Data analytics and the fight against corruption

Until a few years ago, public institutions had to struggle with the problem of access to information. Apparently, the efforts to overcome such situation generated another problem. Nowadays, there is a vast array of information available and, consequently, the need to find the needle in the haystack. In this context, what is the importance and the benefits of data analytics?

This is really an evolutionary issue for public institutions. Although it is indeed the case that there were prior problems, I don't think the advent of Big Data in and of itself will eliminate every access and data quality issue faced by the institution. What has changed is the relative abundance of data, the dramatic reduction of the cost of storage and computation, and new ways of thinking and methods associated with efficiently extracting value from the data. As such, data analytics has rightly become very important to the business of public institutions, especially for those who employ creative, problem-solving approaches to their mission challenges

Timothy Persons

Chief Scientist of the United States Government Accountability Office (GAO)

Dr. Timothy M. Persons is a member of the Senior Executive Service of the U.S. federal government and was appointed the Chief Scientist of the United States Government Accountability Office (GAO) in 2008. In this capacity, he serves to direct GAO's Center for Science, Technology, and Engineering (CSTE), a group of highly specialized scientists, engineers, and operations research staff. In these roles he directs science and technology (S&T) studies and is an expert advisor and chief consultant to the GAO, Congress, and other federal agencies and government programs on cutting-edge S&T, key highly-specialized complex systems, engineering policies and best practices, and original research studies in the fields of engineering, computer, and the physical and biological sciences to ensure strategic and effective use of S&T in the federal sector.

During this interview to ***Revista do TCU*** Dr. Persons talked about the importance of data analytics to the work carried out by an audit institution, especially when it comes to curbing corruption and fraud. In his opinion, a Supreme Audit Institution (SAI) who remains behind on implementing data analytic approaches will face limitations in its effectiveness in finding and fighting fraudulent activities. Although believing that the future of auditing will be almost entirely data analytic in nature, he emphasizes that technology will never replace the collective wisdom of experienced professionals.

coupled with a subtle, but important shift in their view of data as an expense/problem to be managed to data as an asset/opportunity to create value.

Is data analytics just a temporary trend or the key to solve from basic to sophisticated problems?

There's nothing temporary about data analytics. For example, here in America we utilize the word "Google" as a verb (I 'googled' TCU to learn more about it as an institution). Although it will certainly not be the answer or approach to all problems, data analytics done well does broaden the available solution space for certain types of problems which involve goals such as improving efficiency, providing more incisive insight on root causes, and supporting more informed decision making.

Corruption and fraud are major problems faced by Brazil and, therefore, major challenges presented to the Brazilian Court of Accounts. In your opinion, how data analytics may be used by the Court in order to tackle fraud and corruption?

From my perspective, fraud and corruption are effectively cost-benefit analyses done by people who might fall into the temptation of bettering themselves at the expense of others. As such, the rise and proliferation of effective anti-fraud data analytics approaches has increased the risk and cost of being discovered doing nefarious things and subsequently being prosecuted thus effectively suppressing fraudulent behavior. In this same manner, I imagine that the Brazilian Court of Accounts would very much benefit from fraud analytic approaches

"I believe SAIs who are faced with significant corruption-fighting challenges would do well to leverage existing technologies and approaches as part of their arsenal of methods."

such as linking software for social network/tracking funds kind of analyses, geospatial analysis, text mining of large datasets for sense-making purposes, among others.

How important is it for Supreme Audit Institutions (SAIs) of countries in great need of fighting corruption to invest in projects related to data analytics?

SAIs who remain behind on implementing approaches will face limitations in their effectiveness at finding and curbing fraudulent activities. It is indeed possible for a given SAI to preserve a business model which uses conventional methods alone, but I wouldn't expect it to be as effective as those SAIs which add to conventional methods augmented or assisted by newer data analytics approaches (some of which have reported dramatic improvements in the effectiveness of their revised and updated business model). As such, I believe SAIs who are faced with significant corruption-fighting challenges would do well to leverage existing technologies and approaches as part of their arsenal of methods.

Could you mention a case/situation in which data analytics made the difference in curbing corruption and promoting transparency?

The U.S. Department of Agriculture (USDA) saw a dramatic drop in fraudulent claims for crop insurance after they implemented data analytics approaches. Their Crop Insurance Program Compliance and Integrity Data Warehouse utilized multiple datasets to prevent fraudulent claim payments which resulted in billions of dollars in savings. Specifically, USDA data analysts utilized access to 170 data sources - including several terabytes of policy information, 120 terabytes of weather, satellite and other remotely sensed data, and 1.3 million crop insurance policies across 3,200 counties – to look for atypical patterns among insurance claims, cross-checking them with data from high-solution satellite images and weather records. Their approach involved both conventional prosecution activities as well as newer, "softer" approaches where letters of inquiry were sent to claimants who were suspected of fraud. This latter method resulted in a subsequent dramatic drop in claims simply because the participants in the program quickly became aware of USDA's new ability to detect fraud or suspected fraudulent activities.

The use of information obtained by means of data analytics may be limited to either open or shape an investigation or do you believe it could be used as a piece of formal evidence?

I believe that these new methods will need to be court tested in time such that they will evolve to where they can be utilized as

formal evidence in a trial. That said, I believe they can and should be utilized to shape or inform investigations. Once the newer methods are tested and refined through many cases for tipping and cueing kind of tasks, I think we will begin to see them introduced in due time as formal evidence in court cases. (*cf.*, Daubert standards utilized here in the U.S.)

One of the duties of the Brazilian Court of Accounts is to assess whether a governmental body is achieving economy, efficiency and effectiveness in the employment of available resources. How data analytics could be used to the benefit of a performance audit and help to shape public policies.

In my view, performance auditing practice and methods are prime candidates for exploration and piloting of advanced analytics. Here at GAO, we are well known for our conventional analytic methods and we are in the process of considering and piloting some of the newer approaches to see what, if any, things we might be able to do to support the economy, efficiency, and effectiveness of various public sector operations…including our own. As in other problem sets, data analytics methods may or may not add value to various activities of an SAI. Even so, in my experience, when such methods have added value they usually do so in a transformative manner. Once again, I think we'll need to keep track of how SAIs are using new analytics approaches and remain connected and share lessons learned about what works and what doesn't as we walk this path together.

"The rise and proliferation of effective anti-fraud data analytics approaches has increased the risk and cost of being discovered doing nefarious things and subsequently being prosecuted thus effectively suppressing fraudulent behavior."

Could you mention a practical example?

One of our earliest pilots involved text mining of a thousand documents to sort and cluster them according to a pre-defined taxonomy based on the audit objectives. By having an algorithm sort the documents according to topic (for example, some documents of this federal grant program were related to funding, some were technical in nature, and some contained geospatial information). As a result, a large amount of analytic time was saved by using this newer approach since, based on the questions of the audit, the analysts could read only that subset of all the documents that were relevant to them. Our pilot demonstrated that this approach (tested on a previously completed job using conventional methods) resulted in a reduction of weeks to hours of analytic time to achieve the same results.

Is data analytics the future of auditing? Would you define it as a decision-making system or a decision supporting system?

I believe the future of auditing will be almost entirely data analytic in nature. Specifically, data analytics will transform the accountability business wherever "gut" instincts or audit process inefficiencies lay. That said, I don't believe data analytics will ever be a decision-making system in and of itself (i.e., it will never replace the collective wisdom of experienced people), but will remain a decision supporting system (i.e., it will support more efficient and more effective data driven operations.)

What are today's cutting-edge technologies that a SAI must have in order to perform its duties?

From my perspective, SAIs who wish to conduct advanced analytics would do well to utilize data visualization, statistical computation, link and network analysis, geospatial analysis, text mining, document clustering, and optical character recognition software packages. Supporting infrastructure for data analytics involves the use of structured and unstructured databases as well as cloud and/or Hadoop architectures (depending on your data access and management policies). As always, the technologies should serve the mission elements and not vice versa.

Could you mention any limitations that still need to be overcome in order for data analytics to provide better results for a SAI?

Although there are technical (i.e., ICT architecture) and methodological (i.e., data reliability) challenges which would need to be addressed by an SAI desiring to develop and operationalize data analytic approaches, the primary limitation I have noted is cultural in nature. This is to say, institutio-

nal culture might resist new methods because, for example, they are unfamiliar (or may have been developed externally), they often require working dynamic, socially-integrated teams, and are likely to have a high rate of failure when attempting to evolve new analytic methods. The solution to such cultural barriers involves strong, consistent leadership supporting the new methods (i.e., "tone at the top"), the enablement of creative, problem-solving kind of thinking (often found in the more junior staff who are likely to be more open to the new methods), and a willingness to experiment and be wrong at times, but to learn quickly from each failure such that lessons learned are incorporated into the next iteration. Building symbiotic relationships with mission partners within an institution is also a *sine qua non* of data analytics.

Which SAIs are presently more advanced in the use of data analytics? What did they do differently than others that are not so advanced?

Although I'm not familiar with all SAIs, much less the extent to which they utilize data analytics, the ones who participated in the recent joint GAO (USA)/CNAO (China)-hosted event included (in addition to the hosts), Brazil, UK, Canada, the Netherlands, Norway, South Korea, and New Zealand. Indications from the conference were that many of these SAIs are building capacity and creatively enhancing their baseline capabilities to enhance the performance of their institutions. What struck me as a common theme from the proceedings was their willingness to think in an innovative context by essentially asking themselves the following question: "*How might we dramatically improve or add to our baseline capabilities with these new methods and tools?*" Some are succeeding because they have to… they don't have an abundance of staff to employ more manual or labor intensive methods. Others have been allowed to innovate through exploration of methods and a greater openness to what the new methods (including having a high tolerance for trying things that don't work, but adapting quickly and incorporating lessons learned in an agile process of continuous testing and development.)

> *"I have been impressed with what TCU has been doing in data analytics and regard much of their efforts as pioneering for the global accountability community."*

How important is the role of Chief Data Officer (CDO) in a SAI? What should be expected from such desk?

I think the INTOSAI community will see an increase of SAI CDOs over time, especially as the community evolves to recognize, embrace, and deploy the new approaches. In that way, I think the role of the CDO will be seen as increasingly important, even necessary, for more efficient and effective operations of SAIs. One critical distinction will involve making sure that the CDO isn't confused with or subsumed under the CIO. CIOs have a very important role for the management and security and ICT infrastructure and should be integrally partnered with the CDO, but is generally not designed to perform the functions of a CDO in terms of analytics with the data. The CDO should be expected to develop functional, symbiotic relationships with the desired mission elements of the SAI and should be allowed to work in an innovation context which allows them to safely fail, but learn and adapt to find workable operational solutions.

Considering your knowledge of the work performed by the Brazilian Court of Accounts, would you say the Court is in the right path when it comes to data analytics? What would you evaluate as a good move and in what areas the Court is still lagging?

I have been impressed with what TCU has been doing in data analytics and regard much of their efforts as pioneering for the global accountability community. Given the top-level support from the President and the extensive research into developing and deploying capabilities (including travels and interviews with experts), and the success of some early pilots which they shared at a recent conference, I believe TCU is exhibiting the elements of long-term, sustainable success in the use of data analytics in support of their various missions. As I understand the way forward for TCU, I applaud the institution's embrace of the overall datafication of Brazil something that, if leveraged properly through appropriate data analytics, should yield ongoing benefits to the Brazilian taxpayer. The challenge then will be to address the data reliability and methodological verification of the analytics in order to increase confidence, assurance, and efficacy of the new methods.