ASKING THE ORACLE

Kate Crawford

Multispectral satellite image of Delphi and southern Greece

Know Thyself: *gnēthi seauton*
Nothing in Excess: *meden agan*
A Pledge, and Ruin Is Near: *eggua para d'atē*
—The Delphic precepts

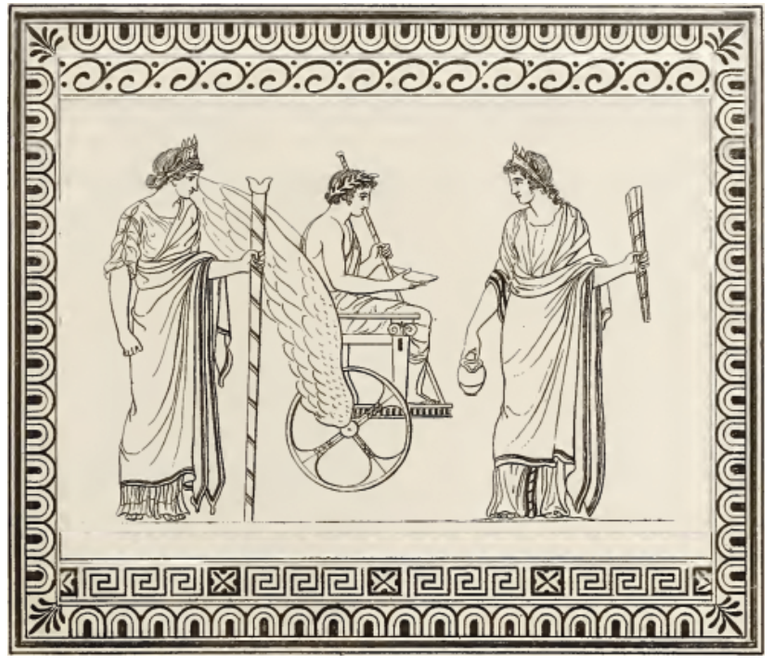
I SIT AT THE DESK and look at the screen. A software program normally used for digital forensics is open before me. This database contains the Snowden archive: all the documents, PowerPoint presentations, internal memos, newsletters, and technical manuals that Edward Snowden leaked in 2013. Well over a decade of intelligence thinking and communication is here, from within the National Security Agency in the United States and the Government Communication Headquarters in the United Kingdom, and reaching out into the international network of the Five Eyes. The immense collection of material captures the era when mass collection metastasized: the black world's gradual evolution of many of the techniques and approaches that we now call "big data." This knowledge is normally off limits, part of a "classified empire" once estimated to be growing five times faster than the public storehouse of knowledge.[1]

For many reasons, this database is a machine for producing anxiety. The interface is query driven: it centers on a search box. Like a highly classified version of Google or Reddit's Ask Me Anything, the only way to make discoveries is to throw some terms out there. There is no easy browsing or pretty visualization of the repository itself. Instead, you must begin by phrasing any questions you have about the military-intelligence complex in the form of Boolean search terms. So I type in words. It feels exhilarating, terrifying: Where to begin? How about something about cryptographic techniques, specific algorithms, or existing NSA programs? Thousands of search results. Days of work just to figure out what will be relevant. Instead, I try increasingly idiosyncratic, unusual combinations. Even then, the result is often dozens of documents, each with its own suggestive paths to follow. This is just the first challenge of the archive.

For the ancient Greeks, Delphi was the center of the world. It was the home of the Oracle, and she possessed the power to explain the present and see into the future. Only a select few could ask her a

question, and for them the Oracle offered a comms channel to the god Apollo. She predicted military attacks, saw into family tragedies, knew when kings would die. The Oracle became a serious force in politics and culture and remained so for centuries. According to archeologists, oracles were based at Delphi from the eighth to the second century BCE, before the temple was destroyed in CE 390.[2] For an information technology, this is an exceptionally long lifespan. And this is due not to simplicity of form or function. The Oracle of Delphi was a complex assemblage of parts that required much maintenance. There was the Oracle herself, known as Pythia, a priestess chosen for her exemplary life, who channeled Apollo's wisdom while in a trancelike state. Priests in turn transcribed her words into poetic hexameters. Then there's the expansive temple, which was located near a chasm that may have been issuing forth gas clouds rich in the intoxicant ethylene. Out of this mixture of elements, rich prophecies would emerge. Some concerned high matters of state: when to go to war, colonize a new city, give pardon or punishment. Others were strictly personal. Over time, this system became a vital part of Greek society. The Oracle knew all the secrets.

Like the divinations of the Oracle, the problem with the Snowden archive is that you never find an easy answer. Documents lead to other documents, one NSA program will point to another. Code names obscure specific companies and technological capacities. Some of these can be unlocked, but it's obsessive, painstaking work. Remember that *New York Times* and ProPublica investigation that finally revealed the identities behind the "Fairview" codename (AT&T's partner program with the NSA) and "Stormbrew" (Verizon's program)?[3] That was the result of multiple journalists working relentlessly for months. There's a reason so many of the articles about Snowden's archive have shared bylines: it takes a complex combination of skills to reverse-engineer just what all the terms mean, let alone how they work. And it all begins with typing questions into little boxes. Then doing it again. Above all, what you find are more questions. Over time, I've come to think of it as a contemporary experience of the Delphic Oracle: you ask something and receive cryptic information that may offer you some answers, but only by raising more questions.



Apollo at Delphi with priestess and unnamed queen. Illustration by Thomas Kirk, from *Outlines from the Figures and Compositions upon the Greek, Roman, and Etruscan Vases of the Late Sir William Hamilton* (London, 1804).

The Snowden documents may be vast in number, but they also have strange consistencies. First, there is the style: every government agency and consulting firm has its own way of laying out documents, a mode of conveying information that is distinctly its own. The NSA house style is most viscerally conveyed in PowerPoint: you are first struck by the banality of bullet points, drop shadows, and Clipart before you are blown away by the magnitude of the information being conveyed. Wizards, crystal balls, four-leaf clovers, alchemists—once objects of power, magic, and prophecy, now rendered completely inert in a wasteland of Comic Sans and WordArt. There's a deep cognitive dissonance in reading business-convention slide decks that look laughable while they simultaneously outline a colossal, terrifying surveillance infrastructure. You get over that. But the sensation of

shaping search queries for an enigmatic system that can tell you the secrets of the classified world? This remains with you. It feels something like vertigo.

If you were given an audience with the Delphic Oracle you could expect an oblique response. Raw Delphic data is meaningless without the work of interpretation. When Lysander, the warrior who won the Peloponnesian War, visited the oracle in 403 BCE, he was told to beware "the dragon, earthborn, in craftiness coming behind thee." Eight years later, he was stabbed from behind by a man with a serpent on his shield.[4] Socrates, of whom the Oracle once said, "No one is wiser," understood her words as a paradox. Indeed, many of the Oracle's responses took the form of epistemic paradoxes—riddles that highlight inconsistencies in models of knowledge while casting light on a common error or misconception. The Oracle's role wasn't just to predict a possible future but to show the fallacies of the present. As a system of information it skewed toward difficult forms of data, often accenting the flaws and limitations of the supplicant.

Embedded in the architecture of the temple were messages counseling restraint. Anyone who entered the temple would face the maxims of Delphi, carved in stone: KNOW THYSELF, NOTHING IN EXCESS, and A PLEDGE, AND RUIN IS NEAR. The first of these, Know Thyself, is perhaps the best known, but the precept has always been double edged. To modern ears it sounds like a prescription for self-knowledge, but as historians and philosophers have argued, this was not its original meaning. It advised knowing one's limits in seeking data: in short, "Don't ask too many questions." In this way, all the Delphic precepts were instructions to be cautious and stay within bounds:

> When you question the oracle, examine yourself closely and the questions you are going to ask, those you wish to ask, and, since you must restrict yourself to the fewest questions and not ask too many, carefully consider yourself and what you need to know.[5]

So the Oracle, as a technology, set up particular restrictions and limitations. The information flow was restricted by the number of people who could visit the Oracle, by how many questions they could ask, and by the cryptic nature of the responses they received. In this sense there is a strange similarity with the Snowden archive. The person seated before the search box must decide what to ask next and try to exercise restraint so as not to be drawn into thousands of documents and stories and systems. But in another sense, when analysts consult the database inside the fortresses of the NSA and the GCHQ, there seems to be little respect for limits beyond the strictures of policy. Everything that can be captured will be. The archive is an epic testament to information acquisition, overreach, and confidence. It's as though the guiding principles of Delphi were reversed. Know Everyone. Everything in Excess. Just keep pledging that all the necessary protections are in place.

Know Thyself

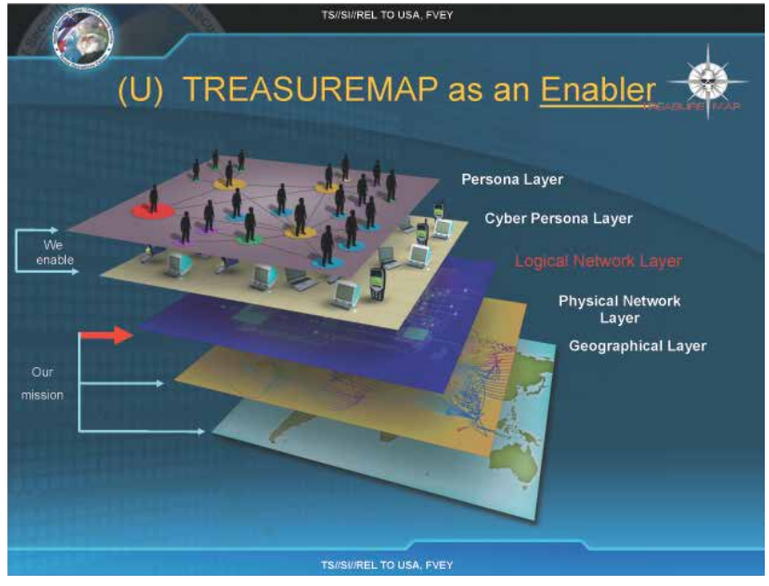
The archive is the ultimate rabbit hole. Days can pass without stopping, without eating: I barely rise from the desk. It feels like I can see into the complete structure of a global system, spread out before me in neat network diagrams. This, of course, is an illusion. The archive is always partial—necessarily incomplete, truncated on the day when Snowden took copies of the documents. But it is also fractured and dispersed by the operating procedures of the NSA and GCHQ themselves. Intelligence work has to be compartmentalized: the work of one department is kept separate from another, and these divisions are reflected in the collection of documents. There are blank spaces, dead ends, and missing parts.

That said, the documents nonetheless offer extraordinary coverage of the core period of the expansion of big data techniques during the early 2000s, up to 2012. Tens of thousands of memos, internal newsletters, and specific investigations. And, of course, the PowerPoint presentations. These have been the preferred documents for most journalistic reporting because they are designed to be dramatic. The PowerPoint decks seek to convey the sheer force of the surveillance systems as simply as possible in order to impress senior military figures, convince analysts at annual conferences, and secure ongoing funding from politicians. Treasure Map, for example, makes for a jaw-dropping presentation—the program builds an almost real-time interactive map of the Internet.[6] That's everything, including the location and owner

of any connected computer, mobile, or router, even those we imagine to be safely located behind private networks or obscured by dynamic routing tables. "Map the entire internet—any device, anywhere, all the time," a slide boasts. A few slides on, "TREASUREMAP as an Enabler" offers up a layer-cake image of signals analysis. Above the geographical layer and the network layer is the "Cyber Persona Layer"—quaintly represented on the slide by jellybean-era iMacs and Nokia feature phones—and then the "Persona Layer." That's us—everyone across the world. We are represented as Clipart men standing on dots, connected by radiating lines in the style of a social graph. Our names, our homes, our viewing histories: all easily searchable ("near real time!") and ready for analysis. It is called a "300,000 foot view of the Internet."[7]

But should we even believe this presentation? It reads a lot like a sales pitch, and there are real infrastructural limits to the claims. There are also deep institutional reasons why agencies would add bravado to their decks: there's internal competition to be seen to be producing the most sophisticated attacks, with funding and institutional support hinging on that perception. But even when the technical claims are exaggerated, there's still much to be learned in these depictions. At every level, the documents instruct us about the extent of the aspirations for big data surveillance, the power that they wish to secure, and the world they want to build.

Like *Powers of Ten* (1977), the experimental film by Charles and Ray Eames that telescopes through views of the universe, the Snowden archive produces dizzying leaps in perspective of both time and space for the reader. It goes big picture, then comes in close. If Treasure Map is the God's-eye view for the NSA, then FoxAcid is closer to home: the snitch in your system. "If we can get the target to visit us in some sort of browser, we can probably own them," a slide explains.[8] Once users have been induced to click on spam or visit a website, the NSA drops files through a browser that will live on in their system, quietly reporting everything one does back to base. On the descriptions go, in a casual, jocular style, with illustrations of foxes drowning in a barrel of acid, and another fox winking on a tin of Spam. One slide describes how analysts "deploy very targeted emails" that require "a level of guilty knowledge" about the target, a technique known to black-hat operators as "spear phishing." Guilty knowledge? Where, in this system, does *that* begin



Slide from NSA "Treasure Map" PowerPoint slide deck

and end? While there are some limits on how the data of American citizens can be collected and used, these are mentioned in the documents as uncomfortable restrictions. One PowerPoint notes that the NSA is working on multiple fronts to "aggressively pursue legal authorities and a policy framework mapped more fully to the information age."[9] Change the laws to fit the tools, not the other way around.

If we zoom back to a more proximate level of resolution, there am I, reading leaked, secret government documents on an air-gapped machine in an old warehouse building. I only read the documents at this location, and I can't copy them, so all research time requires arranging an in-person visit in order to ask further questions. My knowledge also feels guilty and full of risk. It is always being made clear that these documents are highly restricted. Just as the Delphic temple was inscribed with reminders of caution, each page of the Snowden archive is marked with a header noting different forms of classification. TOP SECRET//SI//ORCON//NOFORN. This knowledge has limited access. Only some may pass. Who are *you* to ask questions here?

Nothing in Excess

One day I come across a memo in the archive, drawn from the classified internal network of the Signals Intelligence Directorate. It describes the way analysts can suffer from being drawn into the data, unable to disengage or admit defeat. The author details how mountaineers who wish to summit Everest train for years, becoming obsessed with their goal. But this deep sense of investment also puts them at grave risk: they will push ahead with a dangerous climb despite signs of danger.

Mountaineers call this phenomenon "summit fever"—when an "individual becomes so fixated on reaching the summit that all else fades from consciousness." I think part of this phenomenon is due to the high level of investment (monetary and spiritual) in the project that pushes people to make decisions that are not otherwise supported by objective data:

> I believe that SIGINTers, like the world-class climbers, are not immune to summit fever. It's easy enough to lose sight of the bad weather and push on relentlessly, especially after pouring lots of money, time, and resources into something. From turning off a database or collection site to starting over from scratch on a target set or software code, it's difficult to let go of the dream and your work so far.[10]

There are many symptoms of "summit fever" in the documents. Sometimes it's an offhand remark, as when one analyst jokes about the desire of analysts to collect it all: "Dude! Map all the networks!!!" Other times it's couched in more serious or legalistic terms, in the need to keep acquiring ever more data and greater permissions. One paper in the archive blandly speaks of the NSA's aims to expand all its "capabilities to reach previously inaccessible targets in support of exploitation, cyberdefense and cyberoperations." It is a project with no end. There is no letting go of the dream of perfect information.

This voracious appetite for gathering and connecting information was, in part, enabled and sanctioned by the 9/11 Commission. A key recommendation of the Commission's report was to improve "Information Sharing and Fusion": "The president should lead the government-wide effort to bring the major national security institutions into the information revolution. He should coordinate the resolution of the legal, policy, and technical issues across agencies to create a 'trusted information network.'"[11] But as I've written elsewhere, as the scale of the intelligence network multiplied, so did the anxiety over missing crucial data or not seeing the right connections.[12] It became a cultural imperative that "more data is better," even as analysts were drowning in information. "We in the agency are at risk of a similar, collective paralysis in the face of a dizzying array of choices every single day," an NSA analyst wrote in a memo in 2011.

> "Analysis paralysis" isn't only a cute rhyme. It's the term for what happens when you spend so much time analyzing a situation that you ultimately stymie any outcome. . . . It's what happens in SIGINT when we have access to endless possibilities, but we struggle to prioritize, narrow, and exploit the best ones.[13]

Just as this phenomenon afflicts intelligence analysts, a related sensation comes with reading the Snowden documents. Although it lacks all the "near real-time" search capacity of the systems used by the agencies themselves, the Snowden cache offers the seduction of the archival search, a sensation well known to the investigator, the detective, and the historian. An analyst pores over the data to track a person of interest, but a reader of the Snowden archive shapes search queries in order to piece together a *practice*: how do these surveillance programs work, what technological capacities are in play, what are the broader legal, cultural, political ramifications? It is about tracking a system instead of catching a suspect. If there is a similarity here, it is in the obsessive focus on the data, intently scrutinizing the databases in order to find new meanings and connections and losing a sense of boundaries.

Hence the susceptibility to excess. There are, of course, endless connections and interpretations to be made in any massive archive of data, and the Snowden database is a particularly significant one. What analysts call summit fever, philosophers have described as archive fever: "It is to burn with a passion," Jacques Derrida writes. "It is never to rest,

interminably, from searching for the right archive even as it slips away. It is to run after the archive, even if there's too much of it."[14]

The fantasy is that, if only you look long enough, you will find the truth. If you only had more time, you could find the single document that illuminates the whole collection. But archives are tricky beasts. Derrida argued that the archive produces as much as it records history: the way information is stored, accessed, and transmitted shapes the nature of the knowledge it offers. In other words, our understanding of the NSA is being shaped by the type of access Snowden had as a contractor, by the search interface on top of the database through which journalists and researchers access it, and by the ways newspapers report it. The design limitations of PowerPoint and HTML affect it, too. For example, one frustration of researching the Snowden archive is that the copies of internal web pages are riddled with broken links and inaccessible images. The data encoded in these interconnections are visibly absent. The particularities of hyperlinks and networked documents mean that they don't travel well, in contrast to the self-contained PDFs and training manuals, which work perfectly. As a result, the data they contain become the preferred raw material of history. The disparity between these data formats also serves as a reminder of the immense technical imbalance between the capacities of the agencies and the system containing the Snowden archive.

It's tempting to fetishize archives as providing an unfiltered access to a past reality. But as the historian Dominick LaCapra observes, it's a trap to mistake the archive as a "literal substitute for the 'reality' of the past . . . a stand-in for the past that brings the mystified experience of the thing itself."[15] More realistically, the archive can only ever be a very particular type of reconstruction, a keyhole view. It is not a window into the truth of things. Like the Oracle, it gives us coded answers, told through a technology that changes the very meaning of what is being transmitted.

This is not to say that the Snowden archive is anything less than an extraordinary account of the expansion of Western surveillance in the late twentieth and early twenty-first centuries. But we can only approach it through these attenuated channels: the Boolean search query, the unstructured data file, the document type.

Even so, can we ever truly understand it? Everything in the documents is in some form of code. There are many thousands of code names for programs, for technical capacities, for NSA partner organizations. Beyond this, there are the professional codes of tradecraft, hidden in plain language but indecipherable to people not in the business. We can guess what is being hinted at, but spies have their own forms of speech—they know when a document is full of bluster and overreach or when something is being chillingly understated.

A Pledge, and Ruin Is Near

Just as the imperatives of Know Thyself and Nothing in Excess seem foreign in the context of the NSA and GCHQ, so does the final Delphic maxim: A Pledge, and Ruin Is Near. Translations of this phrase vary, but it means something like "When you consult the gods, do not make vows and commitments that you cannot honor."[16] The Greeks used this as a warning against making promises that would come back to haunt you.

Intelligence analysts that use these extraordinary systems of data harvesting and tracking are bound to follow the law as well as specific policies that are meant to restrict their access. Of course, it doesn't always work that way: analysts have been caught tracking ex-wives and potential love interests (agents call it LOVEINT), or they make errors that result in the wrong people or countries having their data harvested.[17] But even with some safeguards in place, the ruin may be inevitable. Law professor Paul Ohm has described the emergence of a "database of ruin" in the private sector as companies gather potentially devastating information about medical conditions and family histories and then combine their data stores. "Once we have created this database," he writes, "it is unlikely we will ever be able to tear it apart."[18] The NSA has built its own database of ruin, one that contains all those phone records, search histories, Skype calls, location data, network connections, chat logs. These are the phantom bodies of data that stand in for us.

When the first stories from the Snowden archive appeared in newspapers in mid-2013, intelligence agencies experienced a new kind of public scrutiny and pressure. Years later the most substantial legal reform has come through the USA Freedom Act, which ended the bulk collection of Americans' phone metadata records, although it leaves much Internet data collection untouched. That gigantic data set will

Ruins of the Temple of Apollo, Delphi

become a decaying monument to the period before Snowden, a ruin from an earlier time, as analysts come to rely on different tools. The Snowden database will also erode as a gauge of the technical capabilities of the NSA and GCHQ. But its power as a cultural, political, and historical archive will remain for as long we ask questions of it.

When the 9/11 Commission recommended a new era of information sharing, it described the need to "simultaneously empower and constrain officials, telling them clearly what is and is not permitted," because "the policy and legal issues are harder than the technical ones."[19] They remain the hardest problems. The temple at Delphi is a ruin, but the precepts of restraint have endured as long as any of the Oracle's prophecies. The lasting cautions for the era of global surveillance are still being learned.

NOTES

Where possible, I have provided references to published texts using the Snowden documents, or containing similar information.

1. Peter Galison, "Removing Knowledge," *Critical Inquiry* 31, no. 1 (2004), p. 229.
2. Hugh Bowden, *Classical Athens and the Delphic Oracle: Divination and Democracy* (Cambridge: Cambridge University Press, 2005), pp. 6–9.
3. Julia Angwin, Charlie Savage, Jeff Larson, Henrik Moltke, Laura Poitras, and James Risen, "AT&T Helped U.S. Spy on Internet on a Vast Scale," *New York Times*, August 15, 2015, http://www.nytimes.com/2015/08/16/us/politics/att-helped-nsa-spy-on-an-array-of-internet-traffic.html; and Jeff Larson and Julia Angwin, Henrik Moltke and Laura Poitras, "A Trail of Evidence Leading to AT&T's Partnership with the NSA," ProPublica.org, August 15, 2015, https://www.propublica.org/article/a-trail-of-evidence-leading-to-atts-partnership-with-the-nsa.
4. *Plutarch's Lives*, trans. Bernadotte Perrin (Cambridge, Mass.: Harvard University Press, 1916), p. 317.
5. Michel Foucault, *The Hermeneutics of the Subject: Lectures at the Collège de France*, vol. 6, *1981–1982*, ed. Frédéric Gros, trans. Graham Burchell (New York: Palgrave Macmillan, 2005), p. 4.
6. On Treasure Map, see James Risen and Laura Poitras, "N.S.A. Report Outlined Goals for More Power," *New York Times*, November 22, 2013, http://www.nytimes.com/2013/11/23/us/politics/nsa-report-outlined-goals-for-more-power.html; and Andy Müller-Maguhn, Laura Poitras, Marcel Rosenbach, Michael Sontheimer, and Christian Grothoff, "Treasure Map: The NSA Breach of Telekom and Other German Firms," *Der Spiegel*, September 14, 2014, http://www.spiegel.de/international/world/snowden-documents-indicate-nsa-has-breached-deutsche-telekom-a-991503.html.
7. Quoted in Risen and Poitras, "N.S.A. Report Outlined Goals for More Power."
8. Bruce Schneier, "Attacking Tor: How the NSA Targets Users' Online Anonymity," *The Guardian*, October 4, 2013, http://www.theguardian.com/world/2013/oct/04/tor-attacks-nsa-users-online-anonymity; source document available at: https://theintercept.com/document/2014/03/12/nsa-phishing-tactics-man-middle-attacks.
9. Quoted in Risen and Poitras, "N.S.A. Report Outlined Goals for More Power."
10. The full document was made available by Peter Maass on Documentcloud: "'Signal v. Noise' Column: Summit Fever," https://www.documentcloud.org/documents/2088979-summit-fever.html. It served as a source for his article "Inside NSA, Officials Privately Criticize 'Collect it All' Surveillance," *The Intercept*, May 28, 2015, https://theintercept.com/2015/05/28/nsa-officials-privately-criticize-collect-it-all-surveillance.
11. Government Accountability Office, *Summary of Recommendations—The 9/11 Commission Report*, doc. no. B-303692, September 9, 2004 (Washington, DC), p. 33, http://www.gao.gov/decisions/other/303692.pdf; and *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks upon the United States* (Washington, D.C.: United States Government Printing Office, 2004), p. 418, http://www.9-11commission.gov/report/911Report.pdf.
12. Kate Crawford, "The Anxieties of Big Data," *New Inquiry*, May 30, 2014, http://thenewinquiry.com/essays/the-anxieties-of-big-data.
13. Quoted in Maass, "Inside NSA, Officials Privately Criticize 'Collect It All' Surveillance."
14. Jacques Derrida, *Archive Fever: A Freudian Impression* (Chicago: University of Chicago, 1996), p. 91.
15. Dominick LaCapra, *Rethinking Intellectual History: Texts, Contexts, Language* (Ithaca, N.Y.: Cornell University Press, 1983) p. 344.
16. Foucault, *Hermeneutics of the Subject*.
17. Andrea Peterson, "LOVEINT: When the NSA Officers Use Their Spying Power on Love Interests," *Washington Post*, August 24, 2013, https://www.washingtonpost.com/blogs/the-switch/wp/2013/08/24/loveint-when-nsa-officers-use-their-spying-power-on-love-interests.
18. Paul Ohm, "Don't Build a Database of Ruin," *Harvard Business Review*, August 23, 2012, https://hbr.org/2012/08/dont-build-a-database-of-ruin.
19. *The 9/11 Commission Report*, p. 419.