

Making STI Electronically Available at the Savannah River Site

Julianna U. Hearn
Westinghouse Savannah River Company

A paper for InForum '99
May 5-6, 1999
Oak Ridge, TN

Introduction

The Scientific and Technical Information (STI) program at the [Savannah River Site \(SRS\)](#) has endeavored to publish scientific and technical information electronically for a number of years. Since 1993, we have worked to continually improve our processes by incorporating new methods, technologies, and ideas to expand the areas of dissemination incrementally. This paper will show a brief chronology of our efforts to date, beginning with the initial posting of STI abstracts on an internal server, to making full-text STI documents available from the SRS external web page. The discussion will also include some lessons learned and a look at near-future efforts with some promising technologies.

Electronic Abstract Collection

In the past, the STI group at SRS was responsible for compiling, editing, and producing a hard copy abstract compendium for approved documents on a monthly basis. This compendium was comprised of an average of 60 abstracts and was disseminated to 300-400 people across the site each month. In 1993, a Filemaker Pro database was created for the purpose of publishing the abstract collection electronically. The database contained bibliographic information and edited abstract text for each document that had been approved for release off site. The database had numerous built-in search and display menus, as well as reporting capabilities and was accessible by site personnel via the site network.

Electronic publishing of the abstract compendium in this manner eliminated the need for the hard copy distribution, saving time and money. In addition, it made the information about published documents more readily available to site personnel in an easily searchable form. However, the existence of this database was not widely known so it was not widely used. And, it was not available to off-site researchers. One other drawback was that there were no clear-cut criteria for determining which documents should be represented in the database. The database quickly became over-populated with records for anything and everything that had been approved for release off-site, whether they were STI products or not.

SGML-encoded Bibliographic File Generation and Submission

In June of 1994, SRS began electronically submitting SGML-encoded bibliographic files to [OSTI](#). In order to generate these files, the Filemaker Pro database used for tracking documents through the review and approval process was modified to include scripting for extracting the necessary information from the fields, exporting it to another file, and incorporating the required SGML tags. A batch file of SGML-encoded bibliographic records for OSTI-reportable documents approved for release was submitted to OSTI on a weekly basis via FTP.

The idea for this innovative use of the STI tracking database to generate SGML-encoded bibliographic files was developed and implemented in-house without any additional personnel or resources. The database format and structure was shared with several other sites for use in implementing their own processes for SGML-encoded bibliographic file submission. These SGML-encoded bibliographic files provided the capabilities for electronic indexing and searching of the related documents, and were the first step toward realizing the OSTI vision of a “virtual library” of STI products. One drawback to this method is that it required in-depth knowledge of Filemaker Pro database structure, file relationships, and scripting. Any changes in the bibliographic file requirements from OSTI required modifications to the scripting.

Documentum Pilot Project

The Document Management System (DMS) project that was piloted in 1995 proposed a major change for the SRS review process. It was envisioned that this system, which utilized a product called Documentum, would provide the capability for electronic document routing and tracking of full-text documents through the review and approval process, as well as full-text document storage with indexing and retrieval. Though somewhat complex, the system initially appeared to have a lot of promise. However, lack of dedicated programming support necessary for tailoring the product for the specific needs related to STI, and a general reluctance on the part of the reviewers to accept the digital signature idea caused the project to go no further than the pilot stage.

Web Publication of the Abstract Collection

In late 1995, a programmer was added to the staff of the STI parent organization. One of the initial tasks this individual had was to develop and implement a method for the publication of the [Abstract Collection](#) via the World Wide Web. The existing abstract collection was first converted to text and stored in an ORACLE database. Then, a web-based interface was developed to provide search, database interface, retrieval, and display using perl scripts. In conjunction with this system development, a decision was made to store and provide electronic access to only those abstracts related to STI products approved for release.

In addition to the initial posting of the abstract collection to the web, the programmer developed a method for updating the collection with the bibliographic information and abstracts of newly approved STI documents. This method involved an HTML interface with several lengthy and complicated steps and usually required some “tweaking” by the programmer during each reiteration to obtain the desired results. The drawbacks to this method of developing the abstract collection were that it was more difficult to keep the information updated, and that ORACLE programming and other specialized skills were required.

The benefits of making the STI-related abstract collection available to more individuals, both on site and off site, included:

- providing consolidated reference information
- realizing cost savings related to the elimination of hard-copy distribution
- receiving fewer but more specific information release requests
- identifying new business opportunities for the site through technology transfer and partnering efforts

During this time, the STI group continued to generate and electronically submit SGML-encoded bibliographic files to OSTI on a weekly basis.

Submission of SGML-encoded Full-text Documents to OSTI, Posting HTML Full-text Documents on the Site’s External Web Page

With input from the programming support person, STI also developed a plan for implementing a process for electronic submission of SGML-encoded full-text documents to OSTI. In 1996, the OSTI report.dtd was modified slightly for SRS use with conference papers and journal articles. Development was begun on a system that would not only track and store information about documents as they went through the review and approval process, but also electronically convert and store the full-text document as SGML-encoded text. Unfortunately, later that year we lost our programming support due to a site-wide reduction in force and this system never got much further than the initial programming of the web interface.

In order to meet an SRS milestone associated with electronic submission of SGML-encoded documents to OSTI, a number of full-text documents were manually tagged and individually submitted to OSTI beginning in 1997. The HTML versions of these documents were published on the SRS external web page to meet another site milestone. These HTML documents posted to the external web page were the first in the SRS on-line [STI full-text collection](#).

STI Abstract Collection Available via Web-based Interface to Filemaker Pro Database

In early 1998, SRS STI instituted the use of Filemaker Pro version 4.0 for making the abstract collection available directly from the database over the World Wide Web. A web-enabled version of the database that contains the abstract collection was published on a Macintosh server with its own IP address.

The publication of the STI abstract collection directly to the web using Filemaker Pro 4.0 was a vast improvement over the previous method of abstract collection development via ORACLE database files and perl scripts. Since the tracking database being used by STI is also Filemaker Pro, it became much easier to extract the data needed for the abstract collection making updates easier and less time consuming. In addition, the search capabilities inherent in Filemaker Pro provide the users with the ability to search on any field in the database through web-based forms. The templates that come with version 4.0 made it relatively simple to set up the necessary forms and related Filemaker Pro code for the user interface.

However, there have been some drawbacks to this arrangement. The Macintosh-based server has proven to be not as reliable as other servers and has crashed or become “frozen” numerous times. During some of these unscheduled interruptions, the database file has been corrupted as well, requiring it to be rebuilt or reloaded from a backup. SRS Computer Security also required the Macintosh server to be connected to the Internet through an additional layer of security called a DMZ. This meant the actual IP address had to be included in the links from other web pages in order for the server to be accessible by both internal and external customers. The internal customers actually go out of the firewall and then back in to access the collection. One other limitation was that there was no link between the abstract collection and the related full-text document online (if one existed).

Modification of Systems for New DOE Order 241.1 Requirements

In mid 1998, several complex-wide STI Program (STIP) initiatives led to the modification of SRS processes and systems. The STIP Goal Team 4, led by Jeanne Sellers, had identified acceptable formats for electronic submission of full-text documents to OSTI. OSTI developed a new metadata DTD and modified the submission requirements for bibliographic information. And, the STIP community developed the new [DOE Order 241.1](#) and corresponding Guide. In conjunction with these changes, SRS STI also had a milestone to begin publishing the full text of all STI conference papers and journal articles online.

In the summer of 1998, HTML was identified as the means of publishing full-text documents to the web and several documents were chosen as test cases. These documents were converted from their native word processing format into HTML, the related graphics files converted to GIF or JPG format, and the resulting files posted to the SRS external web site as additions to the collection begun in conjunction with the SGML full-text document submissions. This required some additional software for HTML conversions and graphics conversions, as well as training for STI personnel and the development of desktop procedures. Conversion to HTML worked well unless the document had numerous equations and mathematical or scientific symbols. Each equation could be converted into a graphic, but that was determined to be much too time consuming for a production process. It was, therefore, decided that those documents that contained equations or numerous mathematical and scientific symbols would be converted to PDF for publishing on the web. Beginning in October 1998 and continuing to date, the STI group has published nearly 100% of all approved STI conference papers and journal articles as either HTML or PDF files. Overall, the largest impediment has been obtaining a usable electronic copy of the document from the author. The education of the author population has been relatively slow, but the publication of their document in full-text via the web has proven to be a real inducement for gaining cooperation.

The implementation of the new DOE Order and Guide, as well as the new metadata DTD in October meant the implementation of some major modifications to the STI databases. The tracking database was modified to capture the new metadata information, the SGML encoding scripts were modified for the new metadata requirements, and the new 241.1 forms were added to the database for printing. The SGML-encoded metadata batch files were parsed against NSGMLS prior to submission to OSTI, and were initially submitted to OSTI as cc:Mail attachments. Once OSTI implemented [Energy-Link](#), the metadata batch files began to be submitted via that interface, still maintaining a weekly submission schedule.

Since October 1998, SRS has been meeting the OSTI goal of electronically submitting full-text documents by providing the URLs for the full-text files published on the SRS external web site as part of the SGML-encoded metadata file that is submitted. To date, only conference papers and journal articles (including peer-review journal articles) that have been approved for release through the STI review process are being electronically submitted in this manner. However, plans are in place to incrementally expand the process to include electronic submission of additional document types.

In December of 1998, a search engine was added to the SRS external web site. To take advantage of this capability, <META> tags were added to the <HEAD> section of each of the full-text documents published in HTML that already resided on the web site. The <META> tags incorporated include: document number, title, author, date published, site identification, authoring organization, OSTI Subject Category, and abstract text. The search engine can search on the <META> tags as well as the text within the document itself. The <META> tags continue to be included in the newly generated full-text HTML documents.

The STI abstract collection continues to be posted to the web via the Filemaker Pro server. This collection is updated on a monthly basis with records for newly approved STI products. Bibliographic records with abstracts are not generated for those documents published as full-text on the web. In February of this year, the database was modified to include a field for the URL, and the code was modified to provide a link from the displayed abstract to the corresponding on-line full-text document, if one exists.

Future Plans

SRS will continue to electronically submit SGML-encoded metadata batch files to OSTI using E-Link and will continue to publish a portion of the approved STI documents on the site's external web page. Other immediate future plans have been identified as a result of problems stemming from the current methods of capturing and publishing full-text documents. One of the most pressing needs is to obtain the equipment and software for scanning documents and generating good-quality text documents for conversion to HTML or PDF. Another need is to make the PDF files more "user friendly" by incorporating index information when the files are generated; and using section thumbnails and links for large, complicated documents.

Another issue is that of file server space for publishing the full-text documents. This will become more of a problem when additional document types are added as full-text products. STI is currently working with Information Technology (IT) personnel to identify a dedicated server for STI products. Initial evaluation has identified the possibility of utilizing an existing Windows NT server with an existing security plan for external web use. If such a server can be obtained for STI use, the full-text collection and the abstract collection would be migrated to this environment. This may also lead to the conversion of the abstract collection to HTML and the combining of the two collections into one.

The long-term future goal is to develop and implement an integrated document management and publishing system for STI products. This system would include the capability for tracking STI products through the review/approval process, for generating SGML-encoded metadata files, for storing and managing the related full-text products, and for providing user interface for search and retrieval of information. Since SRS will be transitioning to the use of Lotus Notes for electronic mail later this year, STI and IT personnel are examining the possibility of also utilizing Lotus Notes as an application platform. One of this product's strengths is in the area of information management. Full-text documents could be routed electronically for review and eventually stored in a related Lotus database. A powerful search engine, also part of the product, would provide users with the capability to search for and retrieve information in the metadata as well as the full text. STI hopes to be involved in a pilot of this product for our use later this year.

Conclusion

Over the past six years the SRS STI program has progressively published the site's scientific and technical information electronically. Starting with the publication of the STI abstract collection via the site-wide network, to the current publication of STI conference papers and journal articles on the SRS external web page, we've developed and utilized in-house expertise and capitalized on changes in technology. This trend will be continued with future endeavors to develop an integrated information management system and build a comprehensive, searchable, on-line catalog of SRS STI products.