2017

Data Analytics, The Next Frontier: Taking a Byte Out of Corruption

Shlomo Fischer

CENTER FOR THE ADVANCEMENT OF PUBLIC INTEGRITY

Data Analytics, the Next Frontier:

Taking a Byte Out of Corruption

In June 2017, CAPI presented the second installment of our signature conference, Global Cities II, which brought together anti-corruption leaders from government and civil society worldwide, including delegates from Bogotá, Cape Town, London, Melbourne, Miami, Montréal, New York, Paris, Rio de Janeiro, and San Francisco, to discuss important topics such as using data analytics to combat corruption, government transparency, enforcement challenges and victories, and innovations in oversight. Videos and other materials from Global Cities II can be found here.

Original CAPI Publication:
This brief was prepared by the Center for the Advancement of Public Integrity at Columbia Law School. We can be reached at CAPI@law.columbia.edu.

CAPI would like to thank Shlomo Fischer, CAPI summer research intern and Columbia Law School student, for his authorship of this brief.

Introduction

The panel entitled "Data Analytics, the Next Frontier: *Taking a Byte out of Corruption*" featured four speakers: Jennifer Rodgers, the Executive Director of CAPI; Bruno Bondarovsky, the Secretary for Planning in the city of Mesquita, Rio de Janeiro; and Milton Yu and Calvin Lam from New York City's Department of Investigation. The panel was moderated by Elif Ryder from Kroll Advisory Solutions. PowerPoint presentations and videos of the Data Analytics panel can be found here.

Jennifer Rodgers, Executive Director of CAPI

Rodgers spoke first, describing CAPI's recent publication entitled *Taking a Byte Out of Corruption*, a roadmap for cities to implement data-driven strategies for identifying government corruption. The project, supported by the Laura and John Arnold Foundation, involved assembling a working group of experts, who met over the course of 15 months. Comprised of leading practitioners, scholars, engineers, and civil servants, the group focused on framing the issue, identifying key risk areas ripe for application, describing analogous fields in which data analytics have been employed effectively, and setting forth best practices for data management. The working group's deliberations were supplemented by field interviews with additional industry experts as documented in the report.

Rodgers then highlighted the publication's "Top Ten" list of recommended starting points for data-driven approaches to fraud and corruption risks using readily available resources. These areas included:

- **Fraud by inspectors**: City employees with discretionary enforcement powers are particularly susceptible to bribes where lucrative business opportunities are involved. A trend toward electronic data storage for inspection records has allowed for outlier analyses to be conducted on actions by individual inspectors, such as violation downgrades and the volume of reported code violations.
- **Human resources**: Timekeeping data in many cities has transitioned to electronic storage, making it easier to parse data for suspicious trends by conducting outlier analyses, especially in relation to reporting overtime hours.
- **Benefits fraud**: CAPI recommended searching for benefit card transactions outside of the zip codes of beneficiaries, and performing outlier analyses on emergency benefit disbursements.
- **Campaign finance violations**: Cities with campaign finance regimes can root out "straw donors" by examining matching programs/campaign limits which give publicly available donation data, and by looking for donations from out-of-district donors.

- **Petty theft of resources and inventory**: CAPI suggested analyzing purchase-order data for volume deviations.
- **Procurement fraud**: Cities can identify prohibited collusive bidding on government-issued contracts and opportunities using specialized software programs.
- **Fraudulent legal claims against the city**: CAPI recommended minimizing fraud by inspecting possibly fraudulent claims immediately below the city's threshold for settlement/litigation.

Bruno Bondarovsky, Secretary of Planning in the city of Mesquita, Rio de Janeiro

The next speaker was Bruno Bondarovsky, who spoke about the successful, data-driven overhaul of the city's antiquated business license issuing regime. Under the old system, paper-based applications were processed in a lengthy series of steps, including evaluation by six or more agencies. The new process was designed to filter digitized applications through a streamlined workflow system that automated and thus depersonalized several parts of the approval process. Among the major victories achieved by the new system are an estimated $6 million in savings from reduced paperwork expenses, an 80% increase in license issuances, and perhaps most significantly, a much quicker process in which 50% of consultations are automatically processed (whereas licensing under the old regime would take anywhere from 45 days to 3 months). The online portal was designed by a data specialist and integrates registration with outside entities to create a more user-friendly experience while increasing visibility by displaying system-wide data on the dashboard. The success of the upgrade has been two-pronged: the streamlined approval process has catalyzed a surge in the region's economic activity (while consequently providing greater tax income for the local government), and has also furthered efforts to diminish corruption vulnerabilities by centralizing licensing data, allowing for new opportunities for corruption identification, and discouraging misbehavior through transparency.

Milton Yu, Inspector General with New York City's Department of Investigation
Calvin Lam, Senior Data Analyst with New York City's Department of Investigation

Milton Yu was the final speaker and was joined by Calvin Lam. Yu profiled DOI's efforts to develop an in-house data analytics capacity using massive tranches of client data mainly from the Human Resource Administration (HRA), the entity that administers public assistance as well as Medicaid to city residents. HRA is an important agency in which to foster anti-corruption efforts due to its size: 14,000 employees, a $9.7 billion budget and over $500 million in annual contracts. The digitized data collected by DOI is a gold mine for investigators hoping to enforce integrity measures and root out fraudulent activity.

DOI opted to cultivate an internal data management system rather than outsource the technology through SaaS (Software as a Service) vendors because they hoped to establish sustainable infrastructure with iterative functionality; that is, they hoped to develop a consistent methodology with repeatable test configurations that could be passed down to future officials working in similar capacities. Yu outlined three challenges the DOI faces, which remain the goals of the data analytics initiative:

- **Obtaining a global view of the issues**: The first issue that arises when employing a data analytics strategy is establishing a controlled environment. This involves regulating the flow of information, determining which data sets are useful and connecting various data points to enhance the investigation. This allows DOI to identify principal players, the extent of activity, and additional parties that may be involved in a fraud scheme. Seemingly innocuous complaints may uncover broader schemes when processed through the system. Similarly, data records on computer logs, phone records, and payment records may be symptoms of corrupt activity. For example, DOI uncovered a $1.5 million scheme in December 2015 involving both food-stamp "fencing" (reselling stolen goods) to local groceries, and repeatedly issuing fraudulent rental subsidies to the same address. The department analyzed computer logs of employees who had interacted with clients who issued complaints, and localized the investigation to 6 employees whose computers could

have been used in the scheme. Enhancing a little further, they cross-referenced these computer logs with payroll records to see when these employees clocked in. Upon discovering that only one of these employees (a supervisor) was present when the benefits were issued, the department concluded that the other 5 employees' identities were stolen by their supervisor. Further investigation into employee phone records during the time frames marked by suspicious activity yielded the discovery of an outside recruiter who appeared to be facilitating the scheme.

- **Speed of investigatory processes**: As data sets increase in volume, efficient filtration functions must be implemented to overcome resource and manpower limitations. DOI uses automation techniques to narrow the scope of investigations where appropriate. An analysis of city vehicle misuse by employees was greatly enhanced by eliminating all data collected within New York City, because vehicle use outside of the City is much more likely to constitute misuse. This decision resulted in narrowing the results from 136 vehicles with over 3 million GPS coordinate data points to 6 vehicles and 38,000 data points – much more manageable figures from which to launch an investigation.
- **Institutional memory**: Recurring fraud indicators are inserted into a "dictionary of fraud," which provides a helpful framework from which to jumpstart investigations. In one significant case, an employee responsible for renewing Medicaid certifications was found to have been extending benefits for certain clients for unusually lengthy periods of time; the beneficiaries were all found to be family members of the employee. In order to target future instances of self-dealing with family members, analysts paved the way for more effective data pivoting by codifying this information into an algorithm designed to detect this specific pattern of fraudulent activity and thus accelerating the speed with which these investigations are conducted.

Yu also noted that from a statutory perspective, the data received from the HRA belongs to the state and therefore cannot be exchanged freely with city government entities. Yu suggested that investigatory bodies work together with agencies that are primary users of this data in order to have full access to the data in the context of an inter-agency partnership.

For governmental entities that balk at the idea of building data management systems from the ground up, Ryder noted that SaaS providers recognize that out-of-the-box solutions do not serve clients' interests and that the industry has prioritized the ability to integrate data processes into the client's long-term plans so they can leverage applications for future use.

Breakout Session

The Data Analytics Breakout Session brought nearly thirty city delegates together to continue to discuss this important and challenging area. Delegates touched on the following themes:

1. **The importance of recognizing data analytics as an essential anti-corruption instrument**. One delegate suggested that criminal investigation bureaus view data analytics with hesitation because it is esoteric in nature and difficult for non-experts to understand. One solution to that is to utilize CAPI's report, which hopes to clarify the field. Others noted that with respect to data-driven solutions, it is often difficult to quantify successes so as to justify substantial resource allocation. Nevertheless, many conference attendees noticed a change in behavior in corruption-sensitive areas as a result of training sessions associated with data analytics initiatives.
2. **Standardizing data to promote synergies.** It is important to map out data borders before beginning a query so the process remains relatively confined. The data tables need to match up in order to derive meaningful analysis from the information. One delegate suggested that governments should impose contractual provisions and requirements on a citywide basis to have data submitted in a reportable format.
3. **Honing in on specific issues.** A conference attendee specializing in data management explained that investigations usually begin with a complaint or an allegation. Once a preliminary analysis is conducted, a strong indication of fraud will lead to further verification. Ultimately, the goal should be for the analytics to

drive and lead investigations, but for now the query system prevails. However, with improvements in cross-data set analysis, one can run substantially more simulations than before and test theories with greater speed.

4. **Resistance to change.** Some delegates noted that when city governments feel overburdened by existing systems of data documentation, they may not have to appetite for additional data-heavy work. Others indicated that the implementation of sophisticated anti-corruption tools may be met with resistance from civil servants and high-ranking officials who may feel that their job security is being threatened. Persistence, clarity of purpose, and reference to notable successes elsewhere may help overcome this obstacle.

Integrity in Brief Series | laura and john arnold foundation

This publication is part of an ongoing series of contributions from practitioners, policymakers, and civil society leaders in the public integrity community. If you have expertise you would like to share, please contact us at **CAPI@law.columbia.edu**.

The series is made possible thanks to the generous support of the Laura and John Arnold Foundation. The views expressed here are solely those of the author and do not necessarily represent the views of the author's organization or affiliations, the Center for the Advancement of Public Integrity, Columbia Law School, or the Laura and John Arnold Foundation.