Snowden Edward, *Permanent record,* Macmillan, 2019.

The report’s full classification designation was TOP SECRET//STLW//HCS/COMINT//ORCON/NOFORN, which translates to: pretty much only a few dozen people in the world are allowed to read this. I was most definitely not one of them. The report came to my attention by mistake: someone in the NSA IG’s office had left a draft copy on a system that I, as a sysadmin, had access to. Its caveat of STLW, which I didn’t recognize, turned out to be what’s called a “dirty word” on my system: a label signifying a document that wasn’t supposed to be stored on lower - security drives. These drives were being constantly checked for any newly appearing dirty words, and the moment one was found I was alerted so that I could decide how best to scrub the document from the system. But before I did, I’d have to examine the offending file myself, just to confirm that the dirty word search hadn’t flagged anything accidentally. Usually I’d take just the briefest glance at the thing. But this time, as soon I opened the document and read the title, I knew I’d be reading it all the way through.

 Here was everything that was missing from the unclassified version. Here was everything that the journalism I’d read had lacked, and that the court proceedings I’d followed had been denied: a complete accounting of the NSA’s most secret surveillance programs, and the agency directives and Department of Justice policies that had been used to subvert American law and contravene the US Constitution. After reading the thing, I could understand why no IC employee had ever leaked it to journalists, and no judge would be able to force the government to produce it in open court. The document was so deeply classified that anybody who had access to it who wasn’t a sysadmin would be immediately identifiable. And the activities it outlined were so deeply criminal that no government would ever allow it to be released unredacted.

[…] Whereas the unclassified version merely made reference to the NSA being ordered to intensify its intelligence - gathering practices following 9/11, the classified version laid out the nature, and scale, of that intensification. The NSA’s historic brief had been fundamentally altered from targeted collection of communications to “bulk collection,” which is the agency’s euphemism for mass surveillance. And whereas the unclassified version obfuscated this shift, advocating for expanded surveillance by scaring the public with the specter of terror, the classified version made this shift explicit, justifying it as the legitimate corollary of expanded technological capability.

 The NSA IG’s portion of the classified report outlined what it called “a collection gap,” noting that existing surveillance legislation (particularly the Foreign Intelligence Surveillance Act) dated from 1978, a time when most communications signals traveled via radio or telephone lines, rather than fiber - optic cables and satellites. In essence, the agency was arguing that the speed and volume of contemporary communication had outpaced, and outgrown, American law — no court, not even a secret court, could issue enough individually targeted warrants fast enough to keep up — and that a truly global world required a truly global intelligence agency. All of this pointed, in the NSA’s logic, to the necessity of the bulk collection of Internet communications.

 [..] THE TERM “ MASS surveillance” is more clear to me, and I think to most people, than the government’s preferred “bulk collection,” which to my mind threatens to give a falsely fuzzy impression of the agency’s work. “Bulk collection” makes it sound like a particularly busy post office or sanitation department, as opposed to a historic effort to achieve total access to — and clandestinely take possession of — the records of all digital communications in existence.

 But even once a common ground of terminology is established, misperceptions can still abound. Most people, even today, tend to think of mass surveillance in terms of content — the actual words they use when they make a phone call or write an email. When they find out that the government actually cares comparatively little about that content, they tend to care comparatively little about government surveillance. This relief is understandable, to a degree, due to what each of us must regard as the uniquely revealing and intimate nature of our communications: the sound of our voice, almost as personal as a thumbprint; the inimitable facial expression we put on in a selfie sent by text. The unfortunate truth, however, is that the content of our communications is rarely as revealing as its other elements — the unwritten, unspoken information that can expose the broader context and patterns of behavior.

 The NSA calls this “metadata.” The term’s prefix, “meta,” which traditionally is translated as “above” or “beyond,” is here used in the sense of “about”: metadata is data about data. It is, more accurately, data that is made by data — a cluster of tags and markers that allow data to be useful. The most direct way of thinking about metadata, however, is as “activity data,” all the records of all the things you do on your devices and all the things your devices do on their own. Take a phone call, for example: its metadata might include the date and time of the call, the call’s duration, the number from which the call was made, the number being called, and their locations. An email’s metadata might include information about what type of computer it was generated on, where, and when, who the computer belonged to, who sent the email, who received it, where and when it was sent and received, and who if anyone besides the sender and recipient accessed it, and where and when. Metadata can tell your surveillant the address you slept at last night and what time you got up this morning. It reveals every place you visited during your day and how long you spent there. It shows who you were in touch with and who was in touch with you.