

# **“Conquering Text: An Achievable Vision for the Scientific and Technical Information Program”**

**Walter L. Warnick, Ph.D., Director  
U.S. DOE Office of Scientific and Technical Information (OSTI)**

**Presentation at InForum '99, May 5, 1999**

## **Introduction**

Welcome to InForum '99!

We have here a core group of the STI professionals of the entire Department of Energy. I will review where we as a group have been over the last year. Collectively, over the next two days we hope to set the stage for where we are going as we enter the new millennium. I think we are poised to do great things.

All of us in the information business, which I think is about everyone in this room, are living through a *revolution--the digital revolution*.

Here is evidence of the size of this revolution:

- The number of items sent by e-mail outnumbers the items sent by postal mail ... by a factor of 10!
- Through the first few months of this year, the average number of visits to the Drudge Report (Matt Drudge's news Web site) exceeds the circulation of Time magazine - the highest circulation news magazine in America - by a factor of 5!
- Internet advertising doubled in the last year. It now exceeds revenues from all outdoor advertising, which includes all billboards.

The revolution is very new, yet is already very big.

This revolution is not of our making. But we did have a profound choice to make. Should we resist the revolution? Or, should we embrace it? What you will hear here today shows that we have [embraced the revolution](#).

We really had no choice. If we make peaceful revolution *impossible*, we make contentious revolution *inevitable*. By embracing the revolution, we are making it work FOR us.

Tremendous advances are possible using the new technology. But, it is not enough that a thing be possible for it to be achieved. Sharing a vision is but the first step. Then we must all take concrete steps to realize the vision.

The [Scientific and Technical Information Program \(STIP\)](#) is doing just that.

STIP has achieved significant milestones within the past year. We are making swift and revolutionary changes in the way we manage STI. One might ask, where are these changes taking us?

Briefly, we can imagine that in the next few years STIP will conquer text. By conquering text, I mean that searching and retrieving text will no longer be a challenge to users or to us. Our making text information available on the Web will be routine.

It is within our grasp to make a vast comprehensive [collection of scientific and technical text](#) immediately searchable and retrievable. Text comes in various forms: report literature, journal literature, preprints, and books. Of these, the most timely information is report literature, journal literature and preprints. So let's look at the prospects for conquering text in each of its most timely forms.

For report literature, the tide of battle is clear. Shortly, the conquest of report literature will be completed by the [DOE Information Bridge](#). The decisive armament was when we all adopted electronic formats that rendered paper transmission unnecessary.

Journal literature is about to be conquered by PubScience. I will talk more about PubScience later.

Preprint servers and e-print servers, the fastest growing sources of text, are up and running in physics, but have not yet arrived in many other disciplines. Preprints ought to be the business of STIP. While the sponsor of the leading e-print publisher has called upon OSTI to assume a significant role, OSTI's funding has so far not allowed for that prospect. Neither can OSTI afford to launch new preprint servers. But just because OSTI is not positioned to move need not mean that STIP cannot. Sharon Jordan will have more to say about preprint servers later at this meeting.

And as we conquer text, other formats - such as non-text images and video - will be addressed by our collective talents.

Our purpose is to get maximum mileage out of information. That this is important is appreciated by everyone, including [Secretary Richardson](#). In a speech two weeks ago, he said:

“For science to rapidly advance at the frontiers, it must be open. And shared knowledge is the enabler of scientific progress.

“The community of science is an international community that transcends political borders. Scientific disciplines link peoples and Nations. Each member of the scientific community builds upon the work of others and provides the foundations of knowledge for future generations.

“The foreign scientists who come to our shores are the best and the brightest of their countries. [R&D results,] published in open literature, add to the store of knowledge that increases wealth, health, and standards of living.”

This “shared knowledge” to which Secretary Richardson refers is what STIP is all about.

The STIP Partnership started two years ago. Our vision was, and is, that the contribution of DOE’s STI would be recognized not only within the U.S., but by the world. The STIP community agreed on our mission. The STIP mission is: to create, collect, and share scientific and technical information through a collaborative and innovative Program forged by all Departmental elements. We intended to use state-of-the-art technologies.

We agreed to work together on four goals:

1. To make STI more easily accessible, principally through electronic means;
2. To work via voluntary collaboration;
3. To promote consensus agreement and use best practices, not command and control; and
4. To produce STI products and services that our users want and that meet their needs.

Two years later, there is much evidence that we in the STIP community are making a difference. In brief, STIP works. I will talk next about achievements stemming from our STIP partnership. Later, I will talk about achievements produced by OSTI and achievements produced by individual Labs and Offices. This is a truly extraordinary collection of achievements.

## **STIP Accomplishments**

STIP partners - that’s our laboratories, operations offices, programs, other contractor facilities, and OSTI - have [shared in several accomplishments](#) since last year’s InForum meeting. The most remarkable deals with process:

Via collaboration, we revised our STI Management directives, resulting in the issuance last August of the new Order and Guide ([DOE O 241.1 and DOE G 241.1-1](#)). These signaled a revolutionary change in the way DOE’s STI is managed by marking the transition to an electronic environment. Many of you have worked hard to help achieve this electronic transition. Your work has been critical to our shared success. This is the single most revolutionary change in STI management in the entire 51-year history of STI management at DOE and predecessor agencies.

There have been a number of other noteworthy accomplishments by the STIP partnership.

The Goal 1 working group, which focuses on STI access, helped make site information resources electronically accessible. New links to site collections were added to [EnergyFiles: Virtual Library of Energy Science and Technology](#). A set of metadata was developed, agreed-upon, and implemented for the DOE announcement record.

The Collaboration Working Group (Goal 2) has been leading the e-journal initiative. It has pulled together 15 DOE sites interested in using consortium purchasing for online scientific journals. Using collective buying power, many sites were able to subscribe to *Science Online* at a favorable rate, and *Science's Next Wave* has been made available throughout the DOE complex. Currently, the group is collaborating with LANL in a test arrangement with Elsevier, whereby LANL's experience is enabling full-text access to over 1,100 Elsevier titles for a 60-day trial period. DOE Headquarters now has more titles available than at any time in the history of DOE or predecessor agencies.

Another successful e-journal pilot has been the access to *ISI Web of Science* for headquarters for a 7-week test at no charge. A break-out session this afternoon will provide more information about the e-journal activities. Our e-journal efforts have been very well received.

Additionally, various publishers have agreed to provide electronic journal citations at no charge to be added to a new product under development, the DOE database PubScience. I'll mention this again in a moment.

The Best Business Practices team (Goal 3) produced much of the content for the new Guide that we issued last summer. Lately, this working group has planned the transition to a decentralized electronic environment for software, similar to what we have accomplished with text. The proposed changes in software management will be discussed in one of the break-out sessions tomorrow afternoon.

The customer-focused products group (Goal 4) developed the list of acceptable electronic formats for transmitting full-text reports. I'd like to especially acknowledge the accomplishment of this group and the very capable leadership of the late Jeanne Sellers of Savannah River. She was a role model for many in the STI community and is sorely missed by all the STIP members.

These working groups have made collaboration a common approach for achieving our goals. The STIP representatives yesterday agreed to take this concept a step farther -- by setting up special task groups or ad hoc working groups as needed to focus on key initiatives or to help resolve particular problems. This is evidence to me that our STIP partnership is growing.

Another STIP group active the past year is the R&D Visibility Working Group. Formed last year, it has been exploring various ways to increase the visibility of DOE's R&D output. Improving the public image of the Department is one objective. As *Secretary Richardson was quoted* in the March issue of *Inside Energy*, "I want this agency to reconnect with the public, something it doesn't do well except on things you don't want it to connect on." The working group met Monday here in Oak Ridge and has produced a concrete proposal for how a DOE distributed information resource would be organized. The proposal was accepted by STIP on Tuesday, subject to further comment.

OSTI and Los Alamos National Laboratory (LANL) have agreed to jointly develop a new

product called PubScience, and GPO has indicated its interest as well. [PubScience modernizes OSTI's traditional role](#) providing access to DOE R&D outcomes published in journals. PubScience, which deals mostly with the physical sciences, is modeled after [PubMed](#), a hugely successful product of the National Library of Medicine for the life sciences. Already, the [prototype](#) is online, and we have committed to have a developed product yet this year. PubScience will allow searches of STI citations and will provide hyper-links to publishers' servers where full text may be retrieved. We estimate that over 80 percent of STI in disciplines of concern to DOE is available electronically in full text. Other STIP members are welcome to join with LANL, OSTI, and GPO to develop PubScience.

OSTI and the [Electronic Resource Library at Amarillo College](#) in Texas are nearing the end of a successful 2-year collaboration. The Library was interested in digitizing historical documents on plutonium. OSTI provided the original paper documents to the Library, which in turn scanned and OCR'd them. The product is an historical collection of digitized plutonium information. The original documents are returned to OSTI, which also receives images and text that can be added to the Information Bridge.

So, we have just reviewed a few of the accomplishments coming from the [STIP collaboration](#) in the last year. One lesson we have learned in STIP, if we did not know it already, is that voluntary collaboration is not easy. Working with others means that any one individual's views and preferences are unlikely to always prevail. This can be frustrating. As multiple people need to be convinced about the wisdom of any contemplated action, progress can seem insufferably slow.

Collaboration is also impeded by what I perceive as a sort of cultural divide between Federal staff and Lab staff. There is still too little trust within STIP. I can offer no special insight about the causes of the cultural divide, or what to do about it, except to keep trying, as I think we are all doing.

Despite these problems, however, the STIP collaboration is clearly working. In fact it is working quite well. I am not aware of any other collaboration in DOE that works as well as STIP. If anyone here is aware of a voluntary collaboration in DOE that works as well as STIP, please call it to my attention. We should all be proud about how well STIP is working.

## **OSTI Undertakings**

In addition to these STIP accomplishments, OSTI has been busy pursuing several projects. Here are a few accomplishments that I would like to highlight:

We developed the [DOE Energy Link](#) system (called "E-Link") for you to electronically announce new STI reports (rather than submitting a paper Form 241.1). Reports are made available by your sites hosting them at a unique URL, or by uploading electronic formats to OSTI. As E-link becomes fully operational, full-text STI can be located and searched regardless

of where it is hosted. Because E-Link is being developed on new hardware and software architecture, it has been a challenge to all of us, but we are about to get there! There's a break-out session tomorrow for those who provide input to OSTI. We'll be demonstrating this system at the Exhibits.

We increased the content of the DOE Information Bridge to almost 50,000 full-text documents. We celebrated a milestone last September when the 2 millionth full-text page was added to the collection. Now we're at 3 million full-text pages. And I am pleased to tell you about the [awards](#) the Bridge has won!

The Information Bridge received the IM Council Technical Excellence Award in October 1998. This is the second year in a row that we've won awards from that group. OSTI is the only repeat winner, and we have won each time we have applied.

The Information Bridge received a [Hammer Award](#). We were invited to receive the Hammer Award at the Virtual Government Conference in February, which was attended by about 300, including the Chief Information Officers of many Federal agencies. We were the only organization in all of government to be so invited. The Bridge was featured in the inaugural issue of V.P. Gore's Access America On-line Magazine. We received a commendation from a GPO advisory committee and were hailed by Senator Warner as a model for GPO to follow with other Agencies. The Bridge was also a Yahoo Pick of the Day and Pick of the Week.

The Information Bridge packaging was nice, but without the content provided by this audience, there would be no Information Bridge.

In addition to the Information Bridge, two other DOE systems were featured in the October 1998 inaugural issue of Vice President Gore's [Access America Online Magazine](#): EnergyFiles: the Virtual Library of Energy Science and Technology and [DOE R&D Project Summaries](#). Such promotion of our efforts can only help the STI Program and the visibility of DOE's R&D results.

These same three systems are National Partnership for Reinventing Government (NPR) Hammer Awardees, plus a fourth Hammer for the overall STIP partnership. (You'll hear more about these awards a bit later on this morning's agenda.)

[Another new product](#) is the [DOE R&D Accomplishments Database](#). It is a Web-based forum about the outcome of past DOE R&D that has had significant economic impact, improved people's lives, or been widely recognized as remarkable advance in science. When friends ask, "What has DOE done?" send them to this Web site. It is one way we can get the public to reconnect with DOE, as Secretary Richardson discussed.

R&D Accomplishments is distinct from R&D highlights. An R&D accomplishment is the outcome of *past* DOE R&D whose benefits are being realized *now*. In contrast, highlights provide information about *current* R&D whose benefits are expected in the future. DOE has a

number of sites that present highlights.

Items are added to the R&D Accomplishments Database just as reports are added to the Bridge. R&D Accomplishments provides *solid evidence* that the taxpayers' investment in government research has paid huge dividends. Dr. Krebs transmitted a memorandum throughout DOE noting the tremendous potential to showcase the Department's R&D - if the DOE community participates and provides information for the database. In response to the Krebs memo, we have already received new input from two Assistant Secretaries. We invite Labs, Ops Offices, and program offices to submit reports of accomplishments. Instruction for submitting reports are on the Web site.

[EnergyFiles: Virtual Library of Energy Science and Technology](#) now has a [distributed search](#) capability up and running. It is being demonstrated here at the InForum exhibits. You may recall that, last year, subject pathways were a new feature added to EnergyFiles. I think the addition of the distributed search capability to EnergyFiles makes it a true virtual library.

The major challenge of a virtual library is how to search across heterogeneous databases and Web sites. By heterogeneous, I meant that there is no standardization of data, information resides in multiple forms, it is in a variety of unrelated systems, and it is in widely dispersed facilities. The goal is to achieve distributed search capabilities despite these obstacles.

EnergyFiles now contains over 450 collections. The need for distributed search is clear. OSTI has developed an exciting two-pronged approach. It portends a major breakthrough in how information is retrieved and used.

The first prong is distributed searching of databases of other agencies, not just titles or bibliographic information but full text, key words and anything else that might be contained in that database. To our knowledge, this is the first full-text interagency distributed searching in government.

It is not dependent on extensive data mapping, common data elements or bibliographic standards, or commonality of systems. In fact, data systems do not have to be standardized in any way. The mechanism capitalizes on the existing features of each database, so it is less expensive and requires no retrofitting of existing systems.

The second prong is distributed searching across Web sites. EnergyFiles uses [Webinator](#) to search the EnergyFiles collections. This distributed search option provides both breadth and depth.

1. Breadth is achieved by searching across information collections residing at multiple sites. It enables users to search both HTML and PDF.
2. Depth is achieved by searching a prescribed number of levels down into each site. That means users are not just searching all the sites on EnergyFiles, they are searching all of the

links off of each of those sites. Access to information grows exponentially beyond the 450 “base” collections in the virtual library.

Distributed searching was made available to the public for a subset of databases and Web sites of EnergyFiles on April 21.

Next steps for EnergyFiles include: (1) make more databases subject to the distributed search, (2) combine the two-pronged approach into one so that databases and Web sites can both be searched via a single query, and (3) deal with traffic congestion, if it develops.

### **Partners’ Achievements**

Our [STIP partners](#) have also accomplished a great deal working independently. Later in this Conference, [Beth Perry](#) will review these accomplishments in detail. However, I cannot let this opportunity pass without briefly reviewing some of the highlights.

[Fermi National Accelerator Laboratory \(Fermilab\)](#) is digitizing its entire R&D report collection dating back to 1968 when the laboratory was established. Through collaboration with OSTI, these reports are being integrated into the DOE Information Bridge. Cynthia Crego leads the project. Fermilab has documented cost savings of approximately \$200,000 per year through the electronic publishing initiative. We invite other Labs to follow the Fermilab lead.

LANL and OSTI have entered into an agreement for bi-directional electronic access and sharing of full-text documents at each other’s sites. LANL’s partnering with OSTI will incorporate their huge legacy collection into the DOE Information Bridge. Rick Luce heads this project. We invite other Labs to follow the LANL lead.

The [Bonneville Power Administration \(BPA\)](#) accomplishments for STI were recently reported in their “[Circuit](#)” newsletter, including a picture of Technical Information Officer Rose Anne Ranft. The article highlighted the re-engineering of their paper report process to an electronic document management process, thus saving the cost incurred in handling paper, such as a \$30,000 per month printing bill. We applaud their success.

[Oak Ridge National Laboratory \(ORNL\)](#) has determined that it will move aggressively toward producing technical documents in electronic form. Specifically, as of March 15, 1999, all conference papers, technical reports up to 25 pages, and Foreign Trip Reports must be provided in searchable Portable Document Format (PDF). This is yet another step towards our STIP vision.

[PNNL](#) is developing a system they call ERICA (Electronic Records and Information Capture Archive). With ERICA, they will have a searchable repository of information/records. The prototype is being tested now. Information about this development is being shared here at InForum later this afternoon.



The [Hanford Technical Library](#) has been aggressively moving into cyberspace with an extensive list of URLs containing documents of interest to the DOE and contractor community. Their slogan, “Making the Virtual Reality, ...Connecting People & Information for Results,” is one we all might like to adopt.

During FY 1999, Lockheed Martin Energy Systems has adopted electronic submittal and online access to STI. Their process addresses the STIP goal for the site to maintain the file and provide a URL to OSTI for the electronic document.

[Westinghouse Savannah River Company \(WSRC\)](#) has aggressively supported the transition from paper to electronic document submittal and in making electronic information available to customers. This effort is on the InForum agenda tomorrow.

These are just a few of the achievements that we have heard about and are examples of what all the STIP sites are experiencing: an explosion of changes and advancements in the Information Age. The revolution is being embraced.

### **Partnerships with Headquarters Offices**

The [Scientific and Technical Information Coordinating Group](#) is comprised of representatives from headquarters offices. It meets two or three times a year. At the last STICG meeting, we discussed STIP comments on the draft Work Authorization System. The Office of Field Management incorporated our comments and issued the final order a couple of weeks ago.

Recently, we jointly signed a memorandum with the Office of Procurement Policy encouraging Procurement Directors to join the electronic transition with grantees and other non-M&O researchers. These researchers likely create their STI products electronically and are capable of submitting full text in electronic formats. We are currently working together to make progress in this area.

OSTI has a long-standing partnership with the [Office of Environmental Management](#). OSTI is working with EM on the development of a EM specific digital library.

This was just a summary. By any measure, progress is tremendous. There is a theme to these accomplishments: each one brought us closer to conquering text. Each shows we have embraced the revolution.

### **What Decision-Makers Must Know**

Yes, these achievements are great, but one is often challenged, “Are they cost-beneficial?”

Decision-makers like to see cost-benefit analyses. But performing a cost-benefit analysis for library functions is very tough. In fact, credibly translating the value of the service and information into dollars may well be impossible.

Our costs are real and easily quantified. The benefits we produce are equally real, but are *not* easily quantified. That is why cost/benefit is tough to analyze.

This situation is merely a symptom of the underlying principle that library operations can never succeed on a cost-recovered basis. Those who insist on a self-supporting library operation want something that has never been, nor never will be. Information Age technologies have not changed this principle.

Andrew Carnegie is famous for building libraries. He constructed 2,000 beautiful buildings across the U.S. But he did not OPERATE any libraries. The responsibility for operating libraries was left to the local citizenry. Similarly, for Federal information services, operations require funding.

Decision-makers sometimes have unrealistic expectations about the new technology. Yes, it can dramatically cut costs per user served, but transitioning to the new technology often means that total costs go up.

Further, decision-makers always underestimate the difficulty of applying Information Age technologies. Practically no one outside the field of systems development appreciates how difficult systems development is. Last week, the Department of Interior threw in the towel on a \$400 million effort. We read in the newspapers about multi-billion dollar fiascos at the Internal Revenue Service (IRS) and the Federal Aviation Administration (FAA). People read about these disasters but fail to appreciate the implication: systems development is rife with pitfalls. One of the most profound commitments that I make in Washington is a two word sentence. OSTI delivers. The commitment is profound because systems development is a very tough business.

Decision-makers typically overestimate the power of the Web to find good information. According to John Rumble of the [National Institute for Science and Technology \(NIST\)](#), “You can find anything on the Web EXCEPT good data. ... And, if you should happen to find it, you don’t know it.” For over 50 years, we in the information business at DOE and predecessor agencies have specialized in GOOD data. We need to constantly remind decision-makers of the fundamental difference in utility between data and GOOD data.

While it is difficult to monetize the benefits of library operations, the Information Age revolution is so profound that it is standing the concept of cost/benefit on its head. The stock market faces the same problem of monetizing benefits when it values e-commerce companies. Take, for example, Amazon.com. The company does lots of business, but it has yet to make its first dollar of profit. It has lots of traffic on its Web site. Based entirely on a capitalization of a vague notion of future earnings, the stock market has set a value on Amazon.com in the multi-billions.

Perhaps when talking to decision makers, we should be leveraging this argument. After all, like Amazon.com we do not make a profit either. And we do have lots of traffic on our Web sites. If one would do a straight line interpolation, OSTI's value exceeds the funds we have received from Congress.

## **Consequences**

We cannot predict all the consequences of the revolution. I am convinced that the digital revolution will progress. But I am very unclear about the future of traditional lines of the information business.

Take print for example. Some people presume that print will decline, or even go away. But such a conclusion may well be hasty.

What little evidence we have suggests that the effects of this information revolution on print media can be counter-intuitive. For example, the [National Academy of Sciences](#) recently put its report library online. The demand for print reports shot up by 40%.

We cannot predict all the consequences of the revolution, but we can predict the consequences of passivity.

## **Challenge/Call for Action**

The main limit to our realization of tomorrow will be our doubts of today.

Let us move forward to Conquer Text! Think about it. We can be the first people in all of human history to conquer text. The prospect of conquering text now before us is truly animating and ought to excite the exertion of all of us here in STIP.

[Here](#) is what I think we need to do, and what we can achieve together:

1. We need to finish our electronic transition plans. This is *critical* to our continued success.
2. We need to consider alternatives for making preprints accessible in a variety of subject disciplines; Sharon Jordan will speak more on this topic in a few minutes.
3. We need to aggressively pursue collaborations and consortium arrangements for electronic access to scientific journals.
4. We need to assure that your Web sites and databases are linked to EnergyFiles: Virtual Library of Energy Science and Technology.
5. We need to rapidly expand distributed search on EnergyFiles.

In sum, we need to take steps individually and together as STIP to achieve our shared vision.

Let's commit to each other to have open and honest communication about our Program and our goals. Let's continue to build upon our good collaboration.

Secretary Richardson properly notes that "Shared knowledge is the enabler of scientific progress." When we conquer text, knowledge in text format will be shared instantly. We could hardly do anything more important than that to enable scientific progress.