Metadata and Infrastructure in Internet History: Sockets in the Arpanet Host-Host Protocol

Bradley Fidler
University of California, Los Angeles
Computer Science Department
4732 Boelter Hall
Los Angeles, CA 90095
fidler@cs.ucla.edu

Amelia Acker
University of Pittsburgh
School of Information Sciences
135 North Bellefield Avenue
Pittsburgh, PA 15260
aacker@ucla.edu

ABSTRACT
In this paper we describe the generation and utilization of metadata as part of normal network function on the early Arpanet. By using the Arpanet Host-Host Protocol and its sockets as an entry point for studying the generation of metadata, we show that the development and function of key Arpanet infrastructure cannot be studied without examining the creation and stabilization of metadata standards. More specifically, we use the Host-Host Protocol's sockets as an example of something that, at the level of the network, functions as both network infrastructure and metadata. By presenting the function of sockets in tandem with an overview of the Host-Host Protocol and a key application built atop it, Telnet, we illustrate the necessity of studying infrastructure and metadata in tandem. Finally, we draw on Esveld (1990) to reintroduce the concept of *infradata* to refer specifically to data that locates data throughout an infrastructure and is required by the infrastructure to function, separating it from established and stabilized standards. We argue for the future application of infradata as a concept for the study of histories and political economies of networks.

Keywords
Arpanet, infradata, infrastructure, metadata, protocols, sockets.

INTRODUCTION
Recent histories of infrastructure and metadata have increasingly come to the foreground in many areas of information science, including infrastructure studies, science and technology studies, and history of technology, amongst others. However, in many examinations of infrastructure and information and communication technologies (ICTs) the stabilization of metadata, including its generation and points of origin, are left out. Instead, many histories focus on the establishment of standards through the consensus of experts, and how metadata is generated from such. The reverse can be said of metadata histories as well—often these stories are told with an eye towards large-scale standards bodies and their organizational development, highlighting the political and cultural economy of experts creating standards, but undercutting the impact that metadata has on infrastructures before it has become stable and standardized. The separation between studies of infrastructure and metadata indicates that an area of analysis has come to rest too early—we argue that infrastructure and metadata should be studied together when examining network technologies, such as the Internet or mobile telephony. This is because the function of network infrastructure, including its core protocols, cannot be separated from the development of certain metadata standards which, once in place, set conditions of operation for that infrastructure. What is more, development of key metadata standards during the design of network infrastructure – often, as is the case in our analysis, in protocol – has implications beyond the intentions of the designers. The development of metadata standards and the designation of their function(s) in network infrastructure creates a set amount of metadata that is thereby required so that the network infrastructure can function. This specific kind of metadata, which is an integral component of network infrastructure that is generated as a part of normal operations is what we call infrastructural metadata, or *infradata*. It is distinguished by its function: metadata required for network infrastructure to work. Analytically, infradata calls our attention to the design decisions (and their economies, professional practices, ethics and values) that creates specific forms, and not others, of (necessary) infradata in network infrastructure. Infradata is not optional; it will always be present in a functioning network as part of normal operation.

77th ASIST Annual Meeting, October 31 - November 5, 2014, Seattle, WA, USA.

In this paper we reference the development of network sockets in the early Advanced Research Projects Agency Computer Network (Arpanet), the network that served as a key foundation upon which the Internet was developed. Sockets, which link computers in inter-process communication over a network, provide a case for studying the development of metadata standards as part of infrastructural analysis. They also reveal how metadata in turn, shapes infrastructures of networking. If we study infrastructure and metadata separately, we miss the determining power that metadata generation and collection have on future infrastructures and standards that are bootstrapped together (Traugott & Huddleston, 1998), as well as on the kinds of metadata that will always be present in a given infrastructure. We argue that these two concepts, which are often presented as distinct conceptual principles and separate avenues of inquiry in histories of networked ICTs, are indeed not separate because infrastructure needs metadata to function and metadata is always in part encoded into infrastructure (Bowker & Star, 2006). The stabilization and generation of metadata reflects infrastructure and the future impact that it has on the collection, repurposing, and future use of metadata, in addition to the stabilization of infrastructure itself. Understanding that network infrastructure is created in part by the original generation and implementation of metadata as a conceptual strategy allows us to examine the development of capabilities in each through networking technologies over the last fifty years, in what Andrew Russell has called "histories of networking" (2012). There has always been integration between metadata and infrastructure, but this integration is tightening and becoming more indistinguishable with the Internet and next generation mobile networks.

Our study of the Arpanet illustrates how engineers nearly wrote descriptive user data into the network's infradata: in other words, descriptive user data would have been generated and maintained by the network in order for it to function. The understudied and *ad hoc* development of these standards, as well as their historical context, is not well understood, nor is the decision to ultimately remove most but not all of this user metadata generation from network infrastructure before it was standardized.

As we will argue below, the Arpanet is a central antecedent of the modern (IP-based) Internet and provides an important example of the relationship between infrastructure and metadata in a historical perspective. Protocols and metadata standards that governed Arpanet infrastructure carried over, or were used as a jumping-off point, for those that later governed the modern Internet (Abbate, 2000). What is more, the case study we develop below – certain applications and protocols that were barely in place by 1973 – provide insight into early and formative stages of Arpanet, and thus, Internet infrastructure. By choosing an extremely early example, we emphasize that it is not a question of if there is metadata to be found in network infrastructure, but rather, where, how much, for what purpose, and the reasons for its development.

LITERATURE REVIEW

As we have argued, metadata and infrastructure are connected and influence each other, but are often not analyzed in tandem. In this paper we illustrate how they should be historicized together, as mutually constitutive. This literature review provides an introduction to some of the current ways that metadata and infrastructure are defined and conceived in studies of information systems, infrastructure, user-generated metadata, and analysis of early packet networks.

Metadata

According to the National Information Standards Organization (NISO) (2004) metadata "is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data or information about information" (p. 1). Metadata ensures access to information resources in the moment and over time and it can be experienced in systems all around us. From menus in the drive-through restaurant to tax codebooks, metadata organizes things in the world, enables interoperability across systems, and identifies information objects. In addition to describing things for discovery and identification (known as "descriptive metadata") metadata can also be structural. Both descriptive and structuring metadata exists as schemas with distinct elements. Today, large standards bodies like NISO, Dublin Core, or the Internet Engineering Task Force (IETF), are committed to developing and updating metadata schemes ranging from union library catalogs, to packet headers, to open source coding initiatives. As Jane Greenberg (2003) writes, "[t]oday's metadata activities are unprecedented because they extend beyond the traditional library environment in an effort to deal with the [Internet's] exponential growth" (p. 1876). In library and information science (LIS), metadata has been historically concerned with access to information resources like books and union catalogs, however the universe of documentation and information access has changed with the Internet. While a fair amount of effort is still spent on developing and maintaining metadata registries, there is little work that looks at the development of *metadata as a social process*. We know that the creation of metadata and structuring schemas is both an objective and subjective process (Duval et. al, 2002). Many scholars acknowledge that metadata is context specific, and shaped by the systems, people, and customs for which it is located. For example, Bowker and Star argue that standards and classification are essential to making infrastructures work, that infrastructure is "standards all the way down" (2006, p. 234). However, the origins of metadata—its emergence, and the ways in which it is yoked to infrastructures are often obscured by its stabilization and maintenance. Instead, the analysis of metadata focuses upon its reliability and issues in communicating its authority (Bowker & Star, 2006, p. 238). Often sociohistorical analysis of metadata will focus upon an organized standards body's development including how its authority is agreed to, created, supported and imbued in communities of practice. This can be seen in many recent histories of standards, digital formats, and telecommunication development (Hillebrand, 2002; Pelkmans, 2001; Russell, 2006; Sterne, 2012) instead of the origins of metadata as it is initially conceived or first generated. The standards organizations that create and circulate metadata schemes such as Dublin Core, NISO, IETF are essential, however, often, the initial creation of schemas can be unwieldy and appear without reference to their stabilization or historical ontology (Ribes & Polk, 2012).

Standards and standardization development, which include the generation and stabilization of metadata, are cultural practices that influence and shape how our society communicates and creates culture. The relationship between the individual and metadata has changed rapidly in the last decade. In 1970, computing infrastructure was largely stationary, and users could move around from computer to computer, but the user-generated metadata had constraints based on the grounded nodes of the Arpanet (as we will see below, a system that accounted for user mobility was proposed in 1971 and nearly developed, but ultimately discarded). Forty years later, users have access to wireless data networks where metadata is generated all the time as part of being connected to the network with mobile devices. Recent work from forensic scientists, privacy researchers and computer scientists has found that user generated metadata created with mobile devices and Internet connected mobile networks leave little room for anonymity or obfuscation (boyd & Crawford, 2012; Brunton & Nissenbaum, 2013; de Montjoye et. al, 2013; Glisson, et. al, 2011; Mayer & Mutchler, 2014; Willassen 2003; Willassen, 2005). Mobile and wireless data networks create swaths of user metadata that institutes a new kind of subjectivity with networked ICTS. Historians of technology and information scholars who examine ICT standards show that these hidden standards structure our experiences because they "shape not only the physical world around us but our social lives and our even our very selves" (Busch, 2011, p. 2). The historical ontology of metadata as an artifact is, as we wrote earlier, yoked to the creation and stabilization of infrastructure—including technology, standards, material things, and practices of communicating with groups of people. In the next section we provide an introduction to the literature on metadata and infrastructure studies.

Infrastructure

Recently, a turn towards the study of infrastructure has been seen in information science, science and technology studies, and history of technology. Infrastructure is made up of the devices, standards, technical architecture and network elements that make ICTs work. This includes the global information and communication technologies that transmit information across national and regional boundaries (Bowker et al., 2010, p. 98). Because standards and metadata structure our experience of ICTs, histories of their development, infrastructure, and cultural possibilities are important for information science research. Infrastructure studies aims to do this contributing to a comprehensive historical understanding of the development of their information infrastructure, including the development of standards. Information infrastructure studies examine large-scale ICT infrastructures ranging from scholarly communication with cyberinfrastructure (Borgman, 2003), the proliferation of Internet (Bowker et al., 2010) to studies of privacy, and surveillance contexts (Shilton, 2009). A variety of scholars from history of technology, science and technology studies and information science have coalesced around a variety of descriptive large-scale examples of infrastructure and communities (Edwards 1996, Lee et al. 2006, Traweek 1992). Infrastructure studies allow us to examine the nexus of histories of networking. Such histories of networking infrastructures bring into relief how technical development, policy, standards and international cooperation and large-scale networks and the groups of people that use and communicate across networks. In the next sections we discuss our method and a case where the generation of metadata and the establishment and stabilization of infrastructure is mutually constitutive. We show how the concept of infradata is useful for characterizing and analyzing the relationship between metadata and infrastructure.

BACKGROUND AND METHOD

The development of the network protocols, applications, and metadata standards was a highly decentralized process. The Advanced Research Projects Agency's (ARPA) main contractor for the network, Bolt Beranek and Newman (BBN), designed and implemented the subnetwork – the part of the network that shuttled data between the network's computers or, as they were called, hosts. This left much to the *ad hoc* Network Working Group (NWG). It was composed of graduate students from the network's early sites and BBN staff, and took on the task to design and implement the protocols that would provide for a general way for applications to connect with each other across the network: a "Host-Host Protocol." Until this protocol was in place, the Arpanet provided the technology to allow networked computers to communicate with each other, but no general way with which to do so. Differences between computers, such as different character sets, communication protocols, and operating systems, had to be manually accounted for, and as such, the Arpanet was not initially useful as a computer utility. The process of developing host-host protocols was documented officially in the Request For Comments (RFCs) series, which emerged as an informal way to discuss the development of these protocols (Abbate 2000) by the NWG. Still in use today, albeit in a more formalized manner, RFCs provide insight into the rationale behind design decisions, as well as paths not taken, in the development of the host-host protocol. We rely on RFCs as primary documents, as well as published articles documenting these applications and protocols, to elucidate the development and design decisions of the protocols, applications, and metadata that we document herein. We interpret these primary sources using historical analysis to tell this story of sockets on the ARPANET. Historical interpretation of documents is the systematic and disciplined analysis of traces or primary sources to make sense of the past. This is asking the who, what, where, when, and how to a collection of documents or artifacts created in a certain time and for certain purposes, in this case, for the generation of metadata in support of networking infrastructures. We also use techniques from infrastructure studies to examine and interpret these primary sources related to the development of sockets, in addition to the origin points of metadata generation. In the case study that follows we examine where the generation of metadata and the establishment and stabilization of infrastructure is mutually constitutive.

HISTORICAL AND TECHNOLOGICAL PERSPECTIVES OF THE ARPANET

Explorations of the social and technological origins of the modern Internet often begin with the Arpanet, the experimental and later operational research network developed by the Advanced Research Projects Agency (now DARPA). By 1975 it was considered operational and transferred from DARPA to the Defense Communications Agency (DCA), and continued to serve a largely university, government, and industrial community until it was decommissioned in 1989 (Abbate 2000). It played a central role in creating social, institutional, and technological foundations upon which the Internet was built. The protocols and applications developed on the Arpanet had a great influence on those used on the modern Internet: either as a foundation that was improved incrementally, or a foil that revealed shortcomings to be corrected. Crucially, Arpanet was the major test bed for the development of the protocol suite that underlies our modern Internet, TCP/IP. In 1983 the Arpanet's Host-Host Protocol, which we discuss herein, was replaced with this Transmission Control Protocol / Internet Protocol (TCP/IP). Any network running the Internet Protocol (IP) on its computers could now seamlessly "inter-network" with the Arpanet, and as they did, the modern "Internet," a massive inter-network, emerged around it. Many of the tools, techniques, and practices that evolved on the Arpanet remain with us on today's Internet, or served as important predecessors to their subsequent iterations today. TCP/IP, for example, was in part a response to the shortcomings of the Host-Host Protocol, and many of its components, such as sockets, were improved versions that nonetheless owed inspiration to these earlier versions.

In September 1973, the Arpanet consisted of 41 sites (see Figure 1), of which 23 were (DARPA-funded) academic research institutes, 11 were military R&D, and 9 non-Defense Department R&D (Lukasik, 2011, p. 13). It connected sites across the continental United States, as well as sites in Hawaii, Norway, and the UK. Each Arpanet site connected between one and four of their own computers (called "hosts," roughly analogous to today's "servers") to the network and, through these, between one and dozens of terminals, composed of a keyboard and either a Teletype-like printer or Cathode Ray Tube (CRT) display. Unlike earlier computers, these were "time-shared" systems, meaning that multiple users could access the system simultaneously. From these terminals, users could use the Arpanet to access any host on the Arpanet as if they were physically proximate (i.e. next to the machine on a local terminal), an ability that is taken for granted today but was a major innovation in the early 1970s.

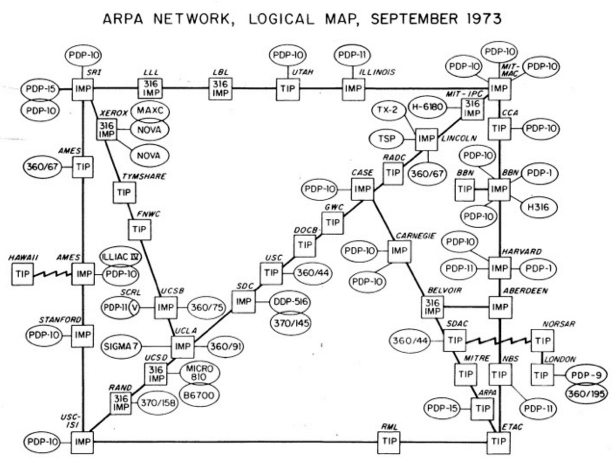

Figure 1. 1973 ARPANET Logical Map. ARPANET Completion Report (BBN, 1978)

A REMOTE CONNECTION OVER THE ARPANET

Before the Host-Host Protocol, the Arpanet provided the underlying technology through which host computers could send information to one another across the network. Each computer was connected to its local Interface Message Processor (IMP), which were responsible for getting information to and from its origin and destination host computer. The IMP subnet, as it was called, was able to route data from origin to destination (McQuillan et al. 1979). Atop this protocol was a protocol for how host computers would package information to send to their local IMP, ensuring that the hosts would provide enough information to their IMP so that the IMPs could shuttle it across the network, and relay it to the destination host (BBN 1978). This protocol also provided the way for the IMP to take this information and pass it off to the destination host in a way that the host would understand. What these technologies did not provide was a general purpose way for specific *applications* within different hosts to communicate with each other across the network. As mentioned, any connection between hosts had to be set up on a custom basis to account for differences in operating systems (such as character sets). It was the ability to connect between applications that would make the network useful; once this ability was developed as a protocol it would become a central component of the network infrastructure.

The Host-Host Protocol that was required would create a set of communication standards that would be implemented on each connected computer so that they would have a common language with which to 'talk' between one another's applications. The Host-Host Protocol was implemented on each Arpanet computer as the Network Control Program (NCP), so that each (heterogeneous) machine could communicate with others in the same language. The Host-Host Protocol allowed the development of applications that could make use of its common language to communicate with other systems: much like today, rather than having each network application specify everything about how it would communicate with another, each application draws on lower-level protocols (such as Host-Host) in order to take advantage of an existing set of communication standards. Once in place, Telnet was quickly developed as the primary application through which users would access remote hosts: Telnet provided a common command-line interface for communicating between different operating systems, atop the Host-Host Protocol (Davidson, et. al, 1977). The user would use Telnet to access a remote machine, log in (with credentials, if necessary) and run programs local to it (Davidson, 1977). Thus when we discuss the significance of sockets in the Host-Host Protocol, it was frequently Telnet that would draw on them to make connections. As mentioned, our study is of the process prior to logging in – it does not include any personal data that the user would volunteer in making the connection. Instead, we use this example to focus on the metadata that is generated automatically, as required for the network to function as it opens a connection.

In order to explore the importance of metadata in Internet history, we turn to this example of a user login to a remote host. Today, this is analogous to connecting from the local computer (i.e. a laptop) to a remote system (i.e. an email service). The importance of the Host-Host Protocol in this process is further explained by exploring the nature of these connections. Connecting two host computers over the Arpanet was not a matter of two computers being already connected via the lines of the network maps. Due to the design decisions embedded in the Arpanet subnet (McQuillan, 1977), connections between computers needed to be negotiated and established on a case-by-case basis, and were ended after they had carried out their function (for example, after a user's remote login session had been closed). What is more, time-shared systems in particular had many programs being run simultaneously by many users – there needed to be a way to distinguish between which *part* of the computer was going to communicate with which part of the other computer before the connection could be negotiated and established: all this needed to happen before the remote computer could, for example, provide the user with a login prompt. In creating a common means by which computers could link specific applications for computer-to-computer communication, the Host-Host Protocol specified the structure and function of *sockets.*

Sockets, which exist in a contemporary iteration on the modern Internet, were 32-bit numbers created by a host computer that specified an exact process (such as an application) to which it was connected. By sending data to a socket number on a remote computer, the sender would know that any data it sent to that socket would be directed to a specific process. In other words, by knowing the socket number of the receiving computer, the sending computer would know where to address the data of the process that was to send it. What is more, in generating the socket number, the receiving computer would know where the incoming data was supposed to go. Of course, each (different) operating system had a different way of keeping track of their various processes. Sockets ensured that different operating systems did not need to understand each other's specifics: instead, when an operating system generated a socket using the Host-Host Protocol, it would remember the process to which it referred; the other computer would only need to know the number. In order to negotiate a connection between computers, an application called the Initial Connection Protocol (ICP) would automate the process of creating and exchanging socket numbers between two computers, so that inter-process communication could commence.

The socket number, then, was precisely specified to locate data across the network. It contained both the network address of the host computer, as well as a specified identifier of the process (usually an application) that had been initiated by a user (for an explanation of sockets in the Host-Host Protocol, see McKenzie 1972). If we consider this communication between computers at the level of the network, sockets were data about the location of other data on the network infrastructure. We can illustrate their functions by walking through a common task on the Arpanet in 1973: logging into a remote host using the Telnet application. In order to establish an initial connection between computers, so that a user could log in remotely, Telnet (as well as other applications) made use of the Initial Connection Protocol (see Postel, 1971). The ICP's job was to establish that early connection between computer processes by establishing that one computer would send and another receives to and from a specific process. It sat 'between' Telnet and the Host-Host Protocol by automating the creation and negotiation of connections with the use of sockets. When a user opened the Telnet application and instructed it to connect to a remote computer, ICP would invisibly create the necessary socket to commence the connection. After these steps were completed, the user would then be faced with a login prompt from the remote system. Simply by requesting a connection with a remote computer using Telnet, then – and before logging in or running programs – a user would trigger the ICP to draw on the Host-Host Protocol to negotiate socket connections between the two computers. It was also designed to be invisible to the user – it would run automatically when called upon by Telnet. As mentioned, Telnet was the application most commonly used for accessing remote computers on the Arpanet. This level of detail is provided to emphasize that, for decades, the simplest connections over a network generated and required metadata of various types.

Communication between applications on the Arpanet required a very specific standardization of the data about the location of other data on the network. Accessing a specific process on a remote machine was, after all, a practice of accessing software resources, databases, and data. Decisions and debates regarding exactly how much data to include in these standards were being published in RFCs from 1969 until the early specification of the Host-Host Protocol in 1972; the earliest standard form of Telnet appeared in 1973. Through those debates and discussions we can see some of the paths not taken for these protocols and applications, as a process of stabilizing and standardizing these metadata.

Early socket designs in 1970 and 1971, for example, assumed that sockets would carry significant user metadata as a matter of course. In 1970, the socket was to be much more tied to a specific user than, as was later settled on in the final Host-Host Protocol specification of 1972 (McKenzie, 1972), a specific process such as an application (that was nonetheless initiated by the user). Indeed, rather than just specifying a connection between two processes, sockets would be linked to users via metadata and could even follow them to multiple (host) computers:

> "Each user is assigned a 24-bit user number, which uniquely identifies him throughout the network. Generally this will be the 8-bit HOST number of his home HOST, followed by 16 bits, which uniquely identify him at that HOST. Provision can also be made for a user to have a user number not keyed to a particular HOST, an arrangement desirable for mobile users who might have no home HOST or more than one home HOST." (Crocker et. al, 1970)

This system would function so that, upon logging in to a remote computer, each process run by the user would be "tagged" with a unique user number. What is more, when the user would log in at a remote computer, the user number would also be used to tag processes at that remote host (Crocker, et. al, 1970). In other words, user actions would be logged in tandem at their local and remote machines – wherever they might be – as part of the metadata requirements for the network to function.

Elsewhere, this potential design decision was further encouraged, alongside other options for how a socket would function. In one scheme proposed a year later in 1971, users would select their own socket number before initiating a connection, thereby making the socket process highly visible to each user accessing the network, in contrast to the final socket design, which was completely invisible to each user. In another, "administratively assigned user identities" similar to Crocker's earlier (1970) suggestion were proposed, wherein a user is "permanently assigned a user identity" by the user's host computer, which broadcasts the identity of who is attempting to connect with a remote system, even before logging in or running a program (Harlsem & Heafner, 1971). Similarly, another suggestion repeated the necessity of recording user metadata, and expanding on it, because the Host-Host Protocol:

> "[S]hould log each connection made and record the time the connection was made, the time the connection was closed, the number of messages and number of bits transmitted over the connection, the sending and receiving hosts, and the sockets at the sending host and receiving host which participated in the connection." (Winett, 1971)

In this case, the reason for this user metadata generation was collection, in particular, for network accounting that would allow for sites or users to be billed for their network usage. Another interest was to create powerful tools for distributed computing, a broad interest in BBN at the time (see Schantz, 2006, and Walden, 1970), wherein a user would have the power to link any number of remote processes at different hosts into a distributed system.

Nonetheless, when the Host-Host Protocol was standardized in 1972, there were no user directories, and the user information contained in a socket number was limited to only a process being run *by* a user (McKenzie, 1972). Throughout each of these iterations of the Host-Host Protocol's sockets, the discussion was equally about what kind of user metadata to include, and how exactly it would be coded as metadata into sockets, and how it would function in infrastructure. For example, how many bits of the socket number would be allocated to the user identity, and how would it be encoded? How would that information be used to locate further user information in a database? Would users be able to have a role in generating their identifying metadata, or would it happen automatically, transparent to them? These are questions not only about metadata, but how core infrastructure of the network was to function.

More research is required on the social and technical histories of Internet protocols, such as the Arpanet's Host-Host Protocol, but it is nonetheless clear that the standardization process is at once the development of the network infrastructure. Determining the metadata structure and infrastructure functions of sockets was a fundamental problem, without which the Arpanet would not have been generally usable.

CONCLUSION: THE SIGNIFICANCE OF SOCKETS AS INFRADATA

Understanding the necessary conditions for a simple remote login over the Arpanet with Telnet, including the negotiation of sockets by the ICP over the Host-Host Protocol, alerts us to a category of infrastructural metadata – infradata – that must exist for network infrastructure to function. Visual and even narrative metaphors for the Internet describe computers "connected" if they are plugged in or connected to a network. While in many senses this is true, it is equally the case that connections at the application level were being constantly negotiated, formed, and terminated: this process is absolutely central to the network infrastructure, as without it, host computers (and users) cannot connect with each other.

In this paper we have shown an instance of how the analysis of networking infrastructure and metadata generation should follow histories of metadata generation, including how it is linked to early design decisions and path dependencies that shape later infrastructures. We analyze the structure and functions of metadata in network infrastructure as *infradata* – metadata that functions as part of infrastructure to locate data at both ends of a network connection. Sockets, for example, are data about the location of other data – endpoints in the remote operating system of an inter-process communication – that are used to form network connections as part of an infrastructure. This concept was used by Esveld (1990) to denote data about the kinds, locations, and characteristics of infrastructure throughout railway track systems. For Esveld, the physical layer of the infrastructure's requirements point to reliability (compared to, say, a higher transport layer of scheduling trains), but not the potential for data in endpoints, as this is not required for track reliability. However, for the Host-Host "transport" layer that we describe, infradata includes data structure endpoints, as located through the use of sockets. In the case of sockets, infradata is a type of metadata standard that is created based on the necessity of having data about data as it exists and is communicated across a network infrastructure. For every type of infradata, it is worthwhile asking what kinds of information are generated and recorded as required by the normal operation of the infrastructure.

We point to infradata in an early case of metadata and infrastructure in Internet history. We hope that it provides a way for researchers to ask how much metadata is generated in the course of basic network operation. Furthermore, we might ask, what are the earlier decisions – such as the necessity of creating a system of connection-based links between host computers, as was the design goal of the Arpanet – that create different amounts and kinds of infradata. We see, for example, that sockets were nearly created with user metadata built in, and, in the form they took on the Arpanet, they did link to processes initiated by users, which, in the host operating system, was typically linked to a user login. We have argued that infrastructure and metadata should be examined together, and that the impacts of infradata influence the questions we ask of not just historical situations but new, present, and unfolding experiences of networked ICTs.

ACKNOWLEDGMENTS

The authors would like to thank Alex McKenzie, Steve Crocker, and Ray Tomlinson for their help in our understanding of the Host-Host Protocol and sockets.

REFERENCES

Abbate, Janet. Inventing the Internet. MIT Press, 2000.

Bärwolff, Matthias. End-to-End Arguments in the Internet: Principles, Practices, and Theory. self-published with CreateSpace, 2010.

Borgman, C. L. (2003). *From Gutenberg to the global information infrastructure: access to information in the networked world*. Cambridge: The MIT Press.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in a networked environment. In *International handbook of internet research*, 97-117.

Bowker, G. C. & Star, S. L., (2006). *How to infrastructure. Handbook of new media: Social shaping and social consequences of ICTs,* 230-245.

boyd, d., and Crawford, K. (2012). "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon," *Information, Communication & Society* 15 no. 5: 662-679.

Brunton, F., & Nissenbaum, H. (2013). "Political and ethical perspectives on data obfuscation." In *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology,* 164-188.

David, P. A., and W. E. Steinmueller. (1994). "Economics of compatibility standards and competition in telecommunication networks." *Information Economics and Policy* 6.3: 217-241

Davidson, J., W. Hathaway, J. Postel, N. Mimno, R. Thomas, and D. Walden (1977). "The Arpanet Telnet Protocol: Its Purpose, Principles, Implementation, and Impact on Host Operating System Design." In Proceedings of the Fifth Symposium on Data Communications, 4.10–4.18. SIGCOMM '77. New York, NY, USA: ACM.

Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata principles and practicalities. *D-lib Magazine*, *8*(4), 16.

Edquist, C. (2002). *Internet and Mobile Telecommunications System of Innovation: Developments in Equipment, Access and Content*. Edward Elgar Publishing, Incorporated.

Edwards, P. N. (2003) “Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems.” *Modernity and Technology* eds. Thomas J. Misa, Philip Brey, and Andrew Feenberg. Cambridge, MA: MIT Press: pp. 185-225.

Esveld, Coenraad (1990). “Computer-Aided Maintenance and Renewal of Track.” In Railroad Conference, 1990, Technical Papers Presented at the 1990 ASME/IEEE Joint, 165–70.

Glisson, W. B., Storer, T., Mayall, G., Moug, I., & Grispos, G. (2011). Electronic retention: what does your mobile phone reveal about you? *International Journal of Information Security*, *10*(6), 337-349.

Greenberg, J. (2003). Metadata and the world wide web. *Encyclopedia of library and information science*, *3*, 1876-1888.

Harslem, E. E. and J. F. Heafner (1971). *A Request for Comment on Socket Name Structure*. Request For Comment 129.

Hillebrand, F. (Ed.) (2002). GSM and UMTS: The creation of global mobile communication, Indianapolis, IN: John Wiley.

Lee, C. P., Dourish, P., & Mark, G. (2006). The human infrastructure of cyberinfrastructure. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*: 483-492.

Lukasik, S.J. “Why the Arpanet Was Built.” Annals of the History of Computing, IEEE 33, no. 3 (March 2011): 4 – 21.

Mayer, J. & Mutchler, P. “MetaPhone: The Sensitivity of Telephone Metadata” *Web Policy* [BLOG]. Retrieved from http://webpolicy.org/2014/03/12/metaphone-the-sensitivity-of-telephone-metadata/

McKenzie, Alexander (1972). *Host/Host Protocol for the ARPA Network*.

McQuillan, John M., Ira Richer, and Eric C. Rosen. “An Overview of the New Routing Algorithm for the ARPANET.” In Proceedings of the Sixth Symposium on Data Communications, 63–68. SIGCOMM ’79. New York, NY, USA: ACM, 1979.

McQuillan, John M, and David C Walden. “The ARPA Network Design Decisions.” Computer Networks (1976) 1, no. 5 (August 1977): 243–89. doi:10.1016/0376-5075(77)90014-9.

National Information Standards Organization (2004). *Understanding metadata*. Bethesda, MD: NISO Press.

Pelkmans, J. (2001). "The GSM standard: explaining a success story." Journal of European Public Policy 8.3: 432-453.

Postel, Jon (1971). *A Preferred Official Initial Connection Protocol*. Request For Comments 165.

Ribes, D., and Polk, J. B. (2012). “Historical Ontology and Infrastructure,” *Proceedings of the 2012 iConference*: 254-262.

Russell, A. L. (2006). "Telecommunications Standards in the Second and Third Industrial Revolutions." *Journal of the Communications Network* 5.1: 100-106.

Russell, A. L. (2012). Histories of Networking vs. the History of the Internet. In *SIGCIS Workshop.*

Shilton, K. (2009). Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection. *Communications of the ACM*, *52*(11), 48-53.

Sterne, J. (2012). *MP3: the meaning of a format*. Duke University Press.

Traugott, S., & Huddleston, J. (1998, December). Bootstrapping an Infrastructure. In *LISA*: 181-196.

Traweek, S. (1992). Big science and colonialist discourse: Building high-energy physics in Japan. In *Big Science: The Growth of Large-scale Research:* 99 -126. Stanford University Press.

Willassen, S. (2003). Forensics and the GSM mobile telephone system. *International Journal of Digital Evidence*, *2*(1), 1-17.

Willassen, S. (2005). “Forensic analysis of mobile phone internal memory.” *Advances in Digital Forensics*: pp. 191-204.

Winett, Joel M. (1971). *The Definition of a Socket*. Request for Comments 147.