Professional Perspective

DOJ Data Analytics Identify & Prosecute Fraud

Julian L. André & Justin P. Murphy, McDermott Will & Emery

Bloomberg Law

DOJ Data Analytics Identify & Prosecute Fraud

Contributed by Julian L. André & Justin P. Murphy, McDermott Will & Emery

The Department of Justice (DOJ) and other federal law enforcement agencies have long discussed using data analytics—the process of examining data sets to draw conclusions and identify patterns about the information they contain—to identify and prosecute criminal conduct.

But there had been little evidence that DOJ's use of data analytics had a meaningful impact on the volume or types of fraud prosecutions pursued, until now. The DOJ and other federal agencies, aided by advancements in technology and robust and useable data sets, appear to have finally cracked the code. Most notably, the DOJ has been using data analytics to rapidly identify suspicious or fraudulent Paycheck Protection Program (PPP) loans, leading to over 100 prosecutions in just over one year.

The DOJ also initiated a program in 2020 to use data analytics to detect potential collusion in public procurement in connection with its Procurement Collusion Strike Force (PCSF). And the DOJ and other agencies are expanding their data analytics and data mining capabilities to combat other fraud schemes, including health-care fraud, securities fraud, and even tax evasion.

The DOJ's expanded and effective use of data analytics is a game changer, and will likely lead to increased criminal and civil enforcement activity across the board. In response, companies need to understand what data is readily available to the government, how the government can use data to identify fraudulent conduct, and incorporate data analytics in their own compliance plans to minimize enforcement risk.

And defense counsel will need to seek discovery regarding the government's use of data analytics, challenge any improper collection or use of the underlying data, and conduct their own data analysis to prepare an effective defense strategy.

Paycheck Protection Program

The government's recent PPP prosecutions provide a concrete example of the DOJ's success with data analytics. The PPP was created as part of the CARES Act to provide forgivable loans of up to $10 million to small businesses. Since the enactment of the PPP in 2020, the Small Business Administration (SBA) and PPP lenders have approved over 11.5 million PPP loans worth close to $800 billion.

The volume of PPP loans issued is remarkable as is the speed in which the DOJ has initiated PPP loan fraud prosecutions. In just over one year, the DOJ has charged more than 125 defendants with PPP loan fraud in over 100 different criminal cases throughout the country.

The DOJ has explicitly touted its use of data analytics to identify and prosecute PPP loan fraud with "unparalleled speed." In September 2020, Acting Assistant Attorney General Brian Rabbitt said,

> To bring these [PPP] cases as quickly as we have, and to sort through the volume of loans made by the SBA, the Fraud Section and its partners deployed the first-in-class data analytics capabilities they have developed and employed to great effect in other criminal investigative areas, such as health care fraud and market manipulation.

And the DOJ Fraud Section has repeatedly referred to PPP cases as "data-driven investigations."

Although the specific details of the DOJ's data analytics program are not publicly available, a vast amount of data is readily accessible to the DOJ and the federal agencies investigating PPP loan fraud cases. It is easy for one to see how the effective collection and analysis of this data could lead to numerous PPP fraud prosecutions and investigations.

For example, PPP loan applications require businesses and their owners to submit extensive information to financial institutions and the SBA. In addition to basic identifying information, businesses must provide ownership information and bank account information, and submit payroll data and tax information. The internet protocol (IP) address used to submit the PPP loan application are also collected by most lenders. See, e.g., Aff. Crim. Compl., *United States v. Redfern*, No. 1:20-

MJ-256, Dkt. 2 at ¶ 31 (M.D.N.C. Aug. 25, 2020). All of this data is then electronically transmitted from the lender to the SBA, and can be easily accessed by the DOJ and other federal agencies.

The SBA data alone is a treasure trove for prosecutors and criminal investigators seeking to detect fraud. Investigators can quickly determine whether companies or individuals submitted PPP loan applications to multiple lenders, or whether multiple loan applications were submitted using the same IP address, business address, email address, EIN, or bank account information. Investigators can also identify more complex patterns, such as multiple loan applications based on identical payroll data and employee information.

The DOJ's data analytics capabilities, however, likely extends far beyond simply mining the SBA loan data for suspicious activity. The DOJ could also compare the SBA loan data against other public and non-public information to develop additional leads. Businesses are not eligible for PPP loans if they have been barred from a federal program or are involved in bankruptcy proceedings. The DOJ can thus quickly identify other fraudulent loans by cross-referencing the SBA loan data against other voluminous data sources such as the System for Award Management (SAM) debarment database or bankruptcy filing data.

The DOJ and other law enforcement agencies also have access to vast amounts of non-public financial information maintained by FinCEN, including Suspicious Activity Reports (SARs) and Currency Transaction Reports (CTRs) submitted by financial institutions. Under the Bank Secrecy Act, financial institutions are required to report within 30 days all cash transactions exceeding $10,000 per day and any suspicious activity that may signal criminal conduct. Information from SARs could be analyzed with the SBA loan data to identify additional patterns and investigative leads.

Importantly, the DOJ can access and analyze all this data without ever having to issue a grand jury subpoena, obtain a search warrant, or interview a single witness. This means their entire investigation can be "covert" or outside of the public's view. Before a PPP loan fraud investigation is formally opened, investigators may already have substantial evidence regarding the potential fraud and be able to effectively and efficiently deploy its investigative resources. And, once an investigation is opened, additional evidence can be analyzed against the initial data sources to develop more investigative leads and identify additional patterns that may inform the direction of the investigation and identify further subjects.

Procurement Collusion Strike Force

The Antitrust Division is likewise using data analytics to boost its criminal enforcement program through the . The PCSF leads the DOJ coordinated national response to combating antitrust crimes and related schemes in government procurement, grant, and program funding fraud at the federal, state, and local government level. The PCSF is active in a broad array of industries that contract with the government and focuses on collusive conspiracies occurring at both the domestic and international level that impact U.S. procurement dollars.

Importantly, the PCSF is using a new resource to detect potential crimes; they are developing data analytics to mine available procurement data to identify patterns that indicate or suggest collusion or fraud.

The PCSF is not going it alone—they are working aggressively to advance their use of data analytics in coordination with other law enforcement agencies. Starting in 2020, DOJ hosted four data analytics webinars with more than 1,000 data scientists, analysts, and auditors in attendance. The PCSF Data Analytics Project—working in conjunction with dozens of other government agency data analytic groups—is identifying procurement platforms that collect and retain certain bid data from current and prospective government contractors, and developing tools to identify patterns that indicate or suggest collusion and/or fraud.

Such an undertaking may not have been possible—or certainly not as effective—even five or 10 years ago. Historically, some federal agencies accepted procurement bids in paper form and preserved records in hard copy or electronic formats that were difficult to search and sort. They also did not always maintain complete aspects of procurement submissions.

Today, more federal agencies collect and maintain procurement data in electronic (hence easier to use) formats—this facilitates the use and effectiveness of analytical tools to detect collusion and/or fraud. Simply put, the data available to the PCSF and other government agencies, as well as the PCSF's advocacy for the thorough collection and retention of bid data across the government, allows data analytics to be a potentially effective investigative tool.

The PCSF data analytics project could be effective in using the red flags of collusion to identify patterns that may suggest or indicate collusion. For example, data analytics could quickly identify patterns in bidding, pricing, or other aspects of

procurement over lengthy procurement periods that would otherwise take significant resources and time to detect. These are often iterative schemes that take years to uncover.

Data analytics may allow the PCSF to identify suspicions at narrow and broader levels; for example, it may direct investigators to a specific, individualized bid, a series of bids for a general type of project, or even bids throughout a broader industry. The data analytics results may tell investigators where their resources are likely to bear more fruit and could point them to opening up costly and extensive grand jury investigations.

DOJ's Data Analytics Use in Other Areas

The DOJ's increasing use of data analytics is not limited to PPP loans or the PCSF. Rather, the agency intends to use data analytics to pursue both criminal and civil fraud cases in a wide range of industries. For example, in February 2021, Acting Assistant Attorney General Brian M. Boynton emphasized the importance of data analysis to "identify potential fraudsters" in the health-care arena, noting that "the Civil Division has been actively using its data analysis for this very purpose."

Boynton also linked the DOJ's use of data analytics to the significant increase in federal False Claims Act (FCA) cases initiated directly by the DOJ, as opposed to via *qui tam* whistleblower complaints. Indeed, in 2020, the DOJ initiated 100 more FCA cases than in 2019, and the most non-*qui-tam* FCA cases filed in nearly 30 years. And the DOJ's Civil Division plans to rely on such data analysis to combat other types of fraud.

Other federal agencies have already implemented their own robust data analytics programs. For example, the SEC's Market Abuse Unit created the Analysis & Detection Center in 2011 for the specific purpose of using data analytics to uncover and investigate misconduct. The SEC then expanded its data analytics use with the ATLAS initiative, which allows the SEC's Division of Enforcement to "harness multiple streams of data" using "sophisticated artificial intelligence software." For example, many—if not most—insider trading investigations are initiated as a result of the SEC's effective use of data analytics.

Similarly, in 2018, the IRS entered into a $99 million contract with Palantir Technologies to improve its data analytics capabilities. The IRS now appears to be using data analytics to identify cryptocurrency owners who have failed to pay their taxes, among others.

Given the success the DOJ and other agencies have already achieved using data analytics, federal law enforcement's use of data analytics will become increasingly prominent. This is particularly true as the volume and quality of available electronic data and technological abilities continues to increase.

Responding to Data-Analytic Driven Investigations

The DOJ's increasing use of data analytics is not limited to PPP loans or the PCSF. Rather, the agency intends to use data analytics to pursue both criminal and civil fraud cases in a wide range of industries. For example, in February 2021, Acting Assistant Attorney General Brian M. Boynton emphasized the importance of data analysis to "identify potential fraudsters" in the health-care arena, noting that "the Civil Division has been actively using its data analysis for this very purpose."

Proactive Response: Update Corporate Compliance Programs

Companies must incorporate data analytics in their compliance programs to minimize enforcement risk. In June 2020, the DOJ updated its guidance on effective corporate compliance programs to specifically require prosecutors to consider whether companies are appropriately using data analysis. Specifically, the guidance directs prosecutors to determine whether compliance teams have "sufficient direct or indirect access to relevant sources of data to allow for timely and effective monitoring and/or testing of policies, controls, and transactions."

Companies should heed this advice. By properly analyzing available internal and external data sources, companies may be able to stay one step ahead of investigators, and be better able to detect potential misconduct, understand any compliance risks the company faces at an early stage, and respond accordingly. Using data analytics in a compliance audit or review may prevent a company from being ensnared in a costly government investigation, whereas a company's failure to do so could lead to increased penalties and, in extreme cases, criminal prosecution.

If Charged, Reactive Responses to DOJ's Use of Data Analytics

Request Discovery Regarding Government's Use of Data Analytics

Companies or individuals facing criminal charges or civil actions should request discovery regarding the government's use of data analytics in connection with its investigation, as well as access to the underlying data. Although prosecutors may argue that internal government documents prepared in connection with investigating or prosecuting case are not discoverable, there are multiple bases upon which such discovery requests could be made.

First, under *Brady v. Maryland*, 373 U.S. 83 (1963), and its progeny, the government is required to produce all evidence that is material to the defense. To the extent specific data sources identified potentially fraudulent conduct, the same data sources could contain or are likely to lead to exculpatory information. See, e.g., *United States v. Bagley*, 473 U.S. 667 (1985).

For example, in a government procurement investigation, data analytics may have targeted a contractor that declined to bid on particular procurements when it often bid on similar projects in the past. However, the underlying data may corroborate that the contractor declined to bid on those procurements as they instead wished to bid for another upcoming project and wished to have their employees available. Indeed, it may be worthwhile to seek access to government data sources in discovery regardless of whether such data was used during the investigation on the theory that such data is within the possession of or accessible to the government's prosecution team. See Justice Manual §9-5.001 (Jan. 2020 ed.).

Second, a defendant generally has a right to challenge the adequacy of the government's investigation at trial. As the Supreme Court explained in *Kyles v. Whitley*, 514 U.S. 419, 432-33 (1995), "[w]hen ... the probative force of evidence depends on the circumstances in which it was obtained and those circumstances raise a possibility of fraud, indications of conscientious police work will enhance probative force and slovenly work will diminish it." Defendants, however, cannot meaningfully challenge the nature or adequacy of the government's investigation without a fulsome understanding as to whether and how the government used data and/or data analytics during the investigation.

Fourth Amendment Challenges Based on Improper Data Use

Criminal defendants charged after "data driven" investigations should consider bringing Fourth Amendment challenges based on how the data was collected *and* whether the use of such data violates an individual's right to privacy. The Supreme Court has long held that a person does not have a Fourth Amendment privacy interest in business records, such as bank statements or telephone records, that are possessed, owned, and controlled by third parties. See *United States v. Miller*, 425 U.S. 435 (1976); *Smith v. Maryland*, 442 U.S. 735 (1979). Although the "third-party doctrine" permits the government to obtain vast amounts of personal information without obtaining a search warrant or other court order, the Supreme Court recently recognized in *Carpenter v. United States*, 138 S. Ct. 2206 (2018), that technology has changed a great deal since the 1970s.

In *Carpenter*, the Supreme Court held that the third-party doctrine did not apply to cell-site location information maintained by wireless cellphone carriers and that the government was required to obtain a search warrant to seize such records. "There is a world of difference between the limited types of personal information addressed in *Miller* and *Smith* and the exhaustive chronicle of location information casually collected by wireless carriers [today]." Although *Carpenter*'s holding was "narrow," the Court nevertheless recognized that when people voluntarily discloses information to a third party, they do not automatically forfeit privacy expectations in that information.

Since *Carpenter*, other federal courts have noted that the aggregation of collected data may raise privacy and Fourth Amendment concerns. For example, in *United States v. Moalin*, 973 F.3d 977 (2020), the Ninth Circuit considered a challenge involving the National Security Agency's collection, aggregation, and analysis of telephone metadata. Although the Ninth Circuit did not reach defendant's Fourth Amendment argument, it noted that the argument had "considerable force."

Among other things, the court noted that the government had failed to "recognize that the collection of millions of other people's telephony metadata, *and the ability to aggregate and analyze it*, makes the collection of [the defendant's] *own* metadata considerably more revealing." As technology continues to advance and the government increasingly utilizes sophisticated data analytics, data mining, and artificial intelligence tools to extract private information from data sources without first obtaining a search warrant, courts—in our new era of awareness of technology and digital privacy—may be willing to reconsider the third-party doctrine and its impact on individual privacy rights.

Justice Sonia Sotomayor stated in her concurrence in *United States v. Jones*, 565 U.S. 400, 417 (2012) that:

> It may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties. This approach is ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks.

Just as new technology has allowed for the collection of expansive new types of data, one could make a compelling argument that allowing the government to aggregate and mine such data using data analytics raises just as serious privacy and Fourth Amendment concerns.