



## 10. Using Predictive Analytics to Prevent Rather Than React and Respond: A Case Study of the Recovery Accountability and Transparency Board

*By Earl Devaney*

### Introduction

Predictive analytics involves extracting crucial information from multiple databases and then using it to help predict future trends, events, and behaviors. Often, sophisticated techniques like data mining and statistical modeling are used to enhance the ability to predict outcomes and to identify risks and opportunities. In turn, real-time decision-making becomes possible. Of course, the usability of the results depends heavily on the quality of the data used and the assumptions made during the analysis. One of the most well-known applications of predictive analytics is the FICO credit score (developed by Fair, Isaac, and Company) used to determine a borrower's creditworthiness.

Intelligence and law enforcement agencies have been using predictive analytics to fight terrorism and crime for years. More recently, the civilian side of government has begun to use predictive analytics to help identify improper payments, waste, and fraud. In particular, predictive analytics have helped to shift the paradigm from simply detecting these abnormalities to preventing them from happening. Clearly, the government can save a significant amount of time and resources by getting payments right in the first place more often, rather than sending money out and then expending administrative and sometimes legal resources to recover money provided improperly. Perhaps the best example to date of this new approach for using analytics to improve government efficiency is the Recovery Operations Center (the ROC), which was set up by the Recovery Accountability and Transparency Board (the Board) to monitor the \$840 billion American Recovery and Reinvestment Act (ARRA) economic stimulus program, including contracts, grants, and loans.

### Case Study: Use of Predictive Analytics at the Recovery Accountability and Transparency Board

ROC analysts use a variety of new technology tools to mine more than 25 government and open-source databases, looking for anomalies and other indicators of fraud or waste. When problems arise, ROC analysts now provide an alert to both agencies and law enforcement that a particular grant, contract, or loan may be vulnerable to fraud or waste. In most instances, these alerts are issued before the funds have been handed out but even in cases where some money has been awarded, the notification comes in time to prevent further distribution of funds. Obviously, preventing improper payments, waste, or fraud is far preferable to the traditional "detect and chase" approach after the monies have been lost. The remarkable success the ROC has had in minimizing fraud and waste is evidenced by the numbers: Less than one-half of one percent of the nearly 277,000 contracts, grants, and loans awarded under the Recovery Act are under investigation. This pales in comparison to the five to seven percent figure normally associated with losses for any large government program.

The Board used multiple avenues to achieve these results, described below.

**Transparency.** The Recovery Act mandated that the Board establish and maintain a user-friendly website that would provide historic levels of transparency on how American tax dollars were being spent. The Board created two separate websites:

- [FederalReporting.gov](#) to collect data from Recovery recipients on how they spent their awards
- [Recovery.gov](#) to display that Recovery Act spending

Together, these websites heralded a new era in government transparency.

Because millions of citizens, reporters, and pundits were able to see this spending so transparently, the “bad guys” stayed away from Recovery Act money. Thus, one takeaway from Recovery’s grand experiment is that predictive analytics are best used in conjunction with transparency to maximize effectiveness. Neither is fully effective without the other. However, when used in tandem, transparency becomes a force multiplier for accountability.

---

*... government can save a significant amount of time and resources by getting payments right in the first place more often, rather than sending money out and then expending administrative and sometimes legal resources to recover money provided improperly.*

---

The Board also carefully mined information from focus groups and stakeholder meetings and one thing became abundantly clear: everyone wanted to know how Recovery money was impacting their own neighborhood. Responding to this insight, the Board ensured that the Recovery.gov website provided users with the ability to simply enter their zip code to see the spending on a map of their neighborhood. For the first time ever, Americans were able to see on a map where their tax dollars were actually being spent. The zip-code search proved to be the number one feature of this award-winning website.

**Technology.** It is also critically important for government agencies to understand and begin to use the many new technologies that ROC and countless other accountability platforms employ throughout government and the private sector.

Today, technology companies are enabling governments and private-sector enterprises to harness and analyze massive amounts of data. This development permits real-time decision-making, greater transparency, and improved business processes. With respect to predictive analytics, the Board’s strategies relating to big data, geospatial services, and cloud technology can help to provide a framework for IT reform within government in future years.

Data are growing exponentially. Experts expect that the amount of data created annually will grow by a factor of 44 between 2009 and 2020. This growth rate will easily exceed the capacity of traditional software tools to collect, manage, and process the data within acceptable time frames. Given this difficulty, it is no surprise that a host of new “big-data” tools have

emerged to meet this challenge. Government IT professionals, in particular, have been scrambling to find these new tools and strategies to address this phenomenon.

The success of the ROC depended entirely on the Board's ability to find the right set of tools to collect, manage, and analyze many datasets, both public and classified. Adding to the challenge was the unprecedented amount of data being collected from the recipients of Recovery funds. Simply stated, the Board needed to practice predictive analytics in ways that the government had never before tried.

**Data mining.** Initially, the Board used key indicators such as fraudulent business addresses, past criminal behavior, and government suspension and debarment proceedings to isolate potential high-risk recipients. When the assembled data began to overwhelm the ROC's analysts, the Board procured a global software company's new big-data tool that enabled analysts to quickly cut their process time down from five days to five hours.

Once the data have been mined and anomalies identified, sophisticated link analysis tools are used by the analysts to uncover non-obvious and/or high-risk relationships between entities. In one example, "Company A" presents itself as a low-risk recipient. By using the link analysis tool, the Board's staff identifies undisclosed ties to known criminal or high-risk entities, an analysis that shows "Company A" might actually be a high-risk entity.

Under the best scenario, these discoveries are made in the pre-award stage and the responsible government procurement official can avoid a potential loss of funds to fraud or waste. Of course, the Board notifies the appropriate Inspector General when potential fraud or waste is detected so that they can follow up.

**Mapping.** Comprehensive geospatial, or mapping, capability served as another critical component of the Board's predictive analytics platform. This technology focuses on both data analysis and the local impact of spending. The Board obtained this capability from the world's largest geospatial services company, which had first pioneered this approach with the City of Baltimore and the State of Maryland.

However, the elegance of geospatial technology goes way beyond simple consumer mapping applications. The Board pioneered its own unique use for geospatial services. It realized that analysts could use the technology to access and evaluate data from a multitude of geo-databases and make that information accessible in intuitive ways. For instance, by mapping past incidents of fraud and waste and combining that information with data from entities being investigated or audited, the Board created a series of predictive maps that displayed trends and even suggested the redeployment of investigative resources throughout the United States.

**Cloud computing.** The rapid growth of data is driving both government and private-sector chief information officers to seek out new levels of efficiency and cost savings by moving quickly toward cloud technology. Cloud computing uses hardware and software that are delivered normally over the Internet. In the cloud mode, the users entrust a remote service to store and operate their data and software. With cloud computing, users can connect from anywhere. Advantages also include cost savings, a reduction in FTE, business continuity during disasters, scalability in real time, and increased security. It is estimated that by 2020, more than one-third of all data will either live in or pass through the cloud. The Board moved Recovery.gov to the cloud in April, 2010. At the time, no other government entity had yet moved to the cloud. Not too long after, the ROC also migrated toward cloud computing.

Moving to the cloud meant that the Board no longer had to manage Recovery.gov's physical data center and related computer equipment. Through a contract with the website's developer

and systems integrator, a private cloud provider was chosen to host the website and provide computing power as needed. While the Board is only a small agency, initial estimates called for a savings of \$750,000 in the first year. One can easily see the significant cost savings for the federal government as larger agencies follow the Board's lead.

The cloud produced more than cost savings for the Board. Users began getting faster service, energy was conserved, and the Board's team of contractors and staffers were able to focus more intently on the core mission of providing rich content on Recovery.gov. And since the Board no longer had to invest in infrastructure, computer hardware and software assets were redeployed to the Board's oversight mission. Finally, by adding the cloud provider's security platform to the Board's own security system, the overall security posture was significantly enhanced.

**Continuous monitoring.** Other new technologies can also play key roles in adding predictive analytics. For instance, when an enterprise is considering obtaining a big-data device they would be well-advised to first ensure that their data are "cleansed" and put in shape to migrate from the legacy system to the new appliance. In addition, the concept of continuous monitoring has helped reduce the role of human analysts to perform predictive analytics. By leveraging the new big data systems, the continuous monitoring process eliminates the need for additional interpretation or analysis before taking action. This concept is now being used successfully at the Defense Department to help reduce improper payments.

## Conclusion and Recommendations

There are now a number of challenges preventing the government from taking full advantage of the potential power of predictive analytics to make the award of federal funds more effective and efficient. The lack of data standards across government, for instance, prevents the maximum use of predictive analytics.

Early on, the Board realized that there was no single numbering system (award ID) that agencies assign to contracts, grants, and loans. Because there is no requirement that award IDs must be standardized across government, it is extremely difficult to harmonize spending data. The various numbering schemes can result in duplications, errors in reporting, and other discrepancies. Until this problem is addressed, it remains the principal obstacle to government spending transparency and accountability.

**Recommendation One:** To effectively track the money and use data to make better-informed decisions, government will need to reevaluate how its databases interact with and leverage each other. The government should now consider adopting the Board's template of a cohesive, centralized accountability framework to track and oversee spending. One government-wide accountability platform would help establish a consolidated solution that would be less costly and more effective than multiple agencies pursuing isolated efforts.

**Recommendation Two:** Many federal agencies throughout government should consider the use of predictive analytics to make smarter decisions up front, and reduce the time currently spent on reacting to problems after award. The potential for predictive analytics to revolutionize government decision-making extends way beyond spending. Technologies like big-data appliances, geospatial services, and cloud computing are just as germane to medicine, transportation, and purchasing organizations as they have proven to be with the Recovery program.

*Earl Devaney is the retired chairman of the Recovery and Accountability Transparency Board (RATB). He served as Inspector General of the Department of the Interior. He also served in the U.S. Secret Service and the Environmental Protection Agency.*