The Interagency Digital Library for Science and Engineering: A Federated Digital Library Pilot for the U.S. Government Scientist

Blaine Baker
National Air Intelligence Center
4180 Watson Way
Wright Patterson AFB, OH 45433
(USA) 937/257-3328
mbb28@naic.wpafb.af.mil

Dr. John Salerno
Air Force Research Laboratory AFRL/IFED
32 Brooks Road
Rome, NY 13441-4114
(USA) 315/330-3667
salernoj@rl.af.mil

ABSTRACT

CENDI [1] is a federal interagency working group of 10 programs from nine agencies responsible for scientific and technical information management. CENDI supports federal science research and the public's investment through sound information management and infrastructure development. CENDI agencies have developed individual digital libraries. However, when scientists need information from other agencies, they must search different digital libraries. To achieve interoperability in this heterogeneous environment, CENDI is piloting the Interagency Digital Library for Science and Engineering for use by federally funded researchers. The federated digital library technology being used in the pilot is Broadsword [2], originally developed by the Air Force Research Laboratory at Rome AFB for use within the intelligence community. It supports interoperability at three key levels: searching across heterogeneous platforms, data structures, and search engines; universal log-on; and government security levels. The CENDI pilot involves databases from three agencies containing over 10 million records. While the initial focus is on bibliographic and full text databases, the system can be extended to include other formats and document types.

Keywords

heterogeneous database searching; interoperable digital libraries; universal log-on; government security levels; government scientific researchers

INTRODUCTION

DL 99 Berkeley, CA USA
1-58113-145-3/99/08-212

CENDI is an interagency working group of federal scientific and technical information managers. Its members include ten programs from nine agencies **C** the Departments of **C**ommerce and **E**nergy, the Environmental Protection Agency, the **N**ational Libraries of Agriculture, Education and Medicine, the National Aeronautics and Space Administration, the Department of **D**efense (Defense Technical Information Center and the National Air Intelligence Center) and the Department of **I**nterior. CENDI's mission is to support federal science research and the federal investment in the intellectual capital created by this research through sound information management and the development of appropriate federal scientific and technical information infrastructure.

In order to increase the dissemination and re-use of the results of the federal science investment across the agencies, several CENDI agencies have historically shared hardcopy or microfiche government technical reports. Most recently, this process was improved by sharing electronic bibliographic records and image files of scanned documents. However, this process has meant reformatting and re-hosting the information. With the advent of intranets and Web-based information products, many of the agencies are branching out into non-traditional products and services, such as video, sound, linked web-sites, software, models and simulations, and multimedia. These non-traditional environments make it more difficult than ever to share resources and the results of the more than $70 billion in federal government research in the traditional way.

CENDI PILOT PROJECT

The agencies have begun developing digital libraries for internal and external use. While the agencies have attempted to use best practices for aspects of their digital libraries (such as metadata schemes and search engines), a truly homogeneous environment cannot be achieved. Interoperability must be sought in a heterogeneous environment, because the programs have differing standards and guidelines dictated by their agencies, variant needs of their

primary users, differing legacy systems that must be accommodated, and variant levels of budget support.

Therefore, CENDI is piloting an Interagency Digital Library for Science and Engineering for the federally funded scientists and engineers that is based on federating heterogeneous digital libraries. The key is interoperability rather than required homogeneity.

A pilot system that supports this multi-agency digital library is being piloted. The pilot involves databases from three agencies, the National Library of Education, the Department of Energy and the Defense Technical Information Center. These databases are the currently available databases that together contain over 10 million records. While the initial focus is on bibliographic and full text databases, the system can ultimately be extended to include other format and document types.

The federated digital library technology being used in this pilot was originally developed by the Air Force Research Laboratory at Rome AFB for use within the intelligence community. The system supports interoperability at three key levels: searching across heterogeneous platforms, data structures, and search engines; logon; and security.

The technology is based on the concept of plug-ins. Each agency builds a plug-in that maps the commonly defined fields to the fields for each database to be searched and also identifies unique data fields in each database. In addition to the data elements, the plug-in maps the definitions for the common search commands, parameters, and syntax. The server software, or the Gatekeeper, located at each site or at one or more network nodes provides an inventory of the databases that are included in the federation. The client software translates the user=s native environment to the common definition. The receiving client translates from the common definition to that required by the database to be searched. The results are provided back to the client for further processing and presentation. This processing may include ranking of the results, visualization, translation, etc.

In addition to providing heterogeneous multi-database searching, the system provides for a universal log-in. A government researcher can sign on to her local digital library and obtain access to the digital libraries of the other agencies. Based on user profiles, the universal log-on supports both variant styles of user ID/password control and variant levels of database access. The universal log-ons are controlled by the Gatekeeper software located at the network nodes.

The system also supports interoperability at appropriate security levels. It handles the full range of government system security, including unclassified, classified, secret and top-secret, and limitations on distribution required for some proprietary and administrative material.

In addition to this pilot, a task group within CENDI is exploring the possibility of adding information to the Gatekeeper to help in selection of databases based on subject categorization. Another group is analyzing the physical or virtual integration of agency controlled vocabularies to support cross-database subject searching.

REFERENCES

[1] CENDI. AWelcome to CENDI: Federal STI Managers Group.@ (www.dtic.mil/cendi/)

[2] Air Force Research Laboratories. AWelcome to Broadsword.@ (www.if.afrl.af.mil.bsword/)