Internet Histories
Digital Technology, Culture and Society

ISSN: 2470-1475 (Print) 2470-1483 (Online) Journal homepage: http://www.tandfonline.com/loi/rint20

Notes from/dev/null

Finn Brunton

To cite this article: Finn Brunton (2017) Notes from/dev/null, Internet Histories, 1:1-2, 138-145

To link to this article: http://dx.doi.org/10.1080/24701475.2017.1307059

INTERNET HISTORIES, 2017
VOL. 1, NOS. 1–2, 138–145
https://doi.org/10.1080/24701475.2017.1307059

Routledge
Taylor & Francis Group

ORIGINAL ARTICLE

Notes from/dev/null

Finn Brunton

Department of Media, Culture, and Communication, New York University (Steinhardt School), New York, NY, United States

ABSTRACT

I will discuss the digital materials that we do not want to archive, or that do not want to be archived, that are particular to Internet history: the trash, cruft, detritus and intentionally opaque hoard of documents and artefacts that constitute our digital middens. Middens are pits of domestic refuse filled with the discards and by-products of material life: the gnawed bones, ashes, fruit stones and potsherds, shells and chips and hair and drippings which together constitute the photographic negative of a community in action and an invaluable record for archaeologists. Using this analogy, I will discuss two from my own research: the archives of spam, which we would all rather forget, and the records of the communities and marketplaces of the so-called "Dark Web," which would prefer to be forgotten. I will also address the challenges of research with other kinds of eccentric, troubling or speculative archives, like blockchains, ephemeral imageboards and doxxes. I will close by discussing ways that we can think of digital historiography, in particular, in terms of these accidental, unwanted, averse archives.

ARTICLE HISTORY
Received 28 November 2016
Accepted 13 March 2017

KEYWORDS
Internet history; digital archives; spam; Dark Web; blockchains

Introduction: strange archives

This is a paper about strange archives: archives that want to disappear, archives that conceal other archives within them, archives that are produced as side effects and accidents and by-products, archives of rejected material, archives that are trying to look like other things, archives that are the negative space of archives that no longer exist, archives that do not permit themselves to be read or try to refuse that permission. The goal of this paper is to itemise and describe some of these strange archives, many of which are particular to the work of studying the history of networked digital media, and to consider questions they raise for us about accounting for and making use of their strangeness.

I would like to put three of these questions to you, each one framed by a section of this paper. There are many more questions, of course, some of which I will bring up in passing below, but these are the three I want to ask in the context of the first issue of *Internet Histories:*

- What can we do with these new "archives" of Internet history that we could not do with prior archival forms?
- What tools, approaches and ways of thinking are best suited to making use of them?

CONTACT Finn Brunton fb42@nyu.edu

– Can we produce typologies, common sets of themes and shared problems among these disparate collections – and what else is out there?

First, to think about the properties of these strange archives of networked computing, I will put them in the context of the richness and complexity of our archival situation in Internet history, and the historiographic approaches that richness makes possible. Second, I will draw on the work that brought me to some strange archival spaces in the history of the Internet to present a way of exploring and using the "middens" I found there: accidental archives, collections of digital rubbish. Finally, I will lay out the start of a list of our strange archives in Internet history and outline their different properties for your consideration. I would like to begin with a seemingly simple, deeply vexing question: When does the history of the Internet start?

Part 1: beginning of the Internet age

The Internet age began behind a biker bar, the Alpine Inn Beer Garden in Portola Valley, about a half-hour south by the freeway from Menlo Park. A small plaque announces this, tucked away among the photos of Little League teams, next to a dartboard and across from a pinball machine ("Attack from Mars"): "BEGINNING OF THE INTERNET AGE," it says. It summarises the trip taken by the SRI International van on 27 August 1976: an interlink between two different networks, from a computer set up on a picnic table in Portola Valley by radio ("PRnet") to the Advanced Research Projects Agency Network (ARPANET), including Bolt, Beranek and Newman (BBN) in Boston and the University of Southern California in Los Angeles as endpoints, using the Transmission Control Protocol and Internet Protocol (TCP/IP). "This event marked the beginning of the Internet Age," says the plaque.

Is that true? It depends on your analysis – on what different historians and students of the history see as the most salient factors with the greatest explanatory power. In some ways, the really significant demonstration of internetworking came in November of the following year: a three-network connection over radio, ARPANET and satellite. The protocol in use in both tests was not what it would become when it was more widely adopted as Internet Protocol version 4 (IPv4), formally specified in a document in September 1981 (which supersedes a document from January 1980) (RFC 760, 1980; RFC 791, 1981). We could date the BEGINNING OF THE INTERNET AGE from when the new protocol was first put into general use: the cutover "flag day" of 1 January 1983, when all the nodes on ARPANET were obligated to switch their systems from the old protocol to TCP/IP, which is when this set of Internet standards became, well, standard.

Or we can push the history farther back. The standard did not materialise in the SRI parking lot in the mid-1970s, after all. We can trace the documents and collaborations back through the Network Working Group and the Request for Comments that introduces the term "internet" (at first a technical adjective, not a noun – "an internet system," as we might say "a robust system") in December 1974 (RFC 675, 1974). But there are still earlier systems which directly influenced the group, like CYCLADES in France and the work of Donald Davies in the UK, so why not start with those? We could write the history of *preconditions* – all those elements which had to be in place for the Internet to happen when and as it did. This could be a study of previous infrastructural layers and ideas that shaped the system (like Tung-Hui Hu's *A Prehistory of the Cloud*) or of the larger social and

economic pressures – the market needs and demands – that set the terrain for the event (like James Beniger's *The Control Revolution*). Indeed, it could be something akin to an indirect history of the US military during the Cold War, with the funding it provided and the institutional challenges it presented: robust and resilient command-and-control that could run by wire or radio, linking airborne and shipboard computers with satellites, radar bases and the buried lines running into Cheyenne Mountain.

I give this litany not only to lay out the richness of Internet historiography, but to indirectly show the richness of our archives – of a piece with the archival richness of studying computing and digital technology generally. We have so many documents: formal papers and publications, memoranda, theses and dissertations, brochures and manuals and patent filings, oral histories, materials for salespeople, plus a galaxy of both hobbyist and professional magazines and journals. We can reconstruct events cinematically and visually, with an astonishing amount of video, ranging from contemporary documentaries (like 1972s *Computer Networks: The Heralds of Resource Sharing,* with Licklider, Kahn and Donald Davies – among others! – at fever pitch) to demos, to records of the software itself in operation. Even in cases not as thoroughly documented, we often have images, photos and screenshots. Finally, of course, we have the technology. We can often tinker with the machines themselves – and, of course, for huge variety of instances, we can emulate and explore the software, read the code and documentation and even reconstruct and run period hardware: it is possible to simulate a network of interface message processors (IMPs) from 1973 (a few years before the drive to Portola Valley) and tune their operation, which is like being able to book a direct flight to Jurassic Park (The IMP Guys, 2013).

There are many angles to assembling Internet histories that are missing from my list of approaches so far, including various kinds of *counter-histories*. We have to consider the space of alternatives against which the project could define itself and come to be defined – from protocols like Open Systems Interconnection (OSI) to other networks entirely, like Usenet, Bulletin Board Systems or Minitel, and networks that emerged in different contexts and did not expand beyond them, like the Soviet All-State Automated System of Management (Driscoll, 2014; Peters, 2016). We must account for the unrealised systems that informed aspects or fantasies of the project: Engelbart's oN-Line System, Nelson's Xanadu or clashing visions of remote time-sharing terminals or personal computers. We need to explore the negative spaces, the failures and dead-ends and missing, excluded, ignored or under-discussed pieces of the network's history that also had their role to play, shaping the whole arrangement like the shadowy planet in the three-body problem, which you can only observe by variations in other orbits. I crash-landed on one of these sunless worlds in 2007 and stayed there for six years.

Part II: = (REAL BANK LOGINS SPAM SUPPLYS) =

During my time excavating the world of garbage that became the book *Spam: A Shadow History of the Internet,* I became convinced of two things. First, that the history of "spam" online, in all its various meanings, was an important component of what the network became – affecting search engines, walled garden systems, email, encryption, distributed denial of service (DDoS) attacks, advertising models, specialised artificial intelligence, the politics of identity, you name it. Second, that it was very hard to properly tell that history, because the very technologies spam had

shaped, had been built to eradicate it. It was a chronicle of notes from /dev/null, the device to which you direct data you want to never be written. For stretches of the period I was covering, I was essentially working on a Gnostic history. Much of what we know about the Gnostics comes through documents created by Church Fathers, theologians and Christian historians devoted to their eradication. Likewise, new spam techniques and technologies would become apparent through their citation and forwarding and copy-pasting by the people trying to evolve policy and software and law enforcement tools to make them stop. Many of the most technically interesting varieties of spam were notable for finding ways to render themselves invisible, unsearchable, un-indexable – dictionary attacks of phrasal salad, with images and HTML tricks to conceal themselves from documentation by their adversaries. They had an intriguing temporality, like *events* rather than objects, meant to get some action and disappear. The most prominent, documented parts of this history tended therefore to be especially memorable, peculiar or egregious examples – circulated, shared and referenced – a trophy case of monsters, no more representative of the larger phenomenon than the wall of a hunting lodge is of the local fauna.

The best things I found for reconstructing this history were *dumps*, literally and figuratively. A key archive for me was the Enron email corpus, a set of about 650,000 emails collected by the Federal Energy Regulatory Commission during their investigation of the company's fraud and market manipulation, which was dumped online in 2003 in a massive tranche of internal corporate media including trading floor phone calls and scanned documents. This collection of messages became one of the benchmark tools for training anti-spam systems – that is, for trying to determine the properties of "normal email." It was a flash-frozen Pompeii of email culture, and it had also become an important scientific artefact, used for many different training and machine learning systems like spam filtering. But it was also the story of Gerald and Lisa, two Enron employees whose disintegrating marriage was documented by their intra-office email exchanges – which introduces the theme of accidental inclusions in the accidental archive of the dump.

That was just the prelude to the archive of Premier Services, a mid-level spam company hacked by someone who called themselves "The Man in the Wilderness." Wilderness Man had a vendetta against Premier, so she, he or they captured several megabytes of data and dumped it all online: chat logs, screenshots of software in operation, budgets, the ins and outs of running the business. It was invaluable for understanding exactly how the spam business worked at that time, after the Web and AOL but before successful widespread automated spam filters. The hack, and the dump, was also a personal attack, including private photographs of the company's founder and employees and plenty of personally identifying information. (Pictures of the founder were used for proto-meme purposes, with mocking text added.)

I took to thinking of these and other haphazard, accreted digital archives as *middens*, the accumulations of by-products and junk and trash and bits and pieces of the working life of computers and communities. It seems like a useful metaphor for this kind of scholarly object. A midden is an archaeological site, the waste dump generated by people: discarded shells (shell middens are a major resource for the study of maritime cultures), potsherds and stone chips ("debitage," from fashioning tools), excrement, husks, peels and rinds, feathers, hair and ashes. When addressing the future, you deliberately leave etched runes, standing stones, red ochre, ziggurats,

grave goods – but, just as important, you also accidentally leave the trash record of how you lived and how you got by, the document of manufacture, preparation, consumption, exclusion and waste. The production of what I think of as informational middens is peculiar to digital technologies, but not unique: textual and archival packrats have always existed, bless them, but the ease of digital information storage and transmission has produced many more of these dumps, with a huge variety of archival information within which a stratum of spam could be found – among many other things. Like middens, these dumps were often accidental, side effects: digging around for material meant to be discarded, material designed to be difficult to observe, index or search, was most rewarding when the "archive" had been created by intentions other than preservation alone, like revenge or legal discovery.

There was a related challenge: to document middens-in-process, as it were, with the data of current working communities and marketplaces, for whom any record was an accidental by-product of their activities. They had no desire to be observed. To understand the spam business, it was necessary to observe the credit card scammers, identity thieves and teams selling slices of botnet capacity to send email in million-message batches. Their Internet Relay Chat (IRC) channels and discussion boards were abuzz with people and bots making sales pitches and doing deals: " = (REAL BANK LOGINS SPAM SUPPLYS) = (SELL BANK LOGINS\PRICE DEPENDS ON BALANCE 10% FROM IT) = (BIG BASE!) = ." They maintained their own systems of records – some users got a +v, "verified," known merchants with good histories who would not rip you off (probably) – but kept no histories, no findable archives. They had to be documented in action.

Spam was an enormous mass of material for which no one wanted to keep records (except email filter developers and the Federal Trade Commission); around it were communities and groups who did not want their records to be kept. Spam had the invisibility of refuse, and spammers and their ancillary businesses had the opacity of deliberate secrecy. Given very different causes and effects, this shares an interesting family resemblance with other major motors of Internet adoption and use, like pornography, online gambling and file sharing – and the drug dens, pre-loaded debit card vendors, covert media libraries and scam sites (hire an assassin!) of the so-called Dark Web of onion sites, neither indexed nor cached nor added to the Internet Archive and generally short-lived indeed.

I approached these projects of accumulation, documentation, mirroring and digging from the perspective of trying to make them into traditionally viable archives, time-stamped and organised. The question I would like to raise now is whether there are other aspects of these middens – things particular to them, properties they have, as the dumps of data in which material meant to vanish is captured – that are worth thinking about. With that question in mind – how to understand a mirror of acropol4ti6ytzeh.onion as a historical record? – I want to turn to our final question. What other kinds of archives are particular and significant (if not always unique) to the work of Internet history? What are their distinctive properties?

Part III: password is lol

There are the archives of hacks and hacking. Not just the materials that the hackers produce, but the processes and media generated *by* them: the Pastebin announcements, the

communiqués and ransom demands and launch screens that tell a company they have been compromised. Consider the pictures – pictures, really, since the computer is now locked, so instead of a screenshot we have an image, taken by a phone, of the screen – of the menacing skeleton announcing the "Guardians Of Peace" hack of Sony Pictures, or the exploding-head manifesto of the Impact Team takeover of Ashley Madison. There are ransomware lock windows, which – among other things – document the adoption of Bitcoin as the extortionist's currency of choice. And there are the archives of the aftermath, from the blackmail messages received by email addresses in the Ashley Madison database, to documents like Andrew "weev" Arnheimer's "Open letter to federal scum," a Pastebin text in which he demands recompense (in Bitcoin, natch) for his prison time, and that he is planning to use the Ashley Madison dump to expose the attorneys who put him away. Finally, there is the *context* of hacks, like the associated Twitter accounts in which hackers address fans, clients and adversaries alike. Consider two tweets from @DotGovs, whose username is "penis" and icon is the head of Buzz Lightyear from the *Toy Story* movies. ("Whose": it is actually a team at work, or so one person claimed; details remain murky.) One tweet immediately follows the other, both on 8 February 2016:

> "watching keeping up with the kardashians"
> "20,000 FBI EMPLOYEES NAMES, TITLES, PHONE NUMBERS, EMAILS, COUNTRY cryptobin.org/78u0h164 password is lol #FreePalestine"

I believe that being able to entirely explain all the details, references and context of these two tweets alone in a few decades will provide a superb window on the Internet history of our time. (The account has been suspended, of course, and is no longer available in official form.) Dumps like the hack of Premier Services are proto-doxxes, but doxxing is now a common strategy: how should we think about and approach these weaponised archives?

There are the archives of the ephemeral and anonymous: words and posts meant to disappear or become unavailable, deliberately or by negligence, and never to connect with any specific identity. Think of 4chan and its many imitators and knock-offs, and the rush to realise apps and platforms for self-destructing or anonymous communication (or, at least, the promise of it). Amateurs and professionals alike have been collecting and saving as much of this kind of material as possible – it is attractive in its very challenge to preservation – but beyond the difficulty of capture, it is interesting to understand an archive of material meant to vanish, or to conceal those using it. I tend to reflexively think of dissident and cryptographic groups in this light, but what about a history of the Low-Orbit Ion Cannon? It was software that exploited a structural flaw in the Internet (the denial of service attack), which was shared and applied by groups of volunteers seeking to aid Anonymous – and it leaked data about its users. Built for ephemerality – a DDoS attack is the flash-mob sit-in of digital activist tactics – and anonymity, it left behind a cache of accidental records that are valuable for themselves as well as for what they reveal.

What about the archives of materials that hope to go unnoticed? There are the detritus of hoaxes, frauds and scams – things which are meant to look like other things, and which seek, above all, to pass without awareness of what makes them significant: phishing messages, faux URLs, ersatz landing pages, alerts from cryptocurrency wallets to trick you into giving up your private key. There are archives of new kinds of marketplaces which hope to draw only the most marginal notice, because they thrive in the corners of far larger platforms which may wish to present themselves differently. There are, or were, Egyptian

shepherds and butchers and Kuwaiti dealers in manga and anime using Instagram as a storefront, and term-paper-writing businesses and gun dealers on Facebook, communities to themselves seeking to avoid wider recognition to dodge moderation or deletion. (Even black markets had subgroups, like the assisted suicide community on the Silk Road, shipping Nembutal amidst all the various white/brown powders and sheets of acid.) A far larger case along these lines is the ad tech and online ad marketplace – not covert so much as *obscure*, the vast bazaar of microsecond analysis, bundling and auctions of browsing data to serve ads that shape so much of the experience of "content" on the web. And there are the archives of attempts to make things disappear. Deindexing, de-listing, burying, deleting, which sometimes leave their own records in the negative space of broken links or create a new kind of archival garbage in the obfuscating materials generated to bury search results: "to expedite the eradication of references to the pepper spray incident," as the "Brand and Reputation Enhancement" memo for the Chancellor of UC Davis puts it. (Stanton & Lambert, 2016; Helen Nissenbaum and I discussed obfuscating data in *Obfuscation*.)

Finally, on a completely different note, it is well worth thinking about new systems like blockchains as another class of novel archives: they are built to act as a shared, collective timestamping mechanism for their contents, unchangeable save by linear additions, an endless list churned out by a system that rewards its members for verifying past details. It is a set of chronicles (in the classic, historiographic sense of a chronologically organised and notated record of events) each of which contains in hashed form, the material that precedes it. Blockchains are rife with metadata and odd, encoded details – the Bitcoin blockchain, in particular, was quickly adopted by people using transactions to store other kinds of data in this accidental, distributed, resilient archive. Indeed, Brewster Kahle (among others) has proposed the broader application of blockchains to make the Internet itself its own backup system (2015).

From a specific event where we (but not our subject) began – with a van in Portola Valley in the afternoon in 1976 that is also, in a sense, on the Internet – to gigabytes of images and accidental caches of millions of messages accrued over decades, to secret marketplaces and timestamped public archives: over all of these hover other questions. What is the nature of the digital archive in particular? (We could start with Matthew Kirschenbaum, with Lori Emerson, with Lisa Gitelman….) How – practically and theoretically – to manage the preservational excess, the terabyte burdens placed on the archival community? (As I write this, the rush is on to protect enormous masses of data endangered by the new Republican administration in the United States – a story familiar from the scale of "data friction" discussed by Paul Edwards.) But, for this first issue of *Internet Histories*, I would like to close with the simple, complex question with which I opened my outline of strange archives: When does Internet history begin?

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributor

Finn Brunton is an assistant professor in the Department of Media, Culture, and Communication at NYU's Steinhardt School. He is the author of *Spam: A Shadow History of the Internet* (MIT Press, 2013)

and, with Helen Nissenbaum, *Obfuscation: A User's Guide for Privacy and Protest* (MIT Press, 2015), and numerous articles and chapters for venues including *Limn, Radical Philosophy, Representations, First Monday, The MoneyLab Reader, Le Monde diplomatique, Payment Objects*, the *Guardian* and *Artforum*. His research focuses on the history and theory of computing and digital media, with a focus on hacking, privacy, cryptocurrencies and other forms of digital money.

References

Beniger, J. (1989). *The control revolution: Technological and economic origins of the information society*. Cambridge, MA: Harvard University Press.

Brunton, F. (2013). *Spam: A shadow history of the internet*. Cambridge, MA: MIT Press.

Brunton, F., & Nissenbaum, H. (2015). *Obfuscation: A user's guide for privacy and protest*. Cambridge, MA: MIT Press.

Driscoll, K. (2014). *Hobbyist inter-networking and the popular internet imaginary: Forgotten histories of networked personal computing, 1978-1998* (dissertation). Los Angeles, CA: University of Southern California.

Hu, T.-H. (2015). *A prehistory of the cloud*. Cambridge, MA: MIT Press.

Kahle, B. (2015). Locking the web open: A call for a distributed web [Blog post]. Retrieved August 11, 2015, from http://brewster.kahle.org/2015/08/11/locking-the-web-open-a-call-for-a-distributed-web-2/.

Peters, B. (2016). *How not to network a nation: The uneasy history of the soviet internet*. Cambridge, MA: MIT Press.

RFC 675. (1974). Specification of internet transmission control program. Retrieved from https://tools.ietf.org/html/rfc675.

RFC 760. (1980). DoD standard internet protocol. Retrieved from https://tools.ietf.org/html/rfc760.

RFC 791. (1981). Internet protocol: DARPA internet program protocol specification. Retrieved from https://tools.ietf.org/html/rfc791.

Stanton, S., & Lambert, D. (2016, April 13). UC Davis spent thousands to scrub pepper-spray references from internet. *The Sacramento Bee*.

The IMP Guys. (2013). The ARPANET IMP program: Retrospective and resurrection (draft of December 2, 2013). Retrieved from http://walden-family.com/bbn/imp-code.pdf.