Which White Pages Service is Appropriate for My Site?

Michael F. Schwartz
University of Colorado - Boulder
Published in Internet Society News 1(4), Winter 1993

A problem of increasing concern in the Internet is locating address, telephone, and electronic mailbox information about network users. In this article I compare several systems capable of supporting this function. I do not consider systems for non-Internet sites, personal "rolodex" systems, or efforts to coordinate/administer directory service activities ([Conklin 1992, National Science Foundation 1992, North American Directory Forum 1992]).

A simple solution is provided by the WHOIS service, which supports a centralized database of registered information [Harrenstien, Stahl & Feinler 1985]. Originally, WHOIS was run by a single ARPANET Network Information Center (NIC). Because of the increasing scale and decentralization of the Internet, there are now many NICs as well as individual sites running WHOIS servers.

Another popular centralized service is CSO (also known as PH, 411, and da) [Dorner & Pomes 1992]. CSO servers often respond to search requests at the finger port [Zimmerman 1990], on machines aliased to the name of a domain. For example, by fingering "schwartz@colorado.edu", you can locate information about all users named "schwartz" at the University of Colorado.

NICs typically populate their directory services with data from registration requests (such as templates filled out by users). In contrast, individual sites typically populate their databases en masse from primary registration sources (such as a university student enrollment database). En masse registrations usually contain more complete and timely data, but this method raises issues regarding privacy and update control. If a user is registered in a NIC database, there is an additional advantage: the user can be located without knowing her/his place of employment. For example, one can query the U.S. Defense Information Systems Agency (DISA) NIC's WHOIS service for all registered users named "schwartz". However, such centralized solutions do not scale well as the network continues to grow.

Neither WHOIS nor CSO supports distributed operation. In particular, neither service coordinates format or content across servers, and neither supports caching or replication. In contrast, the CCITT and ISO jointly developed a distributed directory service standard called X.500, which describes a hierarchical name space with provisions for caching, authentication, and replication [CCITT/ISO 1988]. Beyond support for distribution, another advantage of X.500 is that it supports typed, self-describing fields. This means that users can make more sophisticated queries using X.500, and the database can support more complex information (e.g., bitmap photos of users).

While X.500 offers many advantages, problems in the current implementations have kept X.500 servers from being deployed at more than a few hundred sites worldwide. This lack of "reach" has further eroded sites' interest in installing X.500 software. The principal problems with the current implementations are availability, and the level of machine and human administrative resources required to run a server. A number of efforts are under way to improve these problems. In time, I believe X.500 will become significantly better and more popular.

Netfind provides white pages service by using a number of existing sources of information [Schwartz & Tsirigotis 1991]. Because of the nature of these information sources and the way Netfind cross correlates the information, Netfind can locate users at over 5,000 sites world wide. Moreover, because it searches for users on their home workstations, Netfind can often locate more timely information than the other services discussed here (which depend on registration databases). A downside of this approach is that searches generate more network traffic than the other services. However, this traffic is small compared to other Internet applications, such as remote login, mail/news, and file transfer. A second disadvantage of Netfind is that it can only locate people at directly connected Internet sites (which excludes sites that insulate their internal networks using "firewall" gateways).

The Wide Area Information Servers (WAIS) system [Kahle & Medlar 1991] allows users to deploy, search, and retrieve many different types of information from machines distributed throughout the Internet. As such, one use for WAIS can be for a user directory service. Because each index contains keywords for every word that appears in a document, WAIS supports fairly flexible searches (but not as flexible, for example, as a search system that allows regular expressions). A downside is that the indices require approximately as much space as the data being indexed.

As the number of different directory services available in the Internet has increased, several systems have been developed to provide a uniform user interface to the information. The Knowbot(TM) Information Service (KIS) [Droms 1990] provides a consistent, structured input/output format for accessing multiple user directory services. The Internet Gopher system [McCahill 1992] provides a less consistent menu-oriented interface, but provides access to many more information sources (including user directories and other types of information). Gopher requires the user to select an individual information source before searching. In contrast, KIS sequentially requests information from each of the services it knows about, and allows users to restrict search to a subset of the sources. Human nature being what it is, most users do not restrict searches. This increases search delays and network costs, and in some cases provides users with a large amount of irrelevant data. At present, Gopher supports access to 28 CSO servers, 90 WHOIS servers, Netfind, a WAIS database of USENET user electronic mailboxes, and X.500. KIS supports access to 6 campus directory systems, the DISA NIC's WHOIS service, X.500, and an MCMail directory service.

So, what system should a site choose? Unfortunately, there is no single, ideal solution. Netfind provides the largest colletion of Internet reachable user data, but supports only limited data types, and can only locate people at directly connected Internet sites. X.500 supports the most sophisticated functionality, but its current implementations have availability problems and high resource demands, and there are not enough sites running servers to provide a useful amount of data. Neither Netfind nor X.500 support searches without specifying something about where the person being sought works. WHOIS, CSO, and WAIS can support such searches, but provide no explicit support for data distribution. These problems can be remedied to a certain extent by registering with a centralized server (e.g., the WHOIS server for a national network), but this approach does not scale well.

In the longer term, efforts by cooperative organizations, governments, or commercial service providers will probably improve Internet directory service offerings. In the mean time, sites will probably need to choose a combination of systems to support the needs of their users.

The DISA NIC's WHOIS service can be accessed by telnet to nic.ddn.mil, by logging in as "nic". An ad hoc list of whois servers is available by anonymous FTP from sipb.mit.edu, in /pub/whois/whois-servers.list. The CSO software is available by anonymous FTP from ux1.cso.uiuc.edu, in pub/qi.tar.Z. The QUIPU software is available by anonymous FTP from ftp.psi.com, in wp/. Netfind can be accessed by telnet to bruno.cs.colorado.edu, and logging in as "netfind" (the software is also available). WAIS is available by anonymous FTP from think.com, in wais/. KIS can be accessed by doing "telnet nri.reston.va.us 186". Gopher can be accessed by telnet to consultant.micro.umn.edu, and logging in as "gopher".

References

[CCITT/ISO 1988]
CCITT/ISO. The Directory, Part 1: Overview of Concepts, Models and Services. CCITT/ISO, Gloucester, England, Dec. 1988. CCITT Draft Recommendation X.500/ISO DIS 9594-1.

[Conklin 1992]
J. Conklin. Announcement of CREN White Pages Project. Corporation for Research and Educational Networking, Aug. 1992.

[Dorner & Pomes 1992]
S. Dorner and P. Pomes. The CCSO Nameserver - A Description. Computer and Communications Services Office, Univ. of Illinois at Urbana, Aug. 1992.

[Droms 1990]

R. E. Droms. Access to Heterogeneous Directory Services. Proc. 9th Joint Conf. of IEEE Computer and Communications Societies (InfoCom), June 1990.

[Harrenstien, Stahl & Feinler 1985]

K. Harrenstien, M. Stahl and E. Feinler. NICName/Whois. Req. For Com. 954, SRI International, Oct. 1985.

[Kahle & Medlar 1991]

B. Kahle and A. Medlar. An Information System for Corporate Users: Wide Area Information Servers. ConneXions - The Interoperability Report, 5(11), pp. 2-9, Interop, Inc., Nov. 1991.

[McCahill 1992]

M. McCahill. The Internet Gopher: A Distributed Server Information System. ConneXions - The Interoperability Report, 6(7), pp. 10-14, Interop, Inc, July 1992.

[National Science Foundation 1992]

National Science Foundation. Network Information Services Manager(s) for NSFNET and the NFEN. 1992.

[North American Directory Forum 1992]

North American Directory Forum. User Bill of Rights for Entries and Listings in the Public Directory. Req. For Com. 1295, North American Directory Forum, Jan. 1992.

[Schwartz & Tsirigotis 1991]

M. F. Schwartz and P. G. Tsirigotis. Experience with a Semantically Cognizant Internet White Pages Directory Tool. J. Internetworking: Research and Experience, 2(1), pp. 23-50, Mar. 1991.

[Zimmerman 1990]

D. Zimmerman. The Finger User Information Protocol. Req. For Com. 1288, Center for Discrete Mathematics and Theoretical Computer Science, Nov. 1990.