

THE NSA'S 5-PAGE ENTIRELY REDACTED DEFINITION OF METADATA

In my post on Rosemary Collyer's shitty upstream 702 opinion, I noted that the only known (but entirely redacted) discussions of what constituted metadata were part of the 2004 and 2010 authorizations for the Internet dragnet.

The documents liberated by Charlie Savage (starting at PDF 184) reveal the topic was actually discussed during the resolution of the 2011 upstream fight. In response to a Bates question to "fully describe what constitutes 'metadata'" that can be extracted from Internet transactions, the government defined the term in a footnote that is substantially redacted.

9. ~~(TS//SI)~~ Section 3(b)(5)(b)(4) of the amended NSA minimization procedures allows NSA to use metadata extracted from Internet transactions without first determining whether the metadata was extracted from a wholly domestic communication or a non-target communication. Please fully describe what constitutes "metadata" within the meaning of this provision.

~~(TS//SI//NF)~~ Set forth below are descriptions of the categories and types of metadata¹ the Government extracts from Internet transactions, including single, discrete communications contained within a multi-communication Internet transaction (hereinafter, collectively, "Internet transactions"), acquired through NSA's Internet upstream collection techniques. As the precise metadata transmitted with each particular Internet transaction varies, each type or category of metadata listed below may not be available for extraction by NSA in every particular instance.

¹ ~~(TS//SI//NF)~~ For the purposes of this response, the term "metadata," when used in reference to an Internet transaction, is information about [REDACTED]

That discussion is followed by five entirely redacted pages describing the three (also entirely redacted) categories of metadata.

Metadata Category	Metadata Category Description
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]

So I apologize to the government for suggesting they've never defined the difference between content and metadata in the context of upstream

content collection (the discussion probably closely follows the Internet dragnet discussion, which Bates had had with the government roughly 18 months earlier; that discussion allowed some dialing, routing, addressing, or signaling information that counted as content but didn't convey the message of the communication to be treated as metadata).

That said, what the fuck are you thinking?!?!?

I mean, first of all, Congress is about to reauthorize 702, possibly trying to codify the prohibition on about searches. But most of Congress won't go through the trouble to read this five page definition, much less consult with technical experts to understand if the definition is meaningful and how any draft bill would interact with this language. So it's unclear how closely tested this has been.

As noted, even by the 2010 discussion, it was clear Bates was creating a middle ground for stuff that was technically content but which served a DRAS function – probably something akin to Steve Bellovin et al's definition of architectural content. Given the way NSA asked to and did nuke the existing PRTT data *at precisely this time* (though without letting the Inspector General review their destruction of intake data) it's highly likely they were violating those limits, at least through the processing stage. But legally, using this definition of metadata would all of a sudden be kosher, because the metadata would have been collected under a content standard, so the distinction of it being metadata would matter primarily for the privacy considerations (not least because Americans' metadata collected off this upstream collection could and can be disseminated with a much lower standard than the one in place in the Internet dragnet, and can be disseminated for non-terrorism purposes), not legal ones. In other words, by collecting its domestic metadata using a content collection statute, the legal distinction between metadata and content would no longer matter, after 7

years of mattering.

Except now it does.

If the NSA's five page definition of metadata includes stuff that is legally content, then the promise to avoid "about" collection is probably bogus, because it'd incorporate these definitions of metadata and thereby permit using metadata that actually counts as content as a selector.

Which is probably also why the government is so keen to avoid a prohibition on about searches – because what they're doing, even today, amounts legally to about collection.

I'll have to put some thought to the privacy implications of this (I suspect this explains the utility of upstream collection for cybersecurity purposes).

But if I'm right, there's no way this should be classified, at least not entirely classified, not if the government has claimed to have gotten out of the business of searching for selectors in content.