NASA Technical Memorandum 109718

Application of New Technologies in the Virtual Library

Seminars in Turkey, Portugal, and Spain

Judy F. Hunter
Gladys A. Cotter
National Aeronautics and Space Administration
Wsahington, DC

January 1994

NASA STI PROGRAM
SCIENTIFIC & TECHNICAL INFORMATION

The NASA STI Program ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program plays a key part in helping NASA maintain this important role.

The NASA STI Program provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program is also NASA's institutional mechanism for disseminating the results of its research and development activities.

Specialized services that help round out the Program's diverse offerings include creating custom thesauri, translating material to or from 34 foreign languages, building customized databases, organizing and publishing research results ... even providing videos.

For more information about the NASA STI Program, you can:

- **Phone** the NASA Access Help Desk at (301) 621-0390
- **Fax** your question to the NASA Access Help Desk at (301) 621-0134
- **E-mail** your question via the **Internet** to help@sti.nasa.gov
- **Write** to

  NASA Access Help Desk
  NASA Center for AeroSpace Information
  800 Elkridge Landing Road
  Linthicum Heights, MD 21090-2934

INTRODUCTION

This paper focuses on the technologies that are available today to support the concept of a virtual library. The concept of a "virtual library" or a "library without walls" is meant to convey the idea that information in any format should be available to the enduser from the desktop as if it were located on the local workstation. The physical and geographic location of the information as well as the techniques used to access it are not germane to the enduser. The enduser needs instantaneous access to knowledge and the tools to manage it.

Although advances in information technology are continuing to occur at a rapid pace, the ability to replace legacy systems wholly with new systems is not yet a viable option. First, not all technologies are right for or successful in every environment. Secondly, not all technologies are developed to the degree that it is possible to replace an entire schema. It may be necessary to use a phased approach to integrate new technology into existing systems. Finally, the library or information center may need to evaluate the current processes to determine which should be modernized and which should be discontinued. The introduction of new methods for accessing and using information sources may supersede current methods; for example, process re-engineering as opposed to improving current processes. Even so, there are technologies that are readily available and can be applied today to the virtual library concept. This paper focuses on those technologies and uses the National Aeronautics and Space Administration (NASA) Scientific and Technical Information (STI) Program for specific examples.

BACKGROUND

NASA is one of the major research and development agencies of the U. S. Government. The NASA STI Program was established as part of the Space Act of 1958. Its mission is to identify world-wide sources of scientific, technical, engineering, and related information; develop required policy statements; facilitate authorized access; and manage delivery of the information to NASA and its customer base. The NASA STI customers include NASA employees, NASA contractors, other U. S. Government agencies, other U. S. Government agencies' contractors, the U. S. education system (K-12 and universities), international partners, and the general public.

In addition to the traditional technical reports that are paper documents, more and more of the research is being conducted using electronic methods, such as the use of video to record experiments, and includes "raw data," such as data received from satellites. Even the paper documents are created using word processors to write the technical reports. The digital version of the reports is being stored on file servers. Consequently, the definition of STI within the U. S. Government has changed to include: "(1) basic and applied research that results from the efforts of scientists and engineers (including new theory and information obtained from experimentation, observation, instrumentation or computation in the form of text, numeric data or images) and which may be further transformed, described, evaluated, synthesized and recorded in print, digital, magnetic or other media to enhance its communication and its usefulness and value to a wide spectrum of endusers and uses, and (2) information that bears on business and industry generally, such as economic information, market information and related information, if the agency determines such information would be of value to consumers of the information described in the preceding subparagraph."[1] The definition was changed to reflect the methods being used to create and store information.

TRENDS

With the current emphasis on global competitiveness, successful businesses are organizations that optimize global resources. U. S. Government and private organizations must be able to locate and access information in a timely manner to maintain a competitive edge within the world market. Management of information as a resource will be a critical success factor for businesses in the 21st Century.[2] Success will be based on information literacy; that is, the ability to find and use information to construct a competitive advantage.[3] The trend is to view information as a critical corporate resource along with funding, facilities, staff, products, and services. The Chief Information Officer (CIO) is now at the vice president level and is responsible for the strategic management of information as a resource. Another trend in the information community is to transition away from the massive paper-based services to ones where the users have instantaneous access to the data in electronic formats. Largely due to advances in technology and a user demand for scientific, technical, and engineering knowledge in electronic formats, information organizations realize the need to make the transition. In the 1960s, data was embedded in the program where it was very costly and time consuming to modify. In the 1970s, structured programming methodologies were introduced that allowed patches to current computer programs for modifications. This made it easier, but the data was still contained on punch cards. In the 1980s, databases were developed that allowed the data to be linked to applications. It was at this stage that the data storage and the applications became separate. Today, in the 1990s, open systems architectures are available that allow applications to operate on different vendor hardware platforms. This stage brings flexibility to systems development for the first time.

In most cases, the large information providers, both government and commercial, continue to maintain large legacy systems that were developed within the last twenty to twenty-five years. These paper-based activities are supported by large mainframe computers. At NASA, for example, the online bibliographic database system is operated on a large mainframe computer that must be housed in a computer room where the temperature, humidity, and electrical requirements can be carefully controlled. The trend is to move away from these large mainframe-supported computers towards distributed, client-server computing environments.

In line with moving to a distributed computing architecture, the trend is to decentralize execution, but centralize the management to establish policies and control budgets. Due to the globalization of industry and a competitive world market, organizations demand information that is generated from their internal processes as well as external information, such as reports about research conducted in other countries. It is no longer necessary to put all information on a large mainframe computer. With the introduction of a distributed computing environment, it is now possible to store smaller, more specialized collections of information on local file servers and to use the available networking capabilities to make the information broadly available.

Distributed architectures help to keep the costs of maintaining the information lower. Local Area Networks (LANs) enable applications programmers, who have historically been located within the information/ computer system area with an organization, to move into the functional areas.

Another trend is to leave the care and maintenance of information with the creator or owner. The concept is that the functional area that created the data has a keen interest in assuring that the information is kept up to date and that it is accurate. This means that each functional area will collect specialized information rather than everything in the world. In order to be successful today, organizations must have *integrated systems*—that means they collect information once, then share it across all departments within the organization.

The trend in the 1980s was to implement Total Quality Management (TQM) to make current processes more efficient and cost effective in order to make the corporations more competitive. The trend in the 1990s is to re-engineer the process. Michael Hammer, who is a leader in the concept of "process re-engineering," maintains that continuous improvement does nothing more than maintain the status quo and that the only way to stay competitive and improve government is to *revolutionize* the current processes. Figure 1 on page 4 illustrates the idea that, over time, continuous improvement will cease to gain significant improvements beyond a certain point. To meet and exceed the requirements of the endusers, a revolution in innovation or a re-engineering period is needed to take the organization to the next level of achievement, which then becomes the status quo. At this point, continuous improvement is implemented again to gain another level of improvement for a period of time.

In re-engingeering, organizations are looking at ways to eliminate non-value-added processes. It is no longer a matter of making the current processes better; today, it is important to create a new process to achieve better results. Process re-engineering tries to identify the major changes that will result in a minimum of 80 percent improvement over the old process. Technology is the enabler.

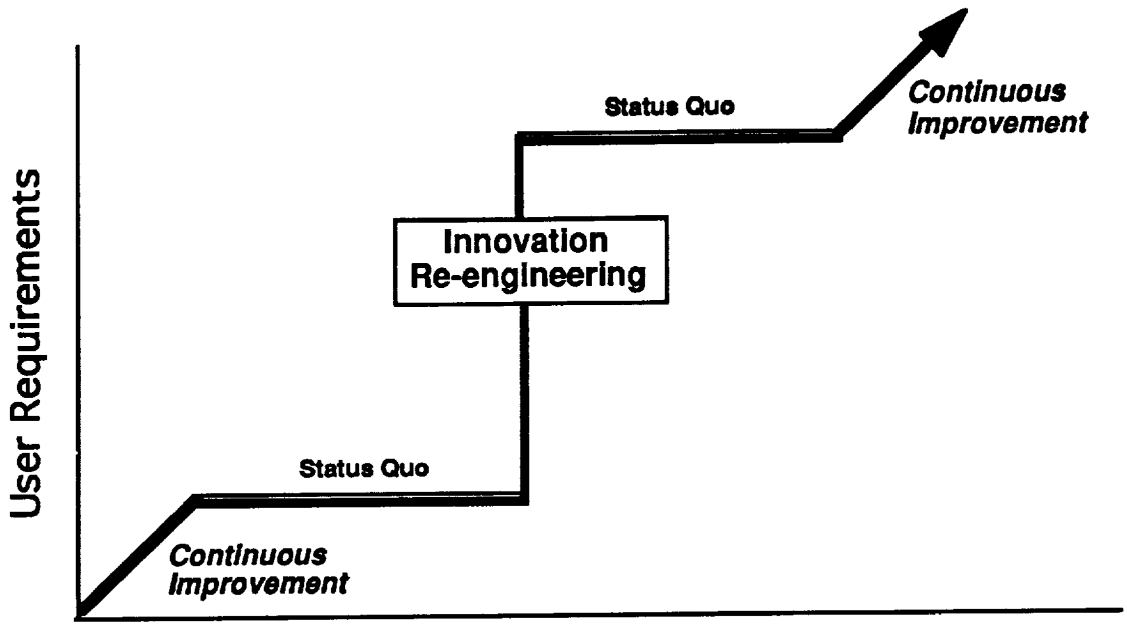

Figure 1: Continuous Improvement versus Process Re-engineering

Another trend is to use transmission technologies, such as Fax (digital facsimile) and electronic mail, to enable cross communications between organizational components. Organizations are using electronic messaging systems to introduce more functionality and flexibility into their communications infrastructure. The idea is to facilitate communications. In the corporate environment, for example, CEOs need to have access to the people who can provide valuable input into their decision-making process *immediately*. Corporate communications are more horizontal today because people are able to exchange information faster and more easily using electronic messaging. Corporations do not have the luxury of waiting for questions to filter down through the chain of command and for the answer to filter back up the chain. Multi-national companies must be able to respond quickly in order to remain competitive.

Electronic messaging systems are dependent on telecommunications functionality. Telecommunications provides the ability to dial into a centralized mail monitoring system to exchange electronic messages; to have a connection to a wide area network (WAN), which can be used as a mail gateway to transfer electronic messages; or to use Fax capability to convert a page of material—printing, writing, drawings, or photographs—into electrical signals and rapidly send it anywhere in the world over a telephone line.[4]

Another trend is to use rapid prototyping techniques 1) to prove the concept of new technologies to the endusers, and 2) to incorporate new developments into existing production systems. Rapid prototyping is a method for applying new technology to a critical user requirement where the user needs a solution fast. In most cases, the system needs to be operational quickly even though it may be a "throwaway." It is less risky to invest in a system that is developed quickly and inexpensively than it is to invest in a large system that is antiquated before it can be put into production.

Finally, large information technology vendors are dedicating a large portion of their corporate resources to developing new information technology products. Realistically, the key to being able to apply new technologies to current problems is dependent on the commitment of information technology vendors. If the vendors are dedicating a portion of their resources to research and development, the technologies will be available for use by public and private organizations. Many large, respected American corporations are incorporating long-range goals into their Information Resources Management (IRM) strategic plans to reflect a commitment to the information literacy that will be required to be successful in the 21st Century. Boeing Computer Services, for example, states that "Computing suppliers are gradually putting together an infrastructure which will allow electronic information to be readily, conveniently, and easily accessed and sent to others. The infrastructure will be the next century's information equivalent to today's worldwide voice telephone system."[5] The Apple Corporation has a video that proposes an integrated information management model for the future.[6] This video was created in 1987, and Apple is moving rapidly towards implementation. In the current version of the Macintosh operating system (System 7), Apple has included a "Play Movie" option that is a precursor to being able to create and render multimedia documents.

TECHNOLOGY ENABLERS

The networking infrastructure is a key element. In the United States, the Internet is heavily used. Internet is the name used for the network of networks that enables computers worldwide to communicate. It was developed by the Department of Defense Advanced Research Projects Agency in the late 1960s to facilitate the collaboration and sharing of information among government agencies, educational institutions, and industry. Internet is a top level meta-network that connects other networks in the U. S., Canada, Europe, Australia, and Mexico. For example, the Internet links to BITNET (U. S.), JANET (U. K.), ARISTOTLE (France), and EUnet (Europe).[4] Though the U. S. is the largest enduser, the international connectivity is rapidly growing. "In January 1993, there are 10 million people from 102 countries connected to the 9,000+ networks that comprise the Internet. The number of people is continuing to double annually."[7] The Internet is the foundation for the National Research and Education Network (NREN) initiative described as the backbone of the National Information Infrastructure being being promoted by U. S. Vice President Gore.[8] It is this backbone that is the basis for today's "glassing of

America," which involves laying fiber optic cable to handle high-speed, high-volume transfer of information across America.

Other networks around the world can be used to facilitate communications and the exchange of information. Examples of such networks are EARN, EUNet, and JANET. The European Academic Research Network (EARN) is a backbone network connecting national and campus networks into a continental network. There are nodes in all Western European countries as well as Cyprus, Turkey, Yugoslavia, and Egypt. EARN has gateways to EUNet and JANET. EUNet is a European-wide cooperative news and mail network used for research and development. Each member country in Europe, including Greece and Portugal, has an EUNet backbone host. The Joint Academic Network (JANET) was created to provide a backbone network for academic and research institutions in the United Kingdom and to provide access to public networks around the world.[4]

Distributed or client server computing is a way of downsizing by moving data off the mainframe computers onto smaller file servers (the server) that allow the endusers to access the server from their workstation (the client). Earlier architectures revolved around large mainframes that were the host computers for all the major applications of the organization (see Figure 2). The central database stored all the files that were needed to run the major applications. Time-sharing allowed individual users to access the central data files for information; however, most of the remote telecommunications were from people involved in providing input or using output from the system. System effectiveness was measured by how many transactions were processed per minute. All user views of data transmitted to enduser screens were formatted on the mainframe host. Ease

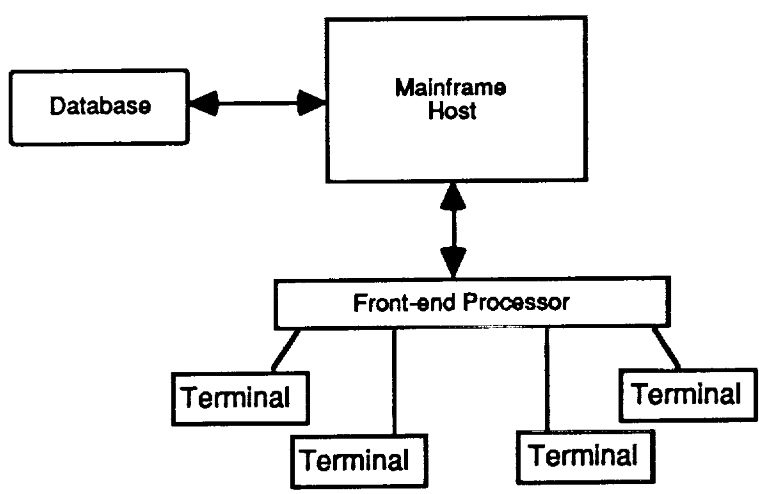

Figure 2: Mainframe Architecture

of use and tailoring the system to the needs of the endusers were secondary considerations. A separate processor, the front-end processor or FEP, was used to prepare communication input and output before sending it to or from the host. The host maintained all information about the database and would seek the files from the database needed by the enduser. Processing speed between the computer and the database and between the computer and the FEP was very fast. Processing speed between the enduser and the computer was very slow. The role of the enduser was to feed the mainframe, not the other way around.2

The client server architecture changes the focus from "feeding the mainframe" to what the user needs. Databases are segmented by subject and assigned to servers. The assignments are made on the basis of which departments use which databases. The user views and the formatting of the data are made on the workstations, not on the server. The server is a specialized system that provides the database engine as well as the communications processor (Figure 3). It is not involved in the multitude of applications performed on the mainframe. The client server is the specialist, rather than the mainframe.2

The significant performance characteristic of the server is no longer millions of instructions per second (MIPS), but what may be called networks instructions per second, or how fast the server can move huge blocks of data through the system. The software that facilitates this process is the new relational and distributed database systems. These systems dramatically change the ways applications are developed. The database software handles the tasks relating to database management, data dictionary, data security, additions and deletions to the file, processing of

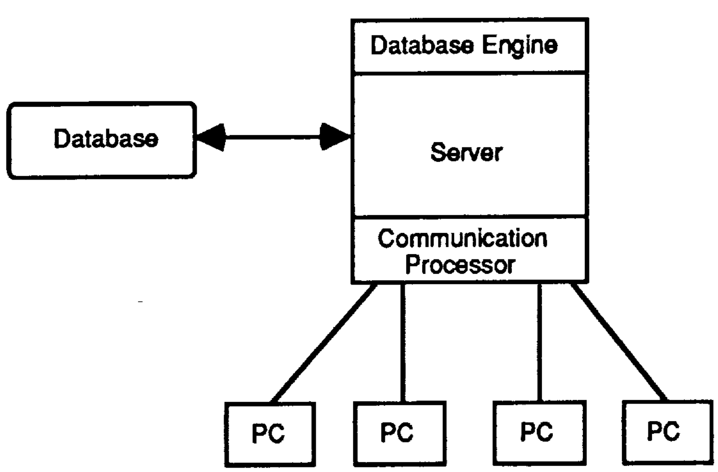

Figure 3: Client Server Architecture

user queries, and all the tasks that translate the requested information from the database into the formats prescribed by the various users. The user defines the screen formats and report layouts in a language that is easy to learn and easy to use. The system is now optimized around the user, and as such is much more responsive and flexible.[2]

One way to envision the client server architecture is to think of an electronic mail application. The server maintains the database, which in this case is the mailbox of everyone using the system. The server also maintains the passwords and mailing lists of the users. The client prepares the message he or she wants to send, and indicates the addressee(s). That is all the client does. The server handles the communications linkages and the necessary database (mailbox) accesses, and completes the transaction. The client can connect to more than one server to integrate data from another database if necessary. Distributed databases facilitate this interserver connectivity.

Client server architecture is a major change in the way information is managed. Because there was a heavy investment in the past in hardware and software, full-scale utilization is just beginning to happen. "Those manufacturers who have a big stake in the mainframe business do not want to see client servers replace their centrally based architecture overnight. The commercial beneficiaries of the client server market are the mini and micro vendors, as well as the telecommunication suppliers, particularly local area networks (LANs). The software players expected to profit are the relational and distributed database vendors."[2]

Multimedia information sources are rapidly becoming more common. Multimedia integration facilitates the electronic capture of data, not only as text, but in graphic, pictorial, and voice format. "Today less than five percent of business data are in electronic form and of the five percent, over 90 percent of the data are text."[2] However, in the research community, multimedia technologies are being used more and more to collect and deliver STI. The information includes data, computer models, scientific visualizations, software, video records of experiments, formulae, text, and graphics.[9] "Key to adequate delivery of such information is its integration into a unified multimedia whole from which the user can easily get the information wanted in the form wanted. Thus, the most obvious benefit of multimedia information presentation is the ability to deliver information in its most effective format so that there are no barriers to information understanding."[9] Multimedia information sources can also provide a high level of interactivity, allowing the enduser to comment back to the publisher using voice or video annotation, to manipulate a live graph or dataset with his or her own numbers, and to enter an interactive, multimedia conference to discuss common issues with other readers of the publication.

There are successful multimedia applications available. One application of note is the Artic Data Interactive (ADI) CD-ROM prototype that was produced by the U. S. Geological Survey to disseminate global change data and information. ADI includes animation and sound, but no full motion video. Another example is "Columbus: A Journey to Discovery," which was produced by the International Business Corporation (IBM). This initiative includes 180 hours of interactive material that provides multiple points of view on its content.9

Graphical User Interfaces (GUIs) and navigational tools are becoming readily available. GUIs are being developed today so that the enduser can "point [a mouse] and click [on the icon]" to select and activate functions. Intelligent Gateway Processor (IGP) technology was developed in the 1980s to provide tools to navigate the telecommunications infrastructure and mask the complexities from the enduser. GUIs that are intuitive and navigational tools will provide endusers with seamless and transparent use of high speed networks to access, organize, and manage information.

THE VISION

What, then, are today's enduser requirements? In 1992, the NASA STI Program surveyed the NASA user community to find out. The users told us that they want 1) "one stop shopping" from the desktop with an intuitive interface; 2) comprehensive identification of information sources, including people; 3) transition from the massive paper-based environment to an optical-based environment; 4) desktop access and delivery; and 5) online translations. Basically, the NASA user community is requesting the development of a virtual library or a "library without walls."

Historically, when an enduser (student, researcher, CEO) needed information from outside his immediate working environment, a trip to a library was essential. A library was a physical building.

The technology is available today, though in various stages of development, to establish the "virtual library" environment shown in Figure 4 on page 10. In this scenario, it is assumed that much of the information needed today is either in an electronic format or that there are electronic records, such as those contained in a bibliographic database, that point to the sources available in a paper format. In any case, most document ordering facilities are electronic.

Most endusers are connected to a LAN within NASA. The functions available on the LAN minimally include a word processing package, a mechanism to send and receive electroni mail, a spreadsheet package, a database management system, and maybe even a fax capabililty. Also situated in the local enviroment are sources of information that are generated from within the organization, such as data on a mainframe computer or on local file servers that are accessible and may be shared by all or by designated segments of the organization.

For information that is required from sources outside of the organization, the LAN and/or the enduser workstation may be linked via

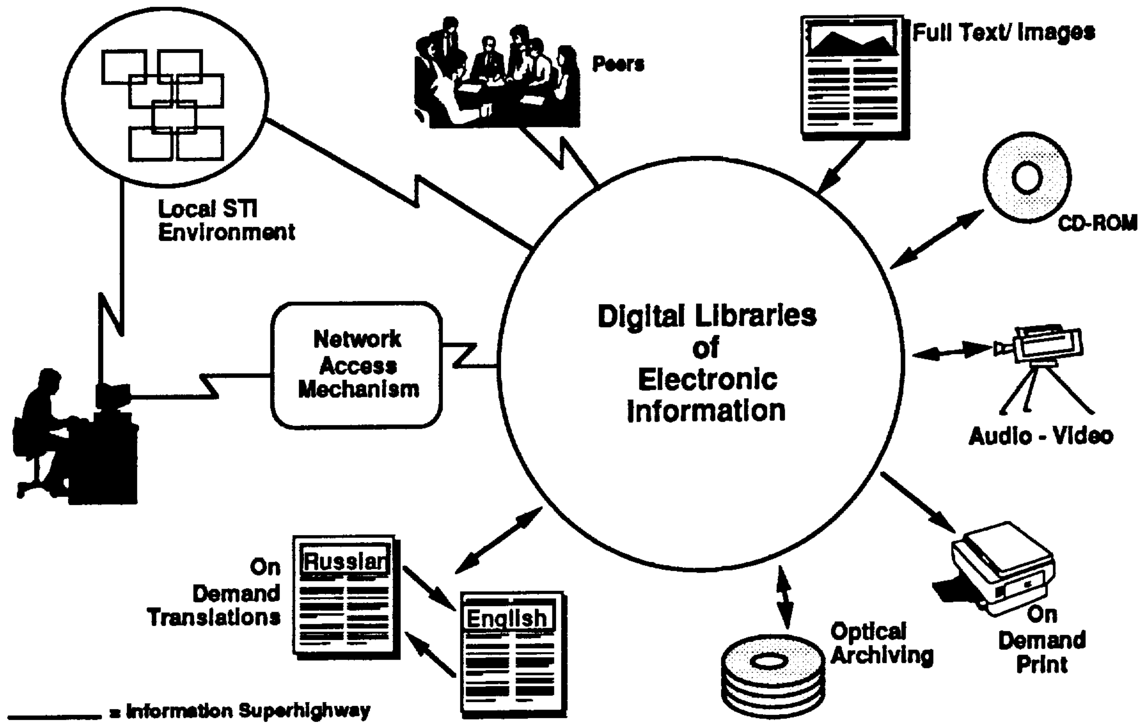

Figure 4: Integrated Information Approach

telecommunications to other LANs or WANs such as the Internet, or via direct dial-up lines to specific hosts.

The collections of electronic information are at the heart of the virtual library. In the virtual library schema, such collections are a part of the information infrastructure. That is, they are connected to the communications infrastructure in some way and are accessible electronically. These collections of information, or "libraries," include people resources, which are being recognized more and more as critical information resources. In the research environment, for example, the scientists communicate on a regular basis with a small, select group of experts in the same discipline. To be accessible through the network, the people require an electronic mail address or a reference to their telephone number or postal address someplace within the network . The idea is to facilitate communications. If this means providing a telephone number, access has been facilitated successfully. The better application is to be able to send the "resource" via electronic mail or even to transfer a full file to that person for review.

The digital libraries include stores of complex documents that contain full text and images—the information located on CD-ROMs, audio and video, and other optical media. Especially in the research environment where the scientists are using sophisticated modeling techniques on supercomputers to perform experiments and conduct research, endusers expect to be able to get data in various formats on various media in foreign languages.

The goal is for the enduser to have immediate access to relevant information in any format from the desktop. It is not enough to give the enduser all the information that is available on his topic. The requirement is to provide *only* the specific information needed by the user. In the past, the enduser was flooded with all the information ever created about a particular subject. The enduser didn't have the time or the patience to sift through massive volumes of information to identify those pieces that were relevant. The enduser needed to be in control of the information as opposed to the information controlling the enduser.

Key to the success of the virtual library concept are tools to identify sources of information, navigate the communications infrastructure to get access to the sources, and analyze the data for specific requirements. The GUI must be intuitively easy to learn, assist the endusers' identification of relevant sources of information, facilitate access to those resources, and provide tools for analysis of the information.

DEVELOPMENT EXAMPLE

The NASA Access Mechanism (NAM) system is one example of how it is possible to apply existing technologies to the client server architecture to logically centralize geographically distributed applications and information. It was developed in the last two years at NASA to provide "one stop shopping" to the NASA user community. Historically, the enduser had to know about a source of information and then be able to figure out how to gain access. If the enduser gained access to information in a database, he then had to learn an often complex query language in order to retrieve data from the system. The enduser then had to put the data in a format that was usable for his application. In the client server architectures and GUIs that are being developed today, it is possible to create an environment for the enduser that uses "point [a mouse] and click [on the icon]" methodologies to simplify functions and mask complexities.

NAM was developed to meet specific requirements of the NASA user community, using the technologies that were available at the time of the initial design in late 1990. The enduser requirements included 1) access to a diversity of information to support research, development, and operations; 2) an intuitive approach to decrease the learning curve; 3) elimination of the need for the user to learn the system/database specific query languages; 4) access to peers and other "informal" information; and 5) a simplified and enhanced presentation of the search results. The initial functionality for the system included access to NASA and non-NASA databases, and SpaceLink; communications access via the NASA Science Internet and dial-up capabilities; a database of databases; search assistance by discipline; peer locator services; electronic mail capabilities; bulletin boards; and other Internet tools. These capabilities are described in more detail in the following paragraphs.

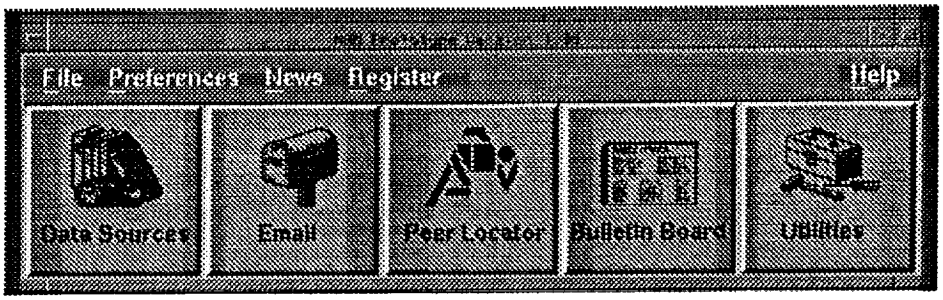

Figure 5: Prototype NAM Main Menu

The NAM functions depicted in Figure 5 include a source locator that searches a database of the available sources to select those sources that cover the subject of interest to the enduser.

If the enduser knows the database he wants to query, he can choose that host as one of the options available within the **Data Sources** button. For the prototype, the available databases include the NASA REsearch CONnection (NASA/RECON), our bibliographic database; the Aerospace Research Information Network (ARIN), which is the online catalog of the holdings for the NASA technical libraries; selected files from the Scientific and Technical Information Network (STN), a commercial information provider; and the European Space Agency (ESA) databases. A sample of this screen is displayed in Figure 6.

On the other hand, the enduser may not know which database contains the information he needs. In this instance, the enduser may enter a particular area of interest, such as "aeronautical engineering," and the system will recommend the database that contains the most information about aerospace sciences. This screen is displayed in Figure 7 on page 13.

It is within the **Data Sources** function that the connections to the host computers occurs—automatically and transparently. The system is designed to "know" which communications paths are available to reach a

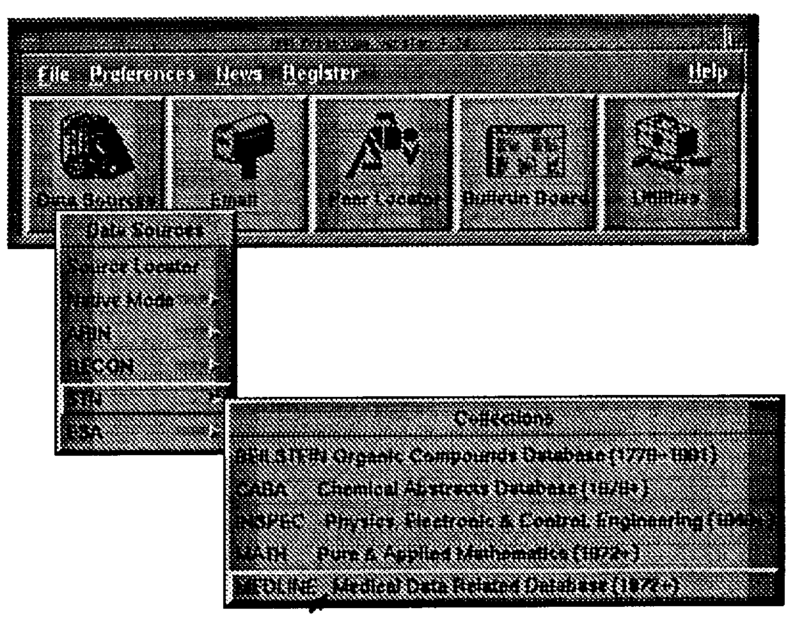

Figure 6: Available Data Sources

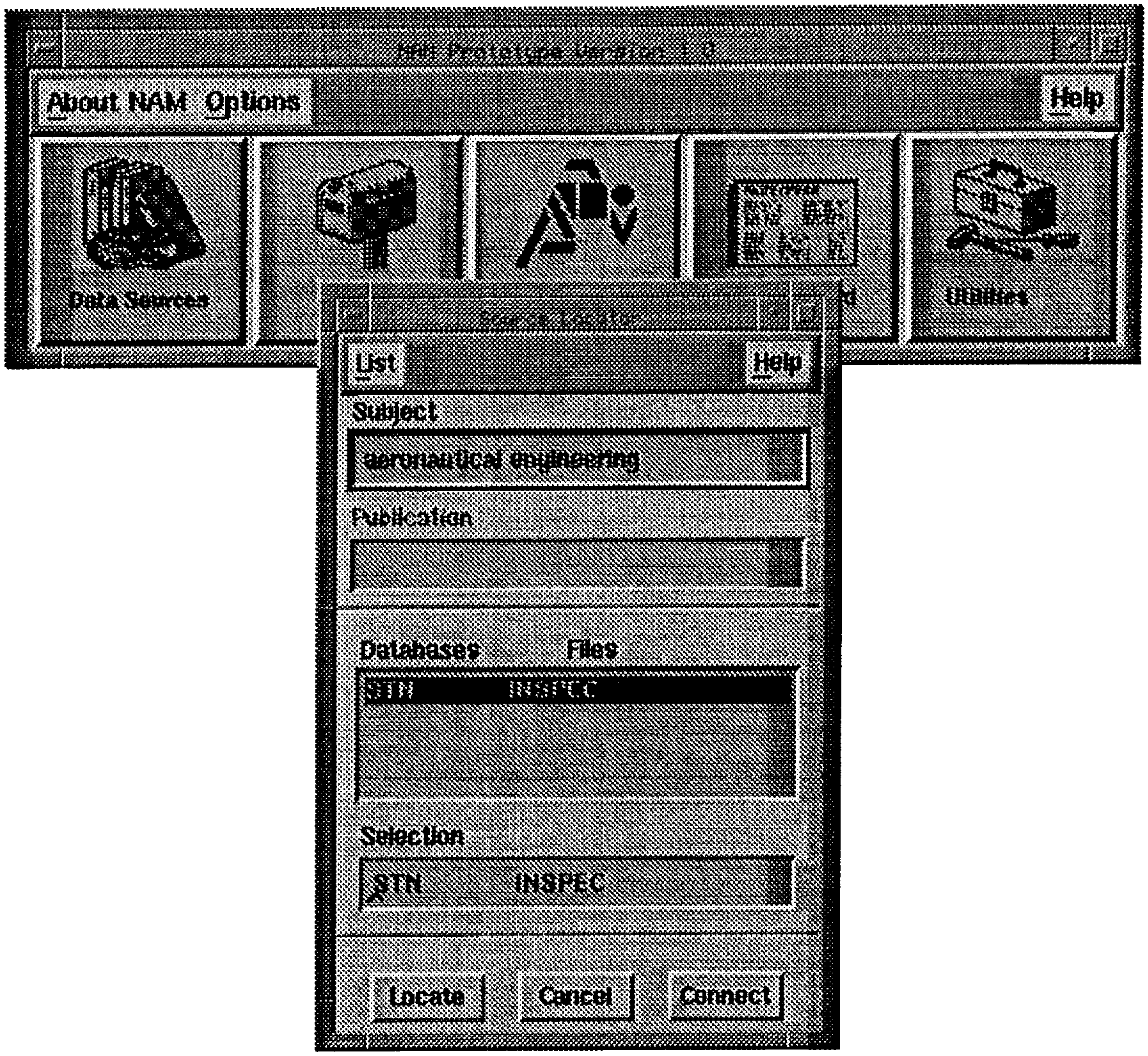

Figure 7: Best Data Sources for Specified Subject Areas

particular host and includes a list of priorities. That is, if the Internet connection is not working, the system will try an X.25 connection or a direct dial-up connection.

These facilities are all provided through an iconic, form-based interface, where the user fills in a form with the keyboard or selects options with the mouse to prepare a query. Query screens are available for novice, intermediate, and expert users. Figure 8 on page 14 illustrates the **intermediate** input screen. The NAM handles connection to the remote system, selection of the appropriate application, and processing of the query at the remote host. The NAM also presents the results of the remote session to the enduser in windows that allow the user to perform further actions on the retrieved information.[10] Figure 9 on page 14 depicts the bibliographic citations retrieved from a database system. The citation includes the bibliographic information along with a full abstract. To the left of the citation, the user has the choice of displaying another citation simply by

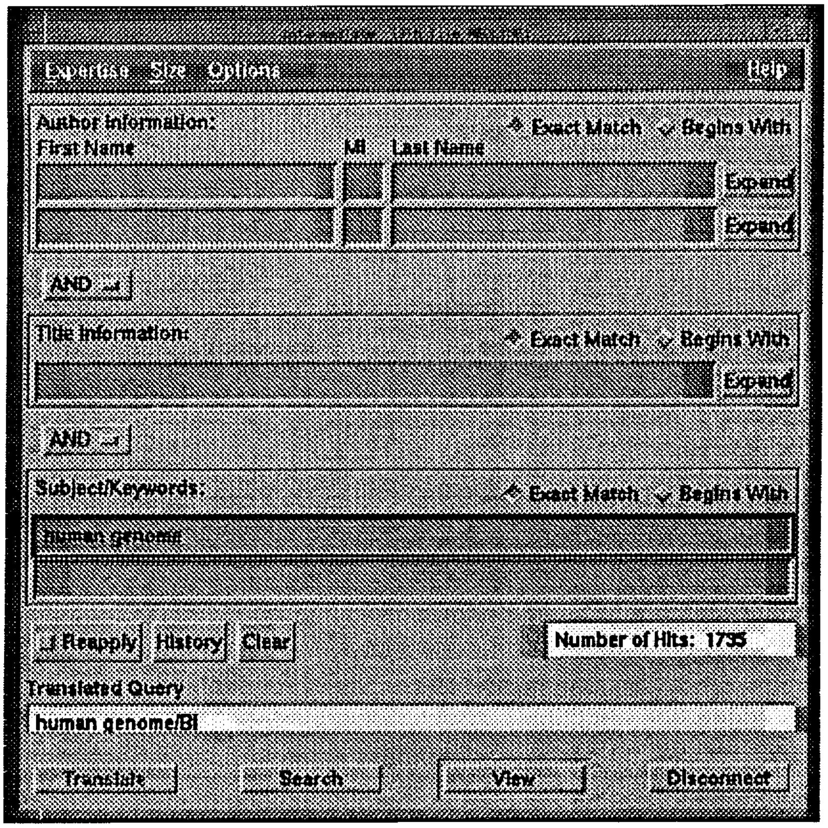

Figure 8: Intermediate Input Screen

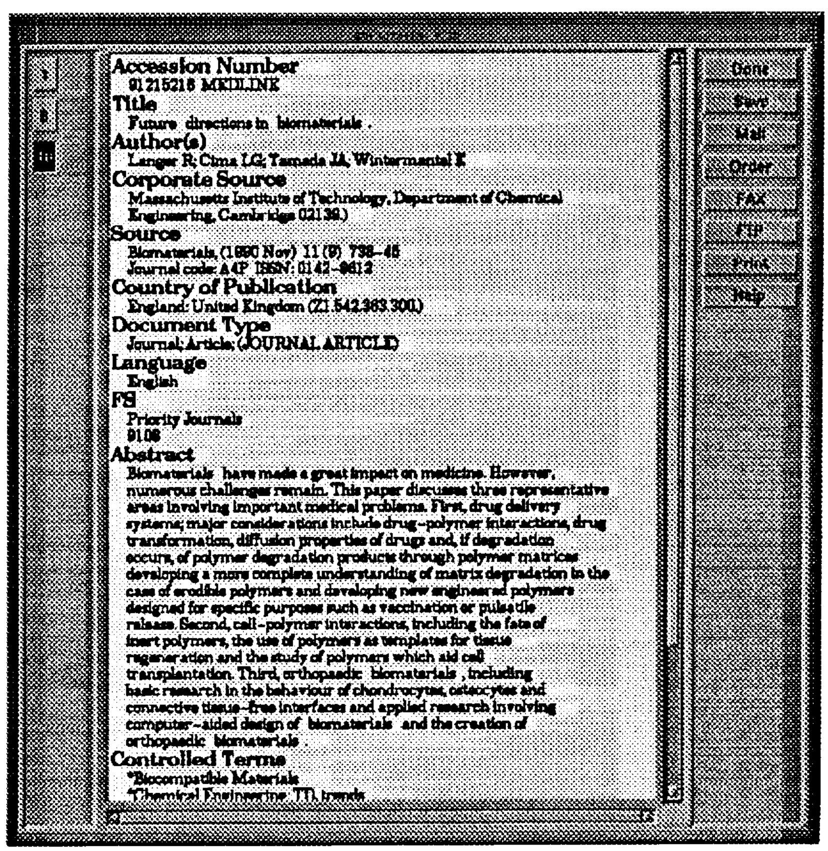

Figure 9: Display Citation

clicking on one of the boxes. To the right of the citation, the user has options for further processing, including the ability to **save** the file, **mail** the file electronically to himself or another user, **order** the full document from his local library, **Fax** a copy of the citation, use the Internet **FTP** utility to send the file to another Internet user/server, or **Print** the citation.

The **EMAIL** function gives the enduser access to an electronic messaging system that he can use to send electronic mail to other users around the world. This function can also be used to mail any information that is downloaded from a host computer to himself or a peer.

The **PEER LOCATOR** function facilitates communications between peers (colleagues). The options available under this button are illustrated in Figure 10. For assistance in locating and communicating with people outside of NASA, an Internet utility called **FINGER** was added that will search specified hosts (the host must be running Finger) to see if the person in question has an account on that system. It typically shows full name, last login time, idle time, terminal line, and terminal location (where applicable). The **NASA Phonebook** is a database that contains the name, address, telephone number, and electronic mail address of NASA personnel located at thirteen NASA Centers and NASA Headquarters. **Whois** is another Internet program that allows users to query a database of people and other Internet entities, such as domains, networks, and hosts, kept at the Defense Data Network Information Center. The information usually includes a person's company name, address, phone number, and electronic mail address. **X. 500 White Pages** is the Consultative Committee on International Telegraph and

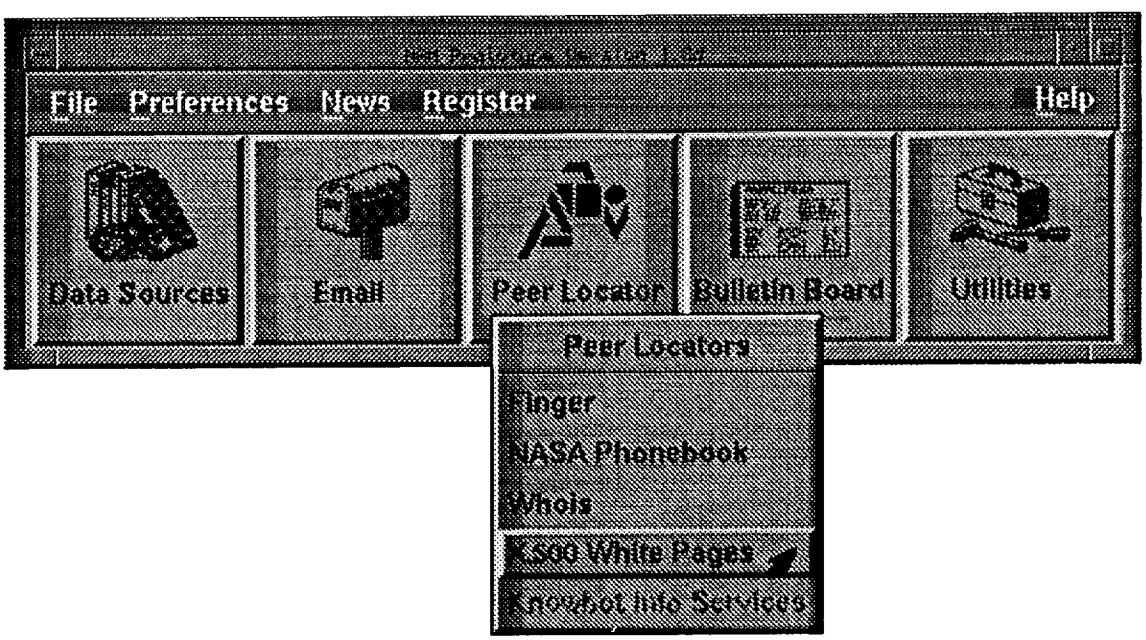

Figure 10: Peer Locator

Telephone (CCITT) and the International Standards Organization (ISO) standard for electronic directory services available on the Internet. NAM automatically inserts the electronic mail address into the electronic messaging system to facilitate sending mail.

In the research environment, scientists rely on information they exchange with their peers informally (first and foremost), on information that has been published in journals and has met peer review requirements for publication (secondly), and, finally, on information they may find someplace that they can validate themselves if it seems relevant.

The **Bulletin Board** button, displayed in Figure 11, includes access to volumes of information located on the Internet on public and private file servers that fall into the third category. In some cases, the enduser wants to get a general idea about any information that is available in a certain area. **Usenet News** is a worldwide bulletin board that is organized into a "news group" hierarchy. There are over 1500 different news groups available. **WAIS** is the Wide Area Information Server that was developed by Thinking Machines, Incorporated that is based on the Z39.50 application layer protocol. WAIS is able to locate and retrieve information using English-like queries and a relevance feedback mechanism. It supports text, image, and voice. There are several hundred WAIS servers accessible over the Internet currently. **Archie** is a system that maintains an index of a large number of Anonymous File Transfer Protocol (FTP) sites on the Internet. The initial implementation provided an indexed directory of file names for all Anonymous FTP archives on the Internet. Later versions provided other collections of information. **World Wide Web** is a hypertext-based distributed information system created

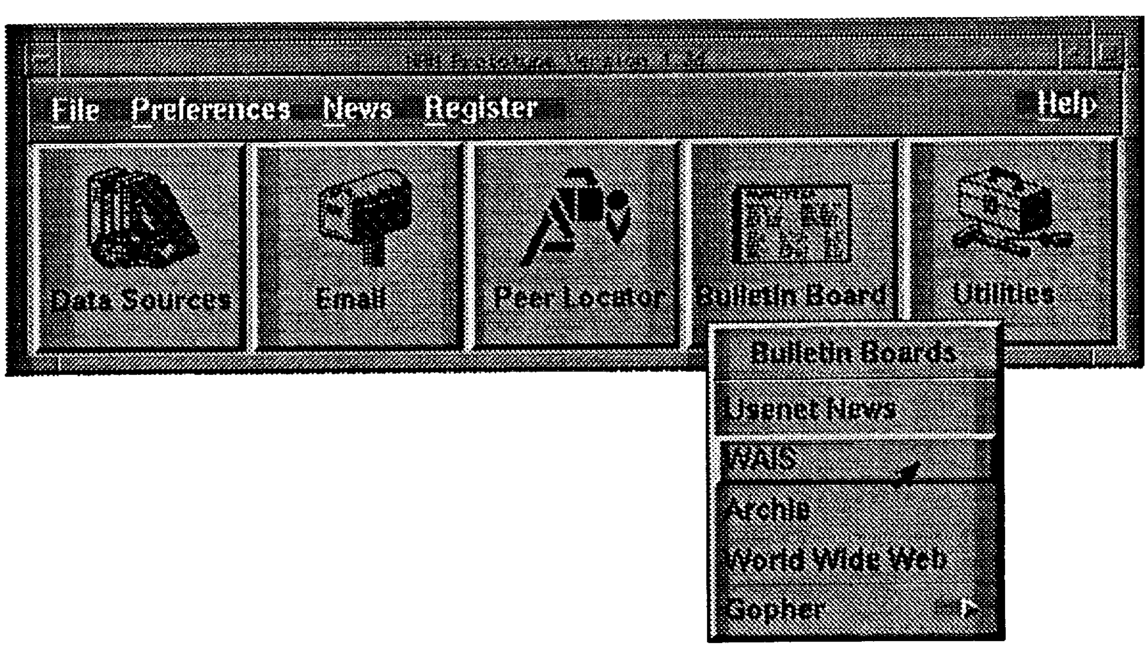

Figure 11: Bulletin Board Functions

by researchers in Switzerland. Users may create, edit, or browse hypertext documents. **Gopher** is a distributed information service that makes available hierarchical collections of information across the Internet.[11]

The endusers need tools to analyze the information. In most environments, some type of 3-D modeling capabilities are required. The **Utilities** button will eventually provide the enduser with tools to assist in the analysis of the information that has been retrieved. The NAM prototype's graphics capability is illustrated in Figure 12.

The NAM prototype was developed in a distributed, multi-vendor platform environment, using available standards. It consists of eight separate modules that communicate with each other using a clearly defined protocol. The underlying mechanism used to provide this support is based on the Inter Process Communications (IPC) abstractions provided by the UNIX Socket Interface, which supports peer-to-peer communications primitives well suited for client server and distributed applications architectures.[10] The NAM Server includes the communications tool kit, the intelligent resource locator, the database of databases, and an electronic mail book.

The NAM client software is a custom built windowing application, tailored to the requirements of the STI Program user community. The GUI is built using the X-Windows system and the

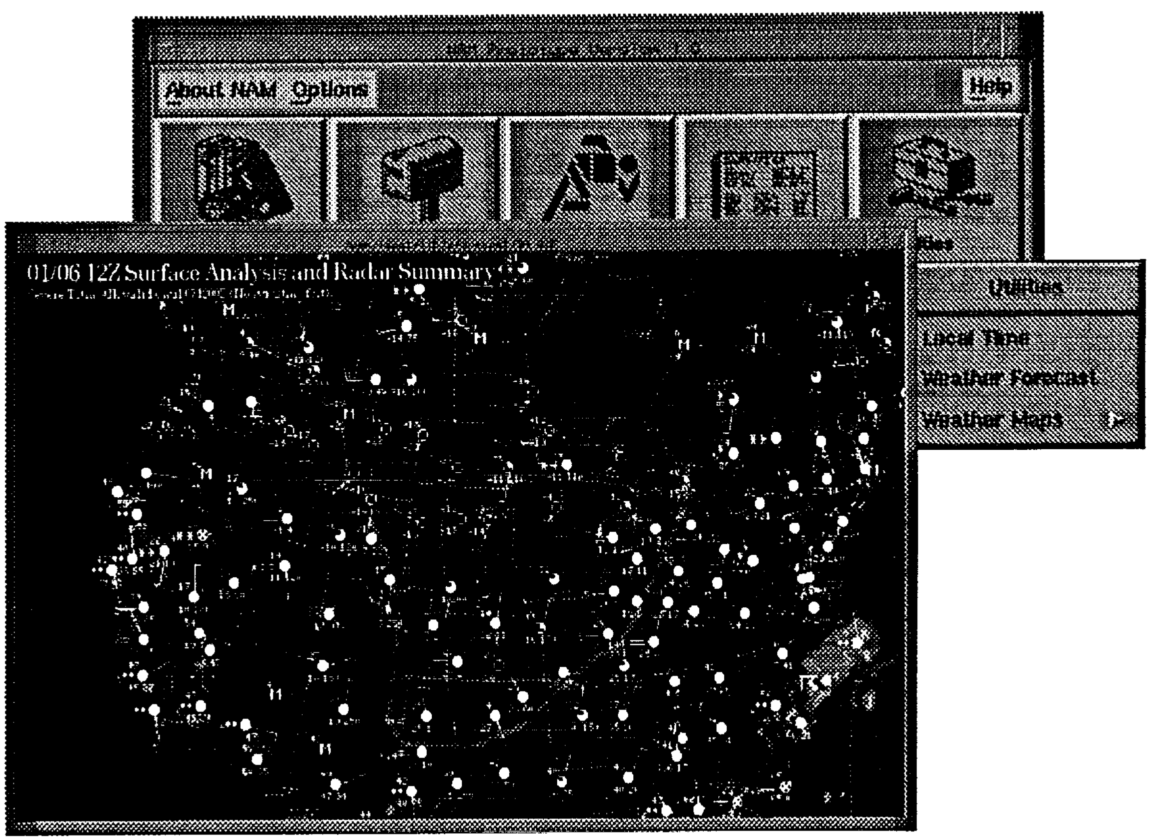

Figure 12: Graphics Capability

OSF/Motif tool kit. Use of standard software for the client ensures that the application will be portable to the variety of platforms found in the user community—DOS, UNIX, and Apple/Macintosh workstations.

BENEFITS

With the synergy of distributed client server architectures, GUIs, and navigational tools that give us "one stop shopping," today's endusers can just "point and click" from their desktops to quickly identify, access, and acquire the information they need, in the format they need, from worldwide sources. They are freed from having to search many sources and volumes of information to find the information they need, and they no longer need to maintain familiarity with various information sources and the complex query language syntax that was needed to access those sources.

In this burgeoning scenario, endusers gain control of maintaining and formatting their information to fit their application. Individual collections of information are specialized, providing an efficient alternative to everyone expending resources to collect everything. Other information is accessed at the point of origin using high-speed networks and navigational tools.

Information is available immediately in a format that is usable to many endusers. In traditional libraries, once a book is borrowed, it is unavailable for others until it is returned. In the virtual library, one "book" can be read by many at the same time.

CHALLENGES

There are significant challenges—critical success factors—associated with implementing a worldwide virtual library. People resist change, no matter how positive. There are cultural changes that will have to be managed across the information enduser community as well as the information development and management communities. The introduction and proliferation of the information age will shake some traditional processes and mindsets.

The development of and adherence to standards will be a critical success factor. International communication and distributed computing standards are essential to the transfer of information among organizations or countries. Standards development is slow; therefore, cooperation among vendors is crucial to ensure interoperability between multi-vendor platforms. Though some flexibility is required to be able to develop outside of standards, the time of "Rambo engineering," where a system is developed quickly without any plan for the future, is past. During the transition period from the old to the new operating architectures, there is a need to put standards into effect with a waiver process that allows some development outside the standards.

The networking infrastructure is another key element. Internet, EARN, EUNet, JANET, NREN, the National Information Infrastructure being promoted in the U. S., and the replacement of "copper" communications links with fiber optic cable are in place or underway today.

These form a strong base for the necessary networking infrastructures, one that will only grow as the timely location and acquisition of pertinent information becomes increasingly critical.

Another major challenge is that of intellectual property rights. How can you control copyright issues on electronic documents? How will authors and publishers receive payment for their work? How will the role of the publishers change in the new environment? These are but a few of the questions being asked, especially by the publishing community, in forums across the world today. Questions regarding intellectual property rights in regards to electronic document transfer and multimedia are just beginning to be addressed. "Every time a new technology comes along, there's a period of 10 to 15 years with litigation. It happened with movies. There was a whole series of cases when TV came along, then cable TV; now with videodiscs and CD."[12]

Finally, how is electronic information archived? Questions have been raised about the life expectancy of digital information. Will it be impossible to retrieve data archived on optical media in twenty years due to advances in computers and operating systems? Ten years? Five years? Rumor has it that some corporations keep a couple of Cray super computers in moth balls to be sure that they will be able to retrieve the engineering data they have archived on optical media. Somehow we must be able to protect our ability to retrieve the massive volumes of information contained on multimedia.

CONCLUSION

In the past, data processing professionals and information retrieval specialists have been concerned with the care and feeding of the data from a "data" perspective. Today, we have moved (are moving) into an environment where information is abundant in massive volumes, and the key is to identify and access relevant information. The emphasis is on the enduser. The focus is on the customer. Technology is the enabler; the customer is the driver. The key is to allow the requirements of the customers (endusers) to push the development and not to allow the technology to push the development. The future will be the time of the "knowledge worker," that person whose primary task is to know and use the information. The current paradigm shift is that more and more applications programmers are moving from the information system to the functional areas. Distributed systems are popular because people have a strong interest in being in control of those things for which they are responsible.

Traditional libraries will become points of electronic access to information on multiple media. The emphasis will be towards unique collections of information at each library rather than entire collections at every library. It is no longer a question of whether there is enough information available; it is more a question of how to manage the vast volumes of information once they are located. The future equation will involve being able to organize knowledge, manage information, and provide access at the point of origin.

BIBLIOGRAPHY

1. "Part 1180—Transfer by Federal Agencies of Scientific, Technical and Engineering Information to the National Technical Information Service," Federal Register, Volume 58, Number 89, Tuesday, May 11, 1993, Proposed Rules.

2. Kanter, Jerome, "Managing with Information," Prentice-Hall, Englewood Cliffs, NJ, 1992.

3. Hunter, Judy F., Cotter, Gladys A., "The Virtual Library: Coming of Age," paper presented at the AGARD/TIP Specialists Meeting, Ottawa, Canada, October 6-7, 1993.

4. Cotter, Gladys A.; "Technology for Libraries and Information Centers: A Seminar in Greece, Portugal, and Turkey;"paper presented at the AGARD Seminar on Basic Documentation Practices; Ankara, Turkey, September 3-4, 1992; Athens, Greece, September 7-8, 1992; Lisbon, Portugal, September 10-11, 1992.

5. Boeing Computer Services, "Long-Range Technology Forecast," February 1993.

6. Apple Computer, Inc., "Knowledge Navigator," Video, January 22, 1987.

7. Markoff, John, "Building the Electronic Superhighway," New York Times, Sunday, January 24, 1993.

8. Clinton, William J., Gore, Albert, Jr., "Technology for America's Economic Growth, A New Direction to Build Economic Strength."

9. Cotter, Gladys A. and Kaye, Karen, "NASA Scientific and Technical Information Program Multimedia Initiative," NASA Technical Memorandum 108982.

10. Hunter, J., Duncan, D., Generous, C., "NASA Access Mechanism—Graphical User Interface Information Retrieval System," NASA Technical Memorandum 108981, January 1993.

11. User Glossary Working Group of the User Services Area of the Internet Engineering Task Force, "Internet Users Glossary," Request For Comment (RFC) 1392, January 1993.

12. Interactive Multimedia White Paper: "Promise and Pitfalls," Knowledge Industry Publications, Inc., 1991.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(leave blank)*	2. REPORT DATE January 1994	3. REPORT TYPE AND DATES COVERED Technical Memorandum
Application of New Technologies in the Virtual Library: A Seminar in Turkey, Portugal, and Spain		5. FUNDING NUMBERS
6. AUTHOR(S) Judy F. Hunter and Glayds A. Cotter		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC		8. PERFORMING ORGANIZATION REPORT NUMBER NASA TM-109718
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546		10. SPONSORING / MONITORING AGENCY REPORT NUMBER NASA TM-109718
11. SUPPLEMENTARY NOTES		
12a. DISTRIBUTION / AVAILABILITY STATEMENT Unclassified - Unlimited Subject category - 81		12b. DISTRIBUTION CODE

13. ABSTRACT

The focus is on the technologies that are available today to support the concept of a virtual library. The concept of a *virtual library* or a *library without walls* is meant to convey the idea that information in any format should be available to the enduser from the desktop as if it were located on the local workstation. Discussed here are the background, trends, technology enablers, enduser requirements, and the NASA Access Mechanism (NAM) system, one example of how it is possible to apply existing technologies to the client server architecture to logically centralize geographically distributed applications and information.

14. SUBJECT TERMS NASA programs, architecture (computers), computer networks, information systems, libraries, information retrieval			15. NUMBER OF PAGES 21
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASS	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASS	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASS	20. LIMITATION OF ABSTRACT UNLIMITED

Available from NASA Center for AeroSpace Information
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
(301) 621-0390

National Aeronautics and
Space Administration
Code JTT
Washington, DC 20546-0001