Proceedings of the 5th Annual
Federal Depository Library Conference
April 15 - 18, 1996

Working Toward a Virtual Library

Laurie E. Stackpole, Naval Research Laboratory
Washington, DC

Introduction

The Naval Research Laboratory, NRL, is the Navy's corporate research laboratory. It is part of the Department of Defense, but relies on sponsor funding for the accomplishment of its research mission. The main NRL campus is located on a 130-acre site on the Potomac River in Southeast Washington. The other primary research locations are at Stennis Space Center in Mississippi and in Monterey, California. All together there are about 4,000 employees and about half of these are scientists, primarily physicists and chemists. The Laboratory also has a large complement of on-site contractors, with about 1,500 engaged in the research effort. The research efforts of the Laboratory are concentrated in 17 broad areas: acoustics, advanced space sensing, artificial intelligence, astrophysics, biotechnology, chemistry, condensed matter science, information technology, materials research, optical sciences, plasma physics, radar and electronics, radiation technology, remote sensing, space science, space systems, and structural dynamics.

The Ruth H. Hooker Research Library and Technical Information Center serves the residents of the D.C. campus and the headquarters staff of NRL's parent organization, the Office of Naval Research (ONR). In addition it provides a number of electronic services to NRL researchers in Stennis and Monterey. The library has been in the forefront in developing services that provide both its local client groups and NRL residents of remote sites with desktop access to information.1 A brief overview of these digital services provide a context for discussing the library's virtual library efforts.

Current Electronic Environment

InfoNet

In response to a 1990 User Needs Analysis that showed that researchers wanted information at their desktop, the library developed a campus-wide information system called the InfoNet. The InfoNet provides users with menu-driven access to a wide spectrum of information resources. These include commercial databases, both on CD-ROM and diskette and on-line through OCLC FirstSearch and other vendors; the library catalog; laboratory in-house databases such as the phone directory and supply store catalog; and selected

Internet resources such as the Library of Congress and University of Maryland Library catalogs. The InfoNet is available to users independent of computing platform both from work and from home. It's up 24-hours a day, seven days a week. The InfoNet went online in August 1992. It currently has over 800 registered users and is accessed remotely over 3,000 times a month and by end-users in the library another 1,500 times monthly. User surveys show that researchers credit the InfoNet with saving them 2 hours per week on average.

InfoNet provides users with access to text-based information. While some of this information is the actual data the user is seeking--for example, flight information from the Official Airline Guide--most InfoNet databases are the equivalent of catalogs and indexes, serving as pointers to information, which may be in a book, journal article, or report. Because of the nature of scientific publications, which contain many equations, formulas, charts, and graphs, the InfoNet only partially meets the information needs of NRL scientists.

TORPEDO

To provide all the information that is in the document, the library has developed TORPEDO (The Optical Retrieval Project: Electronic Documents Online).2 TORPEDO uses commercial software, called EFS (Electronic Filing System), from Excalibur Technologies. The EFS software provides end users with the ability to search the full-text of documents, using a fuzzy search algorithm. It also provides the ability to browse document collections, say all the articles in a journal issue, and to go directly to a particular article if the volume number and page are known.

Document images can be viewed online or printed locally. TORPEDO currently operates as a client-server system with freely distributable PC Windows and Macintosh clients and X-window support. A World Wide Web interface, currently being beta tested by over 50 NRL researchers, will replace the client-server version next month. TORPEDO has been available in the library since January 1995, throughout the DC campus since May 1995, and to remote NRL locations since September 1995.

InfoWeb

TORPEDO is available through the library's World Wide Web pages known as InfoWeb: http://infoweb.nrl.navy.mil. InfoWeb serves not only as a point of entry for TORPEDO, but also for many other library services, some the Web-based counterpart of the InfoNet databases. Furthermore InfoWeb serves as a navigational aid for a wide range of information on the Web that the library has selected for its relevance and content and organized to provide a subject approach.

As an interface to TORPEDO, InfoWeb provides a functional front end that tells the user about the system, provides User Guides in Acrobat format, and launches the Excalibur EFS (Electronic Filing System) software on which TORPEDO is based.

Building Digital Collections

Technical Reports

TORPEDO provides the library with the ability to deliver scientific publications to the desktops of NRL researchers. Technical reports published by DoD, and DoD contractors in industry and academia, are one important source of information for the NRL user community. Since 1988, the NRL Library has been digitizing and storing optically its large collection of technical reports. Begun as a space saving measure, with access from within the library only, the digital collection now consists of 125,000 technical reports (about 7 million pages stored as TIFF images). These images are being OCR'd and added to TORPEDO as a key component of the library's digital collections.

Journals

Journal articles are the other information source with which the library is principally concerned.

American Physical Society Projects

In 1993, the NRL Library and the American Physical Society (APS) agreed to work together to test the electronic dissemination of current journals. For this experiment, the APS provides NRL with copies of two of its journals, Physical Review Letters and Physical Review E, as they are printed. Issues are sent from the APS publisher by overnight mail. The library scans the journal to create the TIFF images and OCRs them to generate ASCII text for searching. This process is accomplished the day the issue is received so that it is available to the user community the next day, well before the paper copy is received in the library. The APS also provides, by e-mail, bibliographic data for each article consisting of title, author, and year of publication. TORPEDO provides access to all issues of both journals starting with January 1994, over 100 issues of Physical Review Letters and 28 issues of Physical Review E.

This effort is a learning process both for the NRL Library and for the APS. Through user surveys and other feedback we're learning what users want and need in a digital information system. By working together to disseminate scientific journals electronically, the NRL Library and the APS hope to determine:

- The attitudes of scientists toward electronic information;
- The attitudes of APS members versus nonmembers;
- The feasibility of disseminating journals in image format over campus networks and the Internet;

Researcher preferences for electronic format options (e.g., images versus page-definition files);

- The desirable features of future electronic journal systems;
- How publishers and libraries can most effectively cooperate in making electronic journals available to scientists, and how they can effectively integrate them with other materials;
- What kind of controls can be used to prohibit unauthorized users from accessing the system.

In November, NRL and the APS signed a Cooperative Research and Development Agreement (CRADA), under the authority of the Technology Transfer Act of 1986, to further experiment with search and retrieval systems for the dissemination of journal information. Under this agreement, the library is assisting the APS in creating an Internet-accessible digital archive of its journals. Called PROLA (Physical Review On-Line Archive), the system is being developed for the APS by computer scientists at the Los Alamos National Laboratory. NRL will scan the journals, starting with the years 1988 through 1993, and FTP the images to Los Alamos. The back files for the two journals already online in TORPEDO will be added to the NRL system as well.

Elsevier Electronic Subscriptions

While the availability of two journals has enabled the library to test TORPEDO and demonstrate its capabilities, it doesn't provide the critical mass of information needed to move researchers beyond casual interest. To provide that critical mass, the library will become one of the pilot libraries, the first in the Federal sector, participating in the Elsevier Electronic Subscription Project. Seventy-seven Elsevier journals are currently licensed for this project and are available for beta testing. The remaining 62 journals to which the library subscribes will be added later this spring. Elsevier Electronic Subscriptions provide TIFF images, unedited ASCII text, and bibliographic data in tagged SGML format, relieving the library of the chore of scanning and OCRing.

The library elected to receive and retain this information on CD-ROMs. The Elsevier journals are shipped to NRL on 4 CD-ROMs per month. A short program has been written to mount the Elsevier journals in native CD-ROM format, point to the defined directory structure, and index the full-text so that it is concurrently searchable along with the other documents in TORPEDO. Once the CD-ROM is received by the library, the new journals are available to the NRL/ONR community at 2 a.m. the next day (even on weekends), after the index has been rebuilt and restarted.

First-time access to an article requires that the appropriate CD be mounted, which, because this is a mechanical process, can take as much as 18 seconds for a single user. However, once an article has been used, it is automatically cached on RAID (Random Array of Independent Disks), so that all subsequent retrievals are essentially instantaneous. This arrangement allows the library to use low cost, but also slower, near-line storage, while assuring rapid access to frequently requested documents.

Planning The Virtual Library

A major thrust of the library's future efforts will be transitioning its electronic information services to the World Wide Web and expanding relationships with publishers and other document providers.

InfoVision/2000

Web access to all library information services, and to journal information in particular, was strongly recommended by a Study Team that met in February to evaluate the libraryþs vision and plans and to recommend to NRL management steps to effect a 21st century NRL

information environment. This study, called InfoVision/2000, was carried out at the request of the NRL Director of Research, by a 13-member team composed of NRL division heads, NRL and ONR researchers, and external experts from academia, other government organizations, and the publishing community.

Vision Statement

In preparation for the Study Team meeting, the library developed a vision and set of three goals for creating a digital, or virtual, library. The vision as set forth by the library and endorsed by the Study Team states:

In an increasingly digital environment, the Ruth H. Hooker Research Library and Technical Information Center will provide the Naval Research Laboratory with a centrally planned and managed information infrastructure to enable NRL and ONR researchers and other knowledge workers to meet their information needs conveniently, consistently, and cost effectively, anywhere and at any time.

Goals

To achieve this vision, the library adopted three goals, which, with minor modifications, were likewise endorsed by the Study Team. As modified these goals are:

- Build digital collections that effectively meet Laboratory scientific and technical information needs;
- Provide uniform, transparent, and convenient access to both locally mounted and world-wide network-based information;
- Provide leadership in creating a shared NRL knowledge base, maximizing the utility of the library's information storage, retrieval, and delivery capability.

Expanding Web Access to Information Resources

To meet Study Team recommendations, the library is working quickly to transition a number of its telnet-based InfoNet services to InfoWeb. In addition to the TORPEDO Web-based interface and the availability of the Elsevier journals discussed earlier, the library will introduce three other new InfoWeb services in May:

- Web access to the library's catalog of book, journal, software, and report holdings;
- Web access to the OCLC FirstSearch databases;
- Web online subscriptions to an E-mail tables of contents service for most journals in the library's collection.

Because it has found the networking of CD-ROMs to be a cost-effective approach for delivering information to the end user, the library is looking for a method to network CD-ROMs, currently only available through the telnet-based InfoNet system, via a Web browser. The alternative, licensing the raw data and providing a local search engine, is also under consideration.

Expanding Electronic Journal Access

Handing Other Formats in TORPEDO

During 1996, the library plans to develop its capability to handle the other formats that publishers are currently providing or plan to provide. Portable Document Files, such as Acrobat, seem to be the most likely alternative to explore at this time as more and more publishers are able to provide journals in this format. The American Physical Society has Acrobat files for its entire Physical Review series and has agreed to provide them to NRL for incorporation into TORPEDO. The library plans to use the corresponding SGML text, which the APS can also provide, as its searchable text database, eliminating the scanning and OCRing process entirely, and positioning the library to work with a greater number of publishers in disseminating journal information electronically.

InfoWeb Links to Publisher Web Sites

While many publishers are making journal information available through their own Web sites, each journal, or the Web-published journals of each publisher, must be searched independently. TORPEDO, on the other hand, allows a researcher to search across all the journals in the library's collection that are available electronically.

While access to many journals to which the library subscribes is currently available through InfoWeb hyperlinks, this is thought to be an interim measure that only partially meets the needs of the user community. In addition to requiring that a researcher know the publisher of a particular journal, it assumes that the researcher has narrowed his search to the point of knowing where to look for the information. Configuration issues, such as passwords and IP addresses, have also been difficult to negotiate with vendors for implementation on InfoWeb. In addition, response time from publisher servers has been found to be inadequate. Negotiations are therefore underway with a number of publishers providing site access to their journals on the Web to add these journals to TORPEDO instead.

Conclusion

The NRL Library is well on the way to providing its research community with a virtual library. With the InfoWeb System, it is providing researchers at their desktops with targeted, organized, annotated, and searchable access to a large number of local and remote information resources. With TORPEDO, it is providing powerful search and retrieval tools that enable end users seated at their computers or workstations to use Web browsers to search, view, and print the contents of large collections of library materials.

1. Atkinson, Roderick D., Laurie E. Stackpole, and John Yokely, Developing the Scientific-Technical Digital Library at a National Laboratory, Digital Libraries: Current Issues, Digital Libraries Workshop DL 94, Newark NJ, USA, May 1994, Selected Papers, Nabil R. Adam, Bharat K. Bhargave, and Yelena Yesha (Eds.), Springer-Verlag, Berlin, Heidelberg, 1995, pp. 265- 279.

2. Atkinson, Roderick D., Laurie E. Stackpole, and John Yokely, "Developing the Scientific-Technical Digital Library at a National Laboratory," Digital Libraries: Current Issues, Digital Libraries Workshop DL 94, Newark NJ, USA, May 1994, Selected Papers, Nabil R. Adam,

Bharat K. Bhargave, and Yelena Yesha (Eds.), Springer-Verlag, Berlin, Heidelberg, 1995, pp. 265-279.