Harvesting your STI:

as easy as 1 - 2 - 3

Background:

Within the Department of Energy (DOE), the Office of Scientific and Technical Information (OSTI) has the responsibility for coordinating scientific and technical information (STI) activities. This includes leading a collaboration to establish a distributed, electronic STI environment that meets Department-wide needs. For over 50 years DOE and contractor organizations have provided their STI documents to the DOE OSTI in Oak Ridge, Tennessee, per DOE Order 241.1A. As a result, the Department maintains one central archive and current record of DOE scientific and technical information from 1948 to the present.

The DOE STI community has transitioned from paper submittals of bibliographic and full-text information to electronic submittals. Now, through collaborations with stakeholders, OSTI is seeking ways to more easily acquire bibliographic information, while allowing the scientific and technical documents to reside electronically at the DOE/DOE contractor site of publication. This goal can be accomplished for unclassified, unlimited documents by "harvesting" bibliographic databases and by accessing electronic documents maintained at the sites.

What is harvesting and what are its benefits?

"My overall opinion of harvesting is that it's very efficient and cost-effective. It saves us staff time, allows us to announce our STI more quickly, and reduces the amount of record keeping that we have to do."

Dennis Elchesen, LLNL

Harvesting is the process for OSTI to retrieve STI bibliographic information in XML format from DOE web servers located at the various sites and organizations. Currently, OSTI harvests only DOE unclassified, unlimited bibliographic information from several DOE/DOE contractor web sites. These participating organizations post their full text on their web sites and make them accessible through a URL that is included as one of the fields in the XML bibliographic information.

Harvesting reduces the burden on the DOE/DOE contractor organization supplying STI to OSTI. The DOE organization will have the one-time effort of writing a program to generate XML from its database; then OSTI will harvest the XML periodically. The DOE/DOE contractor organization will only have to maintain its own database, not make any separate effort to supply the information to OSTI.

"Our IR staff are much happier with the level of effort involved…. All we have to do is make changes in our database, and the rest is automatic."

Pam Novak, PNNL

Harvesting your STI:

as easy as 1 - 2 - 3

What is required for an organization to participate?

1. **Electronic Scientific and Technical Information:** First, an organization must already have or be ready to establish a web site and make its unclassified, unlimited STI documents available through the web with unique URLs.

2. **Database of Bibliographic Information:** Next, the organization must establish a bibliographic database about scientific and technical information. The database should contain metadata about all of the organization's STI including technical reports, journal articles, conference papers and proceedings, etc. The database should be structured with fields such as title, author, research organization, sponsoring program office, date of publication, and either a URL to the full text document if it exists electronically, or information about the availability of the full text, including publisher notices for published journal articles. Two other required fields are: a unique identifier field for each record in the database and a date field showing when the record was added to the database or changed in the database.

3. **An XML Interface:** The organization must create a web accessible interface to its database that generates XML records. The interface would be accessible via a URL with two parameters, a start date and an end date. Records are then selected from the site's database, and XML records are generated if the add/change date of the record falls within the specified date range.

Routine Harvesting

Once the organization's XML interface is completed, OSTI and the organization will set up a harvesting schedule. OSTI will harvest the XML bibliographic information on the regular schedule (daily, weekly, etc.) through the URL set up by the organization. The XML records will be processed through OSTI's internal harvesting programs and become part of the collection of DOE scientific and technical information. Confirmation email with a list of records that have been harvested will always be returned to the organization.

"Benefits of harvesting include:

- **Consolidation of all publication records metadata into the NREL Publications Database**

- **Elimination of redundant electronic files**

- **Elimination of E-Link F241.1 submittals**

- **Reduced chance for data entry errors**

- **Automatic author notification**

- **Improved coverage"**

      Mary Donahue, NREL

Contact: Rita Hohenbrink at hohenbrinkr@osti.gov or (865) 576-1773

4/03