1

THE HISTORY OF WHOIS

Through the development of the Internet, in general, we see the development of WHOIS and its concepts as a necessary component. While the need for a clear record set for the network seemed a fundamental technical requirement, it was not simple to construct and manage. Throughout history, questions and discussions about the meaning and use of these resource records began to emerge. It is clear that various policy issues were on the minds of the early RFC authors, which sometimes portend future conflicts.

1.1 IN THE BEGINNING

In 1982, this dry sentence launched the Internet's model of record access for the next 30 years and beyond:

> *The NICNAME/WHOIS Server is an NCP/TCP transaction based query/response server, running on the SRI-NIC machine, that provides net-wide directory service to ARPANET users.*[1]

Where the SRI-NIC machine sits or what "SRI" stands for is not explained or footnoted in the document. Anyone reading it at the time would have common knowledge of its meaning. "NIC" of course stands for Network Information Center or Controller. Understanding what is behind these acronyms opens a door to the history of the Internet. SRI stands for Stanford Research Institute. In 1982, SRI-NIC, and its related machines, was the Internet. Many readers may be more familiar with the ARPANET as a precursor to the Internet.

[1] http://tools.ietf.org/html/rfc812

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

The ARPANET was a government-funded initiative to connect networks at the Massachusetts Institute of Technology (MIT), Harvard, Xerox, the RAND Corporation, The Pentagon, and a dozen other entities. However, we see from this memo that the location and coordination of the record set for this nascent network was at Stanford. The machine referenced would hold the contact information for all the hosts and directories on the ARPANET and respond to requests for that information. So what is the real difference between NICNAME and WHOIS, as they are used synonymously starting with the title of RFC 812? In the Unix services file (/usr/etc/inet/services), different ports are assigned for different network traffic. Port 43 lists "whois" as the service name and "nicname" as the process or program.[2] This is a common snapshot of that file with the Port 43 lines highlighted, compared to the entries for FTP and Telnet that have no alternate identities:

```
ftp          21/tcp
telnet       23/tcp
smtp         25/tcp       mail #Simple Mail Transfer
whois        43/udp       nicname
whois        43/tcp       nicname
...
```

It is in this context a subtle distinction. The *whois* accepts requests through Port 43 for *nicname*. The RFC from 1982 is often marked as the beginning of WHOIS by researchers like Milton Mueller,[3] a professor at the Syracuse University School of Information Studies and one of the major figures in the WHOIS policy debate. However, here we can push the origin back several years and may be even more.

1.2 THE SANDS OF TIME

In our introduction, we made a brief reference to lighthouses and the role they have played from ancient times, not just in warning ships of the coastline but also in the self-identification of the information source. The concepts in play in computing and networking have a long lineage. We often take our advance technology for granted, not understanding that generations past worked at these ideas long before they became real in our time. Our modern communication technology is an amalgamation of human achievements from prehistory, just out of reach, until now, due to a collision of mechanics and electricity in the last century.

The idea of building a network and passing information across the network did not spring into being 50 years ago. Humans have been tackling this problem since ancient times without computers or even electric power. The need to identify sources of information that could be passed through a network became a challenge as soon as the ancient networks began. Two of the best examples come from the Roman Empire and can still be seen (and even used), namely, roads and aqueducts. The Romans were distinguished from other ancient civilizations by the permanent lines linking cities and settlements. The testament to the Roman road was not just in its construction, but more so in its regulation, maintenance, and use. Roads had to be up to a certain size and standard and separated for specific use. Like modern network technology, the Roman road consisted of *layered* construction materials each with its own function.

[2] http://www.informatica.co.cr/linux-kernel/research/1993/0218.html
[3] http://faculty.ischool.syr.edu/mueller/Home.html

1.2.1 Seals

The Roman roads were of course used for travel, commerce, and messaging. Just like the Internet today, where any host can pass traffic, messages carried on ancient roads could come from anywhere. How would a recipient identify the source of a message? Since ancient times, systems of seals or impressions have been used. Older seals were made from clay and more recently wax. The sender would have a signet ring or special cylinder with an official mark impressed in the seal, which would serve as authentication.[4]

1.2.2 From Signal Fires on the Great Wall to Telegraphy

Another great construction feat of antiquity that can still be touched is China's Great Wall. Stretching over 8000 kilometers along China's northern border, its military defensive and border control are well known, but its use as a network is not. A system of fires, cannon, drums, and flags were used to pass information rapidly, not only up and down the wall but also to and from watchtowers outside the wall. Beyond simply warning of an impending attack, variations in the signals indicated enemy troop strength and position.[5]

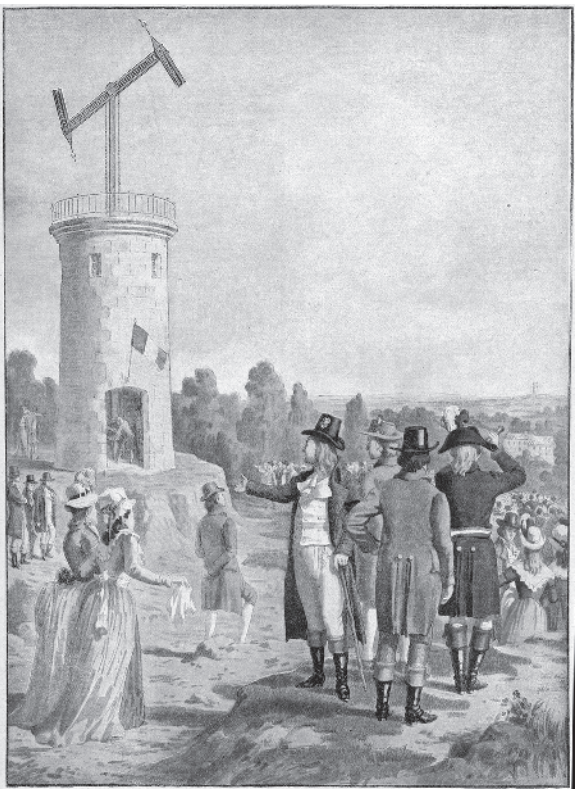

FIGURE 1.1 Chappe optical telegraph from http://farm3.static.flickr.com/2174/3666825198_a7ab2e6270_m.jpg. Mary Evans Picture Library.

[4] http://www.artofmanliness.com/2013/02/13/wax-seals-a-history-and-how-to/
[5] http://greatwall.shanghaifinance.com/greatwallmilitary.php

FIGURE 1.2 Telegraph Hill, San Francisco, California. http://www.superstock.com/stock-photos-images/1885-2819. Travel Library Limited.

FIGURE 1.3 Telegraph Street, South Boston, Massachusetts. Photo by Author.

The use of optical signaling (telegraphy or semaphore) continued centuries on for message transmission until replaced by the telegraph. The term *telegraphy* itself was used initially to refer to long-distance optical signaling.

Many modern cities have locations or streets called "Telegraph Hill," which actually used to house these stations.

The eighteenth-century inventor of modern telegraphy, Claude Chappe [1], was inspired by the writings of the ancient Greek historian and politician Polybius [2].

Polybius is credited with creating one of the earliest coding systems by converting the Greek alphabet to numeric values and representing them with different numbers of torches on the top of a tower. Instead of fire, Chappe used mechanical arms at the top of towers to signal news across France. Prior to the twentieth century, the use of *heliographs*, mirrors reflecting the sun and transmitting the Morse code, was a concept also mentioned in ancient texts [3]. So, yes, the Internet, in thought, can be traced back over 2000 years.

1.2.3 The Eye of Horus

While we have established methods used by the ancients for conveying messages over long distances, there is still the matter of message compression and encoding. However, this was not much of a problem for our ancestors either.

The ancient Egyptians were obsessed with fractions, and one of their most interesting fractional sequences is wrapped up in a critical myth [4]. The god Horus lost his eye during a battle with his evil uncle Set. The broken pieces of the eye were collected and rebuilt. The Eye of Horus or *Wedjat* is a symbol of protection and royal power, which most would recognize. Few, however, know that the broken pieces of the eye each represent individually the fractions 1/2, 1/4, 1/8, 1/16, 1/32, and 1/64—each is one half the previous one. This fraction set was only used for measuring grain, which was a sacred resource.

This sequence appears again in representing our binary values 1, 2, 4, 8, 16, 32, 64, 128, 256, and 564—each one is twice the previous one. This sequence has played a role in limiting IP ranges (which are capped in their segments at 256) and the size of certain digital values (domain names have had a 64-character limit). Why? Because these values are mapped to the literal binary switches in a computer that enable them to "do math" by recognizing whether the switches are either active or inactive. With 10 binary operators each assigned with one of the values in the sequence, it is possible to combine them to create any number. The idea of binary numbers in particular could have originated over 2000 years

FIGURE 1.4 Wedjat, reproduced courtesy of The MIT Press, from Richard J. Gillings, *Mathematics in the Time of the Pharaohs*.

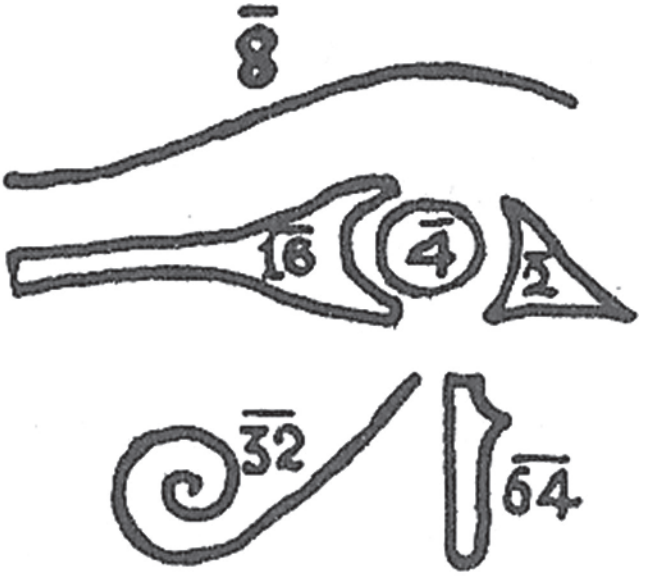

FIGURE 1.5 From mathisgoodforyou.com. http://www.mathsisgoodforyou.com/topicsPages/egyptianmaths/horusfractions.htm. Reproduced courtesy of The MIT Press, from Richard J. Gillings, *Mathematics in the Time of the Pharaohs*.

ago and was experimented with by mathematicians throughout the centuries searching for methods to compress or encode information.

1.3 1950s: ON THE WIRES AND IN THE AIR

Everyone knows there was no Internet in the 1950s. Or was there? It could be said that the Internet was *almost* there. The lines, coding, and terminals were all in place and had been since the previous century, as explained later. WHOIS, the concept anyway, was there too. There were a few pieces missing including ways to store and move large sets of data as well as connections between real client and server networks. The Internet needed a push.

1.3.1 Sputnik Changes Everything

The innovation drive that resulted in our Internet was sparked by the 1957 launch of the USSR's Sputnik satellite.[6] US President Dwight Eisenhower pressed for the immediate creation of a group advancing various technologies in the interest of national security. The Defense Advanced Research Projects Agency (DARPA) was authorized in 1958 to expand research and development beyond the existing military labs. Much of the work was focused on communications and information processing as well as on military hardware like missiles. One of the first projects released was TRANSIT, a satellite navigation

[6] http://tools.ietf.org/html/rfc2235

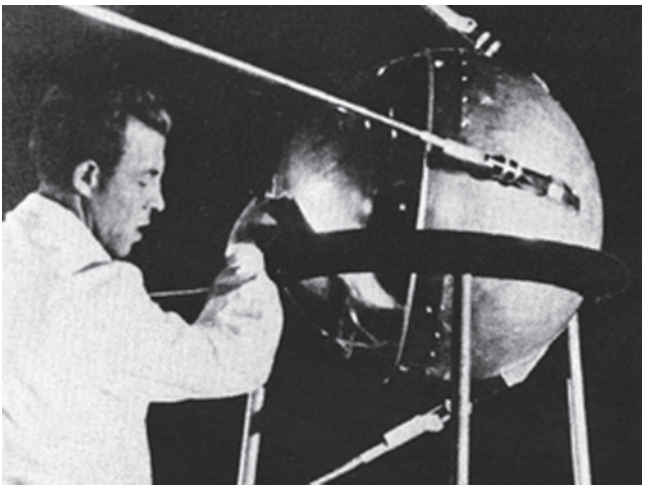

FIGURE 1.6 Sputnik, Courtesy of NASA.

system used for tracking US Navy ships and submarines.[7] This system was a precursor to our current Global Positioning System (GPS). The idea for this system came directly from the efforts to track Sputnik through *Doppler shifts*, which is in essence the change in waves between stationary and moving objects. So we see, wrapped up in the early development of the Internet, a need to identify a remote communication source. In fact, one of the things the US government was concerned about was the detection of nuclear explosions.

It was in part this need to get computers small enough to fit into satellites and submarines that required new thinking in the way computers functioned. For example, one of the early TRANSIT satellites began employing the concept of loading software into memory while orbiting the earth. Better computing and remote communication within DARPA projects of course led to the Internet, but not for another 10 years. Also, we indicate earlier that all of this occurred after *1957*. What happened before?

1.3.2 Telegraphs, Radio, Teletype, and Telephones

In *Victorian Internet*,[8] author Tom Standage explains how an electronic global nineteenth-century network spread news, delivered letters, and was even used for spam.[9] Text messages were converted to Morse code,[10] which consisted of varying electronic pulses representing letters of the alphabet and formatting codes circa 1837. Soon there was even competition in the coding from inventors like Jean-Maurice-Émile Baudot[11] whose code eventually replaced Morse. The text to code conversion was manual as was the retranslation on the receiving end. The telegraph cables ran over land and under the sea transmitting

[7] http://techdigest.jhuapl.edu/td/td1901/danchik.pdf
[8] http://www.smithsonianmag.com/history-archaeology/bookreview_jan99_a.html
[9] http://taint.org/2012/05/24/112415a.html
[10] http://www.history.com/topics/telegraph
[11] Baudot systems.

between North America and Europe in a matter of minutes.[12] The deployment of telegraph was often alongside another type of network, the railroads. The phone company SPRINT actually began as part of a railroad. The name SPRINT stands for Southern Pacific Railroad Internal Network Telecommunications. Telegrams continue to be an important part of global communications even in the Internet age.[13]

Of course, all of this wired technology got competition from the emerging *wireless* technology of the nineteenth century in the form of radio. Both telegraph and radio stations were identified by two-letter *call signs* that became longer as more stations started broadcasting. Since all stations in a telegraph network received all messages, the station codes were included to show who the transmission was intended for. This is not too different from peer-to-peer networks that pass traffic not intended for the intermediate machines. Ships and airplanes as well as ground stations have call signs.

While telegraph operators needed training in handling messages, the transmission and output were already automated, meaning it was not a far stretch to automate the translation of message and print the letters instead of a code. Teletype (TTY) took the existing mechanical typewriter model and connected its operations to the electric input. Pressing a letter key on the TTY would issue the same kind of code tapped manually by a code operator. On the receiving end, the electronic codes were mapped to the TTY keys that printed messages on paper. A 1932 *Popular Mechanics* issue contains this description:

> *[AT&T] for the first time makes available to the public generally this means of transmitting messages electrically over the wires to any other subscriber, so that whatever is typed at one end of a circuit appears practically the instant at the other end, also in typewritten form.*[14]

The first successful TTY transmission occurred in 1904 and commercial sales began in 1910.[15] Its use rapidly expanded for news transmission, law enforcement communication, and even hotel registration. By 1922, the US Navy had successfully used radioteletype (RTTY) to send printing instruction from an aircraft to a ground station.[16]

1.3.3 WRU: The First WHOIS

These devices were connected to the phone system and the combination made for amazing technology at the time. With multiple locations sending and receiving messages, the immediate questions become: where is the message coming from and who wrote it? The TTY machines had a hardcoded HERE IS key, a special code drum identifying the station.[17] Below is an excerpt from a TELETYPE Corporation A Teletype Model 33 ASR manual:

> *Here Is Answer-Back*
>
> *2.1.39 The answer-back will cycle once when the HERE IS key is depressed.*

[12] http://itelegram.com/telegram/Atlantic_Cable_150.asp
[13] http://www.theguardian.com/world/shortcuts/2013/jul/10/final-telegram-to-be-sent-india
[14] http://books.google.com/books?id=g_EDAAAAMBAJ&pg=PA577#v=onepage&q&f=true
[15] http://www.samhallas.co.uk/repository/telegraph/teletype_story.pdf
[16] http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Radioteletype.html
[17] http://www.k7tty.com/development/teletype/model-32/index.html

FIGURE 1.7 HERE IS key. Courtesy of Jessamyn West, flickr.com/photos/iamthebestartist/5559792267.

This encoding could be requested remotely by another terminal by issuing the WRU, which stands for "*WHO ARE YOU?*" This was not a question for the person operating the TTY; it was a question for the machine. Below is the WRU excerpt from the 33 ASR manual:

> *WRU Answer-Back*
>
> *2.1.31 When WRU is sent from the keyboard or tape, the WRU function box mechanism operates at both sending and receiving stations. The answer-back at the sending station is mechanically prevented from responding, while the WRU function box mechanism trips the answer-back at the receiving stations.*

The WRU command still exists in maritime communication.[18] This is the first real WHOIS, the first time an information source would respond automatically to a remote request for identification. The WRU would also be sent along with the end of a TTY message to confirm with the recipient that the transmission was unbroken. This command was also called the ENQ for "ENQuiry."[19] This is an excerpt from a TTY manual showing the location of the ENQ function:

Part of the drum encoding sequence included codes for Acknowledge (ACK), Carriage Return (CR), and Line Feed (LF), which are all part of the online WHOIS transaction that shows up later. Coding the answer drum involved a screwdriver and needle-nose pliers. Encodings were created by removing tiny tines in a sequence to indicate a specific American Standard Code for Information Interchange (ASCII) code.

[18] http://www.polaris-as.dk/wp-content/uploads/downloads/user_manuals/skanti/SKANTI_TRP_1000_Series_Operators_Manual.pdf

[19] http://www.baudot.net/docs/smith--teletype-codes.pdf

FIGURE 1.8 HERE IS key. Courtesy David Gesswein pdp8online.com.

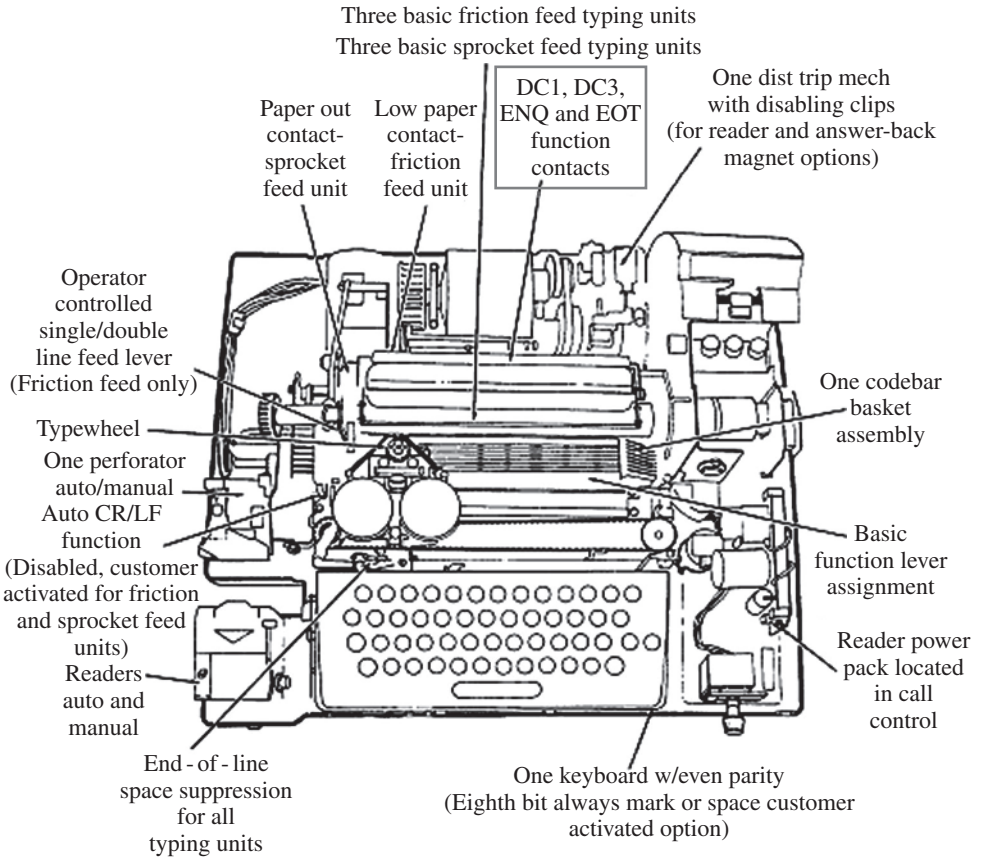

FIGURE 1.9 Teletype Overview - 33 ASR Teletype Manual.

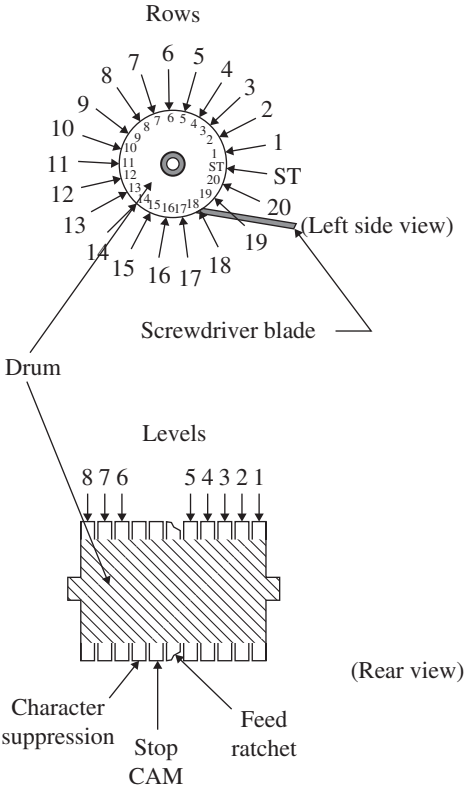

FIGURE 1.10 Teletype answer-back drum—*33 ASR Teletype Manual*. Courtesy of AT&T/Teletype Corporation.

Abbreviation	Key to abbreviation
ACK	Acknowledge
CR	Carriage Return
LF	Line Feed
RO	Rub Out
SP	Space
SUP	Character Suppression

FIGURE 1.11 Key to abbreviations—*33 ASR Teletype Manual*. Courtesy of AT&T/Teletype Corporation.

As telephones were becoming available throughout the United States and elsewhere, one might wonder, why bother? Why not just use the telephone? As the then Governor Ronald Reagan stated later in 1972 in one of the first test electronic email messages,[20] "*All this damned typing… Wouldn't you rather pick up the phone and call?*" Telephones of course were becoming the de facto remote communication standard, which is great as long as you are not deaf. The difficulty of using a telephone is not apparent to those with full hearing, but this actually plays directly into the creation of the Internet and WHOIS with

[20] http://www.arpanetdialogues.net/vol-i/

> 2.10 The number of rows available for actual station identification is less than shown above, because each coded message should begin and end with CARRIAGE RETURN and LINE FEED (this may be altered in specific applications). This assures that the transmitted message will appear at the beginning of a line of the receiving teletypewriter set and eliminates overprinting.

FIGURE 1.12 Section 2.10—*33 ASR Teletype Manual*. Courtesy of AT&T/Teletype Corporation.

USASCII code chart

b_7 b_6 b_5 → / Bits / b_4 b_3 b_2 b_1 / Column → / Row ↓	0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
	0	1	2	3	4	5	6	7
0 0 0 0	0	NUL	DLE	SP	0	@	P	`	p
0 0 0 1	1	SOH	DC1	!	1	A	Q	a	q
0 0 1 0	2	STX	DC2	"	2	B	R	b	r
0 0 1 1	3	ETX	DC3	#	3	C	S	c	s
0 1 0 0	4	EOT	DC4	$	4	D	T	d	t
0 1 0 1	5	ENQ	NAK	%	5	E	U	e	u
0 1 1 0	6	ACK	SYN	8	6	F	V	f	v
0 1 1 1	7	BEL	ETB	´	7	G	W	g	w
1 0 0 0	8	BS	CAN	(	8	H	X	h	x
1 0 0 1	9	HT	EM	)	9	I	Y	i	y
1 0 1 0	10	LF	SUB	*	:	J	Z	j	z
1 0 1 1	11	VT	ESC	+	;	K	[	k	{
1 1 0 0	12	FF	FS	,	<	L	\	l	|
1 1 0 1	13	CR	GS	–	=	M	]	m	}
1 1 1 0	14	SO	RS	.	>	N	^	n	~
1 1 1 1	15	SI	US	/	?	O	—	o	DEL

FIGURE 1.13 USASCII code chart—*33 ASR Teletype Manual*. Courtesy of AT&T/Teletype Corporation.

Ken Harrenstien and Deafnet.[21] Harrenstien and Vinton Cerf are both hearing impaired and worked on various projects to promote text communication. We partially owe thanks for the Internet to people extending services to users who cannot hear.

Manufacturers of TTY introduced many innovations including data storage. Messages could be stored on punched tape and fed back into the device to be sent again. Eventually, TTY became the primary input devices for computers and time-sharing terminals. The WRU or ENQ remained part of the encoding set that moved into the new systems as part of the ASCII table in 1960.

The TTY continued to exist within the Internet even after the device itself was obsoleted by graphic terminals. Terminal software, Telnet, is actually designed to emulate a TTY.

[21] http://www.sri.com/work/timeline-innovation/timeline.php?timeline=health#&innovation=deafnet

To Send	(Explanation)	Also Known as	Push Either	Or	Code Sent is (in hex)	Upon receipt Displayed as
NUL	NULL	^@	[2]		00	
SOH	Start of Heading	[3]^A	LS	[4]CASE A	01	
STX	Start of Text	^B	ATAN	CASE B	02	
ETX	End of Text	^C	LOG	CASE C	03	
EOT	End of Transmission	^D	REFL	CASE D	04	
ENQ	**Enquiry**	**^E**	**[4]**	**CASE E**	**05**	**[5] <ENQ>**
ACK	Acknowledge	^F	UP	CASE F	06	<ACK>
BEL	Bell	^G	DOWN	CASE G	07	<BELL>

FIGURE 1.14 TELNET Character Set.

Telnet actually stands for *Teletype Over Network Protocol*. The earliest RFC on Telnet makes the connection clear:

> *The TELNET protocol is based upon the notion of a virtual teletype, employing a 7-bit ASCII character set. The primary function of a User TELNET, then, is to provide the means by which its users can "hit" all the keys on that virtual teletype.*[22]

The **ENQ** command from ASCII, originally from the TTY WRU response, now becomes part of the TELNET character set[23] and can be invoked with **CTRL-E**, indicated by **^E**:

This function on Telnet is virtually identical to the previous TTY manual operation. Telnet uses ENQ to issue a "Who Are You" request to a remote station identification.[24]

[22] http://www.ietf.org/rfc/rfc0206
[23] See note 22.
[24] http://www.cs.tut.fi/~jkorpela/chars/c0.html

1.4 1960s: SPARKING THE INTERNET TO LIFE

Many of the Internet innovations occurred in the 1960s, especially the concept of *packet switching*, which breaks data into standard sizes for transmission [5]. This method permits large files to move and multiple users to access the network without consuming all the resources for one use. On early networks, the traffic was moved by *circuit switching*, which meant that only one use was permitted at a time.[25] The term "On-Line" also appears for the first time in the paper *ON-LINE MAN-COMPUTER COMMUNICATION* by J.C.R. Licklider and Welden E. Clark of Bolt, Beranek and Newman (BBN) Inc. The BBN continued to play a major role in Internet development, eventually becoming a part of Raytheon.[26] At the same time, the DARPA took a keen interest in networking technologies. This is an important point: many people and organizations were already developing networks and centralizing computing resources. The government did not create the Internet *ex nihilo*; rather, it coordinated existing concepts to expand their reach. In this new space, being able to identify which network was which became more and more important.

1.4.1 SRI, SAIL, and ITS

The proto-WHOIS RFC 742 stated the NAME/FINGER program:

> *Currently only the SAIL (SU-AI), SRI (SRI-(KA/KL)), and ITS (MIT-(AI/ML/MC/DMS)) sites support this protocol.*[27]

This is a list of three labs and seven host machines. SRI, as we saw, stands for Stanford Research Institute and SAIL the Stanford University Artificial Intelligence Laboratory.[28] The MIT system ITS in particular stands for "Incompatible Timesharing System" and was named to differentiate itself from the Compatible Timesharing System. "Timesharing" is something we all take for granted now in our systems and devices. Imagine the early computers that could only respond to one user, command, or process at a time. The idea that multiple processes or users could share a computer resource was a revolutionary concept. ITS was an operating system written in the assembly language MIDAS[29] on a DEC PDP-10 to support the development of artificial intelligence. Along with ITS at the MIT labs came a new concept: letting the public access the network. In order to address potential problems a policy was drafted: the *MIT AI Lab Tourist Policy*.[30] A *tourist* was a nonlaboratory person allowed in during off-hours under certain conditions, but more or less permitted to explore the system. Disruptive, abusive, commercial, and political activities were not permitted. Most importantly, tourists had to register as tourist through the ITS :INQUIRE program.[31]

[25] http://www.packet.cc/files/ev-packet-sw.html
[26] http://www.bbn.com/
[27] http://tools.ietf.org/html/rfc742
[28] http://ai.stanford.edu/
[29] ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-238.pdf
[30] http://www.art.net/Studios/Hackers/Hopkins/Don/text/tourist-policy.html
[31] See note 30.

1.4.2 Doug Engelbart: The Father of Office Automation

On October 29, 1968, the ARPANET was born with the connecting of SRI and the University of California Los Angeles (UCLA). Being that there were only two hosts on the network, there was no need for a WHOIS record set just yet. Less than 2 months later, Doug Engelbart, the director of Augmentation Research Center (ARC) at SRI,[32] used this connection to run the "Mother of All Demos."[33] The key feature of the demonstration is that Engelbart conducted it remotely. In this single presentation, Engelbart did not predict, but showed many of the tools we are used to now including the word processing and hypertext. However, Engelbart did not simply focus on developing new tools; the tools were products of his visionary approach to collaborative work. He employed *bootstrapping* to promote existing resources and talent. With this explosion of innovation, he also saw the need to properly manage it with the right people. Anyone who has never seen the demonstration should be encouraged to watch it here: http://dougengelbart.org/events/1968-demo-highlights.html.

1.5 1970s: OK, NOW THAT WE HAVE AN INTERNET, HOW DO WE KEEP TRACK OF EVERYONE?

A link between two sites is not a network, but it is a private conversation. The ARPANET quickly went from two sites to three, and then all of the major sites within the United States that had their own networks began plugging in. The attendees at this growing gathering needed name tags.

1.5.1 Elizabeth "Jake" Feinler

We often hear about the "fathers" of the Internet, but there are mothers too, several in fact as we will see. Elizabeth "Jake" Feinler had been working at SRI since the 1960s but was asked by Doug Engelbart to draft the first ARPANET Resource Handbook[34] for SRI's ARC. Engelbart, who unfortunately passed away in 2013, may not be a household name but his inventions are, like the computer mouse and the term "On-Line." Feinler had an interesting task: to document this new concept that no one had heard of and few people understood. Having previously authored handbooks on complex chemicals, Feinler was certainly up to the job. Inventing the network was fantastic, but we will see over and over again the need for clear documentation to accompany this complex structure. The cited handbook ends at over 1000 pages. As Feinler shifted into evermore important roles, ensuring that specific files and documents were updated and shared properly became the focus of what made the network function.

1.5.2 The ARPANET Directory as Proto-WHOIS

WHOIS has often been compared to a phone book directory, which is exactly what it started out as a paper phone book called the ARPANET Directory. This hardbound directory was regularly updated and shipped to various project members and contacts. Publishing

[32] http://www.dougengelbart.org/about/cv.html
[33] http://www.dougengelbart.org/firsts/dougs-1968-demo.html
[34] http://www.computerhistory.org/collections/catalog/102714229

The ARPANET DIRECTORY – A directory of users and hosts on the ARPANET. It gives the names, network and U.S. Mail addresses, phone number, and host affiliation of ARPANET users, as well as summary tables of host information.

FIGURE 1.15 The ARPANET Directory. Courtesy of Computer History Museum.

and distributing this ever-larger book was a burden, so as this new online system became available, the distribution of the ARPANET Directory went on it. The last hard copy of the directory was published in March of 1982,[35] after which the directory became completely digital. The following excerpt is from the April 1978 Defense Communications Agency ARPANET INFORMATION BROCHURE [6]:

This clarifies that the directory was for contacts and distinct from the HOSTS table. Morphing at different stages, the ARPANET Directory became WHOIS. While many may think WHOIS was created as a digital record to identify the owners of networked hosts, it was actually a record of persons, office locations, and phone numbers (and later emails) that merged with the new concept of virtual space.

When asked when WHOIS and the Internet became controversial, Feinler easily pointed to the influence of commercial interests. On the early network, there were no advertisements or commercial activity of any kind. Today, Feinler has little interest in the Internet Corporation for Assigned Names and Numbers (ICANN) and the commercialization of the network. She seemed concerned that the deployment of new generic top-level domains (gTLDs) might have negative consequences. The rush for commercialization was not on the minds of early pioneers like Feinler. She and her colleagues prided themselves on their open collaboration and general ethics. Like others interviewed for this book, she expressed a spirit of the time that guided the creation of the Internet, which now seems lost.

1.5.3 The Site Status List

SRI did not immediately provide routing and owner information for a network that was yet to exist. DARPA had contracted Cambridge, Massachusetts, company BBN, now part of Raytheon, to manage the ARPANET Interface Message Processors (IMPs). IMPs were the original Internet routers, packet-switching nodes that interfaced between internal networks and the external network. BBN assigned the IMP port number and collected the information about the entity connecting to the IMP. BBN employee Ellen Westheimer published a descriptive file of the IMPs and their owners called it the *Site Status List*. This was eventually renamed as the *Network Host Status List* [7].

In 1970, the Network Working Group (NWG), led by Steve Crocker, used the Network Control Protocol (NCP) to standardize the ARPANET network interface, which allowed more remote networks to join the new major network. In October 1971, the existing 15 ARPANET networks connected to each other through the NCP at a virtual meeting[36] at the MIT. But what

[35] Feinler interview.

[36] http://www.livinginternet.com/i/ii_ncp.htm

FIGURE 1.16 Interface Message Processor. Courtesy of Computer History Museum.

were the networks and how were they identified? The following is a list of the networks and their assumed hostnames[37]:

Bolt, Beranek and Newman	bbn
Carnegie Mellon University	cmu
Case Western Reserve University	case
Harvard University	harv
Lincoln Laboratory	ll
Massachusetts Institute of Technology	mit
NASA AMES	ames
RAND Corporation	rand
Stanford Research Institute	sri-nic
Stanford University	su-ai
System Development Corporation	sdc
University of California at Los Angeles	ucla
University of California at Santa Barbara	ucsb
University of Illinois at Urbana	illinois
University of Utah	utah

The BBN knew who everyone was because they issued the IMPs, but that was getting more complex as the next conference would connect 40 networks, and so on. The collection and maintenance of this list shifted to SRI-NIC and gradually merged with the paper ARPANET Directory.

[37] http://www.columbia.edu/~rh120/ch106.x08

1.5.4 Distribution of the HOSTS Table

In Andrew Blum's excellent book *TUBES* [8], the author cited Senator Ted Stevens in 2006 describing the Internet as a "series of tubes" as the source of his title. Blum showed us the physical Internet, the wires, and data centers in all of their strangeness. The text showed us that the Internet is real in the sense that it exists in fungible format. However, this is just the medium. The tubes, as they are, only exist to pass data and only have meaning because of the files that explain what each tube should do. The Internet is a series of tubes and a series of flat files that tell where everything is.

One of Feinler's important recruits was a hostmaster. There are an untold number of hostmasters on the Internet today, but for a time Mary K. Stahl was *the* hostmaster for the Internet. Stahl sometimes had the tedious job of manually proofreading and distributing the ARPANET Hosts Table (HOSTS.TXT). This, of course, was the list of hostnames and associated IP addresses on the network that were passed around directly and later available on an FTP site. Today, this file is more to akin the root zone files and the related Domain Name System (DNS) files they point to. Along with the host table, Stahl also distributed what was called the Liaison File (LIAISON.TXT), which had the accompanying contact information for the hosts.[38] Liaisons had a number of duties and responsibilities concerning access granting and information collection.[39] For Stahl, this specifically technical task was a detour from her artistic passions, but when domains later replaced hosts, her job was eliminated, which was a bit of a blessing because working on the ARPANET burned people out. She now works on her painting,[40] having been out of the Internet industry for decades. Regardless, Stahl left her eternal fingerprint on WHOIS in the assignment of Port 43 to WHOIS in the RFC 1060,[41] which is called "MARY" because she had the oversight. The exact line is as follows:

```
43 NICNAME Who Is [55,MARY]
```

She also had the authority of the HOSTNAME server that responded at Port 101:

```
101  HOSTNAME  NIC Host Name Server [54,MARY]
```

1.5.5 Finger

The Finger protocol was developed by Les Earnest in 1971[42] to solve a specific problem while at SAIL. The DEC-10 system Earnest worked on with his colleagues had a utility called **who**, which would show the account names and terminal numbers of people currently logged into the system but none of the offline accounts. Earnest often observed people running their fingers down the output of the **who** command, so he created a database along with the program itself to provide information about all users regardless of their status. Upon request, Earnest added an "out of the office" feature for the database called "Plan," which could be updated to inform other users about vacations and other events. Issues of privacy and security did not yet come into place as Earnest, like many of the early pioneers, cited as the "*comradery of those gentler times*."[43] Finger itself has been the focus of some

[38] http://www.dtic.mil/dtic/tr/fulltext/u2/a482154.pdf
[39] http://www.rfc-editor.org/rfc/museum/ddn-news/ddn-news.n5.1
[40] http://www.marystahl.com/
[41] http://www.ietf.org/rfc/rfc1060.txt
[42] http://blog.djmnet.org/2008/08/05/origin-of-finger/
[43] http://www.djmnet.org/lore/finger-origin.txt

problems as most networks shut off finger now, especially for external users. In 1988, the Morris worm exploited the Finger daemon program, **fingerd**, and its access to the library system function **gets**, which was used to create a buffer overrun and execute the malicious code to keep spreading. The flaw has long since been plugged in **fingerd**, but the incident left a lasting impression on the Internet psyche showing how the network could be crushed by native software that was linked to personal contact information.

1.5.6 Sockets

Around the same time, Jon Postel began to standardize the most basic architecture of the ARPANET, assigning different functions to different *sockets* in RFC 349, *Proposed Standard Socket Numbers.*[44] There were only five specific assignments in this memo, but space was made for where NICNAME/WHOIS would eventually sit. Even more importantly, Postel insisted there be a central management of the system, proposing him to be the "czar" who hands out assignments to others on the network.[45] This power actually remained with Postel through the domain era until his death in 1998.[46] Postel gave out pieces of the Internet including top-level domain (TLD) assignments and even took control back briefly in 1998 from Network Solutions, the so-called DNS Root Authority incident.[47] This incident led to a larger role of Internet management by the US government.

In 1973, L. Peter Deutsch authored RFC 606, Host Names On-line.[48] Deutsch later developed the innovative Portable Document Format (PDF) software, but in 1973 he was part of the NWG. From RFC 606 came one of the best quotes behind the rationale for creating such centralized information services:

> *"Now that we finally have an official list of host names, it seems about time to put an end to the absurd situation where each site on the network must maintain a different, generally out-of-date, host list for the use of its own operating system or user programs."*

The frustration is clear in his tone. Having current and easily retrievable information about other nodes is critical for an interconnected network. He specifically proposed that the NIC (SRI) manage an online file listing names and host addresses. Deutsch insisted that the file be accessible to anyone. The group almost immediately responded in RFC 608:

> *We at the NIC agree with Peter Deutsch's suggestion (in RFC# 606 / NIC# 21246) that the NIC maintain an online ASCII text file of Host names,*[49]

The management went to Jake Feinler at SRI-NIC. In addition to hostnames and IP addresses, the file also included the Technical Liaison, which would be the contact for the hostname. It is from this moment on that the active search for a solution to tackling hosts and their owners began in earnest, but we were still 4 years away from the first WHOIS

[44] http://tools.ietf.org/html/rfc349
[45] See note 44.
[46] http://news.bbc.co.uk/2/hi/science/nature/196487.stm
[47] http://www.computerworld.co.nz/article/517378/internet_veteran_dns_test_causes_storm/
[48] http://tools.ietf.org/html/rfc606
[49] http://tools.ietf.org/html/rfc608

RFC. RFC 620 instructed the participating networks to begin monitoring the HOST table.[50] RFC 623 insisted that a different protocol from FTP be used to distribute the table and the file be stored in redundancy,[51] but RFC 625 pushed back against the notion of not using FTP.[52] The official deployment of HOSTS.TXT was announced in RFC 627.[53]

The need for tracking owners of hostnames was becoming apparent even with this comparatively small number of players on the Internet. Abuse had already started and was so concerning that Jon Postel devoted an RFC to the problem. *On the Junk Mail Problem* detailed the issue and possible solutions:

> "*It would be useful for a Host to be able to decline messages from sources it believes are misbehaving or are simply annoying.*"

Postel proposed building the ability to refuse messages based on reputation right into the network interface (IMP) by "*measuring, per source, the number of undesired messages per unit time*."[54] Rapidly, being able to identify hosts and owners was about more than just management; it was about misuse of a host.

1.5.7 Into the VOID with NLS IDENTFILE

Before gradually and formally becoming WHOIS, the data was stored in Engelbart's oN-Line System[55] (NLS/AUGMENT) known as IDENTSYS. Each NIC user had an entry in the IDENT directory that contained their email address among other information.[56] These IDENTs had multiple purposes including authoring *ARC Journal* articles.[57] The protocol for managing IDENTs was documented in the "The Executive Package" file called EXEC.TXT at the SRI-NIC. The IDENT system was written in the NLS specific language, L10.[58] "*That didn't scale very well and was a beast to maintain*,"[59] related Ken Harrenstien in an email. Harrenstien's description was documented in the *ARC Journal* on November 13, 1972.[60] IDENT was noted in the journal as being a very important database but one that had no real room for growth and lacked controls for access and deletion of obsolete data.

Harrenstien has driven several updates of WHOIS over the years to handle the growing dataset. The first was called VOID DBMS. "*VOID stood for 'Vorxify ID File' which is a meaningless term I made up.*"[61] VOID was developed in the assembly language MIDAS,[62] which was also used to develop the ITS system.

[50] http://tools.ietf.org/html/rfc620
[51] http://tools.ietf.org/html/rfc623
[52] http://tools.ietf.org/html/rfc625
[53] http://tools.ietf.org/html/rfc627
[54] http://tools.ietf.org/html/rfc706
[55] http://archive.computerhistory.org/resources/access/text/2013/05/102724043-05-01-acc.pdf
[56] http://sloan.stanford.edu/mousesite/Archive/Post68/augment-33076.htm
[57] http://tools.ietf.org/html/rfc543
[58] http://www.rfc-editor.org/rfc/rfc694.txt
[59] Harrenstien interview.
[60] *ARC Journal*, 12731 http://archive.computerhistory.org/resources/access/text/2013/05/102724029-05-01-acc.pdf
[61] See note 59.
[62] ftp://ftp.columbia.edu/kermit/dec20/assembler-guide.txt

Some of the data in VOID was entered to keep track of terminal access controller (TAC) data. TACs allowed registered remote users with terminals and modem couplers to dial into the host computers at the ARPANET. Once a user was registered in the database, they were issued a TAC card that provided credentials for logging in through a dial-up terminal. Previously, the phone numbers for remote access became a security risk. With the registration and TAC cards, only authorized users would be able to log on once dialled up.[63] Harrenstien eventually converted the MIDAS VOID program to C code using the TOPS-20 C compiler.

During this same period, Jon Postel wrote about the "Internet" in RFC 675, and some of the first electronic test chats occurred between people not directly involved in the network development, including the then Governor Ronald Reagan calling the technology "Neat stuff."[64] The ARPANET invited a series of people from different disciplines to engage in debates from remote locations using the new technology.

1.5.8 NAME/FINGER RFC 742 (1977)

While not called WHOIS officially yet, RFC 742 NAME/FINGER[65] marked the actual start of what we used for Internet WHOIS in 1977. The main author of 742 is Ken Harrenstien, but he noted in the background of the document that it would not have been possible without Brian Harvey, Les Earnest, and Earl Killian. Harvey in particular provided the spark of the idea that became WHOIS. "*We were just doing what made sense, what we thought had to be done*,"[66] Harvey related in our interview. As with many of the early Internet pioneers, he was remarkably humble about his contribution. It was Harvey's idea to provide a "simple" interface between the existing Name and Finger programs at different network sites.[67] The Finger program had existed since 1971 and provided information about users on a network, while Name identified remote hosts. Combining them actually created the first distributed WHOIS model. The term itself was referenced as a switch within the RFC:

> *the syntax for some servers can be slightly more elaborate. For example, if /W (called the Whois switch) also appears on the line given to an ITS server, much fuller descriptions are returned.*[68]

Harvey was in charge of the electronic mail services at the Artificial Intelligence Laboratory at MIT[69] when he worked with Harrenstien on 742. Managing the mail services between growing numbers of user accounts made the need for tying a person, account, and location important. "*The crux of early WHOIS was about finding someone physically*,"[70] Harvey noted as he recounted a specific story of a student who unleashed a process that

[63] http://www.oac.cdlib.org/findaid/ark:/13030/c8jw8fmx/entire_text/
[64] See note 20.
[65] See note 27.
[66] Harvey interview.
[67] See note 27.
[68] See note 27.
[69] http://www.csail.mit.edu/
[70] See note 66.

would delete every file on the server's disk. "*He didn't think it would actually work! It was curiosity and not malice.*"[71] However, this clearly fits in with our ongoing themes of access, responsibility, and identification. At the time BBN knew who all the hosts were on the network,[72] they had to as a simple technical matter in order for the network to function. This rapidly became a challenge as the methods and sources of network connections changed. As Harvey stated, "*There are merits to equal knowledge of all hosts*"—meaning that everyone else on the network knows who controls all the other nodes. Dr. Harvey continued his contributions to technology as a lecturer at the University of California at Berkeley[73] and the development of SNAP!,[74] a visual drag-and-drop programming language. SNAP! was intended to be a teaching tool for children who can use graphic blocks of programming language to make things happen.

The various pieces are here, but it is not quite WHOIS yet. One of the first distinctions occurred at the beginning of the use instructions:

> *To use via the network: ICP to socket 117*[75]

Instead of our familiar Port 43, we have socket 117. The immediate explanation is that since WHOIS does not yet exist as a named protocol, it cannot be assigned to a specific port. When Jon Postel created the socket list in 1972, the range of sockets from 64 to 127 was generically designated for "host-specific functions."[76] The ICP stands for Initial Connection Protocol,[77] a very basic and early network protocol. This connection should be followed by the specific request and the ubiquitous <CRLF> on the command line. At the time of writing, sending a blank command would return all of the available records, something incomprehensible in today's WHOIS. The list returned would specifically be "*the full names of each user and the physical locations of their terminals.*"[78]

The makeup is also discussed in IEN 103[79] where the experiment of the name server is detailed. The first key detail is that the server will allow access to the data in the official Host Table rather than manually passing the host table around.[80] However, a more complex model emerged:

> *Work is in progress to investigate the feasibility of abstracting host related information from the NIC database management system via direct system calls.*

Well beyond the simple task of serving up the Host Table, this is now about querying discrete information from more complex data structures through a specific protocol.

[71] See note 66.
[72] http://www.sri.com/newsroom/press-releases/computer-history-museum-sri-international-and-bbn-celebrate-40th-anniversary
[73] http://www.eecs.berkeley.edu/Faculty/Homepages/harvey.html
[74] http://byob.berkeley.edu/
[75] See note 27.
[76] See note 44.
[77] http://tools.ietf.org/html/rfc80
[78] See note 27.
[79] http://www.postel.org/ien/txt/ien103.txt
[80] See note 79.

1.5.9 Other Early Models

The need to identify and find contacts for remote resources was universal as the network grew. While one model emerged at the main one, surely other developers were thinking about the issue. RFC 724 stated that "*there are other systems with similar programs that could easily be made servers.*"[81] It turned out that there were a number of concepts emerging to address identification of remote resources across the network. But what were they?

- IPHOST on TOPS-20. As various systems changed with constant improvements, utilities were created to provide information about machines and users. The TOPS-20 operating system was used on the DEC machines running at ARPANET. Many utilities had been developed for TOPS-20 including the NICNAME program,[82] which served WHOIS. Another utility in TOPS-20 was IPHOST, which gave information about ARPANET hosts. If the program had to be invoked on the command line with *IPHOST*, then this prompt would be IPHOST> awaiting commands. The ability to run multiple programs in separate sessions, threads, or windows was further down the road. In this environment, the programs had to be started, used, and exited before other programs could be invoked. The command within IPHOST was *NAME.* Entering *NAME?* would return a list of known ARPANET host. The command *NAME SRI-NIC* would return the IP address of the SRI-NIC server. To convert the other way, the command was *NUMBER* followed by an IP address that would return the hostname.[83]
- Online DIRectory SYStem (DIRSYS). This incremental search system was modeled after the pages in telephone directory. As a user typed a name in an interface like Emacs (Unix extensible text editor), the full pages available, with entries, would appear and become more specific as the user typed more letters. DIRSYS was specifically developed to help locate people at MIT. One interesting aspect was the ability of the users to request updates to their own records through the system, which is how WHOIS is updated now, by the domain registrants. Graduate student Kimberle Koile studied DIRSYS extensively in 1983 with the support of ARPA.[84] DIRSYS was developed with the CLU language (an ALGOG[85] based "CLUsting" code) and deployed two systems: a DEC-SYSTEM 20[86] and a VAX 11/750.[87] Within Koile's technical aspects, policy issues quickly came into play. Who should be in the directory? How will the directory be maintained? Who should access the directory? Specifically, Koile devoted a lengthy section to how access would be permitted for the outside world. The recommendation was to only allow nonincremental search and limit the number of queries permitted.
- Xerox Clearinghouse and Grapevine developed at the Palo Alto Research Center (PARC) [9] was a roaming profile system considered "ahead of its time" and allowed searches not only by name but also by closest printer.

[81] See note 27.
[82] See note 1.
[83] http://bitsavers.informatik.uni-stuttgart.de/pdf/dec/pdp10/TOPS20/arpanet/5221bm.mem.txt
[84] http://bitsavers.trailing-edge.com/pdf/mit/lcs/tr/MIT-LCS-TR-313.pdf
[85] http://www.softwarepreservation.org/projects/ALGOL/
[86] http://pdp10.nocrew.org/docs/ad-h391a-t1.pdf
[87] http://www.old-computers.com/history/detail.asp?n=20&t=3

- Postes, Télégraphes et Téléphones (PTT) Directory System allowed subscribers with terminals in France and public users at post offices to access online white pages. The system allowed reverse searches by telephone number.
- New England Bell Directory System specifically contained numbers in the 617 area code of the United States and used a Dvorak keyboard.
- Bell Labs SOUNDEX was a pattern matching search system triggered by sound.
- Computer Science Network (CSNET) WHOIS allowed online lookups of user email addresses.[88]
- Performance Systems International (PSI) White Pages Pilot Project[89] was an X.500 directory system allowing WHOIS lookups through telnetting to FRED.[90] It was also related to the NYSERNET X.500 Database project.[91]
- Knowbot Information Service (KIS) provided a single point for querying whois, finger, X.500, and other services.[92]

1.6 1980s: WHOIS GETS ITS OWN RFC

In 1982, Ken Harrenstien and Vic White released RFC 812 entitled NICNAME/WHOIS,[93] the first time WHOIS was used in the title of an RFC and the official call to make WHOIS a specific named service. The purpose of this RFC is to describe the service:

> *The NICNAME/WHOIS Server is an NCP/TCP transaction based query/response server, running on the SRI-NIC machine, that provides netwide directory service to ARPANET users.*[94]

This sentence highlights the basic function and intent, which is the same as what we are used to now. WHOIS continued to be a Transmission Control Protocol (TCP) transaction service intended for access across the whole network or Internet. This statement also makes it clear that this is the way to access the formerly paper-only ARPANET Directory:

> *This server, together with the corresponding Identification Data Base provides online directory look-up equivalent to the ARPANET Directory.*[95]

The next section details how users can access it (through programs or clients) and what they can find in it in terms of data:

> *The server is accessible across the ARPANET from user programs running on local hosts, and it delivers the full name, U.S. address, telephone number, and network mailbox for ARPANET users.*

[88] http://www1.chapman.edu/gopher-data/archives/Internet%20Information/cerfnet-users-guide-07-91.txt
[89] http://www.ietf.org/rfc/rfc1803.txt
[90] http://www.lights.ca/hytelnet/dir/dir009.html
[91] http://www.usucan.org/docs/affiliates/New%20York.pdf
[92] http://encyclopedia2.thefreedictionary.com/Knowbot+Information+Service
[93] See note 1.
[94] See note 1.
[95] See note 1.

As part of the introduction, it is made clear that everyone should be able to access WHOIS:

> *DCA strongly encourages network hosts to provide their users with access to this network service.*[96]

The question of who should be in the database is immediately following the introduction:

> *DCA requests that each individual with a directory on an ARPANET host, who is capable of passing traffic across the ARPANET, be registered in the NIC Identification Data Base.*[97]

This makes a clear distinction between hostmasters and simple users. As *hosts* became *domains* a few years later, the concept and intent remained the same. Having a host that can pass traffic on the network was a responsibility that required identification.

The 1982 RFC entitled *WHO TALKS TCP?* began to demonstrate the problems of a growing network. The RFC 834 by David Smallberg was a 13-page list of hostnames and IP addresses.[98] This was a dump of the NIC hostname table showing which ones accepted TCP connections and which nodes were dead. It may be difficult for people today to understand an Internet without domains. The following is an excerpt from the hostname table in 1982[99]:

```
coins-tas          10.0.0.36
src-ccp            10.0.0.39
utah-cs            10.0.0.4
office-1           10.0.0.43
mit-xx             10.0.0.44
collins-pr         10.0.0.46
wpafb              10.0.0.47
afwl               10.0.0.48
bbnb               10.0.0.49
bbnf               10.0.0.5
st-nic             10.0.0.51
ada-vax            10.0.0.52
afsc-ad            10.0.0.53
```

Manually reviewing and routing the growing table was clearly impractical. The memo is merely a list of servers with no contact data; if the contact data were included, this file would be enormous.

1.6.1 The DNS

This was all about the change as Paul Mockapetris drafted the RFC 882, "Domain Names: Concepts and Facilities," in 1983. The document starts off by indicating that the current scheme of mapping between HOSTNAMES and ARPA Internet address through the

[96] See note 1.
[97] See note 1.
[98] http://tools.ietf.org/html/rfc834
[99] See note 98.

ARPANIC HOSTS file on SRI-NIC was rapidly becoming unmanageable. One of the major concerns was mail delivery. Centralizing mail delivery on the expanding network would be impossible, but conversely the variety of emerging mail methods and routes was unwieldy. The fix was to create a consistent name space model that would be used for referrals in distributed authorities. The new model would have three main components: the Domain Name Space, Name Servers, and Resolvers. This RFC also introduces a database model for the DNS, the use Resource Records, and the "Dot" system we are all familiar with for domain names. As far as WHOIS goes, we get some very important concepts as well:

> *There must be a responsible person associated with each domain to be a contact point for questions about the domain…and to resolve any problems*[100]

The RFC also made it clear that data should expire and domains would have to be deleted under certain conditions and references NICNAME/WHOIS as the mechanism for accomplishing these tasks. With the structure documented, the flip from hosts to domain began with RFC 920 in 1984, which made first use of "The Dot" structure we are now accustomed to.[101]

1.6.2 WHOIS Updated for Domains (1985)

Now that the DNS has replaced the hostname system, it seems appropriate for WHOIS change as well. RFC 945 NICNAME/WHOIS was authored by Ken Harrenstien, Mary Stahl, and Jake Feinler who had all been working so closely at the lower levels of managing the practical aspects of the ARPANET. Some of the major differences in this update included dropping NCP from the protocol by strictly using TCP, but the major policy shift introduced the concept of a *registrar* who would handle the domain name entries and collect the WHOIS data. Domain registration was by email:

> *To register, send via electronic mail to* REGISTRAR@SRI-NIC.ARPA *your full name, middle initial, U.S. mailing address (including mail stop and full explanation of abbreviations and acronyms), ZIP code, telephone (including Autovon FTS, if available), and one network mailbox.*[102]

Autovon was a US military phone system,[103] aside from which the requirements for domain WHOIS data have not changed. Registration was required for anyone with a node capable of passing traffic on the network. With this new system, the concept of a Domain Name Administrator arose along with WHOIS being an important part of the toolkit. The Domain Administrator's Guide[104] from 1987 stated:

> *VERIFICATION OF DATA: The verification process can be accomplished in several ways. One of these is through the NIC WHOIS server. If he has access to WHOIS, the DA can type the command "whois domain <domain name><return>". The reply from WHOIS will supply*

[100] http://tools.ietf.org/html/rfc882
[101] http://www.rfc-editor.org/rfc/rfc920.txt
[102] http://www.ietf.org/rfc/rfc954.txt
[103] http://massis.lcs.mit.edu/archives/reports/autovon.instructions
[104] http://tools.ietf.org/html/rfc1032

the following: the name and address of the organization "owning" the domain; the name of the domain; its administrative, technical, and zone contacts; the host names and network addresses of sites providing name service for the domain.

The term NIC can be confusing since it stands for Network Information Center, an entity publishing Internet information, or it could also stand for "Network Interface Card," a specific device on a machine for accessing the network. For the most part in this text, we will use NIC to refer to the former.

1.6.3 Oops! The Internet Goes Public

The Internet started out as a US government-sponsored project with access restricted to the academic and private researchers as well as the US military. As we all know, access was eventually extended to the public. Few are aware that this was not a carefully thought-out process with considerations for identity and abuse. The Internet was opened quite suddenly and, some would say, without authorization. Barry Shein is an interesting character with quite a bit of Internet history under his name including that of becoming the world's first private Internet Service Provider (ISP) by accident.

Shein got his first ARPANET account in 1978 while working at Harvard University using his math skills to support medical research in pulmonary mechanics. "*In those days everyone was a super-user*,"[105] said Shein referring to a complete lack of access levels or distinction of roles, something unthinkable on large networks today. With the help of a grant to purchase computers and equipment, Shein moved to Boston University (BU) and built the university's machine rooms and connected them to what was called "The Triangle," a local high-speed network between Harvard, MIT, and BU. Information technology did not get the respect it does today as Shein used a reclaimed linen closet with a hijacked phone line to do his work. In putting things of that time in the WHOIS or identification context, I asked Shein how miscreants were dealt with. "*I went down to the patch closet and pulled their connection or we called them into the office and had a chat. It's clearly a little different now.*"[106] Shein registered the first domain name for Boston University, BU.EDU, directly from Jon Postel. The backstory of BU.EDU tells an important lesson about who gets to register domain names. Apparently, some of the university fathers resented the abbreviation "BU" in general.

As Shein left the academic world for private sector software development as Software Tool and Die (STD) (std.com[107]), he started getting interesting requests from old colleagues. People who connected to the ARPANET at work wanted to be able to connect from home terminals too. They asked Shein if he could connect them from his new private office in Brookline, Massachusetts. So in the fall of 1989, he bought six 2400 baud[108] modems and installed them in a bookshelf with a router and a created link to the Internet in a nongovernment-sponsored location. This was all fine until the National Science Foundation Network[109] (NSFNET) and the Military Network[110] (MILNET) blocked his

[105] Shein interview.
[106] See note 105.
[107] http://www.std.com/
[108] http://www.linuxjournal.com/files/linuxjournal.com/linuxjournal/articles/010/1097/1097s2.html
[109] http://www.nsfnet-legacy.org/about.php
[110] http://www.computerhope.com/jargon/m/milnet.htm

access. The concern was fundamental to issues of access, responsibility, and identification: how to control abusive users.

After some back-and-forth with NSFNET, Shein was granted *permission* to allow others to access the network through his service, provided that they abide by the NSFNET Backbone Services Acceptable Use Policy,[111] which prohibited "*illegal or specifically unacceptable use*" and "*advertising of any kind.*"[112] Shein also offered to display such terms of use at the login and to drop any users who violated them. STD had contracts with its subscribers, which were enforceable as the point of entry for the Internet. Shein's brief spat with NSFNET became irrelevant as major changes were underway. Beyond the changes in network management, structure, and access, interaction was shifting from text to hypertext. The hypertext interface tuned the Internet into the *Web* and extended it to people not using a keyboard, forever changing its appeal.

1.7 1990s: THE INTERNET AS WE KNOW IT EMERGES

In 1989, domain registration and WHOIS updates were still being done through email,[113] but a major transition was seen at the beginning of the decade that expanded WHOIS drastically. The ARPANET was "decommissioned" in 1990 after 20 years of revolutionary development. The APRANET of course still existed, but the management changed, shifting to the National Science Foundation (NSF) that had been operating the *backbone* of the Internet since the mid-1980s. The backbone consisted of a collection of powerful networks that acted as national Network Access Points (NAP). The NSF filled the financial gap left by the Internet sponsorship shifting away from the Defense Department toward the Department of Commerce.

In 1992, the NSF awarded Network Solutions a contract to manage the DNS and register domain names. Network Solutions created InterNIC to handle the domain services, including WHOIS. The commercial sale of domain names was a disruption in the minds of many early Internet pioneers. Tim Berners-Lee, designer of the World Wide Web, wrote this arrangement:

> *"Network Solutions…made profits but does not have the reputation for accountability."* [10]

This sentiment is reiterated in *Who Controls the Internet: Illusions of a Borderless World* by Jack Goldsmith and Tim Wu:

> *"Another was a widely detested corporation named Network Solutions that had taken over day-to-day administration of Internet domain registration. The community that invented the Net was losing control over its creation."* [11]

Around this same time, web browsers, commercial ISPs, and a new method of traffic routing called Classless Inter-Domain Routing[114] (CIDR) all emerged to bring the Internet closer to general public use.

[111] http://w2.eff.org/Net_culture/Net_info/Technical/Policy/nsfnet.policy
[112] See note 111.
[113] http://tools.ietf.org/html/rfc1118
[114] http://tools.ietf.org/html/rfc1519

In 1994, the IETF Whois and Network Information Lookup Service (WNILS) Working Group reaffirmed the important role of WHOIS in RFC 1689.[115] The memo discusses the most commonly used contemporary search applications, here called *Networked Information Retrieval* tools. The memo explains the status of each along with proposals, spending considerable time on WHOIS. The authors noted that the use of WHOIS is spreading and being modified on different servers noting that the once central service is now *distributed*. It is also acknowledged that Port 43 is no longer the only method for accessing WHOIS records. Gopher, Wais, Archie, and Telnet are being used to get to WHOIS. They also note the growing presence of WHOIS clients in the exploding market of personal computers. The general conclusion of this review is that the service is important and in demand but completely lacks standards. They propose, among other things, to provide consistency for WHOIS and enhance its function. Part of this effort was geared toward the establishment of WHOIS++.

1.7.1 Referral WHOIS or RWhois RFC 1714 (1994)

Referral WHOIS was designed and proposed by the Network Solutions S. Williamson and M. Kosters through RFC 1714[116] in 1994. At this time, Network Solutions had the exclusive contract for registering domain names, so WHOIS services were a big part of their concerns. Within the RFC, the authors encapsulated the problems emerging from the development of WHOIS and proposed a method for addressing these problems. There had originally been one WHOIS database for everything. The database itself became fragmented as new registry authorities assumed control over different portions of the Internet under ICANN distribution. First, the authority over IP addresses that had been delegated to InterNIC was broken into three pieces with RIPE NCC gaining control over European IP space and APNIC for Asia-Pacific. With Network Solutions now registering domain names, there was no longer a single place for all WHOIS records, and someone performing a query would have to know where to look. Referral WHOIS was a plan to accept a query and perform a series of lookups through a hierarchical design until the query could be sent as close as possible to the maintainer of the WHOIS data. This of course is the *referral*. Rather than expecting all Internet users to know where to look or expecting every server to hold all data, we simply tell the client where the data really sits.

1.7.2 WHOIS++ RFCs 1834 and 1835 (1995)

In 1995, the WNILS Working Group and several technicians at BUNYIP Information Systems Inc. proposed an enhanced model for WHOIS services, hence the name "WHOIS-Plus-Plus." The "++" is also a reference to extensibility, something with added features and the ability to grow. The overall goal was to provide more structure to the WHOIS model, because as it was WHOIS had evolved in an ad hoc fashion that was not scalable for the growing Internet. Among other things, the proposal included a plan to create three classes of records—people, hosts, and domains—each with a specific structure. It sounded like a great plan, but it did not take off.

[115] http://tools.ietf.org/html/rfc1689
[116] http://tools.ietf.org/html/rfc1714

In speaking directly to one of the initial developers, Patrik Faltstrom,[117] we got a partial answer as to why. Faltstrom cited some missing pieces in WHOIS including a lack of updates to the existing protocol especially concerning standards.[118] "*XML needed to be born*," said Faltstrom referring to Extensible Markup Language (XML), which would not be released until 1998.[119] One of the key features of XML is that it provides a bridge between human-readable language and machine-readable language. The development of WHOIS++ was also done with BUNYIP Information Systems who may have lacked the level of funding at the time needed to launch a project of this scope. As a result, Lightweight Directory Access Protocol[120] (LDAP) became the preferred model. The concepts of WHOIS++ however were solid, and he hoped the Web Extensible Internet Registration Data Service (WEIRDS) will deliver on the hopes of WHOIS++.

1.7.3 ICANN Takes over WHOIS (1998)

In 1998, the existing functions of the Internet including IP delegation, domain name registration, and DNS management were all reorganized under ICANN through a contract with the US Department of Commerce. This new body partly includes the existing Information Sciences Institute (ISI) and its Internet Assigned Numbers Authority (IANA) operations. The Network Solutions InterNIC operation becomes part of ICANN, and their domain registration monopoly ends with new registrars being given contract opportunities to sell domains. Under ICANN, the existing TLDs were also spun off from Network Solutions to be run by new registries. Much of this was done to expand market competition and bring more private money into the Internet. The authority of ICANN came directly from the US government through the Memorandum of Understanding.[121]

This major change generated an interesting artifact in WHOIS, which still exists today. This advertisement still appears in the WHOIS records returned by VeriSign through InterNIC:

> *Domain names in the .com and .net domains can now be registered with many different competing registrars. Go to* http://www.internic.net *for detailed information.*

For over 10 years, there have been hundreds of different registrars, and this message is a complete anachronism.

1.8 2000s: WHOIS STANDARDS

The ICANN is now in control. Domains are being sold by multiple registrars and new registries are emerging. In this new and growing space, the requirements for WHOIS were reaffirmed through updating the previous WHOIS RFC 954 and contractual obligations for WHOIS.

[117] http://tools.ietf.org/html/rfc1835
[118] Faltstrom interview
[119] http://www.w3.org/XML/
[120] http://tools.ietf.org/html/draft-hall-ldap-whois-01
[121] http://www.icann.org/en/about/agreements/mou-jpa/icann-mou-25nov98-en.htm

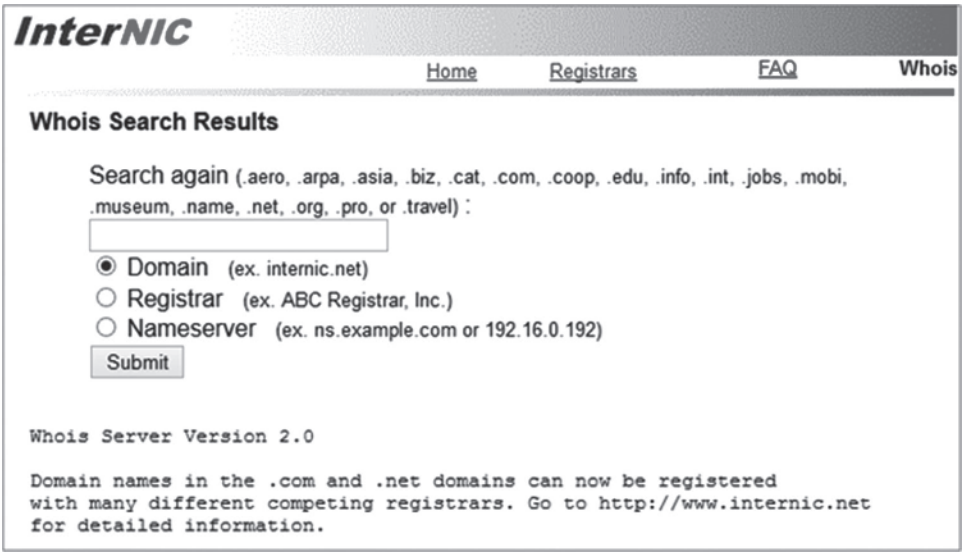

FIGURE 1.17 InterNIC WHOIS.

1.8.1 ICANN's Registrar Accreditation Agreement and WHOIS (2001)

The new nonexclusive registrar scheme included a contract with ICANN, which dictated standards for WHOIS service delivery and accuracy.[122] For all intents and purposes, a fairly large section of the contract is devoted to WHOIS topics and launched a number of controversies in the decade since it was put in place.

The ICANN's Registrar Accreditation Agreement (RAA) WHOIS requirements include the collection of the standard list of point of contact data as has always been part of WHOIS: name, address, email, phone number, etc. However, because the registration and data are now being held by third parties, there are additional obligations on the registrars in terms of WHOIS:

- A service level is expected of the registrar to provide access to WHOIS not only through public Port 43 but also through a webpage.[123]
- Registrars must also sell access to the WHOIS entire record for $10,000 or less.[124]
- And, in a very detailed section, registrars must enforce WHOIS accuracy and terminate domain agreement if the record is falsified.[125]

1.8.2 WHOIS Protocol Specification 2004 RFC 3912 (2004)

Authored by VeriSign's Leslie Daigle, more recently the chief Internet technology officer at the Internet Society[126] (ISOC), RFC 3912 is the Internet Official Protocol Standard for WHOIS[127] and has not been replaced or obsoleted as of 2015. While this document is fairly

[122] http://www.icann.org/en/resources/registrars/raa/raa-17may01-en.htm#3
[123] http://www.icann.org/en/resources/registrars/raa/raa-17may01-en.htm#3.3.1
[124] http://www.icann.org/en/resources/registrars/raa/raa-17may01-en.htm#3.3.6
[125] http://www.icann.org/en/resources/registrars/raa/raa-17may01-en.htm#3.7.7
[126] http://www.internetsociety.org/who-we-are/people/ms-leslie-daigle
[127] http://tools.ietf.org/html/rfc3912

brief and does not change the function of WHOIS, it is a critical part of WHOIS today. While many critics of WHOIS claim it is a dinosaur inherited by ICANN and no longer relevant, this RFC proves otherwise. First, it reaffirms the technical specification of WHOIS as TCP Port 43 transaction-based query-response service. By describing WHOIS in this way, it preserves the expectations of its functionality and the way clients call it and the way servers return information. The schematic of the transaction is spelled out:

```
3. Protocol Example

If one places a request of the WHOIS server located at
whois.nic.mil for information about "Smith", the packets on
the wire will look like:

client  server at whois.nic.mil
open TCP   ---- (SYN) ------------------------------>
           <---- (SYN+ACK) -------------------------
send query ---- "Smith<CR><LF>" -------------------->
get answer <---- "Info about Smith<CR><LF>" ---------
           <---- "More info about Smith<CR><LF>" ----
close      <---- (FIN) ------------------------------
           ----- (FIN) ------------------------------>
```

The meaning of these codes and sequences is spelled out in the chapter on WHOIS use.

Why the WHOIS protocol exists is clearly stated as an information service for Internet *users* to provide information about domain names.[128] There are no other target audiences or purposes stated. WHOIS exists to serve the Internet end user with domain name records and must do so in a "*human-readable*"[129] format. Further, in the RFC, it is even clarified that the data is "*intended to be accessible to everyone.*"[130] One might think this is the end of the argument as to the purpose of WHOIS, but the debate is only the beginning.

Also of critical importance is the acknowledgment of problems within the WHOIS architecture, namely, language support and security. The "*WHOIS protocol has not been internationalised*" and "*lacks mechanisms for access control, integrity, and confidentiality.*"[131] Internationalization and overall system integrity are two very real WHOIS problems that require addressing. However, the ongoing WHOIS debates focus on what data goes into the records with the most common, and unrealistic, recommendation being the elimination of the system altogether. Because of the notable security flaws, domain owners are admonished to only use "*non-sensitive information*"[132] in the records. Keep in mind as we move through the various debates within the text that RFC 3912 is the ruling principle.

1.8.3 Creaking of Politics

What happens next in the history of WHOIS defined its future for some time to come. Beyond the founding technical aspects of Internet development, its structure becomes defined by political difference as it crosses international borders, commercial interests emerge, and more

[128] See note 127.
[129] See note 127.
[130] See note 127.
[131] See note 127.
[132] See note 127.

ordinary people get domains. The potential problems of a global public Internet are no better encapsulated than in the 2002 paper *Tussle in Cyberspace: Defining Tomorrow's Internet*[133] by four eminent computer scientists. The "tussle" referred to is the various disputes destined to emerge between different Internet stakeholders. While the stakeholder disagreements may seem unfortunate, they are seen by the authors as critical to Internet maturity. This is the collision of predictable engineering principles and unpredictable societal movements. The architecture of the Internet passes bits of information that seems straightforward. The issues of what is in the bits, who gets to pass the bits, how many bits people get to move, and who handles the bits along the way make things much more complex. This is a shared space. Keeping the space open for everyone depends on trust, and trust comes from identity:

> *"One of the most profound and irreversible changes in the Internet is that by and large, many of the users don't trust each other… There are parties with adverse interests, and some genuine 'bad guys' out there. This implies that mechanisms that regulate interaction on the basis of mutual trust should be a fundamental part of the Internet of tomorrow."*[134]

This has clearly not improved since, and the authors point to the simple fact that

> *"if communication is to be mediated based on trust, then as a preliminary step, parties must be able to know to whom they are talking."*

The fundamental concerns of identity and trust pervade the real world as well as the digital and are not trivial problems to solve. This debate runs through the heart of WHOIS as a political matter and underlies the Internet development. Politics cannot be separated from technology once the technology crosses out of private space.

REFERENCES

1 Berloquin, P. 2008. Hidden Codes and Grand Designs. New York: Sterling.

2 Thornton, J. 2012. Polybius. The Encyclopedia of Ancient History.

3 Crabtree, J. 1901. The Marvels of Modern Mechanism and Their Relation to Social Betterment. Springfield, MA: San José The King-Richardson Company.

4 Gillings, R. 1972. Mathematics in the Time of the Pharaohs. New York: Dover Publications, P210.

5 Kleinrock, L. 1961. Information flow in large communication nets, http://www.lk.cs.ucla.edu/data/files/Kleinrock/Information%20Flow%20in%20Large%20Communication%20Nets0.pdf (accessed June 20, 2015).

6 Defense Communications Agency. 1978. ARPANET Information Brochure. Ft. Belvoir: Defense Technical Information Center.

7 Feinler, Elizabeth "Jake". 2011. Host Tables, Top-Level Domain Names, and the Origin of Dot Com, IEEE 33:74–79.

8 Blum, A. 2012. TUBES. New York: HarperCollins.

9 Kenyon, T. 2002. Data Networks: Routing, Security, and Performance Optimization. Amsterdam/Boston: Digital Press.

10 Berners-Lee, T. 1999. Weaving the Web. San Francisco: HarperSanFrancisco.

11 Goldsmith, J. and Wu, T. 2006. Who Controls the Internet: Illusions of Borderless World. New York: Oxford University Press.

[133] http://groups.csail.mit.edu/ana/Publications/PubPDFs/Tussle2002.pdf

[134] See note 133.