WHAT QUERIES OF METADATA DERIVED FROM UPSTREAM DATA MIGHT INCLUDE

In this post, I explained that at virtually the exact moment the NSA shut down the PRTT dragnet in 2011, FISC permitted it to start querying metadata derived from upstream collection. After that happened, it started distinguishing between data that was "handled" according to minimization procedures and data that was "processed" before being intelligible.

In this post, I want to talk about what we can learn about metadata derived from FAA 702 from the opinion that authorized it and this document which based on the date, I assume pertains at least to upstream 702 derived metadata (from which the two kinds of MCTs most likely to include domestic communications would be excluded).

First, assuming that this querying document does include upstream, then it means that entirely domestic communications might be included in the querying. The opinion allows,

> NSA to copy metadata from Internet transactions that are not subject tosegregation pursuant to Section 3(b) without first complying with the other rules for handlingnon-segregated transactions – i.e., without ruling out that the metadata pertained to a discretewholly domestic communication or to a discrete non-target communication to or from a U.S.person or a person inside the United States.

This means that after the data comes in to NSA and the two types of metadata most likely to include domestic MCTs are segregated, it can be made available to metadata analysis. The NSA

prevented queries of segregated data via technical means.

> NSA's technical implementation will ensure that USP metadata queries of FAA 702 collection will only run against communications metadata derived from FAA 702 [redacted] and telephony collection.

The document stated that "NSA's Technical Directorate (TD) continues to work to implement this requirement." It's not clear whether that language dates to December 16, 2011, when it was first written, or to August 19, 2013, when it was most recently revised.

Yet even assuming that technical protection occurred, there would still be Americans in the pool. According to John Bates' estimate from the same year, there might be 46,000 domestic communications in there that ended up in the batch because the domestic communication that made mention of targeted selector transited internationally, which led them to get caught in filters supposedly targeted at international traffic.

The opinion mandates that, if after doing the analysis, the analyst realizes she has a completely domestic communication, she has to destroy it (though that requirement would get softer the next year). But a footnote also reveals that the means of determining if a selector was American was not failsafe.

> NSA will rely on an algorithm and/or a business rule to identify queries of communications metadata derived from the FAA 702 [redacted] and telephony collection that start with a United States person identifier. Neither method will identify those queries that start with a United States person identifier with 100 percent accuracy.

Moreover, in an apparent bid to have this querying process interact relatively seamlessly

with Special Procedures Communications Metadata Analysis (SPCMA — a way to query EO 12333 metadata incorporating US person identifiers), the standards were lackadaisical. As with SPCMA, an analyst had to come up with a foreign intelligence justification, but that's just a "memory aid" in case the analyst gets questioned about it "long after the fact" in a fact check. Analysts don't have to seek approval before they use a particular selector to query and they're not required to attach any supporting documentation for their justification (this was in 2013, so requirements may be stronger in the wake of the PCLOB report). And SPCMA training is considered adequate to query metadata derived from 702.

In other words (again, assuming this pertains to upstream querying), there are several risks: that US person data will get thrown in the mix, that it won't get identified by an algorithm as such, and so that that query result will lead to further spying on a US person without getting destroyed.

Still, as made clear, the alternative is SPCMA, which offers even fewer protections than 702 querying.

One more thought: the NSA report on the aftermath of Bates' upstream decision (and the implementation of the 2012 certificates) revealed the PRISM providers incurred cost with the transition between certificates. It's actually quite possible that the upstream metadata queries would come to constitute a critical part of the targeting process, effectively identifying what Goole or Yahoo content might be of interest at the metadata stage, only then to submit that to the provider for the content. If that's true, it would be somewhat easy to end up targeting a US person for content collection via such upstream searches (though that presumably would be captured in the post-targeting process).