THE IRS AND BIG DATA:
THE FUTURE OF FIGHTING TAX FRAUD

By JASON B. FREEMAN, JD, CPA | Column Editor

Welcome to the world of big data, a world where an estimated 2.5 quintillion bytes of data are being produced every day, a world where 90 percent of the world's data has been created in the last two years, and a world that creates enormous challenges – and opportunities – for regulators.

The IRS receives and processes more than 250 million tax returns a year – data that is statutorily protected and that comprises a vast reserve of information. In recent years, it has been tasked with increasing responsibilities, such as implementing tax reform, the Affordable Care Act (ACA) and Foreign Account Tax Compliance Act (FATCA) to name a few. Yet, it is operating on a budget that has been steadily cut since 2011 and facing unprecedented workforce attrition as it fights a tax gap estimated to exceed $450 billion annually.

The solution? It's to work smarter and more efficiently. The IRS is making an investment in big data analytics. And it is seeing a return on that investment. In a recent report, for instance, the IRS' Criminal Investigation Division reported that, despite significant workforce cuts, it had identified approximately 400 percent more tax fraud than in the prior year and over 1,000 percent more in proceeds from other financial crimes compared to the prior year. The IRS credits the prioritization of data, including its use of data analytics, algorithms and "predictive policing," as drivers behind these major strides.

A Thumbnail History

The IRS first began using computers to select tax returns for audit in 1962 and soon thereafter developed the Taxpayer Compliance Measurement Program (TCMP), a program based on in-depth audits designed to obtain data for developing audit-selection strategies. By 1969, it was employing the automated Discriminant Function Analysis (DIF), a computerized, statistical method that rates tax returns based on their so-called "DIF" score and selects them for audit based on the probability that they contain an error or evasion. The DIF system has been refined over the years and is currently the IRS' primary statistical method for selecting tax returns for audit.

In 2002, the IRS began its National Research Program (NRP), which replaced the discontinued TCMP. The NRP represents an effort to comprehensively measure compliance across different types of taxes and taxpayers. Rather than rely on the line-by-line audits that were conducted under the TCMP to gather taxpayer data, NRP audits generally focus on specific sections of the tax return. The IRS applies the data gained through its NRP to update its DIF model.

In 2011, the IRS formed the Office of Compliance Analytics (OCA), which was charged with creating a more data-driven and analytical culture through the use of advanced analytics programs; it represented a modern approach to tax enforcement. In 2016, the OCA was reorganized and merged with the Office of Research, Analysis and Statistics (RAS) to form the Research, Applied Analytics and Statistics (RAAS) organization. The current mission of the RAAS "is to lead a data-driven culture through innovative and strategic research, analytics, statistics and technology services in partnership with internal and external stakeholders." This office serves as a key driver behind the IRS' adoption and implementation of big data analytics.

What is the IRS Currently Doing?

Data mining is the process of extracting information from large sets of data to analyze relationships in the data. Data mining models come in two general forms: predictive and descriptive. Predictive modeling creates a model based on data in order to make predictions. Descriptive models summarize patterns and properties in a data set.

The IRS has access to unprecedented amounts of data, although its collection and maintenance of data can raise constitutional and other concerns, such as privacy considerations. The

IRS is able to mine data from various public sources, including social media outlets such as Facebook, Twitter and Linkedin, as well as other public internet data from sources such as Google Maps.

Reports have indicated that the IRS has even used "spiders" – automated computer programs – to review social media sites. Reports have also indicated its use of phone tracking technology, including a cell-site simulator known as Stingray. In addition, the IRS maintains vast troves of data through its more traditional means, such as the NRP and Individual Master File database. The IRS, in other words, has access to many data sets.

The IRS is cross-referencing and using these data sets to run pattern recognition algorithms to identify trends and understand the relationships in its data. It has employed a number of advanced techniques and tools in this effort, such as anomaly detection, advanced clustering and even neural networks. These tools are aimed at improving case selection and coordination among IRS divisions. Ultimately, data analytics and "predictive policing" will help the IRS identify tax-reporting anomalies and identify tax evasion on a much larger scale.

Areas of Likely Focus in the Future

Developments over the past decade point toward several areas where the IRS' investment in big data analytics is likely to generate a particularly good return. International tax enforcement is one such area. For instance, the IRS has obtained unprecedented amounts of data through international tax enforcement efforts, such as the Offshore Voluntary Disclosure Program, the Swiss Bank Program and numerous high-profile data leaks, such as the "Paradise" Papers, Panama Papers and Offshore Leaks. This data has led to massive databases that the IRS is still unpacking.

Information reporting through the FATCA and information-sharing agreements have led to remarkable structural changes in the global exchange of tax-related information. These developments will be further strengthened in the future by initiatives such as the Joint Chiefs of Global Tax Enforcement, known as the "J5," a collaborative effort by the IRS and several countries to combat the threat of international tax crimes and cryptocurrency.

Those developments hint at another area that is ripe for development: Cryptocurrency. The IRS believes that cryptocurrency-related tax compliance is abysmal. In the past year, the IRS successfully forced Coinbase, the largest domestic virtual currency exchange, to turn over data on several thousand taxpayers' cryptocurrency transactions. During the Coinbase proceedings, the IRS alleged that "only 800 to 900 taxpayers reported gains related to bitcoin" between 2013 and 2015, even though it estimates that perhaps millions should have. The IRS is now actively mining this newly received data.

The IRS has also partnered with third parties, such as Chainalysis, a contractor that utilizes data pulled from public forums, the dark web and other sources to trace cryptocurrency to its owner and remove the perceived cloak of anonymity. Enforcement in these areas and others is likely to see particular benefit from the IRS' investment in data analytics.

Tip of the Iceberg

The IRS has entered the future of fighting tax fraud. It has embraced big data analytics – and we have likely only seen the tip of the iceberg. With a reported year-over-year increase in its detection of tax fraud of more than 400 percent and its more than 1,000 percent increase in the identification of proceeds from other financial crimes, the IRS is likely to up the ante on its bet on big data. ❄

ABOUT THE AUTHOR:
JASON B. FREEMAN, JD, CPA, is the managing member of Freeman Law PLLC, based in the DFW Metroplex, and an adjunct professor of law at Southern Methodist University's Dedman School of Law. He can be reached at Jason@freemanlaw-pllc.com.