

Glossary

archie. Archie was developed at McGill University by Peter Deutsch and Alan Emtage, who have since established a private corporation (Bunyip) to further develop and commercialize the technology. In essence, archie automatically creates, manages, and offers access to a database of the contents of the major FTP archives worldwide and thus permits users to locate FTP archives holding files of interest to them. Note that archie is with a small "a".

Ariel. Ariel is a system developed by RLG (which see) that permits documents to be faxed using the Internet rather than the dial telephone network as a transmission medium. Essentially, documents are scanned into a file on the sending host, transmitted across the network using FTP (which see) and then printed (or viewed on screen) at the receiving ARIEL node. ARIEL is based on IBM personal computer hardware platforms.

ARL. The Association of Research Libraries is a not-for-profit organization representing 119 major research libraries in the United States and Canada. ARL supports a wide range of activities, including studies of the function and costs of the research library system, the development and articulation of policy positions on legislation and other government activities that are of interest to the research library community, and development of resource sharing agreements and policies among its members. ARL is also one of the three parent organizations that created the Coalition for Networked Information (CNI).

BRS. BRS, which is now a part of the Maxwell communications empire, is another commercial search service very similar to Dialog (which see), although somewhat less extensive.

CARL. The Colorado Alliance of Research Libraries (CARL) is an organization based in Denver Colorado that started out building an online catalog for a number of Colorado based libraries, but has recently expanded nationally, offering a linked collection of catalogs for major libraries from Hawaii to Maryland. In addition, CARL creates and offers access to a database of journal tables of contents, a companion document delivery service, and a number of commercial databases.

Center For Research Libraries. The Center for Research Libraries (CRL) is an organization set up by a group of research libraries that essentially serves as a central depository for very rarely used material that is considered to be important to retain within the research library community but which is not used enough to justify any single library in this community retaining as part of the local collection; in these cases the material is sent to CRL where it can be available to the entire research library community but can be stored in a relatively inexpensive facility.

C/./. The Coalition for Networked Information (CNI) is a joint project of EDUCOM and CAUSE, two associations concerned with information technology in higher education, and the Association for Research Libraries, an association of the 119 largest research

libraries in North America. Its purpose is to advance scholarship and intellectual productivity through the use of networks and information technology, and it has played an active role in promoting the development of networked information resources and the underlying technologies necessary to implement and use them.

CONTU. CONTU was the National Commission on New Technological Uses of Copyrighted Works which was established by Congress to make recommendations on copyright legislation in light of developments in technology. CONTU issued its final report in 1978. The work carried out by CONTU helped to define community standards for the interactions between technologies such as photocopying (xerography) and the copyright law, and to help define appropriate practices and standards for the acceptable use of these technologies in contexts such as Interlibrary Loan. CONTU also studied computer software and database intellectual property issues. CONTU made a number of recommendations for legislative changes; some were implemented and others were not. While the CONTU guidelines do not have the force of law, the represent a very real community consensus on acceptable behavior. In some cases they have been used by courts to help to interpret the copyright law, although other courts have ignored the CONTU recommendations as having no legal status.

Dialog. Dialog is a commercial online service that dates back to the late 1960s; originally a subsidiary of Lockheed, it is now owned by Knight-Ridder. Dialog essentially acts as a service bureau to database producers, mounting their databases and providing interactive searching access to them. It also handles billing and training, and typically pays royalties back to database producers who make their databases available through Dialog. Dialog is well known both for its very high prices, which have kept its use within the academic community to a minimum, and for its very complex user interface, which, while offering very powerful search capabilities, is really intended for use by trained searchers and not by end users. Dialog offers access to hundreds of databases; some of these are exclusively accessible through Dialog; others are also available through other channels (for example, CD-ROM, tape licenses direct to institutions, or other competing online services).

Fair Use. Fair use is a provision of the US copyright law which permits copying of copyrighted material for specific purposes, such as personal research. More generally, the term "fair use" is used to refer to the entire set of specific copying exemptions in the law, which include some provision for making a copy of an out of print work that is deteriorating for preservation purposes (if no other reasonable alternative is available), satire and criticism, and other exemptions.

FTP. FTP is the file transfer protocol within the TCP/IP protocol suite and widely used on the Internet to copy files from one host to another. It is the access mechanism used with FTP archives, which are simply large collections of files that are stored on various hosts on the Internet and available for copying. FTP supports an anonymous access mode that allows users to list and copy public files from many hosts on the network without any need for pre-registry with the archive host manager.

Gopher. Gopher is a system that was developed at the University of Minnesota in the early 1990s. It is a distributed system consisting of client programs for a variety of hardware platforms (including PCs, UNIX, and Macintoshes) and servers (again for

various platforms) which allow a server manager to establish a database of menus. The menu entries can point to other menus, either on the local Gopher server or any other Gopher server on the Internet, to documents (stored anywhere on the net), to interactive services, or various other types of information objects or resources. Essentially, Gopher offers a fairly simple, low cost means for an institution or even an individual to provide menu-based access to a variety of networked information (including locally stored information).

IETF. The Internet Engineering Task Force is a self-selected volunteer group that meets quarterly to develop standards for the Internet. While it is not an officially sanctioned national or international standards development body, it operates primarily by consensus and has been tremendously effective in managing the standards needed for the Internet environment. Standards developed by the IETF are called Requests for Comments (RFCs) for historical reasons. The IETF has a number of active working groups dealing with various aspects of standards and architectural models for networked information resources.

LISTSERV. *LISTSERV* (an abbreviation for list server) is a program that was originally developed for the IBM CMS operating system; more recently imitation programs with very similar functionality have been developed for the UNIX operating system environment. Essentially, the *LISTSERV* program permits mailing lists to be established with a wide range of parameters. Typically, users can join or leave a mailing list by sending commands to the *LISTSERV* program by electronic mail. Once they have joined a given mailing list, they can send a mail message to the list which will be echoed to all other subscribers of the list. The *LISTSERV* program also supports a variety of maintenance functions, such as maintaining and offering access to archived messages, permitting people to get lists of the people subscribed to a given mailing list, and the like. *LISTSERV* lists can also be defined as moderated, which means that the list moderator must approve all postings to the list before they are sent out to other subscribers.

Mail Reflectors. A mail reflector is essentially a simple, manual form of a *LISTSERV* (which see). Mail reflectors simply re-transmit mail to a list of users who are tabled as part of that mail reflector; in this sense a mail reflector can be viewed as simply a shorthand for sending mail to a list of people. Maintenance of the list is typically manual (perhaps assisted by computer programs) as distinct from the completely automated list management of *LISTSERVs*. In addition, there is usually no easy way to find out who is signed up on a given mail reflector.

Multicast. Multicast technology is a network facility which permits information to be broadcast to a specific subscriber group of machines attached to the Internet. Rather than having the source send one copy of the information to each recipient in the multicast group separately, multicasting permits the source machine to simply transmit one copy of the data, addressed to the multicast group address, onto the network; the network routing services take care of duplicating it as required so that every machine in the multicast group receives its own copy. Network hosts can join and leave a specific multicast group at will. Currently, multicast technology is used primarily on an experimental basis in the Internet (although it is a basic network service within some types of local area networking technology, such as Ethernet) to carry audio and/or

video traffic to groups of interested recipients in real time. Only a portion of the Internet (called the MBONE, or multicast backbone) supports multicast service at this time.

NNTP. The Network News Transfer Protocol is the protocol that is used by news readers that wish to access Usenet news groups from a newsgroup server. It is defined by RFC 997 [Kantor 1986].

OCLC. The Online Computer Library Center (formerly the Ohio Computer Library Center) is a not for profit organization based in Dublin, Ohio which provides a wide variety of services to the library community. These include interlibrary loan requesting and tracking, shared cataloging using an enormous copy cataloging database, and recently access to a variety of online databases (in a sense competing with commercial firms like Dialog (which see)). Recently OCLC has also partnered with the American Association for the Advancement of Science (AAAS) to mount the electronic journal *Current Clinical Trials*, and is currently working on several other electronic journals in conjunction with various other professional societies. The model for *CCT* is that of a database, where OCLC makes available the necessary user interface software to read articles, view them, and, optionally, print them, rather than the more common model used by other electronic journals where the contents of the journal are often physically transmitted to each subscriber.

PostScript. PostScript is a language developed by Adobe Systems to communicate with printers, particularly high-quality laser printers. Typically, word processors "print" a document by converting it into PostScript form; this is then sent to a printer for interpretation which produces the actual printed page. Programs also exist to preview PostScript files on display devices. It is very difficult to go backwards from PostScript to a revisable form document, or even one that permits reasonable full text searching; in addition, all semantic level markup (and even most syntactic level markup) is lost when a document is converted to PostScript form. In this sense, a PostScript document representation can be viewed as quite similar to a bitmapped image, although it is more compact.

Project *Gutenberg*. Project Gutenberg, managed by Michael Hart, creates (either by scanning or keyboarding) ASCII versions of out-of-copyright books and other works, and makes these materials freely available over the Internet via anonymous FTP.

RLG. RLG is the Research Libraries Group, a consortium of research libraries in north America. One of the major activities of RLG is the operation of RLIN (the Research Libraries Information Network), a service similar to OCLC (which see) that provides libraries with access to a large shared cataloging database, an interlibrary loan requesting and tracking system, and a number of scholarly databases targeted for end users in academic institutions.

SGML. Standard Generalized Markup Language is an international standard for defining markup tagging for text. This can be at a relatively superficial level, where only syntactic structures such as headings and paragraphs are marked, thus facilitating the reformatting of a document for presentation in multiple environments (for example, in print or on a display terminal), or the markup can define very deep semantic meaning,

as is being doing in the Text Encoding Initiative project, which addresses the needs of the humanities computing community to encode text for subsequent computer analysis.

TELNET. TELNET is a protocol within the TCP/IP protocol suite and widely used on the Internet to conduct a terminal session with an interactive service on a remote host. It can be viewed as the network equivalent of dialing up a remote host with a character based terminal.

TOPNODE. This is a project of the Coalition for Networked Information (CNI - which see) to experiment with the description and cataloging of networked information resources.

Usenet *Newsgroups*. These are collections of electronic mail messages that are distributed throughout the Internet and beyond using the Usenet news system. There are many hundreds of such newsgroups, with more being established all the time. Total Usenet traffic is now measured in tens of millions of characters daily. Usenet news groups include the infamous "ALT" newsgroups, which are unmoderated and which deal with rather controversial issues such as sex and drugs, and which have been subject to censorship from time to time by various institutions. One subscribes to and reads Usenet news groups using a news reader. For a more extensive description of Usenet, see [Quarterman].

WA/S. WAIS stands for Wide Area Information Server; this is a system that is based on the **Z39.50** information retrieval protocol (which see) that was originally developed in the early 1990s by Thinking Machines, Apple, Dow Jones and KPMG. The original code was publicly distributed. Since that time a startup company (WAIS Inc.) has been established by several of the original developers to commercialize the technology, while work on the upgrading and extension of the public domain version continues through organizations like the Clearinghouse for Networked Information Discovery and Retrieval in North Carolina. Essentially, WAIS permits full text searching (using sophisticated statistical ranking and matching algorithms) against databases distributed across the Internet using a common user interface.

WWW. WWW is the WorldWide Web system originally developed by Tim Berners-Lee at the CERN center in Geneva, Switzerland. It can be viewed as a network-based hypertext system that allows linkages between arbitrary information objects stored on hosts throughout the network.

Z39.50. This is a US national standard developed by NISO, the National Information Standards Organization, an ANSI accredited standards developing body serving the publishing, library, and information services communities. The standard addresses computer to computer information retrieval and provides a basis for the development of network-based applications that offer a common interface to multiple, autonomously managed information servers.