I CON THE RECORD TRANSPARENCY BINGO (4): HOW 151 MILLION CALL EVENTS CAN LOOK REASONABLE BUT IS BESIDES THE POINT

Other entries in I Con the Record Transparency Bingo:

(1) Only One Positive Hit on a Criminal Search

(2): The Inexplicable Drop in PRTT Numbers

(3): CIA Continues to Hide Its US Person Network Analysis

If your understanding of the phone dragnet replacing the old USA Freedom dragnet came from the the public claims of USA Freedom Act boosters or from this NYT article on the I Con the Record report, you might believe 42 terrorist suspects and their 3,150 friends made 48,000 phone calls last year, which would work out to 130 calls a day … or maybe 24,000 perfectly duplicative calls, which works out to about 65 calls a day.

That's the math suggested by these two entries in the I Con the Record Transparency Report — showing that the 42 targets of the new phone dragnet generated over 151 million "call detail records." But as I'll show, the impact of the 151 million [corrected] records collected last year is in some ways far lower than collecting 65 calls a day, which is a good thing! But it supports a claim that USAF has an entirely different function than boosters understood.

Call Detail Record (CDR) Statistics

Call Detail Records "CDR" – Section 501(b)(2)(C)	CY2016
Total number of orders issued pursuant to applications under Section 501(b)(2)(C)	40
Estimated number of targets of such orders	42

Call Detail Record (CDR) Statistics

Call Detail Records "CDR" – Section 501(b)(2)(C)	CY2016
Estimated number of call detail records received from providers and stored in NSA repositories	151,230,968

Here's the math for assuming these are just phone calls. There were 42 targets approved for use in the new phone dragnet for some part of last year. Given the data showing just 40 orders, they might only be approved for six months of the year (each order lasts for 180 days), but we'll just assume the NSA gets multiple targets approved with each order and that all 42 targets were tasked for the entirety of last year (for example, you could have just two orders getting 42 targets approved to cover all these people for a year).

In its report on the phone dragnet, PCLOB estimated that each target might have 75 total contacts. So a first round would collect on 42 targets, but with a second round you would be collecting on 3,192 people. That would mean each of those 3,192 people would be responsible for roughly 48,000 calls a year, every single one of which might represent a new totally innocent American sucked into NSA's maw for the short term [update: that would be up to a total of 239,400 2nd-degree interlocutors]. The I Con the Record report says that, "the metric provided is over-inclusive because the government counts each record separately even if the government receives the same record multiple times (whether from one provider or multiple providers)." If these were phone calls between just two people, then if our terrorist buddies only spoke to each other, each would be responsible for 24,000

calls a year, or 65 a day, which is certainly doable, but would mean our terrorist suspects and their friends all spent a lot of time calling each other.

The number becomes less surprising when you remember that even with traditional telephony call records can capture calls *and texts*. All of a sudden 65 becomes a lot more doable, and a lot more likely to have lots of perfectly duplicative records as terrorists and their buddies spend afternoons texting back and forth with each other.

Still, it may mean that 65 totally innocent people a day get sucked up by NSA.

All that said, there's no reason to believe we're dealing just with texts and calls.

As the report reminds us, we're actually talking about session identifying information, which in the report I Con the Record pretends are "commonly referred to" as "call events."

> Call Detail Records (CDR) – commonly referred to as "call event metadata" – may be obtained from telecommunications providers pursuant to 50 U.S.C. §1861(b)(2)(C). A CDR is defined as session identifying information (including an originating or terminating telephone number, an International Mobile Subscriber Identity (IMSI) number, or an International Mobile Station Equipment Identity (IMEI) number), a telephone calling card number, or the time or duration of a call. See 50 U.S.C. §1861(k)(3)(A). CDRs do not include the content of any communication, the name, address, or financial information of a subscriber or customer, or cell site location or global positioning system information. See 50 U.S.C. §1861(k)(3)(B). CDRs are stored and queried by the service providers. See 50 U.S.C. §1861(c)(2).

Significantly, this parenthesis — "(including an originating or terminating telephone number, an International Mobile Subscriber Identity (IMSI) number, or an International Mobile Station Equipment Identity (IMEI) number)" — suggests that so long as something returns a phone number, a SIM card number, or a handset number, that can be a "call event." That is, a terrorist using his cell phone to access a site, generating a cookie, would have the requisite identifiers for his phone as well as a time associated with it. And I Con the Record's transparency report says it is collecting these "call event" records from "telecommunications" firms, not phone companies, meaning a lot more kinds of things might be included — certainly iMessage and WhatsApp, possibly Signal. Indeed, that's *necessarily* true given repeated efforts in Congress to get a list of all electronic communications service providers company that don't keep their "call records" 18 months and to track any changes in retention policies. It's also necessarily true given Marco Rubio's claim that we're sending requests out to a "large and significant number of companies" under the new phone dragnet.

The fine print provides further elements that suggest both that the 151 million events collected last year are not that high. First, it suggests a significant number of CDRs fail validation at some point in the process.

> This metric represents the number of records received from the provider(s) and stored in NSA repositories (records that fail at any of a variety of validation steps are not included in this number).

At one level, this means NSA's results resulted in well more than 151 million events collected. But it also means they may be getting junk. One thing that in the past might have represented a failed validation is if the target no longer uses the selector, though the apparent failure at multiple levels suggests there may be far

more interesting reasons for failed validation, some probably technically more interesting.

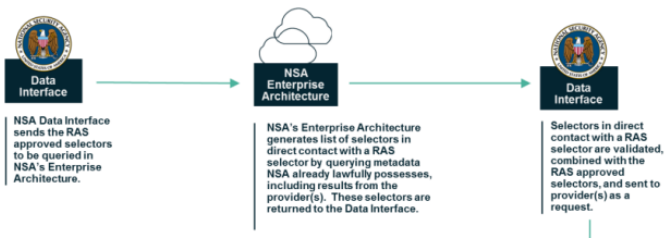
In addition, the fine print notes that the 151 million call events include both historical events collected with the first order as well as the prospective events collected each day.

> CDRs covered by § 501(b)(2)(C) include call detail records created before, on, or after the date of the application relating to an authorized investigation.

So these events *weren't* all generated last year — if they're from AT&T they could have been generated decades ago. Remember that Verizon and T-Mobile agreed to a handshake agreement to keep their call records two years as part of USAF, so for major providers providing just traditional telephony, a request will include at least two years of data, plus the prospective collection. That means our 3,192 targets and friends might only have had 48 calls or texts a day, without any duplication.

Finally, there's one more thing that suggests this huge number isn't that huge, but that also it may be a totally irrelevant measure of the privacy impact. In NSA's document on implementing the program from last year, it described first querying the NSA Enterprise Architecture to find query results, and then sending out selectors for more data.

> Once the one-hop results are retrieved from the NSA's internal holdings, the list of FISC-approved specific selection terms, along with NSA's internal one-hop results, are submitted to the provider(s).



In other words — and this is a point that was clear about the old phone dragnet but which most people simply refused to understand — this program is not only designed to interact seamlessly with EO 12333 collected data (NSA's report says so explicitly, as did the USAF report), but many of the selectors involved are already in NSA's maw.

Under the old phone dragnet, a great proportion of the phone records in question came from EO 12333. NSA preferred then — and I'm sure still prefers now — to rely on queries run on EO 12333 because they came with fewer limits on dissemination.

Which means we need to understand the 65 additional texts — or anything else available only in the US from a large number of electronic communications service providers that might be deemed a session identifier — a day from 42 terrorists and their 3150 buddies on top of the vast store of EO 12333 records that form the primary basis here.

Because (particularly as the rest of the report shows continually expanding metadata analysis and collection) this is literally just the tip of an enormous iceberg, 151 million edge cases to a vast sea of data.

Update: Charlie Savage, who has a really thin skin, wrote me an email trying to dispute this post. In the past, his emails have almost universally devolved into him being really defensive while insisting over and over that stuff I've written doesn't count as reporting (he likes to do this, especially, with stuff he claims a scoop for three years after I've written about it). So I told him I would only engage publicly, which he does here.

Fundamentally, Charlie disputes whether Section 215 is getting *anything* that's not traditional telephony (he says my texts point is "likely right," apparently unaware that a document he obtained in FOIA shows an issue that almost certainly shows they were getting texts years

ago). Fair enough: the law is written to define CDRs as session identifiers, not telephony calls; we'll see whether the government is obtaining things that are session identifiers. The I Con the Record report is obviously misleading on other points, but Charlie relies on language from it rather than the actual law. Charlie ignores the larger point, that any discussion of this needs to engage with how Section 215 requests interact with EO 12333, which was always a problem with the reporting on the topic and remains a problem now.

So, perhaps I'm wrong that it is "necessarily" the case that they're getting non-telephony calls. The law is written such that they can do so (though the bill report limits it to "phone companies," which would make WhatsApp but not iMessage a stretch).

What's remarkable about Charlie's piece, though, is that he utterly and completely misreads this post, "About half" of which, he says, "is devoted to showing how the math to generate 151 million call events within a year is implausible."

The title of this post says, "151 Million Call Events Can Look Reasonable." I then say, "But as I'll show, the impact of the 131 [sic, now corrected] million records collected last year is in some ways far lower than collecting 65 calls a day, which is a good thing!" I then say, "The number becomes less surprising when you remember that even with traditional telephony call records can capture calls *and texts*. All of a sudden 65 becomes a lot more doable, and a lot more likely to have lots of perfectly duplicative records as terrorists and their buddies spend afternoons texting back and forth with each other." I go on to say, "The fine print provides further elements that suggest both that the 151 million events collected last year are not that high." I then go on to say, "So these events *weren't* all generated last year — if they're from AT&T they could have been generated decades ago."

That is, in the title, and at least four times after that, I point out that 151 million is not that high. Yet he claims that my post aims to show that the math is *implausible*, not totally *plausible*. (He also seems to think I've not accounted for the duplicative nature of this, which is curious, since I quote that and incorporate it into my math.)

In his email, I noted that this post replied not just to him, but to others *who were alarmed by the number.* I said specifically with regards the number, "yes, you were among the people I subtweeted there. But not the only one and some people did take this as just live calls. It's not all about you, Charlie."

Yet having been told that that part of the post was not a response to him, Charlie nevertheless persisted in completely misunderstanding the post.

I guess he still believed it was all about him.

Maybe Charlie should spend his time reading the documents he gets in FOIA more attentively rather than writing thin-skinned emails assuming everything is about him?

Update: Once I pointed out that Charlie totally misread this post he told me to go back on my meds.

Charlie Savage
@charlie_savage

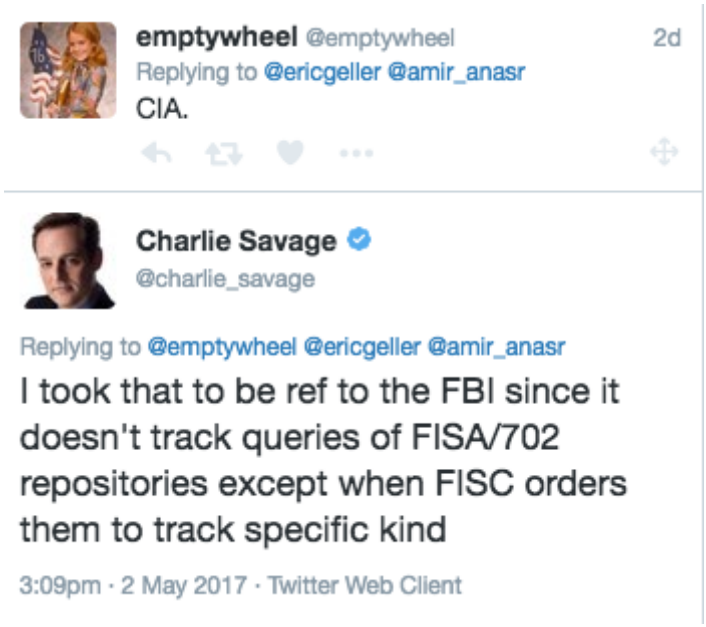
Replying to @emptywheel

That phrase was always there. The 1st 400+ words of your piece are premised on a year=130/65 per person per day. Go back on your meds Marcy.

4:38pm · 4 May 2017 · Twitter Web Client

Since he's being such a douche, I'll give you two more pieces of background. First, after I said that I knew CIA wasn't tracking metadata (because it's all over public records), Charlie

suggested he knew better.



Here's me twice pointing out that the number of call events was not (just) calls (as he had claimed in his story), a point he mostly concedes in his response.

Here's the lead of his story:

> WASHINGTON — The National Security Agency vacuumed up more than 151 million records about Americans' phone calls last year via a new system that Congress created to end the agency's once-secret program that collected domestic calling records in bulk, a report disclosed Tuesday.