

# NCSA Mosaic and the World Wide Web: Global Hypermedia Protocols for the Internet

Bruce R. Schatz and Joseph B. Hardin

Network information systems reached the public consciousness this year as a result of the phenomenal growth in the use of the Internet. In particular, the software constituting NCSA (National Center for Supercomputing Applications) Mosaic and the World Wide Web have made global hypermedia a widespread reality for the first time. The technology underlying this software is described to explain the protocols behind information spaces. These include the historical predecessors, the current protocols with examples, future directions for the software, and discussion of research systems with different architectures. Reasons for its popularity are given, with the goal of illuminating successful services for the National Information Infrastructure.

The Internet, which has been in use by scientists since the late 1960s, hit the big time last year, not only among scientists but among the general public. Part of this was attributable to the greatly expanded scope of the Net: Rather than tens of thousands of users as in the 1970s, there are now tens of millions in the 1990s. In addition, the National Information Infrastructure has become a major national priority, promising to bring the Net to every home and business in America.

Another, perhaps even more significant part of the impact of the Internet, is a result of a fundamental change in its use. Originally intended as a distributed network of computers, it is increasingly viewed instead as a distributed space of information. Rather than transferring files between computers, a user navigates an information space of distributed items of information. The users concentrate on the logical structure of the interconnection of information and data items rather than on the underlying physical structure of computer and communications systems. Formation of the Internet relied critically on standard network protocols and simple universal services. Similarly, new protocols and services have evolved to support this new logical structure of interconnected information spread across the worldwide network, a global hypermedia system.

This article examines some of the issues behind a global hypermedia system by describing the workings of a software program called NCSA Mosaic, developed by the National Center for Supercomputing Applications (NCSA) at the University of Il-

linois in Urbana-Champaign and built on the World Wide Web (WWW) protocols. This is a currently popular information service on the Internet, whose great and sudden success has embedded in the popular consciousness the notion of a worldwide information space. Examination of the protocols and services that constitute Mosaic and WWW will make the current technology underlying a global hypermedia system clearer. Here, "hypermedia" refers to a collection of items of multimedia information with relationship links between the items, and "global" refers to a system that can transparently retrieve these items and navigate these links without regard for their physical location. The presentation is historical in nature and intended to be illustrative of the technology, rather than describing the detailed functionality of the systems.

In addition to examination of the present, there will be a brief examination of the future. The short-term evolution of Mosaic/WWW itself will be described, as well as the longer term evolution of other architectures that may supplant it. Finally, the sociology behind the success of Mosaic is examined, with the hopes of shedding light on the broader question of what services will be successful in the age of information superhighways.

## Gopher: Client-Servers for Network Resources

The Internet was founded upon the TCP/IP (Transmission Control Protocol/Internet Protocol) protocols, which provide universal access for data transmission (1). These protocols support transparent interconnection of data stored in machines spread across the network. Today there are literally millions of machines connected to the Internet, which can exchange packets of data using TCP/IP (2, 3).

One of the primary services of the Internet is file transfer. In its simplest form of directly copying files from one machine to another, it was one of the original services in the 1970s and still remains among the most popular. Later, file transfer evolved into services, such as electronic mail and bulletin boards, which helped to form the Internet community and widely demonstrate a new form of communications media. The evolution of network information systems has thus far centered around browsing and retrieval services, which are surveyed in this article. However, it is already clear that the next stages will center around sharing and publishing services, as touched upon in the future sections at the end.

The simplest form of file transfer is FTP (file transfer protocol). The services built on FTP, and used in the Internet for many years, provided a way of connecting to a specific machine, locating a specific file, and transferring it to the local machine. Services such as "anonymous FTP servers" became common mechanisms for publishing and retrieving information, whereby a user could access a standard location on almost any machine to retrieve files that had been placed there for public access. As the Internet reached more locations at higher bandwidths, a wider range of information could be effectively fetched, such as documents from collections and items from databases. This was one of the primary bases for the development of information services in the Internet.

A network information system mediates between a user and some information across a network. A common terminology for describing the components thus divides the system into the user side of the network, the client, and the information side of the network, the server (4). A client handles the user interaction: It processes commands and displays their results. In a network system, some commands are processed locally, but many are transmitted across the network to a remote server. A server handles the information interaction: It retrieves items and transmits the results to requesting clients. Users typically invoke a client when they wish to use some information service; this program then communicates across the network with the appropriate servers, which are already running.

Gopher, developed in the early 1990s at the University of Minnesota Computer Center, was the key program in demonstrat-

B. R. Schatz is with the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, and the National Center for Supercomputing Applications, Beckman Institute, University of Illinois, Urbana, IL 61801, USA. E-mail: bschatz@ncsa.uiuc.edu. J. B. Hardin is at the National Center for Supercomputing Applications, Beckman Institute, University of Illinois, Urbana, IL 61801, USA. E-mail: hardin@ncsa.uiuc.edu.

ing the client-server approach to file transfer and other network resources (5). It presented the Internet as a hierarchy of servers, from which files could be transparently transferred, rather than as a hierarchy of machines. A user is presented with a list of files and services that can be retrieved from menus by interactive selection, without regard for the physical location of actual storage as with anonymous FTP. Gopher encompasses multiple services, or resources, on the Net, including FTP, Telnet, and Archie, giving users a simple, consistent interface to the Internet's multiple resources (6). Gopher was thus a significant milestone in showing that the Net could be effectively viewed as an information space. There is a set of established protocols by which a file can be plugged into Gopherspace, for example, by providing a simple text packaging for it on a suitable server.

The simple Gopher client was propagated to many machines, and many sites put up Gopher servers with a wide range of textual information. For example, a user can retrieve a document from the National Science Foundation (NSF) by selecting menu items for North America, Washington, D.C., NSF, and a specific form. Gopher spread rapidly across the Internet and opened the eyes of the research community, showing them how to fetch and how to publish information in the Net. Today, there are more than 7000 servers across the Internet on a wide variety of topics.

### WAIS: Network Search Servers

In a similar time frame, another popular Internet information service was developed, from a different paradigm, by a team led by Brewster Kahle from Thinking Machines. The Wide-Area Information Server (WAIS) was inspired by research prototypes in network-based information retrieval, such as the Telesophy system (7), which enabled multiple sources of information to be stored in multiple servers across a network along with other servers that supported associative search. WAIS was the first well-packaged system that supported full-text retrieval and was freely available within the Net (8).

The software can be decomposed into the client, server, and search engine. The client accepts queries and passes them to the server. The server passes queries to the search engine, which searches a full-text index. Results from the search are passed back through the server to the client, which then displays them. The server thus stores the actual documents, whereas the search engine contains an index of the words from those documents for efficient full-text search. There are protocols for how to index documents and how to connect to remote servers. Although the search for the public

domain WAIS is primitive by professional information retrieval standards, its easy availability and simple functionality have resulted in widespread use.

Hundreds of WAIS servers have appeared across the Internet on a wide variety of topics. Originally, WAIS existed in isolation as a separate search server with its own interface client. It is now most commonly accessed via a gateway built into one of the network information services. For example, Gopher has a gateway through which it can pass strings into WAIS and receive matching files to provide a complete query and fetch service. Subsumption of various information services under one client, as in NCSA Mosaic/WWW, a "multiprotocol client," is an established method of integrating the Internet information space.

### WWW: Documents and Links

The WWW began as an attempt to link researchers at CERN, the European Laboratory for Particle Physics, in Geneva, Switzerland. Around the same time that the Gopher and WAIS software were being developed, a small group of developers led by Tim Berners-Lee at CERN proposed a network-based hypertext system for use at the institute. An early proposal requesting funds for this project is dated October 1990 (9).

The web combines two ideas which alone are useful and together have proven to be extraordinarily powerful: networked information and hypertext. By the time of its development, a number of systems that provided users access to networked information were available, anonymous FTP servers being the earliest and most used, Gopher being extremely approachable and opening up the world of networked information to many new users, and WAIS introducing many users to a full-text search engine for a networked retrieval system. All of these systems also allowed users who were knowledgeable enough to place materials on the Net, by putting them on FTP or Gopher servers or by building a WAIS index to a corpus and serving it. However, none of these systems provided a method of linking one piece of information directly to another within the body of a document. Such links are called hyperlinks and are a distinguishing feature of hypertext (10, 11). Indeed, hypertext can be defined simply as a system of links and nodes, where the nodes may be documents that contain such links (12).

The WWW (13) is a set of protocols that (i) allow for the location of any document on the Net through a naming system based on Universal Resource Locators (URLs); (ii) describe a way of placing links using URLs within text documents, called

Hypertext Markup Language (HTML); and (iii) specify a way to request and send a document over the network, the Hypertext Transfer Protocol (HTTP). With these standard protocols in place, one can set up a server and construct hypertext documents with links in them that point to the documents on that server. Selecting the link from within the display of a document sends a request to the server, which in response sends the document back. The retrieved document might have links in it, perhaps pointing to documents on another such server, and thus, a user might browse through an extensive corpus of material distributed around the global network simply by following the hyperlinks, "jumping" from one point of interest to another.

In addition to directed navigation from one document to another, an information space must also support search. In WWW, a retrieved document may be an index of information and thus be searchable by specification of a query string. The results of a query against an index can be automatically composed into an HTML document with short summaries of the returned items containing embedded links that can be followed to retrieve the complete documents. The Web, thus, combines two underlying functions of browsing information spaces: the presentation of information and the method of choosing which information to see next. The combination of this easy method of browsing and a growing base of information to browse on the Net set the stage for an explosive growth in WWW usage.

From its inception, the WWW project has sought to encompass all existing methods of network information navigation and retrieval. The underlying model of hypertext allows for the essentially hierarchical searches of FTP and Gopher directories by presenting these directories and their contents as lists of links that one can follow up or down. By providing the ability to search indexes, the WAIS servers can be searched by providing them with a text string that they attempt to match against their full-text index. The HTTP servers provide the advantage of making hypertext documents available to users. Finally, the last common method of information exchange on the Internet, the Usenet news groups, a globally available bulletin board system with millions of users, can be enfolded by communicating with the Usenet servers and presenting the discussions as nested lists of links.

These protocols, or gateways to them, were provided in the early library of WWW communications software (libWWW), so WWW browsers were able to communicate with all of these information sources and present them as a unified information space

to the users. The WWW became a superset of all of the available networked information services: FTP, Gopher, WAIS, news, and WWW itself. This meant not only that users were provided with a single entry point to the resources of the Net, but also that the critical mass of the Net, that point at which it would become self-catalyzing, with information providers putting more information on the Net because they knew more people would see it and more people using the Net because of the increase of information, could be reached much more easily. This positive feedback relation, and its turnover point that has led to explosive growth, was pushed by the release of an easy-to-use and functional browser in the winter of 1993.

### Mosaic: Multimedia Hyperdocuments

The early implementations of the WWW system were line-mode browsers, which presented the hyperlinks as numbered choices in a menu that users could choose from. Very soon though, graphical user interfaces were developed on NeXT and X Windows system machines, which displayed the hypertext in a window with the links in color or underlined, allowing the user to move through hyperspace by clicking on links with a mouse. The Viola and Midas systems were examples of these. In early 1993, work began in the Software Development Group at NCSA on a graphical browser, which became known as the NCSA Mosaic system (14). This software provides a client to the WWW protocols and an interface to the Web itself.

Among the features introduced in the Mosaic system was the placement of images on the hypertext page, thus creating multimedia hypertext, or "hypermedia" documents. This meant that hyperdocuments could contain pictures and graphical icons. Both the pictures and icons could also be links, and clicking on them would bring the user another hyperdocument, further information, or an expanded image. This capability was soon extended so that the mouse pointer location could be tracked over any of these pictures telling where on the picture a user was when they clicked. This meant that the pictures could be maps, and pointing to a portion of the map and clicking would "take you there."

The short session shown in Fig. 1 illustrates the navigation of the global hypermedia space using Mosaic/WWW. The Web is represented by a graph of interlinked documents. The most common entry is by way of a local home page, which is a document with a collection of "interesting" links with descriptions, such as the NCSA Home Page on local exhibits (Fig. 1A). Figure 1B gives

the internal HTML format of a section of this document. The tags in angle brackets give formatting and structure information. In particular, the HREF tags specify links to other documents. Note the HREF that gives a URL link pointer attached to the text for "The Krannert Art Museum". The URLs typically look like a combination of an Internet domain name, to specify the physical machine, and a hierarchical file structure, to specify the physical file containing the information document. The linked text is colored and underlined in the document display (Fig. 1A), and when selected and clicked, the link is followed to the corresponding document (Fig. 1C). Note that the "Document URL" is the same as the previous HREF, indicating that the URL has been followed and retrieved across the Net. This document contains an embedded picture, which is specified within the HTML by a reference to a GIF (Graphics Interchange Format) file. Mosaic can display this multimedia document because its client software contains a display for HTML text and for GIF images.

External viewers can also be called from the Mosaic browser, which enables types of data that are not supported by the client itself to be displayed. For instance, if a link points to an image that is not a GIF file, then the image data can be passed to a program that can display this format, and the image can be displayed in a separate window alongside the Mosaic window. To the user, this means that a wide range of data can be made easily accessible. For example, documents can contain embedded audio and video materials, which can be displayed by invoking an appropriate external viewer for the local user computer. The ability to incorporate external viewers, coupled with the accompanying standardization of formats of the actual materials, means that most of the materials on the Web can be displayed by most of the users of Mosaic.

The audience for this system was greatly expanded by the introduction of such graphically based hypertext browsers across the three most popular computer operating system platforms: Apple Macintosh, Microsoft Windows, and the X Windows system. This meant that the general public, as well as scientific communities not using Unix, could move easily around the Net with a hypertext browser. This led Anthony Rutkowski, executive director of the Internet Society and longtime watcher of developments on the Net, to say that a "digital cannon" had been fired on 12 November 1993, the date of the first full release by NCSA of Mosaic browsers on all three platforms (15). Already by then, the first version of NCSA Mosaic, for X Windows, had been widely distributed, and WWW byte

traffic on the Net had grown by three orders of magnitude during the previous 10 months that included the Mosaic release (Fig. 2). This growth continued, and in the next 5 months, the Net saw WWW traffic increase by another order of magnitude. The number of users is hard to judge because the software was released free to the general public and placed on the NCSA anonymous FTP server and others around the world, where anyone could have access to it. However, by many estimates, there were more than one million copies of the NCSA Mosaic client alone by spring of 1994. The WWW server at NCSA has seen a constant growth in use, with over a million and a half connections per week currently.

Further developments of the NCSA Mosaic system increased its functionality. The ability to display "forms" like those on a database request, and thus prompt the user to fill in a complex request, was a major step. This allowed sophisticated queries to be specified from a hypertext document and complicated searches of indexed information and databases to be issued. Users filled in the forms—typing in the open fields, clicking on the button choices, or choosing a menu item—and built up a complex query, which was then sent to the requesting search engine and resolved, and the data was sent back to the user. The combination of forms in HTML documents, HTTP for transport, and "gateways" to other information servers made for customized network information search systems, available through the standard Mosaic interface. By taking advantage of the client-server architecture of the system, developers can construct filters that run on the server side, integrating information sources like relational databases with SQL (Structured Query Language) queries into the hypermedia information environment.

This is an example of the idea of "Open Information Systems," systems that allow for the easy integration of existing information sources and that can be extended and expanded by users in ways that were often unanticipated by the original developers. This is one of the strengths of the NCSA Mosaic system and the underlying WWW structures. This also leads to a "bottom up" notion of how such network information systems evolve. Given a malleable framework within which specific information structures may be embedded and the commonality of the universal interface, advanced functionality comes from the distributed enhancement of the system, often at the server side, through increasingly sophisticated filters and scripts that can translate between the built-in viewer and complex data structures.

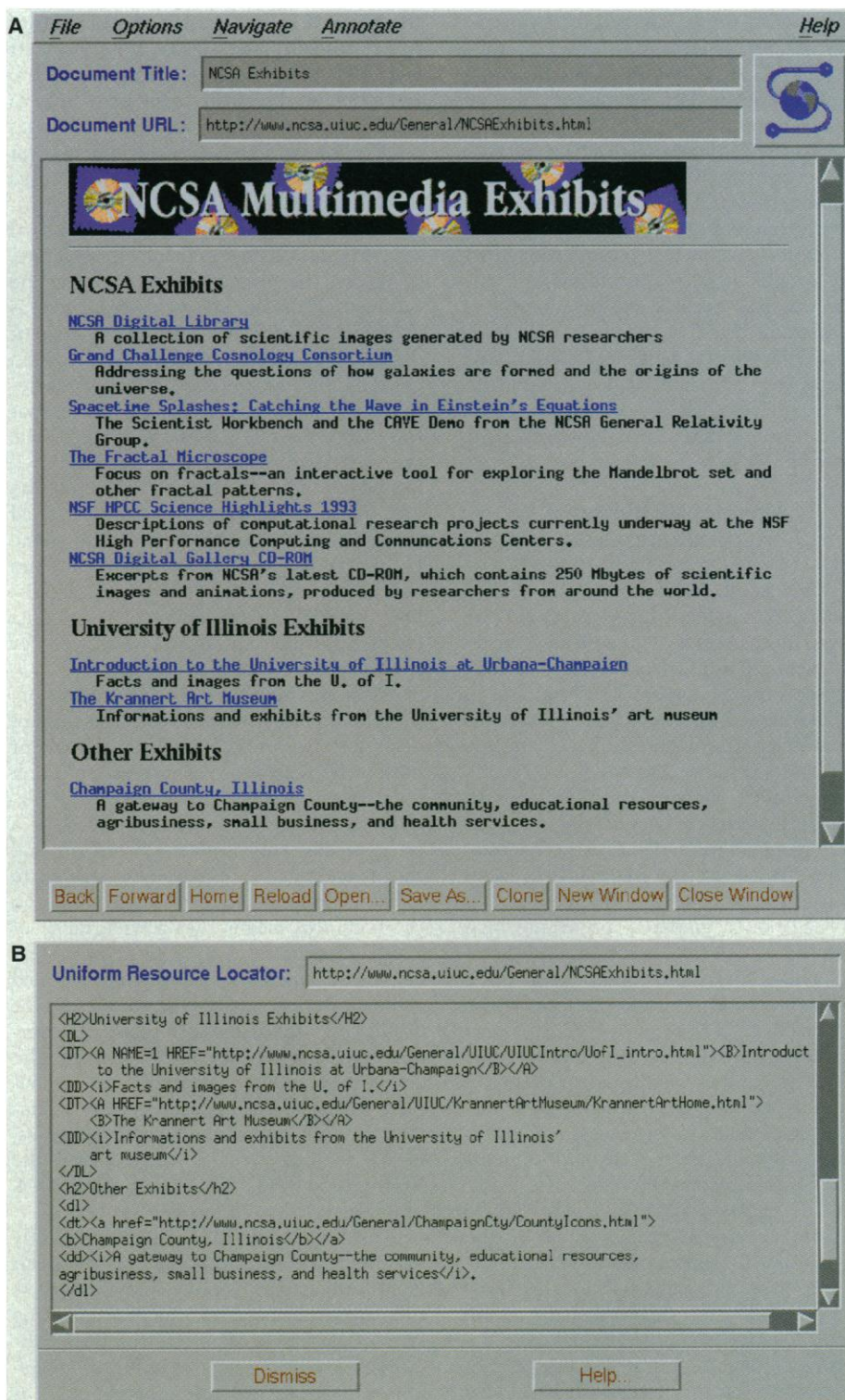
## Evolution of Mosaic/WWW Toward Interaction

The primary concentration of Mosaic and the Web so far has been on information connection: the following of links and the retrieval of documents. Now that there is a large and varied collection of information in the Net, the primary problem increasingly becomes interaction rather than connection, enabling the user to manipulate information. Issues include, for example, how to integrate viewers, provide publishing, and support searching.

Currently, NCSA Mosaic can be passed the type of a retrieved data item, launch a viewer of that type if available on the local machine, and pass this viewer the data for display. A major limitation of this approach, however, is that once in the associated viewer, the ability to use hyperlinks is lost. Once launched, the viewer exists as a separate program, outside of the global hypermedia environment. If viewers could recognize hyperlinks of any sort (text, images, or icons), in their own frames, and pass those links across to Mosaic for resolution and retrieval, this would make it possible to better integrate external programs.

One solution is to provide for interprocess communication between the Mosaic client and third party viewers. Initially, this Common Client Interface could be extremely simple, just notifying Mosaic that a URL is coming and passing it. Eventually, much more in the way of generic coordination and control capabilities could be built into the interface, allowing for more choice in where the resulting document is passed, for instance, whether Mosaic appears on the screen at all, or integrating the two windows more intuitively into the interface and the user's environment.

The key feature of the external integration is to enable two-way communication between the user's client and the various external programs, including servers and viewers. This implies that viewers can become editors and enable user entry as well as data display. For example, a full commercial implementation of the Standard Generalized Markup Language (SGML) could be used as both a document creation and a document display tool, while still retaining the capability of embedding and following links within the document. Any created documents could be sent to a server that supports their indexing and searching. [SGML is a specification language for the structure of a document, including headers and references, that has been widely adopted throughout the publishing industry (16, 17). The HTML is a subset of SGML, specialized for simple inter-



**Fig. 1.** Session with Mosaic/WWW following a URL across the Internet and displaying a hypermedia HTML document. (A) An NCSA Home Page displayed in Mosaic. (B) The HTML format underneath this document. [(C) located on facing page] The result of following a URL from Home Page to display another document.

active displays with embedded links.]

Related to the provision of better integration of external programs is the provision for better search capability. First, this requires modification of the architecture of WWW. The URLs provide a

universal naming scheme for documents, which gives a unique but absolute name. To support scaleable information retrieval, a data item must instead be assigned a unique permanent identifier that is completely disassociated from its



location information, as well as a description bound to that identifier. Otherwise, when an item changes its physical location and thus its absolute name, links to that item can no longer locate it. Once an identifier is obtained, it must be resolved into this location information, which may then be used to retrieve the item for local manipulation.

These permanent identifiers, known as Uniform Resource Names (URNs); their associated transient locators, called Uniform Resource Locators; and their content descriptors, Uniform Resource Citations (URCs), have been described in a series of Internet drafts (18). These drafts are under active review by the Internet community as the standards for the description and location of information resources. By binding a permanent identifier to any data item, one can establish its current location on the network and its preferred access mechanism with a simple query to a central resource locator server.

Once every data item has some associated classification, the clients can be profitably modified to provide gateways to

more sophisticated search servers. A content descriptor such as a URC provides a classification mechanism for the data items on the network, which is very similar to the bibliographic information contained in a library collection. The standard protocol Z39.50 (19, 20) was designed to communicate with information retrieval systems that search collections of bibliographic materials, including fields such as title and author and operators such as booleans and phrases. Searching also differs from browsing in that a "state" is kept of previous queries and results, which may be reused in new search requests to further constrain the information filtering. Most commercial search systems use some variant of Z39.50, so these protocols are the most likely to be incorporated into the gateways to Internet search servers built into Internet information clients such as Mosaic. The search capability of the commercial systems is far more powerful than the existing Internet servers. The introduction of functional search to the Internet coupled with standard classification of materials

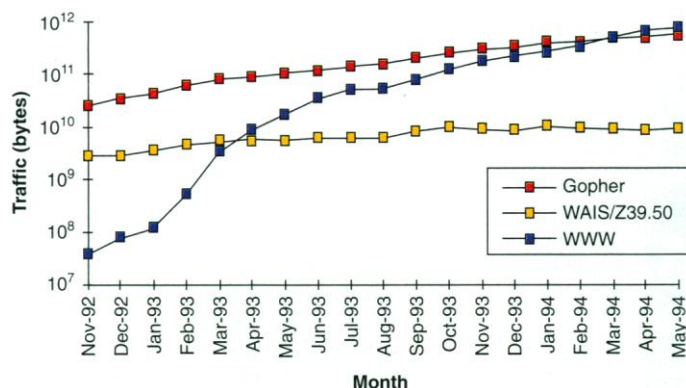
will likely change the character of the navigation of the Net by making digital libraries a widespread reality for the first time.

## Objects and Community Systems

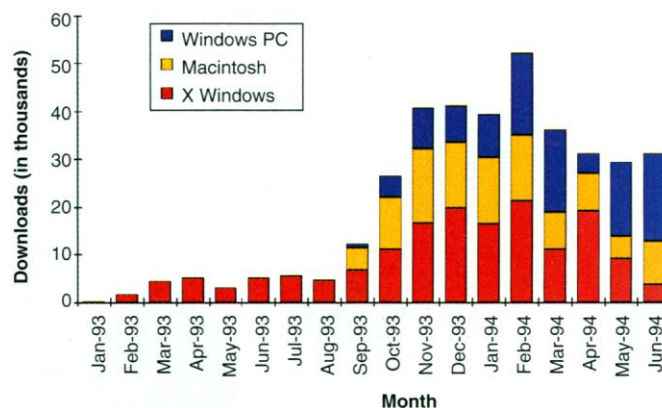
In the longer term future, different architectures may be necessary to meet the continued user demand for new functionality in network information systems. Examination of the evolution of research systems may shed light on the future of Internet services because historical antecedents either directly or indirectly influence current developments.

The previous generation of research systems for network information, in the period 1985 to 1989, preceding the development of WWW, also focused primarily on connection, namely retrieval of documents across the network and the navigation of links between documents. For example, the Telesophy system, developed by the first author while at Bellcore, was deployed at 40 sites around the Internet before the WWW project was begun and provided protocols for the use of documents as a unifying concept for associative search, link navigation, and document authoring, all across a wide variety of multimedia materials distributed across the network, with significant consideration of scaling up (7, 21). Like WWW, Telesophy had a "component" model. That is, there were different displayers (viewers) and searchers (servers), with the client using types to route data to the appropriate software.

The current generation of research network information systems, in the period from 1990 to 1994, has a different architecture, which is centered around structured objects with type checking. This is often referred to as an "object" model and draws on many years of computer science research into object-oriented programming environments (22). In this model, the client is an interpreter rather than a router and has operations specific to each type of object executed within the client itself. An object is a data item with an attached set of operations according to its type. Unlike the typing in the component model, true objects are encapsulated, and their operations have inheritance. Encapsulation means that the objects are always bundled with their operations, so that there is no other way that the data can be accessed. For example, this implies that publication mechanisms, such as privacy control and quality checking, can be strictly enforced by the object handling within the network information system. Inheritance means that object-type operations are arranged in a type hierarchy, so that any



**Fig. 2.** Usage of Internet information services around the time of release of Mosaic.



**Fig. 3.** Propagation of NCSA Mosaic as gauged by the number of downloads from the NCSA FTP server.

operations within a type will be automatically enhanced by those from types above in the hierarchy. For example, this implies that navigation mechanisms, such as link following and link creation, can be guaranteed to be available for any object by defining the operations at the base object-type level.

A crucial difference between the component and the object models is thus the level of guarantee (23). In a component system, facilities are available but not enforced. So, for example, when the interaction facilities become available in Mosaic/WWW, it will be possible for a link in an external viewer to be selected and followed. However, there is no guarantee that this particular viewer will support link following. In an object system, every viewer must conform to the object interaction standards so that link following is always supported. The same is true for creation of new materials as well as retrieval of existing materials. For example, when authoring facilities become available in Mosaic/WWW, the editors or servers are responsible for the correctness of the new materials. There is no guarantee that an item created by one editor cannot be accessed by another in a way that violates its structure, such as adding data in an incorrect format. In an object system, an object of a certain type can only be accessed by operations for that type.

Research prototypes of network information systems exist that provide interaction with structured objects. One example is the Worm Community System (WCS), developed by the first author while at the University of Arizona (24), which was featured in last year's special issue of this journal (25). This system is an experiment in supporting an electronic community where publishing is as important as browsing. It has a client-server architecture with the clients running on workstations in biology labs connecting to servers for genome databases across the Internet. The client knows the type of the data and provides specialized

displays for each type. Thus, it provides an interactive object display, where sets of objects can be displayed graphically and then interacted with, for example by following links. In WCS, these include molecular displays for physical maps and cellular displays for developmental lineages, in addition to documents and forms. The system is symmetric, so that any object that can be retrieved can also be created, directly from the client in the scientist's laboratory. During data entry, the entire object structure is checked for correctness; for example, the system checks if a new gene has a valid name and if its clone field links to an existing clone object. When an object is published, its privacy level of who is permitted to view or modify it is guaranteed by the system.

The object systems guarantee a higher level of complete interaction, at the expense of greater system overhead. Practically, thus far, the component systems have dominated in both the Internet services and the commercial market. The future will judge whether the level of interaction demanded for network information systems can continue to be supplied by component systems through the adoption of features for scalable objects or whether the revolution of new architectures with object systems will displace this evolution. In either case, existing research systems point the way toward future Internet services.

As the need and ability for interacting with many objects simultaneously become greater, the need for guaranteed structure will become greater. In particular, the next generation of research systems, in the period 1995 to 1999, is likely to focus on analysis environments, which provide "complete" support for information, computation, and communication (26). That is, the system is a central controller client that can interact with servers to retrieve items, call programs, and store items. For some operations, the controller can be a router, much as in Mosaic, passing requests from one

external program to another. For other operations, the controller can be an interpreter, much as in WCS, retrieving external objects and executing the operations on these itself. To handle all of the various data types that will be available on the Net in the future, the controller will need to negotiate with the server about the needs of the objects versus the capabilities of the client and then dynamically load an appropriate set of programs to handle the data as well as possible. A multimedia document might thus appear in one format and style on one user platform and in another more appropriate format on another computer platform.

A new branch of science may appear when these complete systems become widely available, supporting interlinked structured objects in a dynamic generic environment. Any user will be able to perform interactive analysis, to discover patterns within the interconnecting web of the global information space. They will be able to individually publish interesting connections for consideration by other users. The goal of this new style of science will be to cross-correlate information within the "dry lab" of the Net.

### Mosaic as an NII Model

One of the most intriguing features of NCSA Mosaic is its sudden acceptance. Its usage has grown on a curve uncommon in scientific circles, gaining over a million users in little more than a year. Figure 3 shows the increase in the number of downloads from the NCSA FTP server, which still underestimates the number of copies because there are many other sources around the world that also distribute this software.

Why did this happen now and why did it happen to Mosaic? The answers are highly relevant in this era of National Information Infrastructure (NII). An examination of the sociological factors behind the rapid adoption may shed light on

what the popular services in the NII will be. These are, of course, somewhat speculative and only a sample of the complicated evolutionary paradigm.

**Access.** The network fabric itself has grown at an exponential rate. In the same year that the Mosaic phenomenon occurred, the Internet itself reached mass popularity, growing to 2 million hosts with an estimated 20 million users (2). In the past few years, the Internet has changed from an esoteric tool for the scientific and research community to an important national topic, the subject of many covers of national magazines and among the highest priorities of the new administration in the federal government. In addition to greatly increasing the connections, the speed also increased, with the bandwidth enabling the transmission of new media across the networks to many users for the first time. Thus, the timing is significant. In the last year, the Net was inundated with new users, largely unfamiliar with computers, seeking something interesting to do and finding Net surfing (27).

**Availability.** The Internet community is accustomed to research quality software, which is easily available but often lacking in support and usability, especially for novice users. That Mosaic was developed and supported by NCSA played a significant role in its adoption because the technology at the point of introduction was ready for widespread deployment but not yet ready for commercial systems. The NCSA is one of the NSF supercomputer centers but is unique in having a large software development group, headed by the second author, specializing in the implementation of "commercial quality" scientific and network software across multiple platforms, which are distributed free via the Internet. The introduction of a native version of Mosaic for Macintoshes and PCs was the turning point of the adoption.

**Features.** The user interface that WWW promotes is extremely simple: display a document with embedded links to other documents that can be fetched by pointing and clicking. Easy access to pictorial materials within documents was finally feasible, both from a multimedia display and from a network speed standpoint. The development of Mosaic concentrated on the front-end client, integrating and using the existing and mature framework of servers and protocols from WWW, Gopher, and FTP sites. This meant that there was a wide body of materials immediately available and streamlined existing mechanisms for adding new materials. In addition, the client was extensible from a display standpoint, so that new data types could be added in the servers if appropriate viewers were added at the same time on the local client machine.

**Community.** The most significant sociological change caused by the new network information systems may well be a revolution in how electronic communities are formed and share their results. The above conditions caused a feedback loop among the users, between the clients and the servers. Users would browse the existing materials and realize how quickly and widely the Net could make information available. They then would set up their own server and publish their own materials. This new information being available then created an even greater demand for other people to obtain a client so that they too could become a user. Mosaic/WWW had a combination of availability and features and a timing of access and information that put them "over the top" into the feedback loop that is essential for successful propagation of network information systems.

What implications might this experience have for services in the NII? In the near future, simple connection services such as in the current Mosaic/WWW will likely spread to become a standard feature of personal computer software, including the computers embedded within the forthcoming video-on-demand televisions. However, once the general public is accustomed to connection, they will begin to demand interaction, such as those features emerging in the evolution of Mosaic/WWW and the revolution of new architectures.

A key factor of interaction is support for the establishment of community. In many scientific communities, publishing on the Internet has already become a primary mechanism for rapid dissemination of methods, discoveries, and knowledge. A significant groundswell of documents and databases have simply appeared in the global information space over the past few years from a wide variety of sources. This is despite the fact that the current information services have little support for composing and publishing, but primarily facilitate browsing and fetching.

The functionality of network information systems determines the medium of electronic communication. In past generations, the rapid communication required for scientific discussion was greatly facilitated by electronic mail and bulletin boards. In the present generation, the technology supports browsing and sharing of text and data. For example, the scientific community is using clients on their desktop computers to access images and animations of the impact of Comet Shoemaker-Levy on Jupiter from servers around the world, placed there directly from the processed telescope data only hours after collection (28). In future generations of network information systems, increased support for individual publishing

and information discovery will provide further new communications media. The revolution of the Net is just beginning.

## REFERENCES AND NOTES

1. V. Cerf and R. Kahn, *IEEE Trans. Commun.* **22**, 637 (May 1974).
2. A. M. Rutkowski, *Internet Soc. News* **2** (no. 4), 6 (1994).
3. J. S. Quarterman, *The Matrix: Computer Networks and Conferencing Systems Worldwide* (Digital Press, Maynard, MA, 1990).
4. D. Comer and D. Stevens, *Internetworking with TCP/IP, Vol. III: Client-Server Programming and Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
5. M. McCahill, *ConneXions: The Interoperability Report* **6** (no. 7), 10 (1992).
6. M. Schwartz, A. Emtage, B. Kahle, B. Neumann, *Comput. Syst.* **5**, 461 (1992).
7. B. R. Schatz, *Proceedings IEEE Globecom '87* (November 1987), pp. 1181-1186; internal Bellcore document proposing Telesophy (August 1984).
8. B. Kahle et al., *Internet Res. Electron. Networking Appl.* **2** (no. 1), 59 (1992).
9. T. Berners-Lee, internal CERN document proposing WWW (October 1990).
10. J. Conklin, *IEEE Comput.* **20** (no. 9), 17 (1987).
11. F. Halasz, *Commun. ACM* **31**, 836 (1988).
12. "Hypertext" was defined by T. Nelson in the 1960s. See *Literary Machines*, 1981 (the Distributors, 702 S. Michigan, South Bend, IN 46618) and *Computer Lib/Dream Machines* (Microsoft Press, Seattle, 1987).
13. T. Berners-Lee, R. Cailliau, J. Groff, B. Pollermann, *Internet Res. Electron. Networking Appl.* **2** (no. 1), 52 (1992); T. Berners-Lee et al., *Commun. ACM* **37**, 76 (1994). See *Internet Res. URL* <http://info.cern.ch/hypertext/WWW/Bibliography/Papers.html>.
14. M. Andreessen and E. Bina, *Internet Res. Electron. Networking Appl.* **4** (no. 1), 7 (1994).
15. A. M. Rutkowski, *Internet Soc. News* **1** (no. 4), 2 (1993).
16. J. H. Coombs, A. H. Renear, S. J. DeRose, *Commun. ACM* **30**, 933 (1987).
17. E. van Herwinjnen, *Practical SGML* (Kluwer, Boston, MA, 1994).
18. T. Berners-Lee, "Internet Draft Standards on URNs/URCs" (1993).
19. C. Tomer, *J. Am. Soc. Inf. Sci.* **43**, 566 (1992).
20. C. Lynch and C. Preston, *Annu. Rev. Inf. Sci. Technol.* **25**, 263 (1990).
21. B. R. Schatz, in *Proceedings of the Fifth IEEE Conference on Data Engineering* (Institute of Electrical and Electronics Engineers, Piscataway, NJ, 1989), pp. 188-197.
22. A. Goldberg and D. Robson, *Smalltalk-80: The Language and Its Implementation* (Addison-Wesley, Reading, MA, 1983).
23. J. Udell, *Byte* **19**, 46 (May 1994).
24. B. R. Schatz, *J. Manage. Inf. Syst.* **8** (no. 3), 87 (winter 1991-1992).
25. R. Pool, *Science* **261**, 841 (1993); *ibid.*, p. 842.
26. B. R. Schatz, plenary talk, Third Keck Symposium on Computational Biology, Houston, TX, November 1992; abstract in *Cell Motil. Cytoskel.* **24**, 286 (1993).
27. For example, the Doonesbury cover of Net surfing, *U.S. News World Rep.* (6 December 1993).
28. Information on the comet is being collated on the Jet Propulsion Lab home page at URL <http://newproducts.jpl.nasa.gov/sl9/sl9.htm>.
29. B. Schatz is supported by an NSF Young Investigator award in science information systems. The NCSA is supported by NSF, the Advanced Research Projects Agency, other federal agencies, corporate partners, the University of Illinois, and the state of Illinois. Thanks are due to all those who provided code and support as NCSA Mosaic developed and to those who continue to support this work and that on the World Wide Web. Finally, thanks to the global Internet community itself for paving the way toward the world of the future by browsing and publishing on the Net.