research as well.

The first speaker, Dr. Bill Arms of Cornell University, presented an interesting depiction of the digital library landscape and reviewed current trends.[6] In regard to primary information, an underlying trend is that every year sees an increase in the proportion of important information that is available with open access and there is an increase in the proportion of important information that is available online. He cited several examples illustrating the changing landscape, including web sites for approximately 1/3 of the courses taught at Cornell and sites such as the Physics Preprint which has transformed research in physics. Within the government there have been marked changes as well. For example, the National Academies' work is now widely known and even the Library of Congress has reached new audiences. Indicative of the trend is the Public Library of Science initiative, which calls for the establishment of an online public library that would provide the full content of the published records of research and scholarly discourse in medicine and the life sciences in a freely accessible, full searchable, interlinked form. Clearly, the forces for open access are strong, and increases in available information will continue to increase. A key question now is, "How can we make use of it?" Lending support for this important question is a recent study by JSTOR that showed there is wide use of the Internet not only in physics and computer sciences as one might expect, but increasingly the Internet is used to access

information pertaining to the humanities and social studies.

While access to primary information has increased, Dr. Arms also noted trends dealing with secondary information. One trend dealing with the information discovery process is that web search services are the most widely used information discovery tools in universities today. Certainly one key reason for this is the speed that is possible. Yet, while speed is indeed a factor, it is recognized that the open access information is sometimes a poor substitute. In addition, clearly much good information is not available with open access.

One dilemma that has emerged deals with economics and the overall ability to compete with a free good, which in turn may impact vulnerable library budgets and publishers' revenues. Dr. Arms presented four economic models which come into play, and he concluded that in regard to scholarly information, the dominant force is author pressure, which emphasizes open access rather than closed access. As a result, a mixture of economic models will coexist, and eventually, there will be open access to most scientific, government and professional information. The most common economic model has information published by the producing organization.

Dr. Arms concluded with the observation that before the web, few people had access to scientific, medical, government and legal information; and now, with the web, much high quality information is available with open access and low costs services can organize this information and provide open access to it.

Discussion at the Workshop then examined the context for infrastructure development. Brooke Dickson, Information Policy and Technology, Office of Management and Budget (OMB), provided insight into current e-government initiatives and the new Administration's commitment to e-government[7] and moving government online, as evidenced by the April 9th release of the U.S. budget. Specifically, there were four major e-government related areas in the budget: FirstGov.gov, the Government Paperwork Elimination Act (GPEA), E-government Fund, and the Public Key Infrastructure (PKI) Bridge. Ms. Dickson provided details on these four areas and the current status of each. She concluded by stating that while e-government may be the wave of the government's future, there remains the key areas of security, privacy and the Government's information policy as stated in OMB Circular A-130. These remain important in the electronic as well as the paper world.

Tom Freebairn, General Services Administration (GSA), provided additional information and background on the establishment of FirstGov and discussed its role in providing an essential government service. FirstGov was funded in the year 2000 by the Federal Chief Information Officers Council and 22 other federal agencies. President Bush's proposed 2002 Budget recognizes the role of FirstGov in cross-agency electronic government. The Budget Blueprint calls for the use of the Internet to create a citizen-centric government and provides a fund that will grow to $100 million over 3 years to support electronic government. FirstGov.gov is recognized as one of the "essential building blocks" for projects that will operate across boundaries. Currently, the FirstGov initiative is focused on the development of cross-agency portals that address specific user groups and topics, such as seniors.gov and workers.gov.

With the current context of infrastructure development examined, several speakers then addressed institutional strategies. Lee Zia, National Science Foundation (NSF), provided an overview of the National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) program.[8] To stimulate and sustain continual

improvements in the quality of science, mathematics, engineering, and technology education (SMETE), NSF launched NSDL. The resulting digital library is intended to serve the needs of learners belonging to a broad user audience. Discussion focused on LEARNS: a Learning Environment and Resources Network for SMETE. As the name implies, LEARNS is designed to meet the needs of learners, in both individual and collaborative settings. It is actively managed to promote reliable anytime, anywhere access to quality collections and services, available both within and without the network, and is constructed to enable dynamic use of a broad array of materials for learning. LEARNS connects users, content and tools and thereby supports learning communities, customizable collections and application services. The various NSDL program tracks that accepted proposals were reviewed. These are:

- Core Integration System Track - projects are expected to focus on the coordination and management of the library's core collections and services and to develop the library's central portal;
- Collections Track - projects are expected to aggregate and manage a subset of the library's content within a coherent theme or specialty;
- Services Track - projects are expected to develop services that support users, collection providers; and
- Targeted Research Track - projects are expected to explore specific topics that have immediate applicability to one of the other three tracks.

Zia highlighted various issues and questions being addressed throughout the process.

Walt Warnick of DOE's Office of Scientific and Technical Information (OSTI) provided a concept for achieving universal access to science information,[9] thereby moving toward the science part of the PITAC (President's Information Technology Advisory Committee) vision of "universal access to human knowledge."[10] He began by confirming the shared mission among many of the participants to disseminate scientific and technical information. He noted the shared premise that the web is today's tool of choice and that interagency collaboration is needed. One underlying fact is that science is not bounded by organization or geography. Dr. Warnick reviewed the overwhelming support for an interagency science portal, which is not only evidenced by the PITAC reports and the Administration's expectations for e-government, but also by many of the ongoing initiatives being discussed at this Workshop and other historical studies including the 2000 Trivelpiece Report. Examples of interagency collaborations that came about as a result of the Trivelpiece Workshop in 2000 and broke new ground regarding interagency collaborations were highlighted. These collaborations, **GrayLIT Network** (http://www.osti.gov/graylit) and **Federal R&D Project Summaries** (http://www.osti.gov/fedrnd), provide improved access to scientific and technical research information across several Federal agencies. A key benefit resulting from these tools is the capability to search documents with a single query across databases of many Federal agencies to find and combine information regardless of where it happens to reside. Therefore, with these new tools, it is no longer necessary for a user to know which agency is working in a particular area or discipline. **GrayLIT Network** provides a portal for over 120,000 full-text technical reports located at the Department of Energy, Department of Defense, Environmental Protection Agency (EPA), and National Aeronautics and Space Administration (NASA). **Federal R&D Project Summaries** includes more than 300,000 research summaries for three of the major sponsors of research in the Federal government. The Federal databases

available via this tool are the Department of Energy R&D Project Summaries; the National Institutes of Health (NIH) CRISP (Computer Retrieval of Information on Scientific Projects) Current Awards; and the National Science Foundation (NSF) Award Data. Dr. Warnick introduced the DOE-developed model for science.gov through which universal access to science information could begin taking shape.

Building the SMETE.org Alliance was the topic of the next presentation.[11] Brandon Muramatsu, SMETE.ORG/UC Berkeley, highlighted efforts to build a successful NSDL for deployment in Fall 2002 to focus on science, mathematics, engineering and technology, and most importantly education. Towards this end, a team has been developed to not only overcome various challenges, but to cover target audiences and disciplines and to share in the development efforts. The result is SMETE.ORG (http://www.smete.org/), an alliance of over 20 partners - industry and other collaborators - which provides seamless access through a tightly coupled federation of educational digital libraries. Design principles are in place to address information organization, labeling, navigation and searching; and various partnership models are at work in SMETE.ORG. The alliance works well together, but also activities do not detract from what partners have done individually. In fact, the strengths of the partners are a key to the success. The partners with existing collections each have a decade of experience providing digital SMETE resources to their target audiences and disciplines, and most partners each have more than ten years of experience as organizations promoting SMETE reform, and organizations serve the full spectrum of audiences.

Prior to continuation of presentations focusing on specific institutional strategies, Daniel Greenstein, Digital Library Federation, provided a perspective of "The Digital Library as a Community, not a Repository."[12] He described the varied functions of the digital library and provided perspective on key challenges which he views as drivers. Key challenges include technologies and their use, bench marking, digital preservation, mobilizing expertise, institutional obstacles, and issues with collections and users. For each challenge, he highlighted the basic problem area, and elaborated on promising directions and "logical trajectories" pertaining to each. Examples were given for a number of current Digital Library initiatives that are up and operating, representing various program models. It was Greenstein's opinion that assembling the various agencies at the Workshop was promising, since it is important to mobilize the scarce capacity that may exist individually. It was noted that it is important to understand how information is being used, and to not operate in a vacuum.

A second panel of speakers from several government agencies then continued the focus on institutional strategies. Eleanor Frierson of the National Agricultural Library (NAL) discussed Institutional Strategies for Partnerships.[13] NAL is currently involved in a variety of partnership arrangements, including collaboration with land grant universities, and projects and activities with other national libraries, other federal libraries, and international organizations. She shared with the workshop participants lessons learned through these various partnerships and stressed the importance of flexibility, noting that there are many solutions to problems and challenges.

Successful partnerships - for any type of business enterprise - require:

- "Wins" for all partners,
- A good fit between mandates and partnership goals,
- Energetic leadership and an honest broker to execute activities,

• Open communication among all members,

• Stated objectives agreed upon by the partners, and

• Recognition of individual contributions.

These qualities are also key to forming a successful interagency alliance for science information. Interagency collaborations are a model for interdisciplinary or cross-agency research projects, and the lessons learned could well be applied to the development of a science information infrastructure. One successful partnership in place is AgNIC, a voluntary alliance of NAL, land-grant universities and other agricultural organizations, in cooperation with citizen groups and government agencies. AgNIC (http://www.agnic.org/) provides agricultural information in electronic format. There are currently 40 partners, and member participants take responsibility for small segments of agricultural information (including basic, applied, and developmental research, extension, teaching activities, etc.) and develop Web sites and reference services in specific subject areas. Members agree on stated benefits, and the collective AgNIC Web resource benefits all members in ways that the individual members cannot achieve alone, thus justifying the local costs of participation. Each official institutional member of AgNIC commits to certain activities in support of the Alliance and its work.

Gladys Cotter of the U.S. Geological Service also provided useful information on a partnership model in her discussion of the National Biological Information Infrastructure (NBII).[14] The NBII (http://www.nbii.gov) is a broad, collaborative program to provide access to data and information on the Nation's biological resources and tools for integration and analysis. It links biological databases, information products, and analytical tools maintained by NBII partners and other contributors in government agencies, academic institutions, non-government organizations, and private industry. From the outset, the NBII focused on community building for collaborative development of standards, content and technologies. The coalition of partners was built via interagency collaboration, regional nodes, cross-disciplinary support, and involvement by a range of participants. The BioEco Working Group provides a U.S. focal point for interagency, intergovernmental and international cooperation and a forum for targeted activities. Through the NBII, not only is access provided to a great variety of biological data and tools from multiple sources, but participants have a structured way to showcase, share and exchange the biological data and tools they are producing - thereby benefitting both the users and the providers.

"Toward an STI Network" was the title of a presentation given by Kurt Molholm, Administrator of the Defense Technical Information Center (DTIC).[15] DTIC provides scientific and technical (S&T) management support and assists in collaboration, as well as providing S&T results and points of contact to researchers. Through DTIC, researcher efficiency is promoted as well as enterprise effectiveness, and the focus is on delivery of information content. Molholm provided examples of customer-oriented products that serve DTIC users, including the public and the secure STINET service. Other customer oriented tools include the S&T Collaboration Tool, which provides a paperless workflow management site for document creation, review and revision, allowing approximately 200 geographically dispersed users to collaborate online; and the Virtual Technology Expo (VTE) which provides information on emerging technologies and showcases research efforts to a broader audience. Molholm also provided information on the Department of Defense (DoD) sponsored Information Analysis Centers

(IACs). There are currently 13 IACs covering all of the DoD critical technology areas, with the purpose of collecting, analyzing, synthesizing, and disseminating worldwide information in their defined field or subject area. They also promote standardization within their respective fields. Additional information is available at http://iac.dtic.mil/. In closing, Molholm reviewed areas that require ongoing attention and necessitate a balancing act. These are: resource allocation, risk taking, the digital divide, risk management, and citizens versus hackers.

Following presentations on some of the leading scientific information programs in the Federal government, Brain Kahin wrapped up the day's discussion by summarizing some of the opportunities and issues of the day.

DAY 2: INTERAGENCY WORKING MEETING -- Developing an Interagency Strategy for Implementation: The Resources, Opportunities, and Challenges to Develop an Interagency Science Portal

The second day of the workshop was planned as an interagency meeting to identify potential ways to strengthen the information infrastructure within and across existing science agency programs. Certainly science mission agencies have been and continue to be impacted by the quickly changing landscape. Having heard the landscape and issues described by key experts on the preceding day, the interagency meeting provided an opportunity to reflect on and discuss the concept of an national infrastructure in the context of the new Administration's priorities and current calls to further the e-government initiative.

The meeting was kicked off by Bonnie Carroll, Information International Associates, who reviewed the reaction to day 1 and put the information in context across agency lines.[16] Each of the science agencies now offers information and services on the Web. There is increasing pressure on government institutions to provide access through this medium. Most activities are agency-centric, which is proper where organizations were created and funded first and foremost to support an agency-specific mission. However, science is not bounded by organization. Thus the topic or thematic approach, providing borderless pathways to science information wherever it resides without the consumer needing to know where it resides, is a worthwhile pursuit for interagency collaboration. Based on a review of day 1, it was evident that the building blocks are available and the assets that could be rallied are enormous.

While individual agencies have worked together in recent years on various initiatives and specific projects, some under the auspices of CENDI, this Workshop was considered by many as a key turning point in advancing a governmental collaboration for an information infrastructure for the sciences. It was noted by some attendees that this was the broadest recorded representation among the science mission agencies regarding technical information issues.

Several agency representatives provided information on other agency-specific initiatives that demonstrate examples of capabilities to be brought to the table. Each of these capabilities presented a unique area of accomplishment and were presented as examples of agency contributions to an interagency effort.

Karen Holland of NAIC provided an overview of Machine Translations. As an organization with responsibilities in intelligence, there has been a long standing need for rapid post editing and translation in order to feed data to analysts. Focus areas have shifted over the years, but a machine translations process has been established to meet the needs of the NAIC community. There are clearly growing and universal applications for such a capability in the science community.

Wally Finch, Associate Director for Business Development at the National Technical Information Service (NTIS), provided information on the NTIS Science Portal Initiative.[17] NTIS, as part of the U.S. Department of Commerce's Technology Administration, was established to collect and provide permanent access to the Nation's scientific, technical and engineering information (STEI). As with many Federal agencies, NTIS has faced the challenges in the Internet era of ensuring permanent access and searching for and finding relevant information. NTIS is addressing these challenges. Consistent with recent advancements in its operating systems, NTIS has addressed the need for a

persistent digital identifier by implementing the Handle System, which assigns, manages, and resolves persistent identifiers for digital objects in multiple locations. In the upcoming months, a new database will also be launched, www.scitech.gov, which will provide a database of key STEI resources: expertise, facilities, R&D results, government inventions, and photos and images. This is envisioned as a primary contribution to a cross-agency science portal.

In follow-up to Dr. Warnick's presentation on day 1, Karen Spence of DOE's Office of Scientific and Technical Information (OSTI) presented the science.gov model.[18] The model represents the culmination of progress since the Department of Energy hosted the May 2000 Workshop, which endorsed the vision of a national infrastructure for the sciences. Following that workshop, OSTI began a dialogue on the nature of a "science.gov" collaborative site and the potential forms it could take. In order to further the discussion at this Workshop and to offer a forum for feedback, OSTI established a shell site at www.science.gov, which includes a representative sampling of the types of information that could be included. Spence provided a demonstration of the site, highlighting the potential functionality. While the DOE site is considered a model, it integrates the distributed search tool used in other DOE products, which is easy to use, utilizes parallel searching, and retrieves information from heterogeneous and geographically dispersed databases and Web sites, thereby providing a tremendous tool for cross-agency initiatives.

Following these presentations of examples of agency contributions, two groups met in concurrent breakout sessions: one addressing content sources, scope, and coverage with Lee Zia, NSF, serving as moderator; and the second group charged with addressing tools and technologies, and chaired by John Rumble of NIST. The first group was asked to address audiences, technical disciplines and interrelationships, and the types of material/information available. The second group was asked to consider the following topics pertaining to tools and technologies: archives, distributed search tools/deep web applications, intelligent agents, metadata management tools, special services (push technologies; alert services), and web mining. Presentations of the groups deliberations were then made to the larger group.

Throughout the discussions, agency representatives clearly recognized the tremendous value of making science information resources more accessible. When the question, "What if we don't do this?," was proposed to the group, the general consensus was that doing nothing is not an option. The overall value of enhanced access was not disputed. With that question laid to rest, several key points emerged from the agencies' deliberation.

EMERGING KEY POINTS

< Integrate Science Assets

"Individual agencies have enormous information assets"

Agencies individually bring enormous assets to serve as the building blocks for an integrated information infrastructure. In recent years, agencies have and continue to utilize information age technologies and, in some cases have radically changed their information services to bring science information to the desktop. In addition, new tools and capabilities have evolved to facilitate the use of digital information.

< Serve Science-Attentive Citizen

"The 'science-attentive citizen' may be served"

A fundamental question was foremost in the discussion, "Who are we working to serve?" Of course, each agency operates under specific mandates and guidelines, and information organizations work to fulfill their respective agencies' diverse missions. However, while the agencies' missions may be diverse, science agencies share a specific responsibility for disseminating scientific information; i.e., the result of Federally funded research, in easily accessible and useful ways that will benefits the users of the information.

Agencies currently serve the public through various products and services dependent on their mission and guidelines. However, for a large segment of the public, access to much government science information available on the Web is something that requires effort and understanding of the various agencies involved in order to access their sites or information outlets. As the public increasingly is attune to using the web, what is needed is an "intuitive" access point on the Web for government science information. This approach is consistent with the FirstGov approach to move toward subject portals or gateways for information access. By working together under an overarching framework, agencies have an opportunity to effectively conceptualize government science information for the public, and to provide information through a science gateway based on topic areas rather than have information bounded by organizational lines. Simply put, this is an opportunity to say to the science-attentive public, "Your government does a lot of science!"

< Increase Visibility

"Opportunity to increase visibility for agency holdings and contributions to science"

Clearly, working toward a single gateway for government science information presents a tremendous opportunity to increase visibility for agency holdings, and thereby facilitate access to tax-payer funded resources and information. Even among the attendees who were knowledgeable of government science resources, much was learned about several agencies' current capabilities and initiatives. This point alone underscores the value of establishing an infrastructure that will facilitate access and visibility, not only for the public but across the Federal research base.

Thus, an important outcome of an information infrastructure for science is the increased awareness across the Federal government, thereby allowing researchers to take advantage of what agencies already do for their own purposes.

< Meet National Policy

"Responds to call for national policy"

Recent calls for increased collaboration and for furthering access to government information directly coincide with the objective of the Workshop. For example, the February 2001 report of the PITAC Digital Libraries Panel stated, "the Federal government can do much more by creating digital libraries faster, improving the access to digital content by the many people who today cannot avail themselves of it, and adopting the aggressive and visionary goal of providing digital content to every citizen."[10] PITAC has also called for transforming how we conduct research, with scientific and technical information being an integrated part of the research process.[19]

< Serve Agency Constituents

"Agencies will continue to serve their primary audience - their constituents"

The science mission agencies represented at the Workshop encompassed the spectrum of science-related disciplines and are in the forefront of the world's research in areas as diverse as human genome, physics, entomology, agricultural engineering, and a range of other science disciplines. Each organization is funded first and foremost to support their respective agency mission, and this was strongly confirmed among the participants. In addition, these organizations are the experts on their information, with full knowledge and experience on how to structure information for their own constituents. In whatever framework is established, there should not be an attempt to pre-determine for the agencies what information or information tools should be encompassed under a single gateway. Rather the gateway concept will depend on the expertise and the knowledge of the organizations involved to make those determinations. Thus, one of the main points clearly evident among the organizations is that working together to establish an information structure for science should in no way detract from an agencies' primary constituents. Yet it was recognized that, in many instances, a science infrastructure can significantly strengthen the base of information available for their own constituents.

< Deliver Science Resources

"Bring all government science resources to the broader science community"

While it was recognized that an information infrastructure for science can strengthen the base of information available for the constituents of the various agencies, it was also noted that a collective effort can bring all government science resources to the broader science community. This broader science community encompasses not only researchers in the private sector, but universities and other educational institutions. This is especially needed at a time when the boundaries between scientific disciplines are increasingly less defined, in terms of advancing our understanding and solving scientific problems. The potential for cross disciplinary research and discovery is a key driver in supporting a cross-agency collaboration.

< Broaden Science Community

"Use this meeting to build a broader community of agency participants"

Attendees at this meeting represented over 20 Federal information organizations, each tasked with specific responsibilities as well as unique knowledge and skills. Many of the organizations present were active participants in CENDI,[20] which encompasses ten organizations from nine U.S. Federal Agencies: Department of Commerce, Department of Energy, Environmental Protection Agency, National Aeronautics and Space Administration, National Libraries of Agriculture, Education and Medicine, Department of Defense, and Department of the Interior. However, this meeting included representatives from additional organizations, and the phrase "CENDI+" was used to characterize the organizations represented. An important outcome of the meeting was the opportunity to make new

contacts and foster new relationships to begin building a broader community. Also, while the representation was significant, there are certainly other organizations within the Government which an information infrastructure for science would encompass, but were unable to attend. This Workshop was certainly an important step in formalizing next steps, steps that can then be communicated to other key organizations, which in turn will strengthen the power of this collaboration.

CONCLUSION AND FINDINGS

Following the deliberation which resulted in the key points noted above, it was evident that the building blocks exist to begin to form a public information infrastructure for science and that there are clear benefits in this interagency collaboration. The questions which now presented themselves were, "What decisions should be reached at this Workshop," and "What next steps should be taken?"

A principal representative from each agency participated in a focused discussion to work through these key questions.

This focused discussion resulted in the following findings:

- The concept of an interagency science gateway whereby the agencies would collectively serve the science-attentive citizen was unanimously endorsed by each of the agencies present.
- This cross-agency science gateway will be referred to as Science.gov.
- In the near-term, this gateway will begin with the basics and facilitate links to current resources.
- Mid-term goals include additional features and improvements to more fully unify science information of each of the participating agencies.
- Twelve agencies agreed to form the Science.gov Alliance.
- A Core Team of the Science.gov Alliance agreed to meet as soon as possible to outline an action plan. The Core Team includes: Department of Agriculture/National Agricultural Library, Department of Energy/Office of Scientific and Technical Information, Environmental Protection Agency/Office of Research and Development, Department of Commerce/National Technical Information Service, National Institutes of Health/National Library of Medicine, and United States Geological Survey/Water Resource Division.
- An overview of the Workshop and the findings will be presented at the next meeting of the FirstGov working group.

Post Note:

What was accomplished in the day and a half Workshop is considered a tremendous step forward in interagency collaboration, and one with the potential to dramatically change access to government science resources. The planned Science.gov gateway has been described as a unified navigation path to science done by the government, a type of electronic federacy of science mission agencies, responding to the President's Information Technology Advisory Committee recommendations and e-government initiatives. This Workshop, along with the May 2000 workshop sponsored by the Department of Energy which resulted in the report "Future Information Infrastructure for the Physical Sciences"

(available at http://www.osti.gov/physicalsciences/) and future deliberations of the Science.gov Alliance, provide agencies with a practical road map and long-term vision for advancing the science infrastructure on pace with information technology and the Internet.

Since the April Workshop, representatives of the Science.gov Alliance Core Team met twice (on May 18 and June 4, 2001) and also have participated in two meetings of FirstGov. At the May meeting of FirstGov, Science.gov was officially designated as the "FirstGov for Science" web site. A Technical Team met June 7, 2001, to begin the real work of web site design and development. Plans are under way to have a web site ready for public announcement by late summer 2001.

Additional updates from the Science.gov Alliance will be posted at www.science.gov/workshop as they become available.

References

1. Laura Rohde. “Internet Access Is Almost Everywhere.” IDG News Service. February 20, 2001. Available: http://www.pcworld.com/news/article/0,aid,42063,00.asp.

2. “Life on the Internet: Net Timeline.” PBS Online. Cochran Interactive, 1997. Available: http://www.pbs.org/internet/timeline/index.html.

3. Bruce Sterling. “Short History of the Internet.” From *Magazine of Fantasy and Science Fiction*, February 1993. Available: http://www.forthnet.gr/forthnet/isoc/short.history.of.internet.

4. “Workshop Report on a Future Information Infrastructure for the Physical Sciences: The Facts of the Matter: Finding, understanding, and using information about our physical world.” Held at the National Academy of Sciences, May 30-31, 2000, Dr. Alvin Trivelpiece, Chair. Department of Energy, July 2000. Available: http://www.osti.gov/physicalsciences/wkshprpt.pdf.

5. Vannevar Bush. “Science - The Endless Frontier.” U.S. Government Printing Office, Washington, 1945. Available: http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm.

6. William Y. Arms. “The Digital Library Landscape: Looking for Trends.” Cornell University. Available: http://www.science.gov/workshop/barms.pdf.

7. Brooke Dickson. “Update on e-Government.” Office of Management and Budget. Available: http://www.science.gov/workshop/bdickson.pdf.

8. Lee Zia. “The National SMET Education Digital Library (NSDL) Program.” National Science Foundation. Available: http://www.science.gov/workshop/lzia.pdf.

9. Walter L. Warnick. “Universal Access to Science Information.” Department of Energy. Available: http://www.science.gov/workshop/Wwarnick.pdf.

10. “Digital Libraries: Universal Access to Human Knowledge,” Feb. 9, 2001, PITAC Digital Libraries Panel. Available: http://www.itrd.gov/pubs/pitac/pitac-dl-9feb01.pdf.

11. Brandon Muramatsu. “Building the SMETE.ORG Alliance.” SMETE.ORG. Available: http://www.science.gov/workshop/bmuramatsu.pdf.

12. Daniel Greenstein. “Digital Library as Community not Repository.” Digital Library Federation. Available: http://www.science.gov/workshop/dgreenstein.pdf.

13. Eleanor Frierson. “Institutional Strategies for Partnerships.” National Agricultural Library. Available: http://www.science.gov/workshop/efrierson.pdf.

14. Gladys Cotter. “NBII: A Federation Approach.” U.S. Geological Service. Available: http://www.science.gov/workshop/gcotter.pdf.

15. Kurt Molholm. “Toward an STI Network.” Defense Technical Information Center. Available: http://www.science.gov/workshop/kmolholm.pdf.

16. Bonnie Carroll. “Public Information Infrastructure for Science.” Information International Associates. Available: http://www.science.gov/workshop/carroll-kahin.pdf.

17. Walter Finch. “National Technical Information Service (NTIS) Science Portal Initiatives.” NTIS. Available: http://www.science.gov/workshop/wfinch.pdf.

18. Karen Spence. “Science.gov Model.” Department of Energy Office of Scientific and Technical Information. Available: http://www.science.gov/workshop/spence.html.

19. “Information Technology Research: Investing in Our Future,” February 1999, President's Information Technology Advisory Committee, http://www.ccic.gov/ac/report/.

20. “CENDI Federal STI Managers Group.” Available: http://131.84.1.34/cendi/cendi_home.html.

Attachment A

AGENDA
April 18-19, 2001

Day 1: Laying the Groundwork
Collaborative Forum

8:30 am	Welcome to NIST . . . . . John Rumble, NIST
8:40 am	Workshop Overview and Strategy . . . . . Brian Kahin, University of Maryland
8:50 am	The Digital Library Landscape . . . . . Bill Arms, Cornell University
9:40 am	Context for Infrastructure Development
	e-Government . . . . . Brooke Dickson, OMB
	FirstGov . . . . . Tom Freebairn, GSA
10:15 am	Break
10:30 a.m.	Institutional Strategies I (Panel)
	The NSDL Initiative . . . . . Lee Zia, NSF
	Universal Access to Science Information . . . . . Walt Warnick, DOE
	SMETE.ORG . . . . . Brandon Muramatsu, SMETE.ORG/UC Berkeley
12:15 noon	Lunch on Your Own
1:45 pm	The Digital Library as a Community, not a Repository . . . . . Dan Greenstein, Digital Library Federation
2:30 pm	Institutional Strategies II (Panel)
	Institutional Strategies for Partnerships . . . . . Eleanor Frierson, NAL
3:00 pm	Break
3:15 pm	Institutional Strategies II (Panel Continued)
	NBII: A Federation Approach . . . . . Gladys Cotter, USGS
	Toward an STI Network . . . . . Kurt Molholm, DTIC
4:15 pm	Opportunities and Issues . . . . . Brian Kahin, University Of Maryland

5:15 - 7:00 pmNetworking and Partnering Reception at the Gaithersburg Hilton

Day 2: Developing an Interagency Strategy for Implementation
(Federal Agencies Only)

The resources, opportunities, and challenges to develop an interagency science portal.

8:30 a.m. Plenary Discussion:
Reaction to Day 1, Infrastructure Integration and Portal Design
Brian Kahin, University of MD/Bonnie Carroll, IIA

9:00 am Making Partnerships Happen
Machine Translations . . . . . Karen Holland, NAIC
NTIS Digital Object Identifier (DOI) Registry Initiative . . . . . Wally Finch, NTIS
science.gov Model . . . . . Karen Spence, DOE/OSTI

9:45 am Break

10:00 am Concurrent Breakout Sessions

I. Content Sources, Scope, and Coverage . . . . . Lee Zia, NSF
Audiences
Technical disciplines and interrelationships
- Agricultural Sciences
- Biomedical Sciences
- Earth Sciences
- Engineering
- Environmental/Ecological Sciences
- Physical Sciences
- Space Sciences

Types of material/information available

II. Tools/Technologies . . . . . John Rumble, NIST
Archives
Distributed Search Tools/Deep Web Applications
Intelligent Agents
Metadata Management Tools
Special Services (push technologies; alert services)
Web Mining

11:30 am Share Conclusions/Next Steps
12:15 pm Principals Working Lunch (by Invitation Only): Forming Partnerships
- Identify projects and agency champions
- Commit to contributions for: content, tools, technologies
- Lay out time line and actions

2:00 pm Report Writing Team: Voluntary Team Begins Outlining Workshop Report

Attachment B
Workshop Participants

Name	Organization
Allen, Valerie	DOE/Office of Scientific and Technical Information
Andre, Pamela Q.J	National Agricultural Library
Arms, Caroline	Library of Congress
Arms, William	Cornell University
Astley, Allan	Defense Technical Information Center
Bauldock, Barbara	U.S. Geological Survey
Blume, Martin	The American Physical Society
Bold, Thomas	National Technical Information Service
Borchelt, Rick	DOE Office of Science
Brown, Angela	Energy Information Administration
Buffum, Elizabeth	NASA CASI
Burrows, Howard	Raytheon ITSS
Carroll, Bonnie	CENDI
Christian, Elliot	U.S. Geological Survey
Cotter, Gladys	U.S. Geological Survey
Dattoria, Vince	DOE/Office of Scientific and Technical Information
Davis, Richard	U.S. Government Printing Office
Day, Maureen	Environmental Protection Agency
Dickson, Brooke	Office of Management and Budget
Diggin, Denise	DOE Energy Library
Draggan, Sidney	U.S. Environmental Protection Agency
Erwin, James	Defense Technical Information Center
Finch, Wally	National Technical Information Service
Fratkin, Susan	SURA
Freebairn, Tom	FirstGov
Frierson, Eleanor G.	National Agricultural Library
Gibson, Julie	Defense Technical Information Center
Gilmore, Judy	DOE/Office of Scientific and Technical Information
Greenstein, Daniel	Digital Library Federation
Griffith, Jane Bortnick	National Library of Medicine
Hellman, Eric	Openly Informatics, Inc.
Hodge, Gail	CENDI
Holland, Karen	USAF/National Air Intelligence Center
Holland, Mike	Office of Management and Budget
Hunt, Stephen	National Library of Education
Johnson, Richard	SPARC
Jordan, Sharon	DOE/Office of Scientific and Technical Information
Kahin, Brian	University of Maryland
Kase, Kate	U.S. Geological Survey
Lahr, Tom	U.S. Geological Survey
Lal, Nand	NASA
Lanfear, Kenneth	U.S. Geological Survey
Lehmann, Edward	National Technical Information Service
Lesk, Michael	National Science Foundation
McDermott, Patrice	OMB Watch
McDonald, Jim	IWA, Inc.
Molholm, Kurt	Defense Technical Information Center

Muramatsu, Brandon	SMETE.ORG/UC Berkeley
Nemeth, John C.	Oak Ridge Associated Universities
Randall, Carl	Defense Technical Information Center
Rajan, Krishna	Rensselaer Polytechnic Institute
Ridgeway, Roland	NASA Headquarters
Rumble, John	National Institute of Standards and Technology
Ryan, R. Paul	Defense Technical Information Center
Scott, RL	DOE/Office of Scientific and Technical Information
Shepanek, Robert	Environmental Protection Agency
Spence, Karen	DOE/Office of Scientific and Technical Information
Sturrock, Charles	NIST Technology Services
Suthard, Pete	Defense Technical Information Center
Warnick, Walt	DOE/Office of Scientific and Technical Information
Weingarten, Fred	American Library Association
Wood, Fred	National Library of Medicine
Zia, Lee	National Science Foundation

Attachment C

Science.gov Alliance Participants

Department of Agriculture
National Agricultural Library

Department of Commerce
National Institute of Standards and Technology
National Technical Information Service

Department of Defense
Defense Technical Information Center
National Air Intelligence Center

Department of Education
National Library of Education

Department of Energy
Energy Library
Office of Science
Office of Scientific and Technical Information

Department of Health and Human Services
National Institutes of Health
National Library of Medicine

Department of the Interior
United States Geological Survey
National Biological Information Infrastructure
Water Resource Division

Environmental Protection Agency
Office of Research and Development

National Aeronautics and Space Administration
Scientific and Technical Information Program

National Science Foundation

Science.gov Core Team Members
Eleanor Frierson, NAL, Chair
Jane Bortnick Griffith, NLM, Vice Chair
Wally Finch, NTIS
Tom Lahr, USGS
Ken Lanfear, USGS
Bob Shepanek, EPA
Rick Thoroughgood, DTIC
Walt Warnick, DOE
Bonnie Carroll and Gail Hodge, CENDI Secretariat