

Network Working Group  
Internet Draft  
Document: [draft-klensin-tld-whois-02.txt](#)

J Klensin  
MCI  
T Wolf, Jr.  
Dun & Bradstreet  
July 29, 1997

## Domain Names and Company Name Retrieval

### Status of this Memo

This document is an Internet Draft. Internet Drafts are working documents of the Internet Engineering Task Force (IETF), its Areas, and its Working Groups. Note that other groups may also distribute working documents as Internet Drafts.

Internet Drafts are draft documents valid for a maximum of six months. Internet Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet Drafts as reference material or to cite them other than as a ``working draft'' or ``work in progress``.

To learn the current status of any Internet-Draft, please check the `1id-abstracts.txt` listing contained in the Internet-Drafts Shadow Directories on `ds.internic.net`, `nic.nordu.net`, `ftp.isi.edu`, or `munari.oz.au`.

A revised version of this draft document may be submitted to the RFC Editor for processing as an Experimental RFC for the Internet Community. Discussion and suggestions for improvement are requested. This draft will expire before January 30, 1998. Distribution of this draft is unlimited.

Changes from prior draft: Small technical clarifications.

### Abstract

Location of web information for particular companies based on their names has become an increasingly difficult problem and the Internet and the web grow. The use of a naming convention and the domain name system (DNS) for that purpose has caused complications for the latter while not solving the problem. While there have been several proposals to use contemporary, high-capability, directory service and search protocols to reduce the dependencies on DNS conventions, none of them have been significantly deployed.

This document proposes a company name to URL mapping service based on the oldest and least complex of Internet directory protocols, whois, in order to explore whether and extremely simple and widely-deployed protocol can succeed where more complex and powerful options have failed or been excessively delayed.

## 1. Introduction and Context

In recent months, there have been many discussions in various segments of the Internet community about "the top level domain problem". Perhaps characteristically, that term is used by different groups to identify different, and perhaps nearly orthogonal, issues. Those issues include:

- 1.1. A "domain administration policy" issue.
- 1.2. A "name ownership" issue, of which the trademark issue may constitute a special case.
- 1.3. An information location issue, specifically the problem of locating the appropriate domain, or information tied to a domain, for an entity given the name by which that entity is usually known.

Of these, controversies about the first two may be inevitable consequences of the growth of the Internet. There have been intermittent difficulties with top level domain administration and various attempts to use the domain registry function as a mechanism for control of service providers or services from time to time since a large number of such domains started being allocated. Those problems led to the publication of the policy guidelines of [[RFC1591](#)].

The third appears to be largely a consequence of the explosive growth of the World Wide Web and, in particular, the exposure of URL formats [[URL](#)] to the end user because no other mechanisms have been available. The absence of an appropriate and adequately-deployed directory service has led to the assumption that it should be possible to locate the web pages for a company by use of a naming convention involving that company's name or product name, i.e., for the XYZ Company, a web page located at

<http://www.xyz.com/>  
or  
<http://www.xyz-company.com/>

has been assumed.

However, as the network grows and as increasing numbers of web sites are rooted in domains other than ".COM", this convention becomes difficult to sustain: there will be too many organizations or companies with legitimate claims --perhaps in different lines of business or jurisdictions-- to the same short descriptive names. For that reason, there has been a general sense in the community for several years that the solution to this information location problem lies, not in changes to the domain name system, but in some type of directory service.

But such directory services have not come into being. There has been ongoing controversy about choices of protocols and accessing mechanisms. IETF has published specifications for several

different directory and search protocols, including [WHOIS++], [RWHOIS], [LDAP], [X500], [GOPHER]. One hypothesis about why this has not happened is that these mechanisms have been hard to select and deploy because they are much more complex than is necessary. This document proposes an extremely simple alternative.

## **2. Using WHOIS**

The WHOIS protocol is the oldest directory access protocol in use on the Internet, dating in published form to March 1982 and first implemented somewhat earlier. The protocol itself is simple and minimalist: the client opens a telnet connection to the WHOIS port (43) and transmits a line over it. The server looks up the line in a fashion that it defines, returns one or more lines of information to the client, and closes the connection.

We suggest that modifications or add-ins be created to Web browsers that would access a new, commercially-provided Whois server, sending a putative company name and receiving back one or more lines, each containing a URL followed by one or more blanks and then a matching company name (that order was chosen to minimize parsing problems: since URLs cannot contain blanks, the first blank character marks the end of the URL and the next non-blank marks the beginning of the company name). As is usual with Whois, the criteria used by the server to match the incoming string is at the server's discretion. The difference between this and the protocol as documented in [WHOIS] is that exactly one company name is returned per line (see [section 3](#) for details of syntax).

The client would then be expected to:

- (i) If a single line (company name and URL) is returned, either ask for confirmation or simply fetch the associated URL as if it had been typed by the user.
- (ii) If multiple lines (names) are returned, present the user with a choice, presumably showing company names rather than (or supplemented by) URLs, then fetch using the URL selected.

Obviously, while the most convenient use of the services contemplated in this document would occur through a client that was part of, or intimately connected with, a Web browser, a user without that type of facility could utilize a traditional WHOIS client and paste or otherwise transfer the relevant information into the target location of a browser.

## **3. Formats, versions, and international character sets**

Preliminary work with the approach suggested above suggests that some specific conventions about syntax and variations would be useful.

### **3.1 Line sent from client to server.**

These lines may take either of two forms:

- (i) A simple 7-bit ASCII string, containing a "company name"
- (ii) A string in the format (using the ABNF notation of [RFC 822](#)):  
Variation "/" 1\*Octet

```
Variation ::= "0" | ( Non-zero-digit 1*Digit)
Non-zero-digit ::= 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
Digit ::= 0 | Non-zero-digit
```

Where Octet is any eight-bit sequence, representing a prefixed variation number.

The first form will be construed as equivalent to the second form with the leading string "0/". Variation numbers are specified in [section 3.3](#).

In all cases, the interpretation of what "company name" might mean and, in particular, what variations of form or spelling, abbreviations, and so on, might be accepted is strictly up to the interpretation of the server. If rules driving the server lead to the conclusion that a string matches some company in its data, the correctness or incorrectness of that decision is not covered by this specification.

For variation 0 and, by default for all others, any alphabetic text in lines is to be construed in a case-insensitive fashion.

### **3.2 Lines sent from server to client.**

The server is expected to return one or more lines to the client, depending on its interpretation of the input string. In general, each line will consist, as described above, of a URL, a space, and a "company name". This document deliberately does not specify the content or semantics of the "company name" string. It might be a name, or a name and descriptive information such as location and type of business, or other information at the option of the server. The expectation, as mentioned above, is that the information will be displayed by the client to aid users in selecting the appropriate URL.

These lines, consistent with normal Internet practice, will be terminated by a CR LF sequence (rather than one or the other of those control characters).

When and if different variation numbers are introduced, their specifications may include variations on what the server is expected to return.

In lieu of "URL and company name" responses, the Server may also return "error messages". These take the form of lines containing:

```
"///" SP String
```

where the String is 7-bit ASCII with no control characters other than SP, unless the variation associated with the variation number specifies otherwise. For this experiment, all "error messages" but the following two are discouraged:

```
/// Not found
    Indicating that the "company name" does not
    match anything
/// Variation not supported
    Indicating that the variation number supplied
    by the client is not recognized by the server.
```

### **3.3. Registered variations**

The following two variations are established as part of this specification:

0/ Query and response are in 7-bit ASCII, no controls other than SP, "Company name" separated from URL by one or more SP characters.

1/ Query and response are in UTF-8, no controls other than SP, "Company name" separated from URL by one or more SP characters, no specification of language on

either

input or output.

The authors will maintain a registry of additional variations which they hope will be very short (see [section 9](#)). If this specification evolves into a proposed standard after an experimental period, the draft for that standard will propose that the registry be turned over to IANA.

## **4. Alternatives not chosen**

Few comments on the initial draft of this document addressed the basic model or protocol design for the service discussed. Instead, they focused on inquiring about the decisions we didn't make and about beliefs about the protocol specification that were not intended by the authors. The latter have been, we hope, corrected. Questions of the following three types predominated in the first category.

### **4.1. Why didn't you use <insert-favorite-directory-protocol-here>?**

Many notes raised the question of how much more could be done with a higher-powered directory protocol rather than the extremely simple WHOIS. Questions were raised about LDAP, X.500 DAP, CCSO, RWHOIS, and WHOIS++. We had several reasons for avoiding them. The most important has been a strong commitment to see how much can be done with an extremely simplistic approach, and WHOIS represented the most simplistic approach we could find. If it turns out to be too simple in practice, things can always evolve

to one or more of the more advanced protocols. But, if we started with one of them, we would never get that information. Other issues included:

- \* None of the existing directory proposals has really emerged as the "right" solution with a large installed base. The deployed base of WHOIS and WHOIS clients is huge, and using it avoids either having to make a premature choice of "winner" or to become embroiled in the debate.
- \* For the casual user, the mechanisms needed to activate the extensive attribute-based directory searches of the stronger protocols are just too complicated and may actually act as a deterrent to effective use.
- \* Substantially since the dawn of the ARPANET, the Internet experience has been that setting up a directory service is easy, but that maintaining one and keeping the records up-to-date is extremely difficult. The economics of operating an effective directory service and keeping everything up to date may will require a revenue-producing product. Use of a very simple protocol for the basic service creates a situation in which basic service can rationally be given away while more advanced service are operated on a charge or subscription basis.

#### **4.2 And why not use a Web search engine?**

Web search engines are immensely effective and powerful, but address a different problem than this protocol. The protocol model here does involve a directory lookup, using a presumed company name as a key. The quality of the result will depend on the quality of the underlying directory and the editorial and research work that goes into its construction (neither of which are matters for the protocol itself -- we trust that marketplace pressures will separate good servers from poor ones). Web search engines are often more effective at locating information about companies than the specific company-designated web pages.

#### **4.3. Why not return a more highly structured information format rather than a simple pair of URL and "company name"?**

Again, the goal was to keep things extremely simple and, in particular, permit minimal interpretation between the user's input and the query and between the response and a display or action. Some of the inquiries on this subject were due to misunderstandings about the implications of the "company name" field; the semantics of that field have been clarified above. We also wanted to avoid the level of standardization implied by a tagging scheme: highly-structured fields might lead either to interoperability problems or excessive restriction on what might be returned.

### **5. Thoughts on Directory Providers**

There is no technical reason why there should be only one provider of company name to URL mapping services using this protocol, nor is there any reason for registries of such providers. Presumably, servers that provide the best-quality mappings will eventually prevail in the marketplace. However, as with most traditional uses of WHOIS, it is desirable for implementations of clients (or Web browsers supporting this protocol) to allow for user choice of servers through configuration options or the equivalent.

## **6. References**

- [RFC1591] J. Postel, "Domain Name System Structure and Delegation", [RFC 1591](#), March 3, 1994
- [GOPHER] F. Anklesaria, M. McCahill, P. Lindner, D. Johnson, D. John, D. Torrey, B. Alberti, "The Internet Gopher Protocol (a distributed document search and retrieval protocol)", [RFC 1436](#), 03/18/1993.
- [LDAP] W. Yeong, T. Howes, S. Kille, "Lightweight Directory Access Protocol", [RFC 1777](#), 03/28/1995.
- [RWHOIS] S. Williamson, M. Kusters, "Referral Whois Protocol (RWhois)", [RFC 1714](#), 12/15/1994.
- [URL] T. Berners-Lee, L. Masinter, M. McCahill, "Uniform Resource Locators (URL)", [RFC 1738](#), December 20, 1994.
- [WHOIS] E. Feinler, K. Harrenstien, M. Stahl, "NICNAME/WHOIS", [RFC 954](#), 0/01/1985.
- [WHOIS++] P. Deutsch, R. Schoultz, P. Faltstrom, C. Weider, "Architecture of the WHOIS++ service", [RFC 1835](#), August 16, 1995.
- [X500] R. Wright, A. Getchell, T. Howes, S. Sataluri, P. Yee, W. Yeong, "Recommendations for an X.500 Production Directory Service", [RFC 1803](#), 06/07/1995.
- [Z39.50] C. Lynch, "Using the Z39.50 Information Retrieval Protocol in the Internet Environment", [RFC 1729](#), 12/16/1994.

## **7. Security Considerations**

This suggested use of the WHOIS protocol adds no significant security risks to those of traditional applications of the protocol which is one of the most widely-deployed applications on the Internet. As usual, servers should expect to use the string sent to them as an information retrieval key, not as a function to be executed in some way. A more significant risk would arise if the server supporting the translation function were somehow spoofed; in that case, an incorrect URL might be returned for a particular company. As with the possibility of finding an

incorrect page using naming conventions, the best protection against the risks that could then occur is careful attention to certificates, signatures, and other authenticity-indicating information.

## **8. Acknowledgements**

This memo was inspired by a many discussions over the last few years about the status and uses of the domain name system, information location using conventions about domain names, exposure of URLs to end users, and convergence of directory and search protocols. While the people involved are too numerous to attempt to list, the authors would like to acknowledge their contributions and comments.

Martin Hamilton, Keith Moore, and Gary Oglesby made important suggestions that have contributed to the revision of this draft.

## **9. Authors' Address**

John C. Klensin  
MCI Internet Architecture  
800 Boylston St, 7th floor  
Boston, MA 02199  
USA

Email: klensin@mci.net  
Tel: +1 617 960 1011

Ted Wolf, Jr.  
Electronic Commerce  
Dun & Bradstreet Information Services  
3 Sylvan Way  
Parsippany, NJ 07054  
USA

Email: ted@usa.net  
Tel: +1 201 605 6308

Address for purposes of registering variants only:  
url-whois@alpha1.reston.mci.net