WORKSHOP PANEL REPORT ON

Accelerating the Spread of Knowledge About Science and Technology

AN EXAMINATION OF THE NEEDS AND OPPORTUNITIES

Convened by the
Office of Scientific and Technical Information
Office of Science
U.S. Department of Energy

Washington, D.C.

February 27, 2007

Acknowledgement

We gratefully acknowledge the assistance of Paul Uhlir, Director, Office of International Science & Technology Information Programs, the National Academy of Sciences, for hosting this workshop. The Panel also wishes to express appreciation to the members of the OSTI staff who did the hard "back office" work of making the arrangements to get us all to the meeting, providing the kind of information that made the workshop flow well, and dealing with the many revisions. They are a class act.

> "If you have an apple and
> I have an apple
> and we exchange these apples,
> then you and I will still each have one apple.
>
> But if you have an idea and
> I have an idea
> and we exchange these ideas,
> then each of us will
> have two ideas."
>
> *–George Bernard Shaw*
> *Playwright, Nobel Prize, 1925*

About the Workshop

A Panel of experts with distinguished careers in federal agencies, academia, and information organizations gathered for a workshop, February 27, 2007, at the National Academy of Sciences. The workshop was organized by the U.S. Department of Energy *(DOE)* Office of Scientific and Technical Information (OSTI), an element of the Office of Science. *(See Appendices for agenda and participants.)* The Panel met to discuss and develop a roadmap for advancing science and technology by accelerating the sharing of scientific and technical knowledge, referred to by the workshop organizers as "knowledge diffusion." Observations and conclusions were based on presentations made by OSTI staff as well as background knowledge of the Panel, which included individuals with experience in several aspects of the problems associated with the rapid dissemination of scientific and technical information and knowledge. Specific steps to accelerate knowledge diffusion were discussed. The Panel identified major thrusts and information infrastructure needs. This is a report of that workshop.

Workshop Panel Report on

Accelerating the Spread of Knowledge About Science and Technology

An Examination of the Needs and Opportunities

Table of Contents

Preface

Scan through almost any scientific or technical journal and you will see that many articles have authors from more than one discipline—researchers who are not from the same institution and in many cases are working in several different countries.

The proliferation of such research teams is in itself a scientific breakthrough, but it is only feasible when the teams can rapidly find and readily share information across the ubiquitous network, the Internet.

Astronomers today tell their colleagues in remote corners of the world about a new observation in little more than the time it takes to hit the send key on their computers. A generation ago such discoveries could take months or even years to become known.

This process of active dissemination occurs in a variety of circumstances. In the mid-1980s, Alex Müller and Georg Bednorz used the best available means to communicate their discovery of high-temperature superconductivity to as many of their colleagues as they could in the shortest time possible. In their case, the tool of choice was the fax machine. Today they would do this with a group send from their e-mail account, perhaps containing a link to their work in an open online archive.

In both of these examples, the discovery or results would normally and eventually be prepared for submission to the appropriate scientific or technical journals. With peer review processes, this can take from weeks to months. In a few cases, the discoveries are of such universal interest that they might appear in newspapers or electronic media.

A closely related process of knowledge diffusion begins when the information is made available to a wider and more general audience, either the broader scientific and technical communities who may have need of it, or the general public. Knowledge diffusion tends to involve some effort by the recipient to find information rather than the information being provided to a close circle of individuals without initiative action. Knowledge diffusion takes advantage of all the means available to have the information in an electronic format so that it can be discovered and applied as soon as possible.

OSTI has been a pioneer in the development of innovative means to make broad classes of scientific and technical information available and useful to scientists and engineers in an efficient and effective manner. OSTI remains a leader in the transition from strictly printed scientific and technical reports, papers, and textbooks to the present virtually instantaneous access to broad classes of scientific and technical literature via the Internet.

These technical capabilities for sharing scientific and technical knowledge need to be maintained and enhanced so that information consumers have search tools and access to information sources needed for U.S. competitiveness.

During the course of the workshop, the Panel became familiar with the state-of-the-art capabilities of OSTI's projects and programs. The Panel offers conclusions about a roadmap toward superior access to quality content.

Executive Summary

This workshop report includes observations, suggestions, and conclusions arrived at by a group of individuals with experience in several aspects of the problems associated with the rapid dissemination of scientific and technical information and knowledge. Deliberations of the participants were informed through presentations made by staff members of the Office of Scientific and Technical Information *(OSTI)*, an element of the Office of Science in the U.S. Department of Energy *(DOE)*.

Rapid advances in information and communication technologies, often termed the advent of the Internet Age, afford compelling possibilities for those working at the leading edge of scientific and technical disciplines to achieve unfettered access to the information and knowledge they need to remain competitive in their fields. These changes also put responsibilities on the funding agencies to ensure that these possibilities are realized.

The Department has embraced this opportunity brought about by the Internet Age, transitioning from information dissemination by the transport of atoms *(i.e., paper in the form of printed journals, etc.)* to the transport of photons *(i.e., broadband communication on a world scale at the speed of light)*. DOE and many other institutions have struggled to keep up with the rate of change that occurred and continues without apparent limit.

Within DOE, OSTI has innovatively responded to the challenges of emerging new information technologies and has parlayed these into the development of excellent new products and services. OSTI has made the transition from a printing press era to the electronic communication era in those activities for which it has responsibility in the Department of Energy. OSTI has also worked to broaden the ability of scientists and engineers to have access to a wider world of information and knowledge than just that generated by DOE programs. Within the government, OSTI has been both a technical leader, capitalizing on the Internet Age, and an organizational leader, promoting partnerships among agencies and governments to broaden access by researchers and the general public.

This workshop report consists of three main sections: the vision, the current environment, and observations/conclusions. The first section summarizes a vision of fostering science and technology advancement through enhanced sharing of knowledge and information. This vision rests on the conviction that any investigator, institution, or nation that does not provide rapid and unfettered access to knowledge and information will not be able to compete as effectively as those that do. This vision of creating conditions that promote easy access to information can only be realized by capturing knowledge and information and putting them into a form that can be effectively shared with others. In the past, this process involved the use of mail, printing, telegraph, phones, and fax machines. Today this process involves the Internet. Although the Internet provides a means of rapidly transmitting information, it does not provide sophisticated capabilities to adequately drill down to relevant scientific data and information required for research purposes. Therefore the funding agencies have a responsibility to create and maintain information systems that deliver high quality information via the Internet. The essence of this vision for DOE involves its ability to meet its responsibilities regarding national security, energy independence, and economic productivity.

Achieving this vision is central to its responsibility to provide appropriate access to information and knowledge to the scientists and engineers that work on its projects and programs.

To that end, the vision of accelerating knowledge diffusion requires a coordinated effort on many fronts. First, it is necessary to ensure that scientists and engineers make information they develop available in electronic format as soon as possible, taking into account the various restrictions involving intellectual property rights, national security, and publication requirements regarding reasonable embargoes on release. Next, it is necessary to have sophisticated systems that store, organize, and recover relevant information from multiple sources that optimally assists scientists and engineers in their discovery process. As the concluding paragraph of the Vision section states, "The information technologies now exist to accelerate knowledge diffusion. Accelerating the spread of scientific and technical knowledge will advance discovery and benefit our national economy and security."

The second section outlines some of the challenges implied by the vision and the environment in which they must be pursued. Included is a review of the situation created by the growing competition on a world scale and the explosive growth of knowledge that has created both an opportunity and a threat. The opportunity is to provide scientists and engineers with superior access to quality content. The threat is that failing to do so will diminish our Nation's capability to compete globally. A brief description is provided on the current state-of-the-art information technologies with respect to the strengths and weaknesses of the systems that now provide access to information and knowledge.

OSTI's transition from print to pixel, providing searchable access to scientific and technical information important to the Department and many other organizations, resulted from the dedicated work of many individuals who enhanced their skills to meet the challenges of this rapidly changing information environment. The panel applauds OSTI's latest initiative to pursue the concept of the "DOE Science Accelerator," for it represents a new level of service and capability that will better serve information consumers in meeting their information needs. The degree to which this is credible is represented by the recent announcement of the Science.world initiative. The British Library signed an agreement to partner with the Department of Energy on the development of this global science gateway.

Based on what was learned in the briefings and the collective experiences of the panel members, some closing observations and conclusions are made. It is clear that the thrust of the briefings and this report are that the future of this activity rests on being able to provide quality content and superior access to a growing collection of information consumers in a manner that supports their efforts and does not impede their progress. For this to occur within the environment of intense global competition and explosive growth of information, certain steps need to be taken. The most important step is to continue and expand programs that accelerate the sharing of knowledge. Further, the controversial issue of open access needs to be reviewed with a goal of establishing guidelines for the scientists and engineers regarding the mode and manner by which their publishable work becomes available and accessible.

Vision for Sharing Scientific and Technical Knowledge

Advance Science and Technology by Accelerating the Spread of Knowledge

Stripped to their essence, research and development *(R&D)* are the creation, analysis, synthesis, and application of knowledge. Knowledge creation in R&D usually comes through experiments or observations involving the physical world. Knowledge analysis in R&D is the examination of experimental or observational data to find likely patterns, correlations or causes. Knowledge synthesis in R&D is the assembly of analyzed data, often from different sources, to infer laws of nature and develop the theoretical expressions for these laws, leading to new understanding of phenomena. Knowledge application in R&D is use of this understanding, either to create something new and useful, or to inform later R&D. In other words, knowledge is both an input to and an output from R&D. While the above description characterizes most R&D, occasionally a great mind may intuit the results and jump to correct conclusions without all of the intervening steps. Thus Einstein is famously quoted as observing that if the data disagreed with the Theory of General Relativity, then the data must be wrong!

> "Sustained scientific advancement and innovation are key to maintaining our competitive edge."
>
> *–President's American Competitiveness Initiative, February 2006*

Much of today's transformational R&D is being created at the margins between traditional disciplines: physics and biology, chemistry and geology, economics and psychology, etc. To excel at interdisciplinary research requires easy and efficient access to knowledge in fields that are not familiar to the researcher. Francis Crick, who shared the Nobel Prize for the discovery of the double helix shape of the DNA molecule, provides a superb example of this. Trained as a physicist, he decided in 1947 to study biology and by 1953 had completed the work with James Watson that was awarded a Nobel Prize. For this work the critical factor was the knowledge gained by a close working relationship with the biologist James Watson, but only a few researchers have this opportunity.

Because of the centrality of knowledge to R&D, the panel concludes that the vision for accelerating knowledge diffusion should be to:

> **Create an environment in which DOE scientists and engineers, other researchers, educators and the public have superior access to the quality content they need to make discoveries, improve lives and advance the energy independence and economic productivity of the Nation.**

This vision presents a difficult challenge, because many techniques using emerging technologies for capturing, storing, searching, presenting and curating knowledge today are in their infancy. As we move from a print to an electronic environment, a measure of progress is that real-time, full-text searches of documents have become possible. However, organizing the results of those searches is still a research topic, and techniques are just beginning to emerge to integrate other types of knowledge, such as images and numerical data.

Importance of Accelerating Knowledge Diffusion

New discovery is required to meet national and worldwide needs for major advances to power the Nation's economy, develop energy independence, and protect our environment. But advances in science and technology depend, in large part, on the success with which scientific knowledge generated by the Nation's extensive R&D programs is disseminated and re-used. It is only through the use of knowledge that science agencies realize value from their investments. This sharing of innovative ideas and discoveries, or knowledge diffusion, advances science and technology and powers much of our Nation's economy.

It is impractical for researchers to spend time finding and sifting through legions of information sources in various disciplines and still have time for life-altering discoveries of their own. Scientists and science-attentive citizens need time-saving, effective interfaces for accessing the totality of scientific knowledge. They need to explore the deep Web, where specialized databases are beyond the reach of, or invisible to, commercial surface Web crawlers such as Yahoo! and Google. They need transformational knowledge diffusion technologies that enable robust and rapid scientific discovery.

The Panel concurs that, if innovative knowledge dissemination strategies are not implemented, negative consequences could arise, particularly to U.S. competitiveness and security. For example, the Panel pointed to the rapid pace of R&D investments in countries such as China, Japan, and India and the accelerating quantity of the world's scientific literature coming from those countries. The Panel noted that U.S. scientists, researchers, and engineers will be at a distinct disadvantage if they do not have comprehensive access to the world's R&D knowledge. Alvin Trivelpiece, Panel chair, stated that the U.S. cannot afford "technical surprises" from other countries. The Panel suggests that the competitive advantages from increased R&D investment under the American Competitiveness Initiative would be wasted if the scientific knowledge generated from it is not made rapidly available for use by other Americans.

What Is Knowledge Diffusion?

It is the sharing of knowledge; the spread or dispersion of ideas.

Knowledge diffusion includes dissemination plus new modes of communication and information discovery only now becoming possible.

Knowledge diffusion advances science and technology. Accelerating the diffusion of knowledge will accelerate the advancement of science and technology.

Opportunity

The information technologies now exist to accelerate knowledge diffusion. Accelerating the spread of scientific and technical knowledge will advance discovery and benefit our national economy and security.

Current Scientific and Technical Information (STI) Environment

Environment and Challenges

In the 1930s, H. G. Wells proposed a "World Brain" in which all scientific knowledge would systematically be stored and readily accessible to all. Society's mechanism for this so-called permanent world encyclopedia would be the scientific library and information center. At the core of such an institution: a synthesis of bibliography and documentation, all indexed and archived.

Well-ordered schemes for reference and reproduction have undergone huge changes since Wells' time, but the vision of ready, universal access to science and technology has endured. In the 1960s, just before the computer revolution, it was not uncommon for the results of research to take months before reaching the relevant scientific community. Batch processing reduced this to weeks, and modern computer and communication technology, including the Internet, have now made the delivery of information potentially instantaneous.

This progress has brought with it massive increases in the volume of data together with the challenge of extracting relevant information. Also, unnecessary barriers to information sharing have arisen or have failed to fall by the wayside. Today, in our newly digital world, research demands more sophisticated information systems that can comprehend a scientist's specialized need and deliver an appropriate response in the timeframe and format needed. To support the advancement of science and technology, a robust information system must manage and search effectively great volumes of information and provide quality results plus linkages to other relevant information. To address increasingly complex challenges, scientists and engineers must have tools that enable them to make observations and uncover connections that were not feasible in yesterday's analog world.

> "Scientific and technological discovery and innovation are the major engines of increasing productivity—indispensable to ensuring growth, job creation, and rising incomes for American families in the technologically driven twenty-first century. The investment is essential if the United States is to maintain its world-class, scientific leadership and global competitiveness."
>
> *–DOE Secretary Samuel W. Bodman before the Committee on Energy and Natural Resources United States Senate regarding the FY 2008 Budget Request, February 7, 2007*

In this new digital century, strategies for marshaling, sharing, transferring, and understanding huge amounts of quality data are crucial to the advancement of science and technology. Thus, these strategies must constitute a fundamental element of DOE's science mission. Effective use of information has a direct bearing on the Department's ability to address pressing national challenges such as energy security and economic competitiveness. In a world economy where advances in research and development come from an expanding array of sources, it is paramount that U.S. scientists, engineers, researchers, students and the general public have comprehensive access to this growing body of knowledge.

Furthermore, because scientific discovery is a cumulative process, with new knowledge building upon earlier findings, it is imperative that unnecessary barriers to sharing the immediate results of research should be removed. In this regard, the Panel supports and encourages the principle that publicly funded unclassified research should be deposited in stable, freely accessible public archives and made freely available as soon as possible after acceptance for publication. This will clearly advance the return on research investment and foster the rapid diffusion of knowledge.

> "Other nations are increasing their investment in basic science, because they know that those who dominate science will dominate the twenty-first century global economy."
>
> *–Dr. Raymond L. Orbach, DOE's Under Secretary for Science, Oral Testimony on the FY 2008 Budget Request, House Appropriations Subcommittee on Energy and Water Development, March 7, 2007*

World Environment

Investment in information infrastructure in both industrializing and developed nations is rising. Worldwide information technology advances and the concurrent explosion of text and data, combined with the recognition that science and technology are key to maintaining a global competitive edge, have created an urgent need and an opportunity to find ways to accelerate the advancement of science and technology. The U.S. has recognized this, as evidenced by the launch of the American Competitiveness Initiative and the Advanced Energy Initiative, as well as the introduction of the America Competes Act.

The need and opportunity exist now to provide superior access to quality content by creating technologies and systems that support diverse and dynamic communities of scientists, engineers, researchers, educators and a public in need of science and technology information. In this technology-driven 21st century, where innovation will increasingly determine economic vitality, sophisticated strategies must be activated for sharing and transferring huge amounts of quality data. These strategies support the DOE mission that enhances U.S. energy security and helps sustain a vibrant national economy.

Information Technology Today

Increased emphasis on discovery, continued specialization, greater market competition, technology transfer, and other social and economic factors are contributing to a vast growth in the global knowledge base.

Concurrently, scientists and engineers have become more and more dependent upon data-intensive forms of research. Emerging means of authorship and publication, the increase of Web-based networking, and new search capabilities have generated massive amounts of information.

Advancements in information technology have provided a foundation for supporting

> "Knowledge is the only instrument of production that is not subject to the law of diminishing return."
>
> *–John Maurice Clark, American economist,1884-1963*

information infrastructure, yet the technology itself has not been applied to developing the tools and capabilities that can facilitate higher quality information content retrieval and enhanced access options. Researchers need the tools and capability to rapidly retrieve and share the information that otherwise may be trapped in obscure databases or within a plethora of non-relevant search returns. Much remains to be done to apply new information technologies to produce the tools that will accelerate knowledge diffusion.

DOE's STI Model

OSTI's mission is to advance science and sustain technological creativity by making R&D findings available and useful to DOE researchers and the American people. From the inception of DOE's predecessor agencies, the need to diffuse knowledge to advance science and technology has been recognized and supported. Now, with the advent of new information technologies, the opportunity exists to apply them to accelerate knowledge diffusion and thereby accelerate science and technology.

To set the stage for the Panel, OSTI provided background information regarding current information technologies and recent developments for leveraging R&D output *(see Appendix 3, Overview of OSTI)*.

In its early days, OSTI was charged with creating an agency-wide R&D information program from scratch. The Atomic Energy Acts of 1946 and 1954, the Energy Reorganization Act of 1974, the Department of Energy Act of 1977, and the Energy Policy Act of 2005, all call for the dissemination of scientific and technical information (STI) to the public, especially information resulting from DOE and predecessor agency R&D.

Throughout its 60-year tenure, OSTI has capitalized on early adoption of information technology, maximizing the use of computers for information search and retrieval.

In the late 1990s, OSTI led the way among federal science agencies in transitioning to Web-based information tools. Among the immediate benefits to scientists and engineers was the delivery of electronic full text to their desktops. To this day, the volume of DOE's full-text R&D information available electronically far outpaces that of any other science agency.

As agencies began hosting Web-searchable R&D databases, the research community certainly benefited but was still left with an inefficient means to locate and navigate through disconnected, disparate collections. Again, OSTI recognized this and pioneered groundbreaking distributed precision searching technology with the launch of Science.gov. OSTI has brought real-time search capabilities to the deep Web through Science.gov and other Web-based products, including E-print Network, Science Conference Proceedings and Federal R&D Project Summaries.

While technology for knowledge diffusion has advanced, this technology needs to be applied to produce a new generation of tools. Today, OSTI is working to overcome the challenge of scalability to take distributed searching beyond its current capacity limitations and enhance precision searching and Web 2.0 applications, accommodating exponential increases in information while still delivering results in seconds. The Panel endorses the importance of these global search capabilities to the acceleration of science and technology.

DOE's STI Program Today

The Panel lauds OSTI for its success in organizing and disseminating scientific and technical information to the DOE community and the public. Despite its small staff and budget, OSTI has excelled in applying information technology to make DOE-produced information easily searched and retrieved using standard Web browsers. OSTI's "deep Web" search techniques set the standard that other federal agencies have followed.

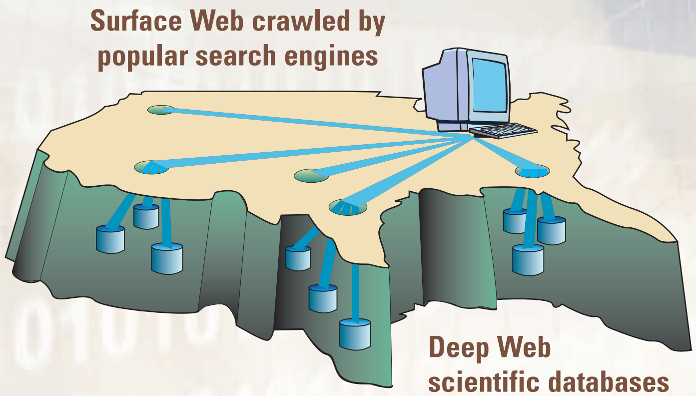
If we think back just a few years, STI was all paper-based. OSTI, and other STI organizations, responded to mail or telephone requests by copying and mailing documents. Response time was measured in weeks. Since then, OSTI has been at the forefront of converting STI to rapid, Web-based technology. Response time is in seconds, and search requests are much more easily and accurately formulated. The Panel is very impressed by the way that OSTI has augmented its internal staff with numerous external partnerships to advance and deploy the information technology to make this possible. Looking to the future, the Panel is impressed that OSTI's technically advanced software architecture will allow users to go beyond the present browser-based interface to more sophisticated user interfaces, or service-oriented architecture *(SOA)*.

This technical leadership has been accompanied by equally impressive leadership in partnering with other federal agencies, private companies, and other countries to integrate the OSTI databases with disparate government repositories and to apply the best information technologies available. OSTI has been a keen leader and participant in CENDI, the federal interagency organization of STI managers. CENDI encompasses all the federal R&D agencies. Through CENDI, OSTI has created Science.gov, which currently makes 30 federal databases transparently searchable with one query. More domestic databases are being added. The user interface may appear simple, but a lot is going on behind the scenes to search the disparate document collections of all the agencies participating in Science.gov, assemble the results, and present them to the customer ranked by relevance. The casual reader might wonder whether this could not easily be done using Google or other commercial search engines. In fact, it cannot be done because these databases are invisible to the search technology used by Google. They are part of the "deep Web."

The Panel has every confidence that OSTI will continue to exercise its technical and organizational leadership in the future. In fact, OSTI has clearly annunciated in its concept of the DOE Science Accelerator that the road to discovery requires a new era in access to scientific and technical information. It specifically calls for the ability to globally search the Web and has begun the creation of Science.world, an effort that the British Library has publicly endorsed. Through a statement of agreement with DOE, the British Library has dedicated its support to making this development a reality. The Panel believes this effort of international leadership is the right direction and is concrete evidence of the continuing forward thinking vision of OSTI management.

> Sharing knowledge advances science and technology
>
> OSTI Corollary
> *Accelerating the sharing of knowledge accelerates the advancement of science and technology.*

Deep Web Content Critical to Science



Commercial search engines, such as Yahoo! and Google, rely upon automated crawlers and are great for finding Web pages, such as www.osti.gov. However, these Web page search engines typically cannot reach information within a database. Instead, database content is retrieved through the search engine of that database. Recognizing this distinction between searching Web page content and database content is important for science and technology, because the bulk of authoritative scientific and technical information resides in databases within the deep Web—out of reach of commercial search engines.

Already, OSTI has pioneered the use of a new class of search engine specifically designed to access distributed resources in the deep Web, enabling a single query to launch searches across a limited number of databases. By using this innovative technology, it no longer matters where the information resides or what format it is in. However, a new challenge has emerged—ramping up to search across larger numbers of deep Web databases. The associated technological barriers need to be overcome.

It is only through the use of knowledge
that science agencies realize
value from their investments

Observations and Conclusions

Thrusts

Developing superior access to quality content is one of the keys to accelerating the spread of knowledge about science and technology. A systematic and continuing improvement in the ability to access information and knowledge is an important effort that supports this concept. Both superior access and quality content must go forward together to create a beneficial outcome.

Superior access means that scientists and engineers have the latest state-of-the-art systems that give them the capability to rapidly locate and acquire the relevant information they need in order to be competitive on a world scale. This includes systems that have an ability to access international and domestic data sources in an effective and efficient manner.

Considerable attention must also be given to the quality, accuracy, and relevancy of the content that the system provides. If a search on a key word returns 100,000 hits, it is unlikely that the information consumer would sort through all these results to find the most relevant. As the various databases expand and the bodies of knowledge in various existing and emerging fields of science and technology continue to expand, it is clear that innovative systems will need to be developed to provide the correct level of superior access to quality content.

> The recent expansion of DOE's R&D programs is expected to multiply the volume of R&D findings *(knowledge)* emanating from this investment.

Strategies for Accelerating the Spread of Knowledge About Science and Technology

To achieve the benefits that would follow from these two interrelated thrusts, a collection of activities needs to be pursued. Some of these are already under way and others need to be added. Taken together they represent a strategic plan or roadmap to bring about the changes that would lead to substantial and beneficial improvements.

Absent an appropriate plan and the determination to see it implemented, U.S. scientists and engineers will be at a growing disadvantage relative to several emerging nations. Some of these nations are producing quality scientists and engineers in numbers that will exceed what U.S. universities can deliver. To offset this disadvantage, U.S. scientists and engineers need tools that enhance their productivity. This plan speaks to one element in a collection of actions that would seem to be needed.

In particular, superior access to quality content information for the Nation's scientists and engineers requires:

- Increasing technical and systems capacity to accommodate escalating amounts of DOE-generated R&D information

- Integrating access to non-textual data and textual information
- Building on the “one-stop” federated searching technology successfully pioneered by OSTI and continuing to expand access to distributed R&D information sources *(including foreign*, *educational*, *numeric and visual sources)* of interest to DOE and U.S. science and engineering communities
- Developing the mechanism to retrieve and make accessible all DOE-sponsored, published R&D literature

Scientists and engineers should rightly expect superior access to the growing body of information and knowledge. OSTI plans to use the following combination of tools to meet this expectation:

- Amplify the power of simultaneous searching with emerging technologies such as grid computing, more powerful servers, and larger bandwidth
- Expand on partnerships with commercial search engines *(such as Google*, *Yahoo!*, *and MSN)* to make DOE R&D results more accessible
- Expand partnerships with the library community and other agencies to encourage interoperability and access to digital data
- Build on “push” technologies such as reference linking DOE reports cited in journals, offering a “more like this” option, linking related information such as projects and results, adding alert services to information products, and preparing to deliver science, technology, and engineering information via evolving communication technologies
- Develop systems that search deeper into content, search more quality sources and improve relevancy of search returns

In any endeavor that makes tools or services available, the matter of branding inevitably surfaces. Given the range of Web tools and online resources, it is suggested appropriate branding be chosen so as to leverage the image and reality of improving superior access to expanding quality content. A properly branded interface would attract those with need for seamless access to the broad range of content.

Sophisticated computational methods with relevant linkages across a variety of information resources should be created, thereby providing users with information they had not planned on exploring, but that would enhance their discovery process.
These might include:

- More advanced full-text relevance ranking
- Enhanced “search-within-a-search” capability to distributed information sources
- Multi-lingual searching and ranking capabilities
- Related document discovery, and searching of charts, graphs, tables and images

Performance measurement and usability testing, as well as embracing basic operational standards and guidance, are suggested as means to enhance the capability to provide superior access and quality content.

Conclusions

Science and technology advance most efficiently when there are mechanisms that enhance the rapid diffusion of knowledge to those who can most effectively use it. OSTI pioneered such mechanisms and tools for scientists and engineers working on projects and programs for DOE and other agencies. To maintain and advance the capabilities and to provide the needed services and support, the Panel offers the following conclusions:

1. Scientific and technological progress will continue to advance through the effective and efficient sharing of information and knowledge.
2. The U.S. will be at a disadvantage if it fails to maintain the capability to access and disseminate information and knowledge on a growing scale of capability and complexity.
3. OSTI and its predecessors have played a key role in establishing systems that result in the rapid and efficient dissemination of scientific and technical knowledge related to the missions of DOE and its predecessor agencies.
4. In the present environment of explosive growth of information and the many diverse science and technology communities supported by such information, it is imperative to have a well-managed, unified information infrastructure.
5. For programs to fulfill their future role in accelerating the spread of knowledge to benefit the U.S. and other nations in their pursuit of developments that advance civilization, increased emphasis and corresponding adequate resources are required.
6. Sharing of information on an international basis is a key component of accelerating knowledge diffusion. Collaborative agreements, such as that completed with the British Library, advance scientific research.
7. The spread of knowledge will be accelerated if DOE-funded unclassified research results are made available expeditiously in a publicly accessible system.
8. Superior Access to Quality Content is key to accelerating science and technology. To make appropriate progress in this regard, some specific activities need to be implemented:

   For Superior Access:

   - Enhance precision searching
   - Develop next-generation algorithms for more sophisticated and faster relevance ranking
   - Utilize enhanced computing power to accommodate increased content and precision searching
   - Overcome barriers to searching a thousand or more databases in parallel
   - Implement distributed access to DOE's electronic educational resources
   - Develop prototypes for analytical tools to support search and retrieval
   - Develop a prototype to apply grid computing technology
   - Expand services to encourage collaboration and sharing of information, including RSS, alerts, forums, blogs, tags, podcasts and other interactive technologies *(referred to by some Internet industry experts as Web 2.0)*
   - Promote cross-discipline communication

   For Quality Content:

   - Discover relevant data with a single entry point
   - Illuminate obscure databases
   - Increase digital full-text R&D collection
   - Expand content to additional sources
   - Preserve and improve access to pre-1990 R&D literature not yet digitized
   - Link to related documents and data

Agenda

Workshop on Accelerating the Spread of Knowledge About Science and Technology

February 27, 2007
National Academy of Sciences Board Room

8:00–8:30	Continental Breakfast	
8:30–9:00	Welcome and Introductions	Walter Warnick
9:00–9:45	Overview of DOE's Program for Scientific and Technical Knowledge Diffusion	Walter Warnick
9:45–10:30	Accomplishments/Milestones	Karen Spence
10:30–10:45	Break	
10:45–11:45	Vision and Thrusts for FY2009 and Beyond	Sharon Jordan
11:45–12:00	Review Charge to Participants	Walter Warnick
12:00–1:15	Lunch and Informal Discussion	
1:15–4:30	Afternoon Sessions	Chair, Alvin Trivelpiece Facilitator, Karen Spence
	1:15–2:00 Consider the Impacts/Importance of Knowledge Diffusion and its Acceleration	
	2:00–3:15 Evaluate, Reconsider, and Revise Thrusts for Accelerating Knowledge Diffusion	
	3:15–3:30 Break	
	3:30–4:30 Prioritize Opportunities/Needs (a Road Map) for Accelerating Knowledge Diffusion	
4:30–5:00	Wrap-up, Address Follow-on Actions, and Adjourn	

Participants

Workshop Panel

Alvin Trivelpiece, Workshop Chair
Former Director, Oak Ridge National Laboratory; Former Director, DOE Office of Energy Research (now called Office of Science); Former Executive Officer for the AAAS Board of Directors

Fran Buckley
Former U.S. Superintendent of Documents; Former Interim Director of the District of Columbia Public Library

Jose-Marie Griffiths
Professor and Dean, University of North Carolina, Chapel Hill; National Science Board Member; Former member of President's Information Technology Advisory Committee and U.S. National Commission on Libraries and Information Science

Rick Johnson
Senior Advisor and Former SPARC Executive Director; Former Senior Vice President of Congressional Information Service and University Publications of America

Bob Marianelli
Former Director, DOE Division of Chemistry; Former staff member of the Office of Science and Technology Policy

Dave Nelson
Former Director, National Coordination Office for Networking and Information Technology Research and Development (NITRD); Former Deputy CIO at NASA; Former Associate Director and Chief of Staff for DOE Office of Energy Research

Kent Smith
Former Deputy Director, National Library of Medicine; Former President of NFAIS and ICSTI; Former Chair of FLICC Policy Group; Former V. P. of UNESCO General Information Program

Workshop Host

Paul Uhlir
Director, Office of International Science & Technology Information Programs for the National Academy of Sciences; Directs U.S. National Committee for CODATA

OSTI Staff

Walter L. Warnick
Director, Office of Scientific and Technical Information for U.S. Department of Energy, Office of Science

Sharon Jordan
Assistant Director, Office of Program Integration for Office of Scientific and Technical Information

Karen Spence
Assistant Director, Office of Information Systems for Office of Scientific and Technical Information

Kristin Bingham
Information Program Specialist, Office of Scientific and Technical Information

Cathey Daniels
Program Outreach Specialist, Office of Scientific and Technical Information

Overview of OSTI

In its early days, OSTI was charged with creating an agency-wide R&D information program from scratch. Today the mission continues, but the means vary.

OSTI's Unique Role

Since 1947 the Department of Energy *(DOE)* scientific and technical information program, managed by the currently named Office of Scientific and Technical Information *(OSTI)*, has been ensuring that researchers and citizens have appropriate access to the Nation's research and development *(R&D)* results. This information spans six decades, from the Manhattan Project to the present, covering DOE and its predecessor agencies, the Atomic Energy Commission *(AEC)* and the Energy Research and Development Administration *(ERDA)*.

OSTI's mission, to advance science and sustain technological creativity by making R&D findings available and useful to DOE researchers and the American people, is met in two primary ways:

- First, for DOE's scientific and technical information (STI), OSTI collects, preserves, and disseminates DOE-funded R&D results. The information derives from work carried out by the DOE national laboratories and other facilities, as well as university grantees. OSTI coordinates an agency-wide program to ensure these results are accessible through world-class e-government information systems.
- Secondly, OSTI provides access to expanded sources of science and technology information from around the world to serve the needs of the DOE research community and science-attentive public. Using innovative tools such as dynamic databases, federated deep Web searching, and relevancy ranking, OSTI advances awareness of a broad array of scientific information related to DOE missions.

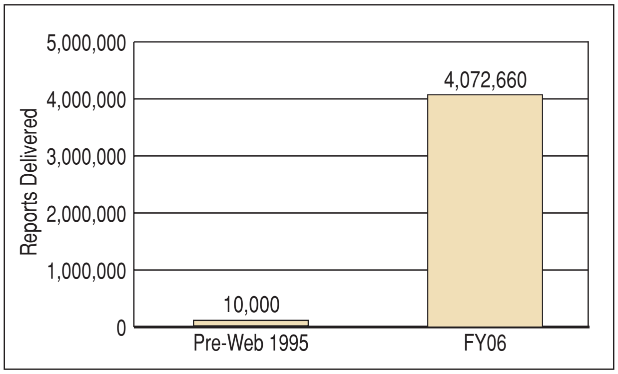
To accelerate scientific progress, it is essential to accelerate the diffusion of knowledge by improving access to science

Statutory Authority

Enabling legislation called for the dissemination of scientific and technical information to the public, especially information resulting from DOE and predecessor agency R&D.

- Atomic Energy Act of 1946 (P.L. 79 585)
- Atomic Energy Act of 1954 (P.L. 83 703)
- Energy Reorganization Act of 1974 (P.L. 93 438)
- Department of Energy Organization Act of 1977 (P.L. 95 91)
- Energy Policy Act of 2005 (P.L. 109-58) The Secretary, through the Office of Scientific and Technical Information, shall maintain within the Department publicly available collections of scientific and technical information resulting from research, development, demonstration, and commercial applications activities supported by the Department. (Sec. 982)

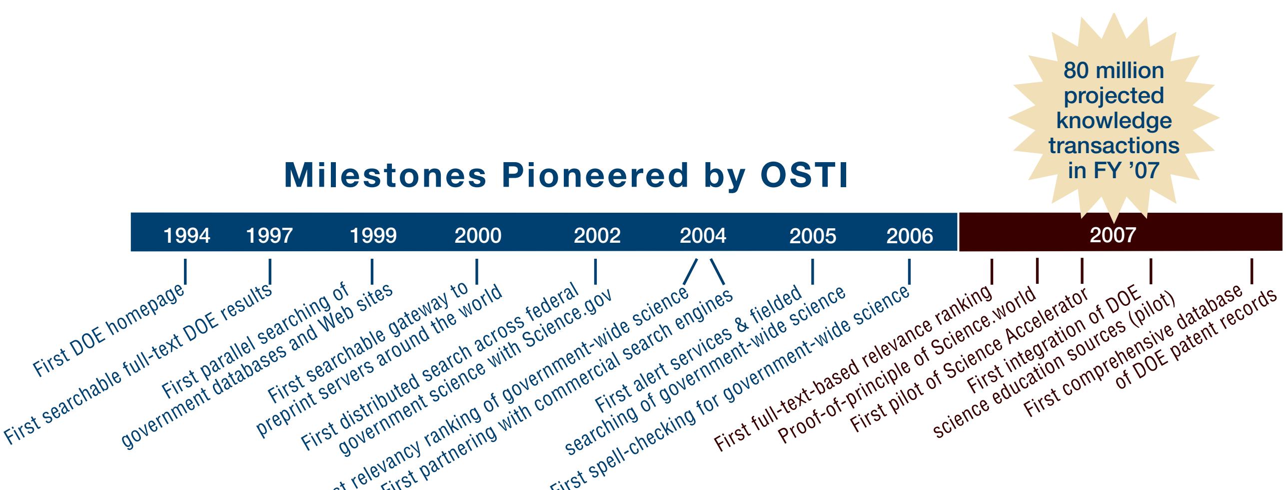
Effect of technology on full-text access



information. OSTI has pioneered a suite of databases and other leading-edge e-government information systems which include the Information Bridge (www.osti.gov/bridge), Energy Citations Database (www.osti.gov/energycitations), E-print Network (www.osti.gov/eprints), Science Conference Proceedings (www.osti.gov/scienceconferences) and Federal R&D Project Summaries (www.osti.gov/fedrnd). Patrons of OSTI collections, including DOE and other federal and contractor researchers, academic institutions, science-attentive citizens, and U.S. industry, use OSTI systems 80 million times annually.

To improve access to U.S. science information, OSTI partners with 12 federal agency counterparts in providing Science.gov, a premier “one-stop” Web system for citizens and researchers to access the government’s R&D collections. Science.gov is an OSTI-hosted gateway to over 1,800 sites and 30 databases of federal R&D information.

Internationally, and in concert with two multilateral information exchange agreements, a current initiative of OSTI includes representation of DOE and the United States in a partnership with the British Library to establish a searchable Global Science Gateway patterned after the model of Science.gov. Throughout its history, OSTI has been instrumental in the international exchange of nuclear information and helped create the International Atomic Energy Agency’s International Nuclear Information System (INIS) in the late 1960s. Today OSTI continues to represent DOE and the U.S. in INIS as well as serving as the Operating Agent for International Energy Agency’s Energy Technology Data Exchange since 1987. Through these two multilateral exchanges and other international partnerships, thousands of foreign R&D records are obtained annually for the U.S./DOE science community.



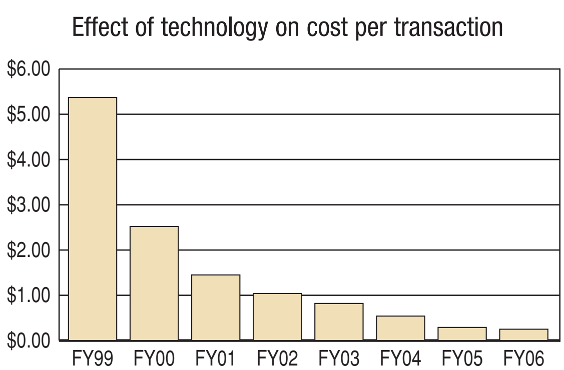
Long before the Internet came along, OSTI was making science information readily available, at times churning out 1 million pages of science documents a day at one of the only federal printing plants in the U.S.

Throughout its 60-year tenure, OSTI has capitalized on early adoption of technology, maximizing the use of computers for information search and retrieval. While technology for knowledge diffusion has advanced, this technology needs to be applied to produce a new generation of tools. Today, OSTI is working to pioneer solutions to the challenges of scalability to take distributed searching beyond its current capacity limitations and enhance precision searching and Web 2.0 applications, accommodating exponential increases in information, while still delivering results in seconds.

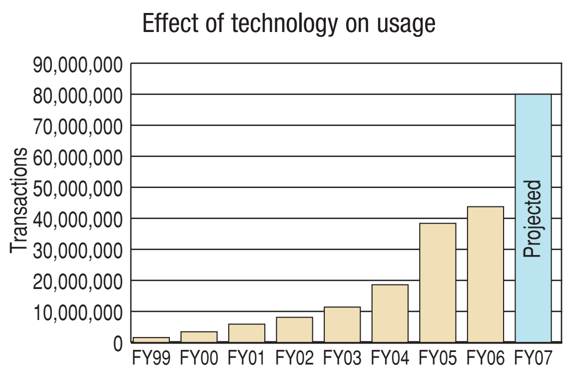
OSTI's unique STI collection today

A repository invaluable to the science community:

- Over 1 million documents, classified and unclassified
  - From the Manhattan Project to present, with daily additions
  - Comprehensive and current
  - Legacy research results not available anywhere else
- Over 146,000 electronic full-text reports, fully searchable online
- Over 4 million R&D citations to research of interest to DOE
- Active research project information



Since 1999, the cost per transaction (page views and downloads) for science R&D information in OSTI's databases has decreased from $5.37 to $0.25



Since 1999, the number of annual transactions (page views and downloads) for science R&D information in OSTI's databases has increased from 1.5 million to 80 million

Suggested Reading

2000 Workshop Report on a Future Information Infrastructure for the Physical Sciences
http://www.osti.gov/physicalsciences

Developing the UK's e-Infrastructure for science and innovation—Report of the OSI e-Infrastructure Working Group
http://www.nesc.ac.uk/documents/OSI/report.pdf

DOE Science Accelerator booklet
http://www.osti.gov/innovation/scienceaccelerator.pdf

DOE Strategic Plan
http://www.energy.gov/media/2006_DOE_Strategic_Plan.pdf

Global Discovery: Increasing the Pace of Knowledge Diffusion to Increase the Pace of Science [Slides from the American Association for the Advancement of Science *(AAAS)* Annual Meeting, St. Louis, Missouri, February 2006]

Notes toward a Nanotech Timeline
by David Kaiser
http://www.osti.gov/innovation/research/diffusion/nanotechdiscussionDK.pdf

On the Transmission Dynamics of Knowledge
by Ariel Cintron-Arias, et al.
http://www.osti.gov/innovation/research/diffusion/epimodeling/transknowledge.pdf

Milestones Web page
OSTI accomplishments for past decade
http://www.osti.gov/milestones

Power of a Good Idea: Quantitative Modeling of the Spread of Ideas from Epidemiological Models
by Luis M. A. Bettencourt, et al.
http://www.osti.gov/innovation/research/diffusion/epimodeling/powerofagoodidea.pdf

Report for the Office of Scientific and Technical Information: Population Modeling of the Emergence and Development of Scientific Fields
by Luis M. A. Bettencourt, et al., Editor's Copy.
http://www.osti.gov/innovation/research/diffusion/epicasediscussion_lb2.pdf

Small Business Innovation Research Grant Opportunities: Discovery, Search, and Communication of Textual Knowledge Resources in Distributed Systems
http://www.science.doe.gov/sbir/solicitations/FY%202007/41.OSTI.Search.htm

The Mathematics of Diseases
by Matt Keeling
http://plus.maths.org/issue14/features/diseases/

Glossary

CENDI
An interagency working group of senior scientific and technical information managers from 12 U.S. federal agencies

Deep Web
The rapidly growing and vast portion of the Web where scientific databases reside and commercial search engines generally cannot reach

DOE
The U.S. Department of Energy (www.energy.gov)

Federated Search
A search capability, often through a portal, that can search multiple data sources with a single query

Knowledge Diffusion
Sharing of information and knowledge; the spread or dispersion of ideas; includes dissemination plus new modes of science communication and information discovery only now becoming possible

OSTI
The Office of Scientific and Technical Information (www.osti.gov), an element in the Office of Science, U.S. Department of Energy

Precision Search
Capability to find research results highly relevant to search queries

Quality Content
A major thrust for accelerating the spread of knowledge about science and technology; encompasses expansion of the body of knowledge for which scientists and engineers have access to enable accelerated discovery

R&D
Research and development

Science.gov
A search engine for more than 50 million pages of government science information and research results from 12 federal agencies, searchable with one query

SOA
Service-oriented architecture, for the purposes of this report, referring to search technologies

Superior Access
A major thrust for accelerating the spread of knowledge about science and technology; encompasses information technologies designed to enable scientists and engineers to access research necessary for accelerated discovery

Surface Web
The small portion of the Web where commercial search engines crawl for information

STI
Scientific and technical information

STI Program
Scientific and Technical Information Program, which OSTI leads and coordinates across the DOE complex

Web 2.0
Moniker for second generation of Web-based services