

# Start small, think big: big data!

**By Ziv Baida,**

BUSINESS DEVELOPMENT DIRECTOR, EUROPEAN GOVERNMENT SECTOR, DUN & BRADSTREET

NINETY-THREE PERCENT OF federal government respondents to a recent survey by Unisys Corp. said that the quality and speed of decision-making improved when applying data analytics; yet nearly 70% reported that they are concerned about their agency's ability to analyse key data rather than simply collect it. One of the reasons for such concern is the 'information explosion,' a term that describes the rapid increase in the amount of available information or data, introducing the challenge of managing the information and making sense out of it.

Giving up is not an option because data is a key asset, or should become a key asset, for every government agency. And while the complexity of managing huge amounts of data should not be underestimated, neither should it be overestimated, because many of the challenges that Customs and border management agencies face can be solved with data-solutions of very manageable complexity. Not every data project is necessarily a 'big' data project.

## Key terms explained

Let us take a step back, and understand key terminology, starting with the distinction between data and information. Data is plain facts. Digitization and the abundance of sensors and 'smart' technologies resulted in an explosion of available data. Yet when people look for more data, in fact, data is a means, not an end.

People seek to make confident decisions based on insights gained from information, which is data that has been processed, organized, structured or presented in a meaningful context. Analytics is the creation of insights from data using systematic computational analysis. Big data refers to analytics undertakings that exhibit complexities along four dimensions that have

been coined by IBM as the 'Four V's of Big Data:'

1. Volume refers to the scale of data;
2. Velocity refers to the analysis of real-time streaming data, for example in a stock exchange, or in vehicles equipped with sensors;
3. Variety refers to the analysis of different forms of data – structured and unstructured data, text, audio, video, sensor measurements, social media, and more;
4. Veracity refers to the need to deal with the uncertainty of data – uncertain data quality, uncertain availability, and completeness and correctness.

**(Big) data for Customs administrations**  
Imagine if you could...

- predict which companies will be non-compliant;
- detect fraud through automated verifications, rather than labour-intensive inspections and audits;
- pro-actively be informed when substantial events occur at traders with authorized economic operator (AEO) status;
- create a helicopter view of everything you know within your agency about a company;
- seamlessly exchange information about traders with other border agencies;
- detect international organized crime networks operating at ports.

Easier said than done? Tackling some of these problems may be less complex than you think.

## What is (not) the challenge?

Pessimists will say that you first need a full-blown (3-year?) information management program before you can start implementing (big data) analytics, because analytics is only as good as the underlying data. While the latter is very true, the former is not the only possible conclusion. An

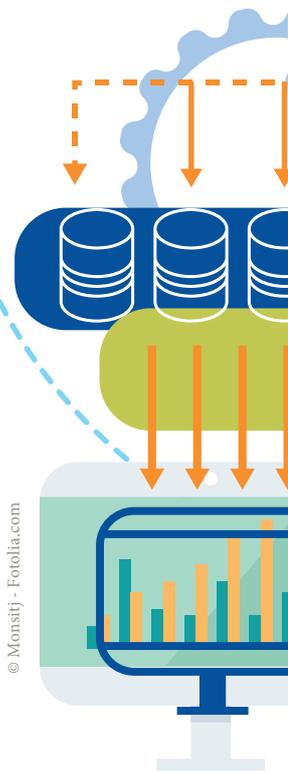
alternative conclusion is that your focus should be on acquiring and using high-quality data, rather than only the data that you have available (but which may not be accessible).

A comprehensive information management program is certainly needed as a long-term investment in the agility and effectiveness of your organization, yet waiting for its completion is no longer an option, given the current pace of change. Quite a few 'low hanging fruits' await your action.

The key to success is to:

- identify specific problems that you want to solve, for example non-compliance with safety regulations relating to fireworks imported before specific celebrations, or value-added tax (VAT) carousel fraud;
- identify the key data required to solve these problems;

© Monsitij - Fotolia.com





- acquire this data;
- embed it in decision points in your information technology (IT) systems.

One pitfall often encountered is that organizations only use data that is already available, either because they are not aware of external data, or because of a preference to keep things 'in-house.' However, availability and quality do not necessarily coincide. If you have not been able to solve the problems with the data that you have available inside your organization, acquire the data from external sources.

#### The role of business information

Customs administrations monitor the flow of goods through their borders to secure and facilitate legitimate trade, thereby stimulating national economies, collecting revenue, and protecting society.

While the role of Customs centres on the flow of goods, it is the companies behind the import/export transactions that are

key to fulfilling the role of Customs. Therefore, Customs requires the richest information possible on the relevant activities of companies involved in international supply chains.

There are large, commercially available databases of companies with worldwide coverage out there, performing millions of updates to their databases daily. Every single entity (company) within a database has its own worldwide unique identifier, allowing Customs to unambiguously identify companies involved in trade.

Once identified, this identifier is the key to unlocking rich value, including the most recent insights about the whereabouts of the company, firmographics (for example, sectors of activity, size, legal status, financials, etc.), corporate linkage, and various risk scores which analytics experts create by comparing the company to all its peers, using this rich historical data.

dun & bradstreet

Dun & Bradstreet (D&B) is the world's leader in business information (i.e. information about companies), holding the largest commercially available database of companies covering the world, with millions of updates daily. Every single entity (company) within the database has its own worldwide unique identifier, the DUNS® Number. Eighty-seven percent of Fortune 500 companies have been successfully using D&B data, integrating it into their core operations to provide critical data and insights, together with many government agencies.

The author is the Business Development Director for the European Government Sector at D&B. Over the past years, Dr Ziv Baida has actively participated in the development of Customs IT solutions in several countries and continents, at both national and international levels. He played a key role in piloting concepts such as Secure Trade Lanes, the Single Window, and AEOs, even before the AEO concept became operational in the European Union (EU). Dr Baida also played a key role in the design of Dutch Customs' import/export declaration management system. In recent years, he has focused on the potential of new technologies for Customs, including big data analytics for Customs risk management, mobile technologies for Customs inspections, and social business for internal and interagency collaboration.

### How can business information services help Customs, revenue, border management, and law enforcement agencies?

Customs typically know some of their local traders very well, while they know most of them somewhat. However, they usually know very little about foreign traders. Large business information publishers fill this gap, because they know these companies, as they have been observing their operations for years, and are therefore able to make a statistically valid statement about their operations in relation to their peers' operations, even if they appear as an unknown first-time importer/exporter to an agency.

Sometimes a simple verification with the service provider can reveal cases of fraud or identity theft, for example when a company trades under the name of a company which a provider knows is out of business. When integrated into government IT systems that process permit applications, VAT refund requests and import/export declarations, such simple verifications will yield substantial benefits.

In all these cases, qualitative company data supports information-based enforcement by providing a means of establishing that an unknown company is likely to be compliant (i.e. reducing the haystack), and by providing freight-targeting officers with signals for potential high-risk cases that should be inspected (i.e. finding the needle).

Another important 'use case' is the Single Window (SW). The company unique identifier number is a means of uniquely identifying companies that form part of a SW environment where multiple government agencies each have their own identification number for a company. Thus, a single unique identifier enables coordinated border management (CBM).

### Short- vs. long-term vision

External company data is available through Application Programming Interfaces (APIs), and can easily be integrated into automated IT systems without major prerequisites in terms of an IT environment. Experience has shown that very substantial benefits can be realized: a reduction in fraud; an increase in revenue collection; improved security; and greater efficiency in utilizing scarce skilled employees.

These solutions can be implemented quite rapidly, to reap short-term benefits, while allowing an organization time to develop data analysis skills, and to undertake a thorough review of its information management program across its various IT systems and databases. An investment in these long-term initiatives is paramount for long-term success in big data analytics.

An information management program will create an information infrastructure, whereby data collected anywhere in the organization – including external data – is available anywhere within the organization. The long-term vision thus explicitly foresees the combination of internal data with external data, because some insights can only be created when these two data sets are combined.

The use cases described above (first-time importer/exporter, identify theft, etc.) require external data. Is it big data? Sometimes it is, but not always. In some cases it suffices to obtain little, yet qualitative, data, and embed it in simple business rules and risk profiles connected to IT systems.

In other cases, a higher degree of analytics skills is required. If one does not yet have these skills, external data is also available on an 'Insights as a Service' basis, until one has developed sufficient skills internally.

Analytics and big data jobs are new to many government agencies, and like any other job, these jobs require specific skills that may not yet be available in an organization today.

Analytics and big data jobs are new to many government agencies, and like any other job, these jobs require specific skills that may not yet be available in an organization today.

Moreover, the use of analytics and big data techniques entails a change in how an organization works: a shift from 'gut feeling'-based inspections to risk-based inspections; and a shift from targeting officers reviewing most shipments to these officers reviewing fewer shipments. Skills development and a change management strategy go hand-in-hand, and their successful implementation will enable an agency to realize the long-term vision, where data has become a core asset in an agency performing information-driven enforcement.

### Lessons learned: how to succeed, or how (not) to fail

1. Start with a clear goal, with a clear business problem. Understand the scope of your problem. Do not engage with big data because 'everybody has to.'
2. Understand the vast potential. Engaging data is a business strategy, not a tactical or operational matter. Executive sponsorship is therefore key.
3. Insist on data quality. Available data is not necessarily qualitative enough.
4. Seek data outside your own organization, guided by the previous bullet.
5. Do not reinvent the wheel. Others like us have implemented what you aim to do.
6. Big is relative. What is 'small' for another agency may be 'big' for yours. Thus it is big!

And finally: Don't boil the ocean. Start small, think big: big data!

**More information**  
[BaidaZ@DNB.com](mailto:BaidaZ@DNB.com)