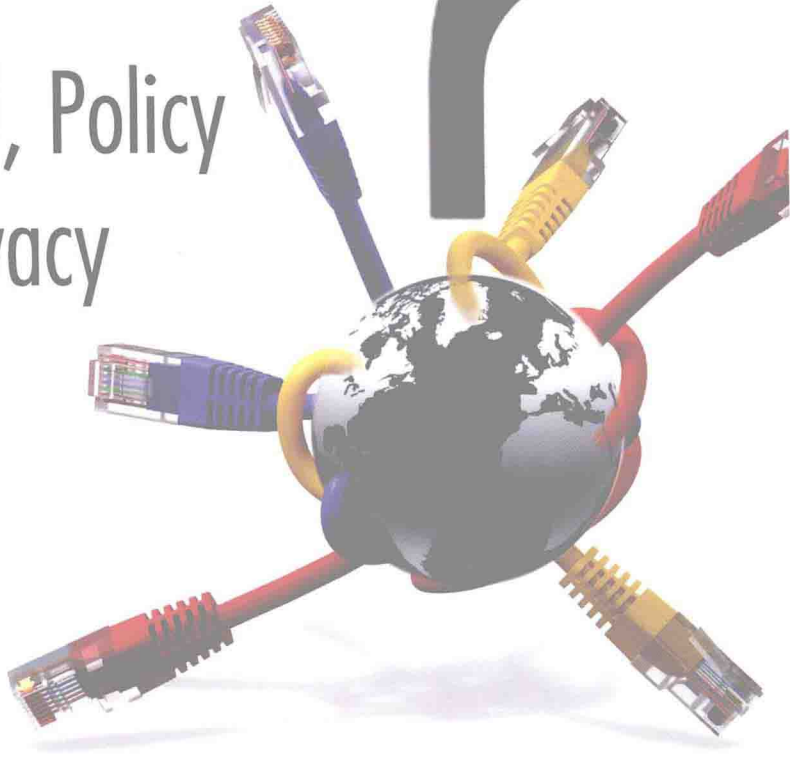
WHOIS Running the Internet

Protocol, Policy and Privacy



Garth O. Bruen

WILEY

WHOIS RUNNING THE INTERNET

Protocol, Policy, and Privacy

GARTH O. BRUEN

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data

Bruen, Garth O.
WHOIS running the Internet : protocol, policy, and privacy / Garth O. Bruen.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-67955-5 (hardback)
1. WHOIS (Computer network protocol) 2. Internet domain names–Government policy. 3. Privacy, Right of. I. Title.
TK5105.5864.B78 2015
004.67'8–dc23

2015020393

Cover image courtesy of *137867102/Jorg Greuel/Getty 455164793/rozkmina/Getty*

Set in 10/12pt Times by SPi Global, Pondicherry, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

1 2016

This book is dedicated to my father, Robert, who taught me all I know and never stopped teaching.

CONTENTS

INTRODUCTION: WHAT IS WHOIS?

WHOIS is a complex topic, as this book explains, but the simplest explanation is that it is a record system for network resources, mostly, but not exclusively on the Internet. WHOIS is one of the most critical and controversial services on the Internet, yet there has been little or no comprehensive documentation. A WHOIS service can be queried to return a WHOIS record, which details who owns or manages an Internet resource. While this service may seem ordinary, WHOIS is one of the most debated issues in Internet policy. In theory, WHOIS is supposed to simply retrieve contact information; in practice, WHOIS varies widely in composition, access, and use. This text covers the universe of topics and issues including the 40-year evolution of the service, policy changes, comprehensive use instructions, service deployment, and advanced coding for programmers. The text is wide in its breadth and attempts to be somewhat deep in each of the major areas, but there are limitations to coverage in a single text.

Unlike computer programming, networking, or hardware development, WHOIS is a disconnected and esoteric discipline. It has many self-taught adepts as well as almost cultish followers. WHOIS is a deep and wide subject without dedicated texts or classroom instruction, a truly strange and hidden world. Welcome, you are about to become a WHOIS sorcerer.

From RFC1177[1] FYI on Questions and Answers to Commonly asked "New Internet User" Questions (1990)

> *WHOIS: An Internet program which allows users to query a database of people and other Internet entities, such as domains, networks, and hosts, kept at the NIC. The information for people shows a person's company name, address, phone number and email address.*

[1] http://tools.ietf.org/pdf/rfc1177.pdf

WHOIS Running the Internet: Protocol, Policy, and Privacy, First Edition. Garth O. Bruen.

Same language in the 1991 version[2]

In its modern usage, "WHOIS" has become a bit of a misnomer. A more accurate term would be "WHOOWNS," "WHOCONTROLS," or "WHOISRESPONSIBLE" since the original WHOIS identified personal accounts or machines tied to a specific person or entity. The one-to-one concept of a resource on the Internet simply no longer applies in most cases, and the WHOIS record will in fact reveal multiple parties with their hands on a domain name or Internet Protocol (IP) address.

Performing a domain WHOIS query lookup on "wiley.com" returns this data:

```
John Wiley & Sons, Inc
Domain Administrator
111 River Street
Hoboken, NJ 07030
US
Phone: +1.3175723355
Email: domains@wiley.com
```

The IP address for wiley.com is 208.215.179.146. A WHOIS query lookup on this address returns this data:

```
Name  John Wiley & Sons
Handle      C00546298
Street      432 Elizabeth Avenue
City  Somerset
State/Province    NJ
Postal Code 08875
Country     US
```

These are two very simple examples of a system, which provokes intense concerns about cybercrime, invasion of privacy, and even the survivability of a single global Internet.

The term WHOIS can refer ambiguously to a service program, a database that stores WHOIS records and the WHOIS record itself. The original reason for having these records and making them publicly available is simple: every node on the Internet is capable of passing traffic to another node, which is what makes the Internet work. If one node has functional problems, it threatens the overall operation of the Internet so other administrators must have the ability to contact the owner of a node experiencing a problem somewhere in the chain.

WHOIS as a protocol concept essentially started in 1971 with the creation of Finger, a program that allowed users on a network to retrieve details about other active users on the network. This was most likely the first time it became possible to remotely create a live "online" connection. An updated version in 1977 called Name/Finger actually introduced the term "Whois" as part of the program function. Being able to see who else is on the network and retrieve information about those persons is a fundamental pillar of the Internet, but one also seen as contributing to the decline in personal privacy. There were so few participants on the early network that sharing contact information

[2]http://tools.ietf.org/pdf/rfc1206.pdf

was not considered controversial. As the network steadily grew, some started to see the public availability of this information as a threat. However, it is generally acknowledged that allowing unaccountable parties onto the public network is just as dangerous. A balance must be found between security and privacy. To address this, a sizable portion of the text is dedicated to this debate.

Following the precise path of the growth of the Internet, WHOIS has experienced changes and even mutations. Unknown to most, there is in fact no single WHOIS database or standard for the Internet. There may be as many as 1500 public WHOIS databases, each with its own rules, formatting, and level of service. The number of WHOIS records currently in existence may exceed 200 million. WHOIS is a massive pile of data with names, addresses, phone numbers, and network resources that explains who owns what is on the Internet.

WHOIS records have long been required for IP addresses and for Internet hostnames. When domain names became available for public consumption, the WHOIS controversy exploded. Criminals began deliberately falsifying WHOIS records, shady marketers exploited the publicly available contact information, and noncommercial domain owners feared for their privacy and safety.

The future of WHOIS is up in the air. There are parties who want to see it banned completely or have access severely restricted. Conversely, the demand and growth of the data is increasing, which calls for better management and more technical tools. Presently, we are at crossroads in the history of WHOIS.

While WHOIS existed in various formats for several decades, the formal documentation used for our current Domain Name System (DNS) was released in 2004 in Request for Comments (RFC) document number 3912.[3] This standards document admits to problems with the security and data formats with the expectation or disclaimer that the data is "*intended to be accessible to everyone.*"

It is important to understand how WHOIS fits in with the overall structure of the DNS.[4] WHOIS records are not "required" for the DNS, meaning there is no technical requirement for the WHOIS record to exist, be reachable, or be accurate for a domain name to resolve. However, a variety of networking services depend on WHOIS, for example, the firewall analyzing program **fwlogwatch**[5] calls WHOIS as one of its functions, the **-W** switch.

What does and does not have a WHOIS record:

example.com—Does
frediessubdom.example.com—Does not
example.com/utils/homepage.html—Does not
ns1.example.com—Does
fred@example.com—Does not

Email addresses do not have WHOIS records, but the domain name that serves the mailbox does. So for each email address, there is one unique WHOIS record for the attached domain, no WHOIS record for specific email addresses. Twitter addresses and Facebook pages do

[3] http://tools.ietf.org/html/rfc3912
[4] http://tools.ietf.org/html/rfc1034
[5] http://linux.die.net/man/8/fwlogwatch

not have WHOIS records but twitter.com and facebook.com do. The raw IP addresses behind domain names have WHOIS records as do nameservers and the major Internet providers who sponsor the architecture of the DNS. Specific services may have internal functions called "WHOIS." For example, Internet Relay Chat[6] (IRC) has the commands WHO,[7] WHOIS,[8] and WHOWAS,[9] which provide information about different account holders; these are not usually considered part of the common WHOIS lexicon. WHOIS has multiple definition and uses, including:

WHOIS *record*
WHOIS *service*
WHOIS *server*
WHOIS *database*
WHOIS *query*
WHOIS *program*

While registration data is casually referred to as "WHOIS," the more accurate term might be Domain Name Registration Data (DNRD), but few outside the industry use this.

I.1 CONVENTIONS USED IN THIS TEXT

All material is intended to be thoroughly sourced with examples and links to additional information or original material—but be warned; the source documents may even be more obscure and difficult to understand. The examples cited are meant to be simple and straightforward. *Italicized* sections are typically literal command strings intended to be typed at a terminal or shell prompt. While the term "WHOIS" is featured in many different ways (whois, WhoIs, etc.), the convention here is to use "WHOIS" for general concepts and "whois" for specific instructions and coding. In some instances, the capitalization may be from the original context of a cited document.

People tend to regard WHOIS as a single system, but nothing can be further from the truth. The results of a WHOIS query are limited by what the specific database has, what the specific server allows access to, the used account's level of access, and the functions of the WHOIS client being used.

The way domain owners are described varies within the industry. The official term is domain *registrant* as no one really owns a domain. Domains are leased for periods of 1 year typically and must be renewed. The colloquial term "domainer" is often used to describe the population of domain registrants in a political context, whereas "registrant" is used to describe their specific relationship with the registrar. Another simpler description is domain *customer*. All term may be used in this text, but generally refer to the same type of person or entity.

The official term describing what a registrar does for a registrant is *sponsorship*. However, domain name registrars do not like this term. "Sponsorship" is what appears in

[6] http://tools.ietf.org/html/rfc2812
[7] http://tools.ietf.org/html/rfc1459#section-4.5.1
[8] http://tools.ietf.org/html/rfc1459#section-4.5.2
[9] http://tools.ietf.org/html/rfc1459#section-4.5.3

the Internet Corporation of Assigned Names and Numbers (ICANN) Registrar Accreditation Agreement (RAA) contract, but registrars are concerned that this term implies a much more active type of oversight than they are required to provide.

Some records returned by WHOIS queries can be exceedingly long. If we have shortened the records for brevity in the text, it should be indicated clearly or terminated with an ellipsis (...). Specific commands list in the flow of discussion are in **bold**. Italicized block citations are typically from documentation, memorandum, or texts. If these italicized blocks are in quotes, they are usually from a single person or attributable to single person. Single italicized lines without quotes are literal command expression to be typed on a terminal or command prompt. Example system responses are indented in a different font.

I.2 FLOW OF THIS TEXT

The goal of this book is to provide a comprehensive overview, with a certain amount of depth through its coverage of WHOIS history and WHOIS use, as well as its greater role the DNS. The full picture is seen in WHOIS programming, WHOIS server details, the complex body of WHOIS policy development, and finally the future of WHOIS. All of these topics are deeply interwoven. The history helps explain why WHOIS has been structured as it is and why some of the problems are a result of those initial decisions. Historical issues have influenced how the services were developed technically and how they are used by various consumers of the data. The WHOIS imprint on the fabric of the Internet's DNS through the servers that implement policy and technical decisions are all dependent factors in the body of WHOIS.

I.3 WHOIS FROM VERSUS WHOIS ABOUT

It is important to understand that it is possible to both query WHOIS *from* a service and *about* a service. Registrars and registries are services that host WHOIS service but also have their own WHOIS records that provide contact information for the registrar or registry company itself.

The term WHOIS can refer ambiguously to a service program, a database that stores WHOIS records as well as the WHOIS records themselves:

- Contact/owner record for an Internet resource
- Database holding Internet contact/owner records
- Query of the database holding Internet contact/owner records
- Server hosting the database Internet contact/owner records
- Service listening for queries of the database Internet contact/owner records
- Client program querying the database Internet contact/owner records
- The entire scope of all services and policy concerning Internet contact/owner records

In the early days, a single failure on the network could stop all the data from moving. The immediacy of having a technical contact in WHOIS has shifted to security and policy needs. With multiple routes available on the Internet, and more coming all the time, this

brings new threats of abuse on the network on even grander scales. The use of WHOIS may have shifted slightly, but its need has become greater.

I.4 ORIGIN OF THE TERM WHOIS

While we can trace the origin of the WHOIS protocol to specific people, events and code finding the exact origin of the term may prove a little difficult. The **who am i** command and related used is familiar to UNIX users,[10] but the use of WHOIS predates even UNIX. Different documents state that WHOIS was already in common use on systems prior to widespread UNIX deployment.[11] The use of whois as a command on Internet Relay Chat (IRC)does not appear until 1988.

Often capitalized, WHOIS is not an acronym. It literally means "who is." At one time, it was possible to type *whois* * (The asterisk "*" is a common wildcard system code, meaning it can be replaced with anything.) and retrieve all the profiles for everyone on the network. But where did it come from? "*Certainly someone coined the term,*"[12] wrote Ken Harrenstien about the origin of WHOIS. Harrenstien wrote the original WHOIS specification, and everyone I talked to said if anyone knew the origin, "*it would be Ken.*" However, at the time, preserving the specific source of the term was not likely a priority. Ken surmised that his "*suspicion is that it first started being used at the MIT AI lab, which is where I first encountered the name.*"[13]

The Artificial Intelligence (AI) Laboratory at the Massachusetts Institute of Technology was famous for the Incompatible Timesharing System (ITS). In the late 1960s, ITS was where great strides occurred in computing. One of the utilities on this system was called **who**. **who** could be used to call up a list of active usernames and the terminal names they were using, but nothing more. For those familiar with Windows NT administration, it would be similar to the **net view** DOS command, which retrieves a list of machine names connected to the network. **who** did not tell you anything about the account holder or even where the terminal was located. In 1971, another program called **finger** was paired with a database to extend the utility of **who** by providing information about the users found with **who**. **finger** would later be combined with the **name** program to create the precursor for today's WHOIS. The name/finger combination documentation in 1977 refers to the term "WHOIS" to describe the function, but the actual command switch was "/W".[14] Since this new process all ran on the ITS system, we must assume it was not new to developers at this point. Over time, WHOIS became the prevailing term for the function of seeing the record previously supplied by **finger**. To follow the logic, if **who** gave us a list active users but no further information, the follow-up question would likely be "who is" a particular user. Some RFCs assigning port number 43 refer to the service as "Who Is,"[15] but obviously the space in the command would cause problems, especially, in earlier systems, so it follows that the term would be contracted.

[10] http://linux.die.net/man/1/who
[11] http://tools.ietf.org/html/rfc742
[12] Harrenstien interview
[13] See note 12.
[14] See note 11.
[15] http://www.ietf.org/rfc/rfc1700.txt

Unlike many other early commands and future UNIX commands, "Who" is pretty straightforward, as compared to **grep**. There are some with the same sort of expected meaning like **which** (shows which version of a program is being used by virtue of the path-name), **whereis** (searches for files related to a utility), and **whatis** (describes a command). The one-letter command **w** combines features of who and finger with some additional features for more powerful searching on the local network. Even more specifically, **whodo** can retrieve a list of processes being run by which user. These commands check the system utmp[16] file (and others), which record user activity. There is also a **whom** command that is used for examining email headers.[17] However, most of these conventions appear long after WHOIS starts creeping into official Internet documents.

It would be difficult to make a direct connection with Internet WHOIS, but the first real use of the term in communication may have come from teletype machines as documented in the chapter on history. Long before the Internet sparked into being on October 29, 1968, remote signals were sent without electronics, and the recipients needed to identify the sender.

I.5 WHY WHOIS IS IMPORTANT (OR SHOULD BE) TO EVERYONE

Anyone who uses the Internet for any commerce or communication needs to understand there is an underlying record set documenting who controls websites and Internet resources. We all share and access the same Internet. How do we identify who controls a resource on this network? Specifically, within the context of a responsible party, for the purpose of addressing technical issues but also in the larger and more subtle context of ensuring a trust relationship on the shared network. Ensuring that a node on the network functions properly and is not passing traffic in a way that disrupts the network is part of that trust foundation. This becomes even more crucial when online transactions come into play. In this world, "transaction" has a few meanings, which need to be clarified. In networking, a *transaction* refers to a very literal transfer of data and has similar use in database programming. However, in the context of our trust relationship, transaction is used to refer to the exchange (sometimes unauthorized) of personal information or money. The fundamental reason for accurate and accessible WHOIS is to offer a layer of protection to users and consumers. WHOIS keeps the Internet democratic.

I.6 WHAT KIND OF USE AND CONTACT IS PERMITTED FOR WHOIS

There are concerns and accusations that WHOIS is being abused, or at least overused, but the records exist for a reason. WHOIS contact details may be used for "any lawful purpose,"[18] which would include research and questions related to online investigations. Registrars, ISPs, registrants, and users engaged in illicit activities may claim that storing or using WHOIS data is a violation of privacy or harassment, but this is merely a tactic. There are limitations on the use of WHOIS data, which includes mass marketing,[19] but this is inapplicable to data gathering in an investigation and contact in relation to the domain name. For example, contacting a domain registrant to ask if they have a valid pharmacy

[16] http://man7.org/linux/man-pages/man5/utmp.5.html
[17] https://www-01.ibm.com/support/knowledgecenter/#!/ssw_aix_61/com.ibm.aix.cmds6/whom.htm
[18] http://www.icann.org/en/resources/registrars/raa/ra-agreement-21may09-en.htm#3.3.5
[19] http://www.icann.org/en/resources/registrars/consensus-policies/wmrp

license for their domain is a completely legitimate use of WHOIS data. Illicit registrants will often accuse investigators of "spamming" them, but routine contact in connection to the use of a domain name is perfectly acceptable.

Registrars will often insert language into the headers of WHOIS records, which contain additional restrictions on the use of WHOIS. However, these conditions are frequently not supported by the registrar contracts. Specifically, the contract states: "*Registrar shall* ***not*** *impose terms and conditions on use of the data provided.*"[20]

I.7 WHERE IS THE WHOIS DATA?

In terms of domain WHOIS data, ICANN does not accept or store WHOIS data. All data is stored in individual registrar or registry databases in addition to the WHOIS escrow at Iron Mountain. The Iron Mountain escrow is not a database that can be queried, and ICANN does not have access to it. The purpose of the Iron Mountain escrow is to provide a recoverable repository of WHOIS data in case of catastrophic failure or if a registrar refuses to turn over their database upon contract termination, which has happened. There is no single WHOIS database. Because of the number of possible office locations, virtual data storage, and off-site backups, the data exists in various states and levels of availability. Some registrar WHOIS servers are even run from small home offices. WHOIS records are not a single record; rather, they are field entries in a database, and in some cases, the results displayed in a query may have come from more than one database. This is why the records will appear different depending on how the record is retrieved or where it is retrieved from. The WHOIS files produced by queries are merely the text output of a database query.

I.8 IDENTIFYING REMOTE COMMUNICATION SOURCES

WHOIS is not a unique or new situation. The problem of identifying persons, devices, or broadcasts on a network predates even the creation of the modern Internet. We can point to the Imperial Wireless Chain[21] and the common telephone system.[22] Consider examples of communication and source identification, which predate even any kind of wired or wireless transmission, namely, lighthouses. In theory, every lighthouse has a different paint pattern for daytime identification and flash lights at different intervals in the dark.[23] While lighthouses keep ships from running aground, they also provide a critical navigational tool; the external stripes, color, or checkers are not just for quaint appearances. This is called a *DAYMARK* in sailor lingo.[24] Compare these two lighthouses from Bodie Island, NC,[25] and Cape Hatteras, NC,[26] respectively. They are very close to each other in terms of location and similar in construction. The variation in pattern distinguishes them for ships in the area.

Communication is not just about transmitting information but also validating the source of that information. The role of lighthouses in civilization stretches back to ancient times. Two

[20] See note 18.
[21] http://hansard.millbanksystems.com/commons/1913/aug/08/new-marconi-agreement
[22] http://www.thefreedictionary.com/Plain+old+telephone+service
[23] http://www.us-lighthouses.com/faq.php
[24] http://pharology.eu/Daymarks.html
[25] http://www.nps.gov/caha/planyourvisit/bils.htm
[26] http://www.nps.gov/caha/learn/historyculture/movingthelighthouse.htm