WHAT KIND OF FISHING TRIP DID THE GOVERNMENT CONDUCT INTO AARON SWARTZ' AMAZON DATA?

Yesterday, privacy researcher Chris Soghoian posted an interesting exchange he had with Aaron Swartz in March 2011.

> But then I wondered about Amazon. Amazon not only has a lot of private data on its own, but they host a lot of other websites with personal data. It seems like everyone is using Amazon EC2 these days Reddit and Netflix and Foursquare and more. Even sites that aren't hosted on EC2, like 37 signals, still use S3 for backup. The "truly paranoid" tarsnap uses both EC2 and S3. (Yes, tarsnap encrypts your data, but [it sometimes has bugs][b] and doesn't protect against traffic analysis.) Hell, even WikiLeaks was hosted there at one point.
>
> What's disturbing is that this means your personal data isn't just accessible by the people who operate these sites it's also accessible by Amazon. And anyone Amazon decides to hand it to.
>
> What are Amazon's policies? I've had several conversations with them about this, but they refuse to comment on the record. Still, I'm in the rare position of getting to experience them firsthand. A couple years ago the government sent Amazon a subpoena for information about an EC2 instance I'd purchased. Amazon handed it over without stopping to warn me. When I asked them about it specifically, they refused to comment. When I asked them about their general

> policy, they refused to comment. The only reason I found out about it was because I filed a FOIA request with the Department of Justice. The DOJ was more transparent about this than Amazon.
>
> As best as I can tell, this is Amazon's policy: When the government asks, turn stuff over. Never tell the people affected. Don't give them a chance to object.

The exchange ends with Soghoian asking if Swartz will publish his piece, to which Swartz says he cannot.

> I thought of that and wish I could, but I can't put my name on it right now personal reasons.

The exchange happened, we now know, in between the time the Cambridge police first arrested him for breaking and entering and the time the government indicted him for a slew of computer crimes. It seems likely that those "personal reasons" include negotiations with the Secret Service about the JSTOR downloads (we know Swartz and his lawyer met with the Secret Service that summer and turned over some hard drives).

As Swartz himself pointed out, this exchange also happened in the wake of news that the government had issued orders to Twitter—basically within a day of the time the Secret Service triggered Swartz' initial arrest—for the communications of people associated with WikiLeaks.

The exchange is notable because of a request Swartz' lawyer made the following year, at the beginning of the pre-trial discovery process. In addition to asking how the government had obtained a bunch of communication involving Swartz and others, his lawyer asked to see everything returned from grand jury subpoenas and orders served on MIT and JSTOR—which makes

sense in this case—but also Twitter, Google, and Amazon.

> These paragraphs request information relating to grand jury subpoenas. Paragraph 1 requested that the government provide "[a]ny and all grand jury subpoenas – and any and all information resulting from their service – seeking information from third parties including but not limited to Twitter. MIT, JSTOR, Internet Archive that would constitute a communication from or to Aaron Swartz or any computer associated with him." Paragraph 4 requested "[a]ny and all SCA applications, orders or subpoenas to MIT, JSTOR, Twitter, Google, Amazon, Internet Archive or any other entity seeking information regarding Aaron Swartz, any account associated with Swartz, or any information regarding communications to and from Swartz and any and all information resulting from their service." Paragraph 20 requested "[a]ny and all paper, documents, materials, information and data of any kind received by the Government as a result of the service of any grand jury subpoena on any person or entity relating to this investigation."
>
> Swartz requests this information because some grand jury subpoenas used in this case contained directives to the recipients which Swartz contends were in conflict with Rule 6(e)(2)(A), see United States v. Kramer, 864 F.2d 99, 101 (11th Cir. 1988), and others sought certification of the produced documents so that they could be offered into evidence under Fed. R. Evid. 803(6), 901. Swartz requires the requested materials to determine whether there is a further basis for moving to exclude evidence under the Fourth Amendment (even though the SCA has no independent

> suppression remedy).
>
> [snip]
>
> Moreover, defendant believes that the items would not have been subpoenaed by the experienced and respected senior prosecutor, nor would evidentiary certifications have been requested, were the subpoenaed items not material to either the prosecution or the defense. Defendant's viewing of any undisclosed subpoenaed materials would not be burdensome, and disclosure of the subpoenas would not intrude upon the government's work product privilege, as the subpoenas were served on third parties, thus waiving any confidentiality or privilege protections. [my emphasis]

Effectively, Swartz' lawyer was indicating that he had seen subpoenas and orders that requested information from—among others—Amazon, but not all of what these providers had returned in exchange was turned over as evidence in the case. He was trying to see what else the government had. He's also making it clear that the government asked for the information in such a form that could be entered as evidence in a trial (meaning the government would not have to call an employee from Amazon or another service provider to certify the authenticity of the data, who could then be questioned by the defense).

And he's suggesting that if the prosecutor asked for these things, then they must be relevant in this case, and therefore discoverable.

I suspect, though, that that last claim is not what the lawyer really thought. I suspect that he believed the grand jury investigating Swartz—during precisely the same period when Swartz was researching how Amazon might respond to a government request for information—had conducted a fishing trip on other issues, and

had done so in such a way that any information gleaned could be used both to prosecute the alleged JSTOR download but also any other crime.

Now I suspect that DOJ's original request to Amazon—the one Swartz mentioned to Soghoian—dated to Swartz' efforts to liberate PACER. It shows up in the part of his FBI file Swartz published on his blog.

> Data that was exfiltrated went to one of two Amazon IP addresses.
>
> Investigation has determined that the Amazon IP address used to access the PACER system belongs to Aaron Swartz.

So it's possible the grand jury was reinvestigating what Aaron had done two years earlier, even though DOJ had earlier declined to press charges, in an effort to criminalize Swartz' efforts to liberate information generally.

But given the timing and Swartz' own tie to the WikiLeaks orders, I also wonder whether there was something else there—whether Swartz believed the government had information pertaining to activities entirely unrelated to JSTOR or PACER.

Ultimately, Swartz didn't get this information. As to the communications, the judge assumed the government's assurances that they had neither used a civil administrative subpoena nor "court ordered electronic surveillance" to get his communications closed the issue (given that the government investigated WikiLeaks as an Espionage case, the government might have claimed access to some of this under the PATRIOT Act simply because of Swartz' ties to the Cambridge hacktavist community). And she refused to turn over the grand jury information on the grounds that the government may use such inquiries to chase down every lead, even if those leads are unrelated.

So it's not clear Swartz ever learned what the government was looking for in its fishing

expedition with Amazon.