# DEVELOPMENT OF THE NIST VIRTUAL LIBRARY

The National Institute of Standards and Technology developed its virtual library as an intranet. The Web-based user interface lets NIST researchers access internal databases, paid subscription services, and CD-ROM-based sources. It also accommodates restricted public access to general library resources.

SANDY RESSLER AND BILL TREFZGER
National Institute of Standards and Technology
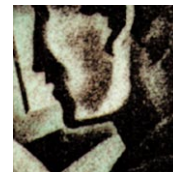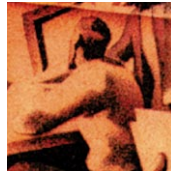
The NIST Virtual Library began with a vision of making the NIST library's services available at researchers' desktops. The library services come in a variety of forms, including CD-ROM databases and subscriptions to various information services. Providing a consistent, easy-to-use interface to these services was a primary goal of the NVL project, and integrating the NVL with existing online systems and databases was a major challenge.

The NVL is an intranet because the primary users are NIST researchers. The NIST staff consists of approximately 3,300 scientists, engineers, technicians, and support personnel, plus some 1,250 visiting researchers each year. However, portions of the NVL are available to the general public as a public service through the Internet at http://nvl.nist.gov. Library services that incur charges are restricted to internal NIST users; to the extent possible, the NVL makes the links to these services invisible to the public. This differentiation introduces complexity not typical of an intranet behind a firewall.

In this case study, we report our experience addressing the organizational and technical issues to meet these requirements. We present representative examples of information services made available by the NVL. In addition, we describe both the first- and second-generation implementations of these services to show how the NVL has evolved.

MOVING TO THE WEB

At the same time NIST began the NVL project, it also established a Cooperative Research and Development Agreement with Adobe
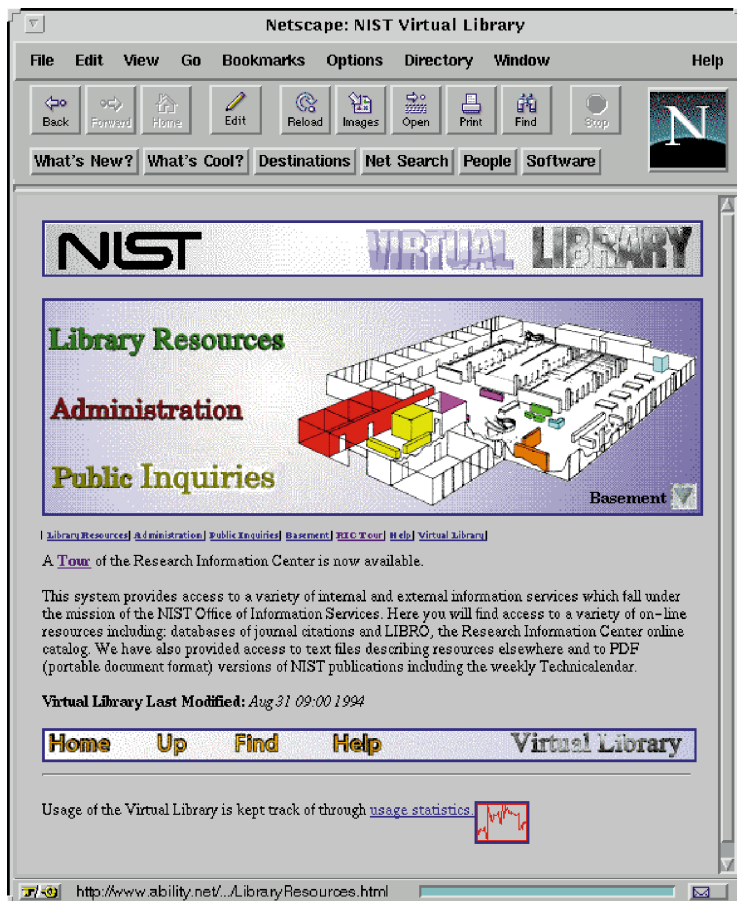
Figure 1. Original NVL home page design.

Systems, by which the company contributed copies of Acrobat Reader and the Distiller program for Mac, PC, and Unix platforms and NIST began to use Portable Document Format as an enterprise-wide document viewing format. This dovetailed nicely with the NVL, allowing us to place PDF documents online and know that users had the software to view them.

From the outset in 1994, we decided to unify the user's view of various information services with a Web-based interface. The classes of information we needed to make accessible included

- CD-ROM databases
- Internally generated databases
- Externally purchased databases
- Online services either internally run or externally purchased

Each information source presented unique problems. CD-ROMs require PC systems that execute access software, a challenge in a heterogeneous computing environment. Internally generated databases are created according to ad-

hoc organizational practices, and the flow of information is often unclear. Each database must be reworked for presentation on the Web—often a laborious task. Externally purchased databases and online information services must be secure from outside users because of the costs incurred, sometimes on a per-query basis.

ACCESS REQUIREMENTS FOR DIFFERENT TYPES OF USERS

Because the NVL is not completely isolated from the wider Internet, we had a number of user and security issues. Although some of these issues were addressed in the initial NVL implementation, more robust and general solutions became necessary as usage scaled up. For example, the initial NVL allowed anyone to access scarce CD-ROM resources, thereby tying up equipment intended for internal NIST use. We needed solutions that would provide adequate security and maintain convenience for the primary users—the NIST research staff.

The initial design forced internal users to go through an extra link to reach the specialized services provided only for NIST staff, and external users often saw links to information they could not access. We wanted to enhance the experience for all users by automatically presenting accessible information. We identified four major user groups:

- NIST users accessing the NVL from their desktops
- External users
- NIST users accessing the NVL from PCs in the NIST Research Library
- External users accessing the NVL from PCs in the NIST Research Library

We decided that requiring IDs and passwords for NIST users would be too large a barrier to NVL use. Network addresses offered a less intrusive way to control access. Accordingly, we wrote several CGI scripts in PERL to this purpose. We describe two examples here.

- *Access to NVL subject guides.* NIST Research Library staff have developed many electronic subject guides in areas such as physics and chemistry. To separate access between internal and external users, we decided to maintain two versions of each guide: one with premium services for NIST staff and the other with open services for the general public. Links to these pages pass through a script that checks the user's IP number (the standard CGI environ-

ment variable REMOTE_ADDR) and delivers a different document according to the IP number. The script checks access for users on affiliated networks at different locations as well as for users on the NIST main campus network.

■ *Access to database search services in the NIST Research Library.* Several of the database search services subscribed to by NIST charge by usage. Access to these services is restricted to NIST staff. However, the library facility is open to the public and includes six public workstations. We needed a mechanism to prevent public users from accessing these often expensive services without inconveniencing NIST staff. Again, a CGI script was written in PERL to generate a password prompt when a user tries to access these services. The script checks to see if a user is on a machine physically in the library, via IP number. If so, the user is instructed to see the reference librarian, who will ask for a NIST ID before giving a password. NIST users sitting at their desktop workstations do not encounter the password requirement.
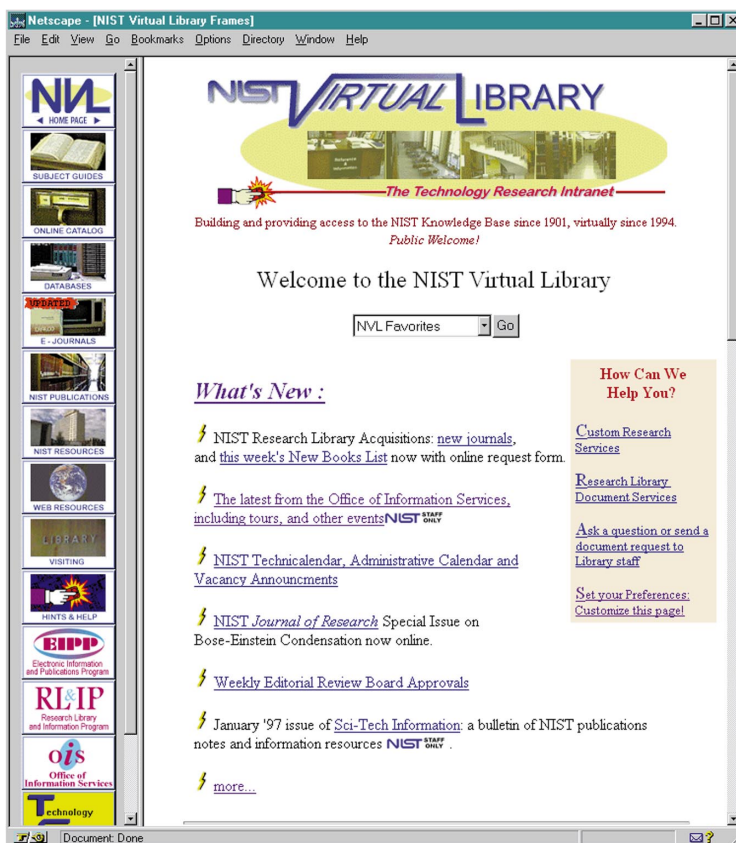

Figure 2. Current NVL home page with frames-based menus.

WEB PAGE DESIGN

The original home page interface displayed a large diagrammatic view of the library as an image map (Figure 1 on the next page). The page itself was visually balanced but gave users no feedback as to the function of areas on the image map. Instead, they had to click on the areas and then visit the help page for detailed instructions. Another problem was that functional areas were associated with the physical library layout—not a useful feature for people unfamiliar with the library. The desire to present a Web page that looked good had overwhelmed concerns about usability.

The NVL home page interface has been revised several times since its origin. In November 1996, a series of enhancements was undertaken to improve the user's overall NVL experience. We looked for ways to help users identify sources quickly and easily, and improve navigation among available services. Simply put, we wanted to emphasize the intranet features of the NVL and make them easier to use.

We started with the interface. Since researchers are likely to look for information across several information sources, we redesigned the interface to facilitate this. As shown in Figure 2, we employed Hypertext Markup Language frames to keep a menu of major information sources on screen at all times, while the changing documents visited by the user appeared in a display frame. This minimized revisits to the home page and use of the "back" button when switching between, for example, physics journals and a physics literature database.

In the spirit of meeting customer needs and not excluding users on the basis of technology, browsers that do not support frames can still use the NVL quite satisfactorily. Also, because frames are considered quite controversial from a usability point of view (for example, see Nielsen[1]), we added features that allowed users to turn off frames, or graphics, along with other customizations.

We also introduced other graphically based interface elements at this time. The three most significant of these were standardized buttons and banners for major sections or information components, an icon identifying premium remote information services available to NIST researchers, and an icon identifying links for "NIST Staff Only."

INTEGRATING INFORMATION SOURCES

We integrated information into the NVL from three types of sources: CD-ROM databases and their access software; off-the-shelf functionally complete databases, many of which are fee-based; and custom databases, some of which are for internal NIST use only. Descriptions of both the initial and second-generation solutions for each source follow.
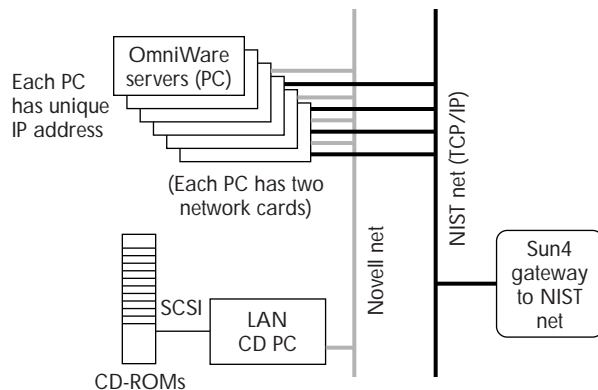
Figure 3. CD-ROM network configuration.

CD-ROMs

Our goal was to allow any researcher on the NIST network to open a "virtual PC" on their desktop workstation that could then execute PC programs with access to the library's collection of CD-ROMs. This would free the researchers from having to go to the library and sit at one of several CD-ROM stations, each of which stored a copy of the CD-ROM data. It would also cut the costs of maintaining separate subscriptions to CD-ROM publications, which are often very expensive.

Remote PCs. Delivering the information on CD-ROMs to researchers' desktops has been particularly challenging. In addition to data, the CD-ROMs contain the software to access it. This software must run on a PC that can access the data. Providing this service required a network of CD-ROMs and PCs. CD-ROM networks are not that unusual, but our requirements, such as delivering the information across a heterogeneous computing environment, complicated the implementation.

NIST had already purchased Logicraft's OmniWare* CD-ROM network system when the NVL project started, so it became another information service to be integrated into the overall NVL architecture. OmniWare had components that function with Unix platforms and PCs, although some of the Unix capabilities were new and hard to use. In particular, the X Windows version of the software had too many configuration difficulties for fonts and keyboards and was eventually dropped.

Currently, to access the CD-ROMs the user opens a telnet window to the virtual library CD-ROM server. The server automatically executes a process to start interaction with the PC CD-ROM network server, configured as illustrated in Figure 3. The server executes the proper sequence because of the telnet port being used. The connection process is invisible to the user, who simply clicks on the appropriate link on the Web page.

For a good overview of remote PC technologies, see Tom Halfhill's article in *Byte*.[2]

Second-generation CD-ROM access tools. The second-generation CD-ROM access tools employ WinFrame* technology from Citrix, Inc., and allow the use of Windows-based CD-ROM interfaces via a Netscape Navigator or Microsoft Internet Explorer helper application or plug-in. The overall architecture of the WinFrame communications is illustrated in Figure 4 (on the next page).

Users of MS Windows can link to specific CD-ROM-based database search systems running on a central applications server (in our case a Windows NT server). At this time, users of other platforms, Macintosh and Unix, must launch a separate X Windows session. Initially, this process invoked a menu of services like that used in the first-generation system; the latest iteration presents a Windows desktop with icons for each CD-ROM application. However, Intrigue Solutions now offers a seamless solution for Unix and Macintosh users with its Ntrigue* product.

Off-the-Shelf Information Services

Meeting the NVL project goal of providing a consistent, easy-to-use interface to a wide variety of information meant integrating online services that were remote to the Web server or off the NIST site completely. Some of them were fee-based. Since the services were functionally complete, integration was the main problem to solve.

We integrated both the NIST card catalog, called Libro, and the Online Computer Library Center's FirstSearch* bibliographic search service. Both services required users to know login procedures, including passwords. This was a major obstacle for occasional users. We needed to develop a simpler access method that was also secure against unauthorized users taking advantage of paid subscription services. The methodology we chose was based on Expect,[3] a Tcl-based toolkit.

Expect as system glue. Prior to the NVL, users would access Libro and OCLC by opening a telnet window and logging in. As with the CD-ROM databases, this required knowledge of a login and password. We used Expect to automate the initial login sequence. It lets you write a script describing the actions normally taken by a user—for example, the sequence performed to log in and start a process on a computer.

The telnet command contains an option parameter, which is the port number for the connection. The Expect script can be glued to telnet via these port numbers. The combination of Expect and telnet is sufficient to allow the user to simply click on a link and go to the selected service.

The connection from telnet to Expect is accomplished by placing an entry in the /etc/inetd.conf and /etc/services files. After these entries are in place, users simply connect to a service using the optional port specification of a telnet command such as telnet mymachine 1000. The actual telnet command is accomplished via gopher or Mosaic, and the user need not be

aware of this mechanism. Each service has a separate Expect script associated with it on a separate telnet port number.

To access Libro and the OCLC service, users simply go to the appropriate NVL Web page and click on the appropriate link, which connects to the appropriate server and opens the application.

Second-generation remote information services. Since the initial rollout of the NVL, it has become more common for technical information service providers, database vendors, and journal publishers to provide access to their products via the Web. Licenses and subscription schemes, along with the technical details to connect to these services, vary considerably. Hiding the access controls imposed by the vendor from users can simplify access. Local CGI scripts, HTML-embedded information, or prior coordination with the vendor all offer means to this end. Some vendors even provide detailed support or scripts for it, but this is the exception.

Access controls implemented by information providers are generally one of three types:

- *IP number protection.* Prior coordination with the vendor provides the network number ranges for which access is allowed. Once the range is set, no additional administration is required.
- *Password protections.* These can be embedded in the HTML links or forms, or in a CGI script, depending on the specifics of the remote system.
- *Registration-required systems.* Some vendors require each user to register, creating a personal account (in some cases, an institutional account is also required). While this is a cumbersome model, it might be appropriate for some types of services, such as online journals.

Access to information services through vendors that provide multiple publications or databases varies considerably. In general, it is preferable to link users to a specific information source. Often, however, you can only link them to the home page or logon level of a given service. Vendors such as OCLC and Cambridge Scientific Abstracts, for example, offer valuable databases in multiple disciplines, but allow access only through their top-level menu. Users must re-navigate menu choices or additional subject hierarchies, often after having already done that on the NVL.

Several vendors now CGI or Java-based clients for Z39.50 databases (Z39.50* is the library standard for machine interfaces to searchable databases). These Z39.50 gateways, which will be made public by the end of the year, offer simultaneous searching for internally developed databases, the Libro card catalog, and external remote information services, such as the OCLC FirstSearch databases or other library catalogs.

A final note here: Access to remote information services has been more a support and administrative challenge than
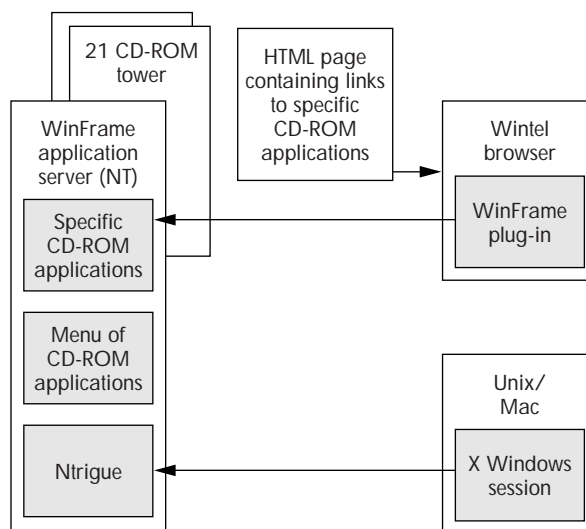

Figure 4. WinFrame CD-ROM application access over a network.

a technical challenge. Currently, the NVL provides access to over 60 remote commercial information services. URLs, access schemes, licensing fees, IDs, and passwords are constantly changing, and NIST users turn to NVL support staff to investigate and resolve access problems.

Internal Databases

We also made some internally developed databases accessible. The initial sets of information we dealt with were an awards database, a collection of Washington Editorial Review Board documents (part of the NIST document approval process), and the Index of Technical Experts. Each database required its own peculiar reformatting and processing to make it accessible. The conversions required substantial effort. These painful tasks can be ameliorated somewhat by documenting the flow of information through the management chain and developing enterprise-wide standards.

Making internal databases accessible. Databases are made accessible by creating indices for fast retrieval by search engines. We used two indexing programs:

- the widely available WAIS text retrieval engine,[4] which provides access to full-text databases, and
- a retrieval engine developed at NIST and known as the PRISE (prototype indexing and search engines) system.[5]

We used the PRISE system to index a technical experts database (now online as the Index to Technical Activities). We wanted to let researchers perform a search and then view an online electronic version of the entire resulting search doc-
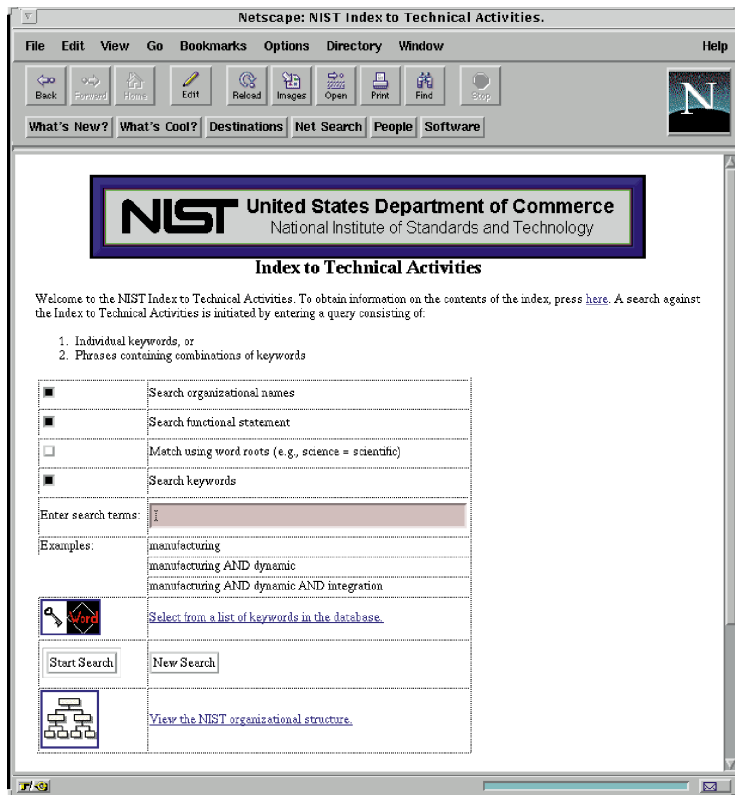
Figure 5. Public page of the Index to Technical Activities.

ument. We demonstrated this successfully in our initial experiment by using original Unix scripts. The search generated a list of documents with links to full-text PDF versions of them. This proof-of-concept was valuable, but too cumbersome to implement on a large scale.

The main organizational issue is having a clearly defined flow of information. Who owns the document? Where does the approval process come into play? Who creates the PDF version, and who updates the database index? How does the index get integrated with the Web, and who is responsible for it? All of these questions must be answered in order to define a process that can work.

Second-generation index. The Index to Technical Activities is an internal database that presents users a form-based front end, shown in Figure 5, to an SQL database.

The Index to Technical Activities is significant because of the administrative infrastructure developed for it. The goal is to have all divisions and groups at NIST perform their own maintenance and updates of the information. Page updates are password-protected by a person designated for the job. Information is entered by accessing the appropriate records through a Web page-updating form.

The goal has not yet been met, but only because the responsibility of performing updates has not migrated into normal working procedures. Further internal promotion is warranted.

Publications as databases. As it evolves, the NVL provides wider access to various NIST-produced databases and publications. Initial efforts to provide Web access to internal databases were successful, but required much customization. NIST publications were made available electronically on the NVL, as either HTML or PDF versions of the print products.

The latest generation of services being developed offer electronic versions of publications that have evolved into databases. The resulting product is both a database and a publication—information generated by the user in a format and at a time needed by the user. Here are two examples:

■ *The NIST Awards Nomination Calendar.* A schedule of technical award nominations, organized by subject areas, is distributed in print monthly by the NIST Office of Information Services. A PDF version has been available on the NVL as well. To provide direct database access via the Web, we employed Allaire's Cold Fusion.* This product incorporates SQL queries (as well as custom functions and logic controls) directly into HTML, producing dynamic HTML pages. NIST users can now produce a list of technical awards based on any of the available criteria—subject areas, schedules, technical society, previous winners, and so on. Continued distribution of the print version is being evaluated.

■ *The NIST Technicalendar.* The NIST Office of Information Services also produces a print publication containing a schedule of technical talks and meetings, at NIST and elsewhere. This popular publication is distributed to over 3,000 individuals each week. An HTML version of the Technicalendar is also on the NVL. A database back end for the Technicalendar is under development. It will be used to produce the current HTML and print versions. It will also add value to the information by supporting on-demand queries by subject areas, speakers, schedules, and so on. It will have none of the space limitations of the print-based version, thus supporting more detailed information, and it will establish a searchable archive of all technical talks and meetings at NIST or given by NIST personnel.

New information distribution methods offer opportunities to improve the production methods as well. The NVL intranet provides an environment for distributing internal

information and for gathering information from many parts of the organization. Rekeying and cut-and-paste can be reduced or eliminated via HTML forms that validate data. Corrections and updates to existing data can also be made more efficient in the familiar Web-based environment.

USAGE LOGGING FOR REMOTE SERVICES

Usage statistics justify the purchase and maintenance of expensive resources, both equipment and labor. Only recently has this information been captured for the NVL in a reliable way. Before the redesign, NVL usage hovered around 15,000 hits a month. After the redesign, it more than doubled the first month to more than 40,000 hits. In July 1997, there were 94,000 hits.

Standard logging provided by Web servers captures detailed information of document usage for documents that live on the Web server. However, a Web server cannot detail which links are being used on one of its HTML pages. Some Web sites address this issue by using an additional HTML page for each remote service they want to log. While accomplishing the goal of capturing usage, it adds an additional layer between the user and the information, as well as additional local administrative overhead.

We applied the same criteria to this problem that we applied to access control. Logging had to be seamless and not interfere with the user interface. This was accomplished with a CGI script that creates a log entry for each connection to a premium remote service. As with the access control scripts also installed at this time, the logging captures the HOST_ADDR environment variable, the user's IP number. This lets us compare usage of these systems from library PCs versus other NIST locations.

CONCLUSION

Fuchs's Law (named after its originator, Ira Fuchs of Princeton University, who cited the law in reference to Internet 2) states: "The time to acquisition is longer than the time to obsolescence." We feel its applicability to developing intranet or Internet services. Successfully addressing organizational issues, such as internal promotion and management, are as critical to the ultimate success of the project as solving technology problems is. Moreover, these issues are often much more difficult to resolve.

Usability studies were lacking in the original NVL design, and subjective analysis proved insufficient. The domain of Web usability is a rapidly developing field of study, and formal measurements of user preferences and task performance will ultimately result in more useful Web sites. While the redesign of the NVL has already increased usage significantly, only time will tell whether it will become an integral part of the NIST library system. Clearly, the NVL provides convenient desktop access to a huge amount of research information, but its success can only be judged finally by a high level of demand. ∎

REFERENCES

1. J. Nielsen, "Why Frames Suck (Most of the Time)," *Alertbox*, http://www.useit.com/alertbox/9612.html, 1996.
2. T. Halfhill, "Cheaper Computing Part 1," *Byte*, Vol. 22, No. 4, Apr. 1997, pp. 66–80.
3. D. Libes, *Exploring Expect*, O'Reilly, Sebastopol, Calif., 1995.
4. UsenetFAQ, "freewais-sf frequently asked questions [FAQ] with answers," http://www.cis.ohio-state.edu/text/faq/usenet/wais-faq/freeWAIS-sf/faq.html.
5. D. Harman and G. Candela, "Retrieving Records from a Gigabyte of Text on a Minicomputer Using Statistical Ranking," *J. American Society for Information Science*, Vol. 41, No. 8, 1990, pp. 581–589.

Sandy Ressler is a member of the NIST Information Technology Laboratory and a project leader for the Visualization and VR for Manufacturing Project. Since joining NIST in 1985, his work has focused on hypertext, document management systems, and virtual reality, primarily as they impact manufacturing. He has published three books with Prentice Hall—most recently, *The Art of Electronic Publishing*.

Bill Trefzger joined the NIST Electronic Information and Publications Program in 1996 as a computer specialist. He is currently developing the next generation of the NIST Virtual Library. He has participated in developing many systems to access government information at agencies such as the National Library of Medicine and the Defense Technical Information Center. He received a master's degree in library science from the State University of New York at Albany in 1985.

Readers may contact Bill Trefzger at william.trefzger@nist.gov.

URLs FOR THIS ARTICLE

*Allaire's Cold Fusion • www.allaire.com
*Citrix's WinFrame • www.citrix.com/winframe.htm
*Intrigue Solutions' Ntrigue • www.insignia.com/ntrigue
*Logicraft OmniWare • www.logicraft.com
*OCLC's FirstSearch • www.oclc.org
*Z39.50 • ds.internic.net/z3950/z3950.html