

Big Data meets Big Govt: New IRS spy software

Exclusive: Daniel J. Pilla explains agency's plan to surveil your finances via Paypal and your email

 [Print](#) [Print](#)

By Daniel J. Pilla

One of the reasons identify theft is considered by the Treasury Inspector General for Tax Administration to be the crime of the century is because of the IRS. The Internal Revenue Service makes growing demands for information about people's businesses and private lives every day. There is no such thing as personal privacy these days. That the IRS sends citizens a so-called "Privacy Act Notice" in all its mailings is a farce. The IRS lays claim to your data without court authority more so than any other government agency. And to make matters worse, they share the data with any other federal, state or local government agency claiming an interest, including foreign governments.

A river of data

In 2019, there will be about 152 million individual tax returns filed with the IRS. There will be roughly another 100 million business tax returns filed. There will be millions more miscellaneous tax returns, including trust, estate and gift tax

returns. On top of that, over 3.6 BILLION information returns (Forms W-2, 1099, etc.) will be filed. There is quite literally a river of data flowing into the agency. The flow cannot be stopped, and as far as the IRS is concerned, they need even more.

For example, one of the six “Strategic Goals” presented in the IRS’ 2018-2022 Strategic Plan is to increase its access to data, and use that data more effectively to drive its agency-wide decision making, as well as case evaluations and selections for enforcement purposes. See: IRS Publication 3744 (4-2018). This is consistent with the IRS goal of becoming a “data driven agency.”

The IRS is awash in data. The 2018-2022 Strategic Plan boasts that the IRS’ volume of data was 100 times larger in 2017 than it was 10 years prior. In 2018, the IRS Criminal Investigation unit alone collected 1.67 terabytes of data from various sources. A terabyte is 1,099,511,627,776 bytes, or 1,024 gigabytes of data. I’m told that approximately 900,000 plain text files can fit into a single gigabyte. The number of users in the IRS with access to that data has increased 23 times (Strategic Plan, p. 19) in the past 10 years.

Managing massive data

How do you manage, process and assimilate such a massive amount of data to the point where it becomes usable? The 2018-2022 Strategic Plan expresses the goal to “invest in analytics and visualization software and tools, and develop processes to support analytics in IRS operations” (p. 20). The end game is presented in these words:

Advancements in how data is collected, stored, accessed and analyzed will allow us to deploy data better. We'll standardize our data processes and protocols and encourage collaboration among all IRS business units. Increased interoperability of data systems and sources will enhance the secure and seamless flow of data to enable greater authorized access to information. We'll invest in training to develop more advanced analytics skill sets across the IRS, and use data to improve our business processes. (Strategic Plan, p. 19.)

The investment in analytics was recently undertaken – in a big way.

Big Government, meet Big Data

On Sept. 27, 2018, the IRS entered into a contract with Palantir Technologies of Palo Alto, California, to handle the task of data assimilation. The contract calls for Palantir to provide hardware, software and training to IRS employees to “capture, curate, store, search, share, transfer, perform deconfliction, analyze and visualize large amounts of disparate structured and unstructured data.” (IRS Contract Proposal, Performance Work Statement, Jan. 11, 2017, p. 1.)

Palantir is to build and train the IRS to use a unified supercomputer to:

search, analyze, visualize, and interact with a wide variety of disparate data sets so users will be able to leverage the platform to perform advanced analytics, such as link, pattern, statistical, behavioral, and geospatial analysis on an

investigative platform that is scalable and interoperable with existing IRS equipment and systems. (Ibid, p. 2.)

What kind of data are we talking about? The contract proposal specifies the following data formats:

- Oracle, MySQL, and PostgreSQL databases;
- Delimited files (.csv, .dsv, .log, or .txt);
- Excel files (.xls, .xlsx);
- GraphML files (.graphml, .xml);
- IVML files;
- Email files (.eml, .pst, .mbox, .msg, .ost, .txt); and
- PCAP files (.pca, .pcap, .pcp). Ibid, pg 20.

Ingesting massive amounts of data

The contract proposal states that the IRS is looking for an “analytical platform with a strong storage and indexing power allowing for rapid integration and analysis of ultra-large scale data sources.” (Ibid, p. 2.) Specifically, the system must meet the following criteria:

- Allow for the rapid ingestion of massive amounts of data.
- Users should be able to immediately use the imported data in the imported format to perform queries, analysis and identify links.
- Allow users to drill down on massive amounts of disparate data to find connections.
- Allow users to visualize connections from millions of records with thousands of links by grouping data visualization by the commonalities and roles. (Ibid, p. 20.)

This would allow the IRS to meaningfully link tens of millions of tax returns, billions of information returns, and trillions of bank and credit card transactions, phone records and even social media posts. For example, if a U.S. citizen moves money from a Swiss bank to some other offshore bank, then uses credit or debit cards to spend the money in the U.S., Palantir's software can link those transactions. It could also flag a person whose tax return shows relatively low annual income but whose social-media posts indicate something entirely different.

This is exactly the kind of data analysis it will take to establish the IRS' so-called "up-front tax system," which I describe in my book ["How to Win Your Tax Audit."](#) Under that system, the taxpayer is essentially removed from the tax preparation process because the IRS knows everything there is to know about your personal, business and financial affairs to the point where the agency prepares the return for you. How's that for tax simplification?

The cost of spying

The IRS began working with Palantir in 2013. The agency spent \$30.8 million on a five-year contract and granted Palantir access to files for more than 1 million people, according to a July 28, 2015, audit report. That contract provides the IRS with access to spy software for use by special agents (criminal investigators) "to generate leads, identify schemes, uncover tax fraud, and conduct money laundering and forfeiture investigative activities." (Case Lead Analysis, PIA ID No. 1120, July 28, 2015, p. 4.)

Under the September 2018 deal, the government will pay Palantir \$98,750,546.94 over seven years to fulfill the contract. My question is, why the extra 94 cents?

If the IRS' \$99 million spy software works as promised, the agency will have unprecedented ability to track the lives and transactions of tens of millions of American citizens.



Daniel J. Pilla is an expert in IRS procedure and advocate of taxpayer rights. He is the author of ["How to Win Your Tax Audit."](#)