

32012
P-8

MO&DSD ONLINE INFORMATION SERVER AND GLOBAL INFORMATION REPOSITORY ACCESS

Diem Nguyen and Kam Ghaffarian
Loral AeroSys
7375 Executive Place
Seabrook, Maryland 20706

Keith Hogie and William Mackey
Computer Sciences Corporation
7700 Hubble Drive
Lanham, Maryland 20706

ABSTRACT

Often in the past, standards and new technology information have been available only in hardcopy form, with reproduction and mailing costs proving rather significant. In light of NASA's current budget constraints and in the interest of efficient communications, the Mission Operations and Data Systems Directorate (MO&DSD) New Technology and Data Standards Office recognizes the need for an online information server (OLIS). This server would allow

- Dissemination of standards and new technology information throughout the Directorate more quickly and economically
- Online browsing and retrieval of documents that have been published for and by MO&DSD
- Searching for current and past study activities on related topics within NASA before issuing a task

This paper explores a variety of available information servers and searching tools, their current capabilities and limitations, and the application of these tools to MO&DSD. Most importantly, the discussion focuses on the way this concept could be easily applied toward improving dissemination of standards and new technologies and improving documentation processes.

INTRODUCTION

The Mission Operations and Data Systems Directorate (MO&DSD) online information server (OLIS) has been established to share the latest trends, technologies, and standards information among MO&DSD organizations. OLIS offers the same information via multiple retrieval methods,

including file transfer protocol (FTP), wide-area information server (WAIS), Gopher, and Mosaic. From their personal computer (PCs), users can electronically retrieve standards information (e.g., Consultative Committee for Space Data Systems recommendations), new technology information (e.g., asynchronous transfer mode), or previous studies and results. This server also provides pointers to other public servers on the Internet.

In support of this initiative, the Systems, Engineering, and Analysis Support technical support group began looking into mechanisms for providing better access to standards and new-technology-related documentation. A standards database on CD-ROM was examined; however, it was found that a local PC-based retrieval mechanism is rather cumbersome and not easily accessible for the more than 4,000 users of the MO&DSD community.

Because Transmission Control Protocol/Internet Protocol (TCP/IP)-based applications are widely available for any platform, the technical support group concentrated on an open systems approach using the Internet as the access mechanism. This approach provides maximum accessibility within Goddard Space Flight Center (GSFC), across NASA, and even to international users.

GOAL

The key issue is giving users timely, reliable, and relevant information. The overall goal of OLIS is to make information available, convenient, and easily accessible to MO&DSD personnel. Users can access and share in minutes information that once took weeks to disseminate. Most significantly, OLIS

focuses on ways to improve awareness of standards, new technologies, and documentation processes. OLIS was established to show what can be done and to encourage other groups to set up similar servers.

INTERNET SERVICES

Many available services require only a basic connection to the Internet, where client applications are used to access information servers. More advanced client/server technologies can access the capabilities of more basic clients and servers (as shown in Figure 1).

Figure 1 shows, on the left, the various types of client programs and, on the right, the various types of servers those programs access. These client/server communications can also occur with both elements on the same system so the same process can be used to access local files, as well remote files.

Also shown are two major types of interfaces. One is a basic textual interface that can be supported on

full-screen text terminals, such as the VT100 series. This type of interface can be supported by most computer vendors and can operate over dialup and Telnet connections. The textual interface works very well over low bandwidth connections as it only uses text and does not include any graphics.

The other type of interface is the full graphical user interface (GUI), which normally provides a more user-friendly interface. However, this type of interface requires an end-user computer system with proper windowing capabilities and is not commonly supported over dialup links.

The support applications (shown at the top of Figure 1) are used by Gopher- and Mosaic-type clients. These applications are activated when the client detects an action or file format that cannot be processed by the client itself. These applications are normally activated after the client has transferred a special-format file to the user's disk drive. Support applications include Telnet applications or special file format processors such as graphics interchange format (GIF) or joint photographic expert group (JPEG) image viewers, QuickTime or MPEG movie viewers, audio file players, or word processor packages.

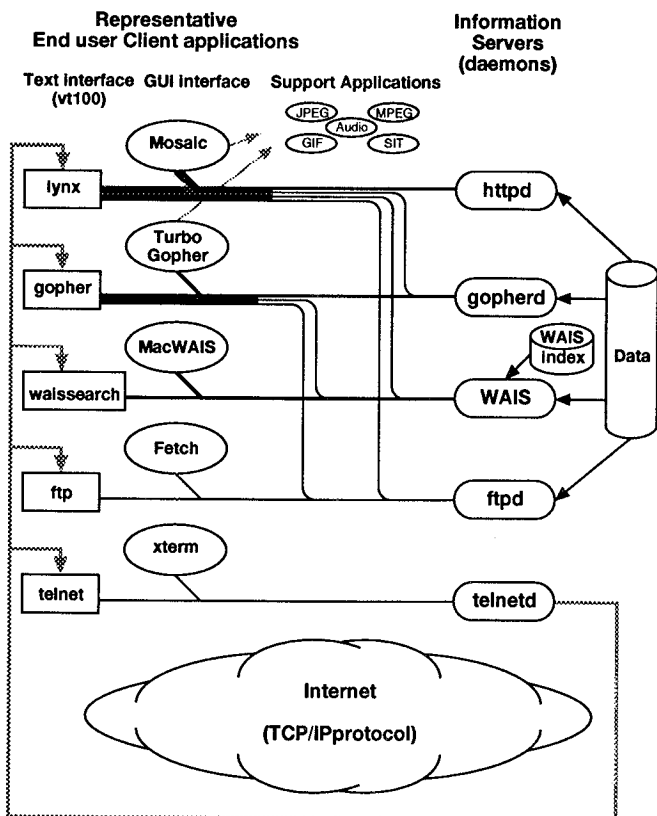


Figure 1. Hierarchy of Client/Server Technologies

Domain Name Service

Remote systems are normally identified by an Internet format hostname address such as "ddwilson.gsfc.nasa.gov." This name must first be converted into a standard four-number IP address such as 128.183.92.144, which can then be used to communicate with the remote system.

A "directory assistance" service, referred to as Domain Name Service (DNS), is available that automatically looks up IP addresses for a given hostname and vice versa. This automated service is silently invoked every time an Internet hostname is given for connection. The lookup is performed by querying a host called a DNS name server. (The identity of this server should be provided at the time the user's computer is set up for Internet access.) In addition to letting user computers resolve IP addresses automatically, the DNS maintains entries, listing the Internet hostname and IP address of the user's computer. This lets remote systems such as anonymous FTP sites determine if the computer is a registered Internet host.

Telnet

Telnet is the standard TCP/IP remote login protocol. To "Telnet" to another system is to run a Telnet client program that establishes a connection to a Telnet server and then logs on to that system with some user ID. This client/server combination provides a basic, text-only access mechanism but works internationally. Connecting to a system via Telnet normally looks exactly like connecting via a dialup line or direct connection. However, with Telnet, a user can support multiple connections to remote systems simultaneously.

On an X-window system, a local window-based program such as xterm is used to provide a scrollable text window from which a Telnet session is then established. On a Macintosh or PC, a window-based application [such as the public domain program, National Center for Supercomputing Applications (NCSA) Mosaic], Telnet, or a commercial package can be used to provide the client end of the Telnet connection.

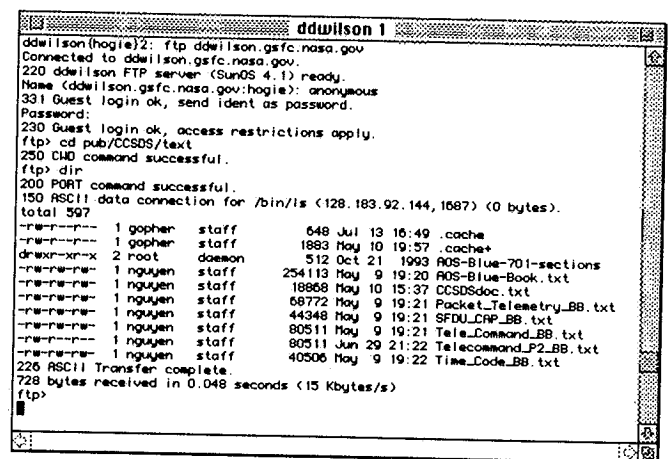
The main limitation of Telnet access is that only one system can be accessed per session, and the user must know and use the appropriate commands to examine information there. Some systems support a "guest" login with no password, but most systems require individual user accounts and passwords. A problem with using Telnet for public information access is that it normally requires a user to execute too many commands on the remote system. The combination of user accounts and extensive command capabilities make this a very poor way to provide a large user community access to many information servers easily and transparently.

File Transfer Protocol

FTP is another information access mechanism that has been in existence since the time of the ARPANET in the mid-1970s. To "FTP" to a system is to run an FTP server, examine file directories, and download or upload files of interest. FTP provides access to a single site at a time and requires explicit commands to disconnect from one site and then connect to another. However, it does provide basic access to anything that can be put into a file. Today, a wide range of information is digitized and stored in files. Users can log on to an FTP server if they have an account and password there. There are also

thousands of FTP servers set up for "anonymous ftp" access. In anonymous ftp, the user can log in with a user name of "anonymous" and a password (generally the user's name and/or email address) and then download files or possibly upload files into a special directory set up for that purpose. An anonymous ftp server was set up as the first information server for OLIS.

FTP initially operated in a command line mode (as shown in Figure 2), where the user entered FTP commands to browse file directories and retrieve files.



```
ddwilson 1
ddwilson(hogie)2: ftp ddwilson.gsfc.nasa.gov
Connected to ddwilson.gsfc.nasa.gov.
220 ddwilson FTP server (SunOS 4.1) ready.
Name (ddwilson.gsfc.nasa.gov:hogie): anonymous
331 Guest login ok, send ident as password.
Password:
230 Guest login ok, access restrictions apply.
ftp> cd pub/CCSDS/text
250 CWD command successful.
ftp> dir
200 PORT command successful.
150 ASCII data connection for /bin/lis (128.183.92.144,1687) (0 bytes).
total 597
-rw-rw-r-- 1 gopher staff 648 Jul 13 16:49 .cache
-rw-rw-r-- 1 gopher staff 1893 May 10 19:57 .cache+
-rw-rw-r-- 2 root daemon 512 Oct 21 1993 AOS-Blue-701-sections
-rw-rw-r-- 1 nguyen staff 254113 May 9 19:20 AOS-Blue-Book.txt
-rw-rw-r-- 1 nguyen staff 18968 May 10 15:37 CCSDSdoc.txt
-rw-rw-r-- 1 nguyen staff 68772 May 9 19:21 Packet_Telemetry_BB.txt
-rw-rw-r-- 1 nguyen staff 44348 May 9 19:21 SFDR_FTP_BB.txt
-rw-rw-r-- 1 nguyen staff 80511 May 9 19:21 Tele_Command_BB.txt
-rw-rw-r-- 1 nguyen staff 80511 Jun 29 21:22 Telecommand_P2_BB.txt
-rw-rw-r-- 1 nguyen staff 40506 May 9 19:22 Time_Code_BB.txt
226 ASCII Transfer complete.
728 bytes received in 0.048 seconds (15 Kbytes/s)
ftp>
```

Figure 2. Command Line FTP Interface

More user-friendly, point-and-click, or GUI FTP client programs have been developed in the last few years. These programs eliminate the need for learning FTP commands, but they still provide only lists of filenames to choose from. On a Macintosh, Fetch is one public domain program that implements FTP. It simply asks for the name of the host to which the user wants to be connected and then provides a point-and-click interface to FTP versus the basic command line mode of operation. Users can browse file directories and, if the filename is descriptive enough, can figure out the content and select files for downloading (as shown in Figure 3).

Although FTP provides a very widely used mechanism for access to file repositories, it still provides only simple lists from a single server at a time. Users need to know how to locate other servers and explicitly connect to them. Therefore, FTP does not really provide the sort of easy access desired for wide use. Most FTP sites ask for some sort of password, even

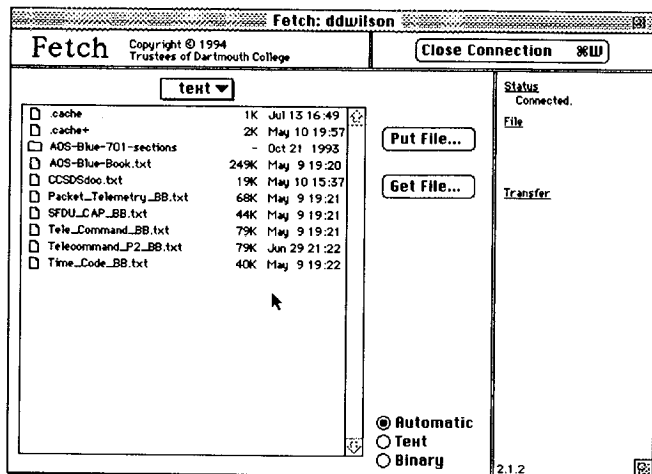


Figure 3 - GUI FTP Interface

if they don't care what it is. The password is a courtesy in the event the FTP site wants to record log-in access. Also, as a security measure, many FTP sites check to ensure that the DNS lists the user's IP address as a registered host.

Wide-Area Information Server

WAIS client/server software began appearing in 1992 and looked very promising for the MO&DSD goals of providing easy user access to documents, as well as full-text search capability on document repositories. A WAIS server was set up on OLIS in 1993. The popularity of WAIS increased rapidly once public domain implementations of the client/server software became stable and available. This was the beginning of the deployment of more advanced clients and servers and associated protocols that allowed users to quickly and easily access more than one server site.

The major feature of WAIS is its full-text search capability. When a document is loaded into a WAIS server, it passes through an indexing application that scans the document and builds a list of all the words in the document and their frequency of occurrence. This information and a pointer back to the original document file are then added to the master indexes on the server. An end user uses a WAIS client program to format a question to be asked of one or more WAIS servers. The question contains keywords of interest and pointers to the servers to be queried. The client program then establishes a temporary connection to a server, asks the question, receives the responses, and then drops that

connection and connects to the next server. After querying all servers, the client displays the resulting filenames to the user. The user can then select any file of interest, and that file is transferred to the user's system.

WAIS clients can be set up to easily query multiple servers with minimal user interaction. The user is not involved in the process of connecting to or disconnecting from each server to be polled. Once a question has been established with keywords and servers identified, the question can be saved for later use. The complete search of the same question can then be executed at a later date with a single user action. The results are returned in a single list indicating filenames, together with a score of how frequently the keywords occur in each file. With WAIS, users begin to see the Internet as a large information repository.

An inherent limitation with WAIS is that the indexer supports only ASCII text as input format; i.e., graphical content must be removed and only the text entered into a WAIS server. This presents a problem in that, today, many documents are prepared in desktop publishing packages and contain tables and graphics, as well as text. To get around this problem, a WAIS server can be used to locate documents of interest, and then the fully formatted version can be retrieved separately. Another limitation is that the user is responsible for identifying all information sites to be queried. One of the files at a site can be a list of other WAIS servers that a user can add to the list of sites to be queried. But the user is still responsible for identifying each site to be used.

With Gopher and WWW, pointers to other sites with information of interest can be followed more easily and transparently, as discussed in the following paragraphs.

Gopher

Gopher servers exploded on the Internet during 1993, with thousands in existence today. Gopher is a popular menu-based information system that integrates access to Telnet, FTP, and WAIS in an easy-to-use interface. Gopher also provides easy access to multiple sites. Users move transparently from one server to another with the Gopher client knowing which protocol to use to access each server. Thus, users spend time searching for information rather

than trying to determine how to use each service and navigate the network.

TurboGopher (shown in Figure 4) is a Macintosh version of a Gopher client that allows users to retrieve the same information accessible via a typical VT100 Gopher client. TurboGopher, however, provides the following advantages:

- Point-and-click graphical interface
- Files saved directly to local Macintosh
- User does not have to first log into UNIX host

When an item is selected from a Gopher menu, one of three things happens:

- Lower-level menu on the same system appears
- Connection is initiated to another server and menu actually stored there appears
- Action is initiated, such as a file transfer or Telnet session

The first case is actually a special instance of the second. In the second case, for menus served by another system on the Internet, TurboGopher automatically connects to the new system and sends the proper low-level commands to retrieve the menu being invoked. As the user browses through the menu hierarchy, the program automatically switches from system to system as needed. In the third case, when a user invokes a menu item, some special action may be performed (e.g., automatically initiate downloading a file). This is implemented essentially by FTP-like functionality built into TurboGopher, while a Telnet session is initiated by activating a support application.

Worldwide Web

The WWW project began at the European Particle Physics Laboratory (CERN). WWW seeks to build a distributed hypermedia system in which all information on the Internet can be accessed consistently and easily. Because WWW has the greatest flexibility and most user-friendly interface, the OLIS effort focuses here.

WWW is geared toward hypermedia that allows selected objects to be expanded at any time to provide additional information (i.e., selected objects are links to other objects, such as text, sounds, images, and animation). The basic building blocks of WWW are the HyperText Transfer Protocol (HTTP) and the HyperText Markup Language

(HTML). HTTP describes the communication protocol between clients and servers; HTML describes the format of the information pages transferred.

WWW capabilities are comparable to Gopher in that WWW is a client/server information system running on the Internet that provides quick and easy access to a wide range of servers. WWW performs the functions of Gopher but also supports hypertext links that permit the creation of more descriptive information for the end user. WWW provides a full screen of textual and graphical information with individual words, phrases, or icons acting as links (shown in Figure 5). The page-formatting capabilities of HTML can be used to create a wide range of point-and-click user interfaces that operate across the network.

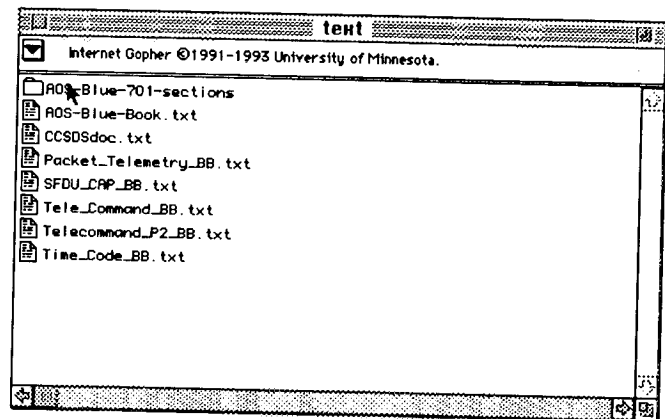


Figure 4 - TurboGopher Interface

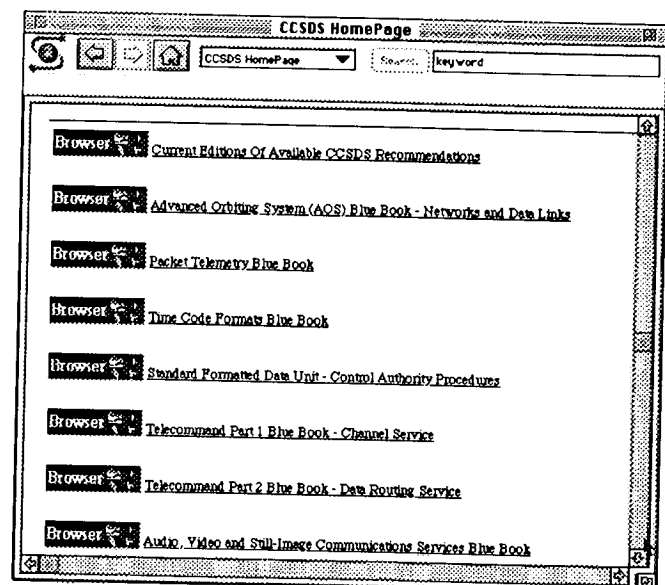


Figure 5. WWW Interface

The WWW model represents everything (document, menu, index, etc.) to the user as a hypertext or hypermedia object. Two navigation operations are available to the user: to follow a link or to send a query to a server. Two powerful features fall out of this hypertext model. One is that almost all other information systems can be represented in terms of WWW documents. The other is that the WWW system has an open but uniform addressing scheme that allows links to be made to any objects on WWW, WAIS, Gopher, FTP, Network File System, or Network News servers. A WWW user can interrogate WAIS indexes and Gopher servers. The hit list returned by a WAIS server (or any other query engine) is treated as a hypertext document with links to the documents found. Gopher menus (or any other hierarchical menu system) are represented as lists of items linked to other objects. The hypertext model also allows the user to put in a hypertext link, when needed, for background information.

NCSA Mosaic

Because WWW supports the functions of other types of clients and servers, this discussion focuses on WWW clients. One of the most popular WWW client applications is the free, public domain software, NCSA Mosaic, which is available for PCs, Macintoshes, and many UNIX workstations. It was developed at the University of Illinois Urbana-Champaign and can be picked up from most major anonymous FTP sites. With Mosaic and the proper support applications, users can explore WWW with full access to multimedia information, including formatted text, graphics, and sound. Through WWW, a user can access Telnet, FTP, WAIS, Gopher, and HTTP systems.

When a user starts Mosaic, the program attempts to connect to a preset known host; the initial host is at the NCSA, but each user can configure their own starting point. Once an initial page of information is downloaded, Mosaic operates in a true client/server fashion. The user sees on the screen a graphical point-and-click hypertext interface (shown in Figure 6). As the user browses through the hypertext tree, Mosaic automatically switches from system to system and protocol to protocol as needed.

Figure 6 shows the Mosaic interface and part of the MO&DSD HomePage. When the user clicks on the

SpaceOps 94 hypertext link, Mosaic automatically retrieves information and displays the SpaceOps HomePage (shown in Figure 7).

If a special file type is referenced on a hypertext link, for example GIF format, Mosaic tries to invoke a GIF viewer to display the file after downloading is completed. GIF viewer software must be installed on the local system, and Mosaic must be configured to use it.

Mosaic displays the retrieved information in the large scrolling window, and links are followed by

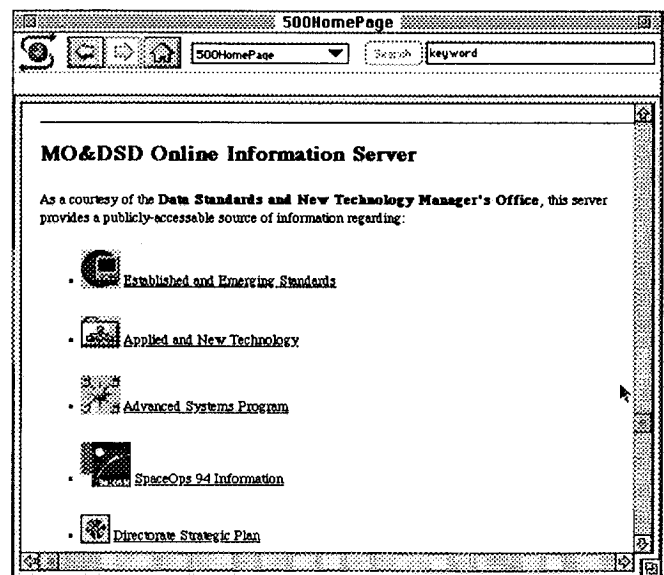


Figure 6. Mosaic Screen

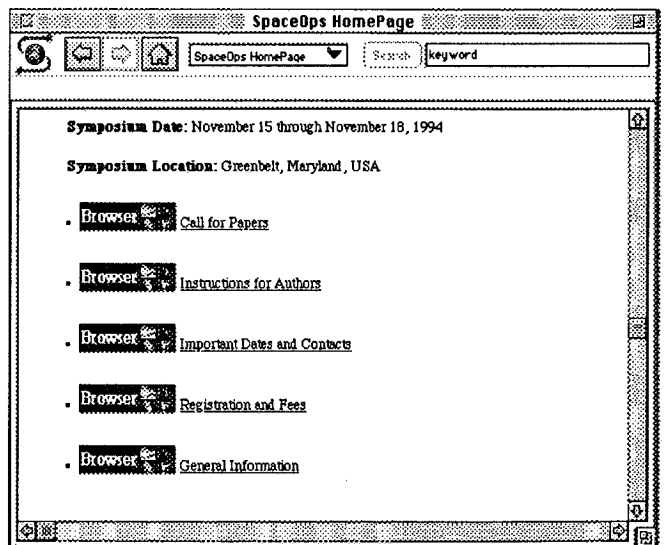


Figure 7. SpaceOps 94 HomePage

clicking on the underlined hypertext items. One of the most important features of the Mosaic client is its ability to save "bookmark" or "hotlist" references to a user's favorite information locations. Thus, the user does not need to know exactly which links were followed to find a particular information site. The user can access these locations again with a single operation using the bookmark or hotlist entry.

DOCUMENT FORMATS

One problem with information servers is the wide variety of possible formats in which data can be available. The MO&DSD server stores all of its documents in ASCII format, as well as in the document's original format. Currently, there are not many bit-mapped images stored on the MO&DSD server. Any images placed there are stored in either GIF or JPEG format.

Plain Text Document

Because ASCII format can deliver textual information to any end user via a wide range of access mechanisms, documents are usually loaded onto information servers in an ASCII format. However, the main problem with this is that all graphics and figures in the original (fully formatted) document are lost. Furthermore, many documents are prepared using desktop publishing packages with paragraph wrap and proportional fonts. When converted to monospace font with individual lines of text, these documents lose their enhanced appearance.

All ASCII documents on the MO&DSD OLIS have been indexed to the WAIS server to facilitate keyword searches. This allows a user to perform keyword searches on located documents of interest, scan the ASCII versions online, and retrieve the original formatted document if desired.

Fully Formatted Document

Original document can be stored in a wide range of formats [e.g., WordPerfect for Windows, Word for Macintosh, Rich Text Format (RTF), or PostScript]. Each of these formats can be used to deliver a fully formatted document. These files are usually converted to a single flat file, with a package such as BinHex, prior to being loaded on the OLIS. When trying to read and process fully formatted documents, the end user must have the proper application

software to reconstruct an exact copy of the original document. Eventually, all fully formatted documents will likely be stored on OLIS in the following three formats:

- Plain ASCII text for searching and browsing (document may not look nice but all the text information will be there)
- Single PostScript file containing the whole fully formatted document including graphics
- Original word processor format

MAKING A MACINTOSH OR PC INTERNET CAPABLE

To use any of the Internet services described, a user must first gain access to the Internet. One way to access these services is through a connection to a computer that is a full host on the Internet. However, in this mode, the user needs to log in to that system and use the proper commands to access these services.

The most powerful way to access these services is to have a direct connection to the Internet. This can be done from most UNIX workstations, Macintoshes, or PCs. The connection may be via a local area network (LAN) or via a dialup phone line. The most powerful dialup access involves the use of Serial Line IP (SLIP) or Point-to-Point Protocol (PPP). These protocols operate between software on the user's computer and similar software on a SLIP or PPP server that the user can dial in to. The protocols then implement an IP connection over the dialup line, and the client applications then operate over TCP/IP, just like when the user is directly connected to a LAN. To get reasonable response with SLIP or PPP, a dialup connection of at least 9600 bits per second is required.

For any system to be connected to the Internet, three basic things are required:

- Connectivity to some point on the Internet
- Unique IP address for the user's computer
- Appropriate TCP/IP software and associated application software

A unique IP address for the user's system is provided by the administrator responsible for connectivity. A basic communication program with VT100 terminal emulation and download capabilities is also required on the computer. Some common packages

are Zterm for Macintoshes and ProComm or CrossTalk for PCs.

Macintosh Platform

For Macintoshes, the major item software required is MacTCP to provide TCP/IP support. GSFC has a site license, so this is readily available to any NASA-owned Macintoshes. Other client and support applications are available from Macintosh servers at GSFC or numerous anonymous FTP sites.

PC Platform

For PCs running Windows 3.1, the main software required is the Windows socket interface dynamically loadable library, "winsock.dll." This provides the interface between most applications and TCP/IP on the PC. The client and support applications are available from PC servers at GSFC or numerous anonymous FTP sites.

SUMMARY

With Internet access, modern information retrieval software, and connection to the MO&DSD OLIS, remote users can

- Easily retrieve well-established standards at a local Macintosh or PC
- Perform online browsing and retrieval of documents published for and by MO&DSD
- Search for current and past study activities on related topics within NASA or other government agencies prior to issuing a task
- Improve standards and technology awareness among peers to ensure a design of interoperable systems for cross-support environments

- Access information at thousands of sites worldwide, and make select information accessible by others

Additional potential benefits could be achieved in areas such as:

- Avoiding duplication of effort by being able to quickly and easily scan information of activities being performed across MO&DSD and, eventually, all of NASA
- Faster development of standards documents via immediate access to huge reference libraries and capability for rapid exchange of documents
- Significant reduction in travel costs through easy exchange of information and documentation over the network rather than by attending meetings.

REFERENCES

- [1] Frank Hecker, *Personal Internet Access Using SLIP or PPP: How You Use It, How It Works*
- [2] Ed Krol, O'Reilly & Associates, Inc., *The Whole Internet Catalog and Users Guide*
- [3] Paul Lindner, editor, *Internet Gopher Users Guide*, University of Minnesota
- [4] Adam Gaffin and Jorg Heitkotter, *Big Dummy's Guide to the Internet*

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. William Mackey, Mr. Steve Harris, and Ms. Sydney Buck for helpful comments on a draft version of this paper.