Title:	**Winning against Tax Fraud with Data Analytics**
Category:	Data Management, Analytics & Visualization
State:	Illinois [Illinois Department of Revenue]
Contact:	Marty Johnson, IDOR, Chief Economist (Acting)
	Debbie Price, IDOR, Chief Information Officer
	Kevin Richards, IDOR, Accounts Processing Program Area
	Kendra Banning, IDOR, GenTax Business Lead

Project Initiation Date: 1/01/2017
Project End Date: 12/31/2018

EXECUTIVE SUMMARY

State government agencies are increasingly asked to "do more with less", as pressures on expenditures and resources require them to use every tool at their disposal to make efficient, effective, and meaningful decisions. Data and advanced analytics have emerged as cornerstones that enable agencies to achieve higher performance across their organizations while reducing operational costs. Effectively using data and advanced analytics allows agencies to make the right decisions the first time by using a sophisticated, data-driven understanding of behaviors and predicted outcomes to guide more accurate decisions. For revenue agencies, the need to enhance operational performance is also coupled with the imperative to increase revenue collections as much as possible.

The Department of Revenue of the State of Illinois [IDOR] is a 1,500+ person agency that services more than six million individual taxpayers and collects over $46 billion in total taxes. The Department administers more than 70 tax types; annually, it processes over 18 million returns, of which 6.1 million are individual income tax returns. IDOR currently uses Fast Enterprises' GenTax system to accept, process, store, and analyze tax returns. GenTax is used by approximately 30 states and US municipal tax agencies.

In 2017, IDOR undertook a major data analytics initiative with the express purpose of increasing revenues and improving compliance with the state tax laws. The vision was to increase efficiency within taxpayer services, to instill a culture of data-driven decision-making. A cross-functional, cross-agency team assembled to convene the effort drawing from the areas of informational technology, business processing, fraud, taxpayer compliance, and research; it involved subject matter experts from IDOR, data scientists and programmers from DoIT, and programmers from our in-house tax administration software firm, FAST Enterprises.

The roadmap for IDOR's data analytics effort was informed by the fact that FAST Enterprises had recently developed an analytics module with the hope that IDOR would be an 'early adopter' of the new technology. In addition, senior leadership at IDOR had expressed commitment to standing up a data analytics practice. The rollout was expected to span three phases. Phase 1 entailed working in tandem with the programming team from FAST to assimilate their knowledge and develop a full understanding of the tool's capabilities and their supporting product offerings. This initial phase, underway in early 2017, extended thru December 2017. Phase 2 – completed in December 2018 – involved supplementing the Analytics Manager tool with algorithms developed by the Data Analytics Practice to augment the tool, including implementing and evaluating the predictive models crafted during Phase 1.

Results from the predictive models developed at IDOR have exceeded expectations: in a press release from February 2019, it was noted that "Since 2015, Illinois has stopped more than $120 million in attempted identity theft and tax refund fraud thanks to increased data analytics and partnerships… 2018 anti-fraud and identity theft detection and prevention efforts led to an increase of $37.2 million in verified savings when compared to 2017."

CONCEPT

Data analytics is a high priority agenda item for the State of Illinois. Sophisticated, leading edge analytical techniques and applications can help increase Illinois state revenues and close the tax gap. Increasing tax revenue can be accomplished by applying advanced analytics to both preventing non-compliance and promoting compliance. These analytical opportunities can be implemented in areas across the value chain - prior to tax return filing, during return filing, and after return filing.

The seeds of this new initiative were planted more than four years ago, sparked by a need for a more data-driven, accurate fraud detection process – a way to root out questionable tax returns and refund requests. Historically, erroneous – and even fraudulent – refund requests were typically detected and acted on only after refund checks had been sent. On top of being time-consuming and costly, efforts to recover those refunds often proved fruitless.

How could the Department catch and rectify such refunds before they went out the door? The answer lay in the sophisticated business rules the Department had already employed for the processing and perfecting of tax returns. Previously, the analytics had only been used well after returns were processed and refunds had been issued.

The use of predictive intelligence to dynamically determine when to process a refund request and when to set it aside for further analysis was enabled by a new component embedded in the GenTax software dubbed the Fraud Manager.

Beginning in 2017 (baselined in 2016), the GenTax Fraud Manager fundamentally repositioned the role of business analytics in the tax return process. Because the Fraud Manager embedded business rules directly into the mainstreamed return process, the Department was able to automatically review each of the 6 million personal income tax refunds it receives annually. The rollout of the Fraud Manager positioned IDOR to identify an additional $19 million in fraudulent or erroneous refunds, which tripled the previous years' fraud prevention efforts.

However, it was recognized early on that, despite the accelerated progress made possible by the automation within the Fraud Manager, manual processes dominated the workloads of the fraud investigators. Data analytics could be better leveraged to simplify the laborious manual processes with both predictive capabilities and automation.

In early 2017, IDOR established within the Research Office a Data Analytics Practice [DAP], charged with working hand-in-hand with the FAST Enterprises data team in the rollout of a new GenTax platform. This platform, called the Analytics Manager, operates within the GenTax environment, and was customized to meet the requirement of IDOR's existing infrastructure and design. The Analytics Manager taps into the SQL Server Analysis Services, providing underlying statistical tools such as regression, decision trees, association rules, and neural nets.

This foray into the world of analytics was truly a partnership involving IDOR, Fast Enterprises, as well as the State Data Practice, which is a bureau within the State's Department of Innovation & Technology [DoIT]. IDOR subject matter experts within the Accounts Processing bureau and the Fraud Prevention unit played a major role in identifying the potential 'quick wins' and business use cases; the newly-formed DAP contributed a data analyst and the Chief Economist; the State Data practice provided the data scientist, as well as programming support. Throughout the spring and summer of 2017, the assembled team met on a regular basis to uncover the most critical business questions, and to identify those areas ripe for the greatest return on investment. This joint effort yielded two key use cases, both involving individual taxpayer noncompliance.

The focus of the first model, dubbed the 'K-12 Education Credit Model', was to identify the predictive characteristics of taxpayers claiming the education credit who do not respond when supportive documentation is requested. By Illinois statute, taxpayers with dependents are allowed a credit against individual income taxes paid, based on a portion of education expenses paid for the taxpayer's dependents. Historically, IDOR tax processors have been expected to manually review the education credits claimed on a tax return, a laborious process subject to much guesswork. In the event of a suspicious return – identified because the credit claimed appeared outside an acceptable range – the taxpayer received a letter requesting support for the claim. The ensuing back-and-forth between the taxpayer and the IDOR employees was time-consuming, stressful, and frequently concluded with no ultimate change to the taxpayer's awarded education credits. The 'K-12' model found that certain taxpayer characteristics correlated with suspected misuse of the education credits; a taxpayer's adjusted gross income, earned income tax credit amounts, and the ratios among the credits vs. claimed education expenses were the most predictive variables.

The Anti-Fraud Team at IDOR worked with FAST, DoIT, and the DAP team to create a second model designed to detect and prevent fraud associated with the earned income tax credit (EITC). The focus was to create a model showing the predictive characteristics of taxpayers and tax preparers who fraudulently report business income or loss in order to manipulate adjusted gross income – which in turn potentially maximizes their credit and refunds. As with the 'K-12' model, the IDOR fraud investigators had, before 2017, pursued these cases manually, reviewing each tax return one-by-one and initiating correspondence with the taxpayer. Through an intense, months-long process, the Analytics Team, using various datasets spanning several tax years, winnowed down the list of predictive variables, which among other indicators included having a professional license, receiving nonemployee compensation, and a taxpayer's adjusted gross income.

Implementation of these models was underway by January 2018, to astounding effect. The 'K-12' model was migrated into the production environment of GenTax, and though IDOR was appropriately cautious in evaluating the impact of false positives (which could run as high as 37% in pre-testing), once in production, the false positive rate was just over 17%. As an additional 'boost' attesting to the efficacy of the model, of the nearly 12,000 taxpayer letters sent throughout calendar year 2018 requesting support for the claimed credit, only 18% of this group could corroborate their initial claim.

Results for the 2018 calendar year were an additional money savings of $4 million from the 'K-12' model, with a corresponding reduced burden on the taxpayer customer services team (in the form of reduced times for telephone conversations as well as reduction of written correspondence).

Regarding the earned income tax credit model: by late 2017 it was determined that this model could not be adequately migrated into the production environment for the 2018 tax season. Instead, the predictive data elements we learned from modelling were incorporated into the Fraud Manager as a series of business rules. These business rules were in place by January 26, 2018 – the start of the 2018 filing season.

The results for the initial 12 months of implementation, once again, exceeded expectations: of the 84,000 letters issued to taxpayers requesting further documentation supporting their business income (or loss, in certain cases), 84% of the notice recipients either did not respond or could not support the originally claimed credit amount. The translated into <u>real</u> savings for the 2018 calendar year: predictive analytics led to an increase of $37.2 million in verified savings when compared to 2017.

SIGNIFICANCE

The Internal Revenue Service has estimated that over the last 30 years, the tax gap, or the difference between total taxes owed against taxes paid on time, has fluctuated between 15% to 18% of total tax liability at the federal level. The tax gap arises through the underreporting of tax liabilities, underpayment of taxes due or "nonfiling" of required tax returns. Taxpayer error (or misinformation) drives a portion of this noncompliance; however, fraudulent activity is the primary contributor, and takes the following forms: deliberately underreporting or omitting income, overstating the amount of deductions, keeping two sets of books, making false entries in books and records, claiming personal expenses as business expenses, claiming false deductions and hiding or transferring assets or income. The latest estimate from the Internal Revenue Services suggests that, at the federal level, the tax gap is upwards of $458 billion.

While the tax gap at the state level is more difficult to pinpoint, curtailing it can pay huge dividends to state coffers, which can then reduce the need to raise taxes. One important method for reducing the tax gap is data mining, which is the process of extracting information from large sets of data in order to analyze relationships in the data. Data mining models come in two general forms: descriptive models summarize patterns and properties in a dataset, and predictive models create a model based upon data in order to make predictions.

The impetus for undertaking the data analytics initiative at IDOR was three-fold:

1. Make better decisions based on the data IDOR already has.

The Department of Revenue has invested in information systems that now contain a significant volume of digital data going back over thirty years. The data analytics initiative has pushed IDOR to take the next step by leveraging data collected within its tax administration platform, GenTax, as well as synthesizing data from legacy systems, to enable agency-wide strategic decision-making through data mining to reveal information that leads to new collections and revenue opportunities.

2. Do more for less.

Insight derived from analytics promotes greater impact with fewer resources, targeting resource allocation on what the data can tell us, enabling a potential recalibration of investment priorities based on specific program impact and effectiveness.

3. Minimize Illinois' exposure to fraud, waste and abuse.

The opportunity to perpetrate fraud is a reality that IDOR is confronting directly, armed with the proper resources: people, processes, technology, and data. Tax fraud in Illinois is of special concern because of the potential scale of the problem and ability of those intent on perpetrating an abuse of the system can operate if countermeasures were not in place to prevent such activity.

The result is an undermining of the public trust and a substantial loss in revenue, exacerbating the state's budget problems and further compromising the state's ability to effectively serve its citizens.

IMPACT

The operationalization of Fraud Analytics ignited a watershed transformation for the Illinois Department of Revenue. It provided cross-functional recognition, proving that data can be used to drive real dollar savings, meaningful changes in policy and procedure, and ultimately that data can be used as a foundation to better serve our citizens.

Real Year-over-Year Dollar Impact
Instead of sending 'refunds' on fraudulent claims, the following dollars were stopped before leaving the department:

- 2016 Anti-Fraud Results (baseline, ahead of full-blown analytics initiative) Verified Money Savings: $19,164,454

- 2017 Anti-Fraud Results: Total of the 2017 Verified Money Savings: $28,815,559

- 2018 Anti-Fraud Results: Total 2018 Verified Money Savings $68,461,457

Improved Modeling Effectiveness
The Year-over-Year outcomes are a tribute to the improved effectiveness upon which the data analytics models were applied. The complex modeling includes the MeF (modern electronic filing), uncashed/undelivered past refunds, ID theft alerts, business income, earned income credit, business loss data, questionable dependents, education, property, and other data points. The initial model was successful, but the team continued to evolve the model. It was operationalized in 2018, producing the substantial real dollar return above.

Cross-functional Team Growth
The data points utilized in the Fraud Analytics model are managed by multiple areas within the Agency. To create and evolve the model, data discussions and sharing broke down many silos and created new understanding and teamwork throughout the Agency and even outside of the Agency, including: the data warehouse team and the Department of Innovation & Technology.

Enhanced Taxpayer Experience
The implementation of Fraud Analytics embraced a deliberate effort to move from an enforcement culture to a service-oriented culture. 'Return Correction Notices' (RCNs) did not emphasize legal language and cite law, instead, they simply stated that the refund was on hold until further information was received. Compliant taxpayers could easily follow-up with the documentation. Non-compliant or fraudulent taxpayers could not – halting dollars and leading to the annual verified savings above.

Foundational Impact - Modeling the Future
Using data to improve outcomes at the Illinois Department of Revenue, especially reducing fraud and abuse of taxpayers,, has proved fruitful and promising. The success of the original model, now operationalized, along with an enlightened staff who thoroughly embrace data-driven decision making, has created an optimal foundation for the State of Illinois to drive outcomes that all its citizens deserve.