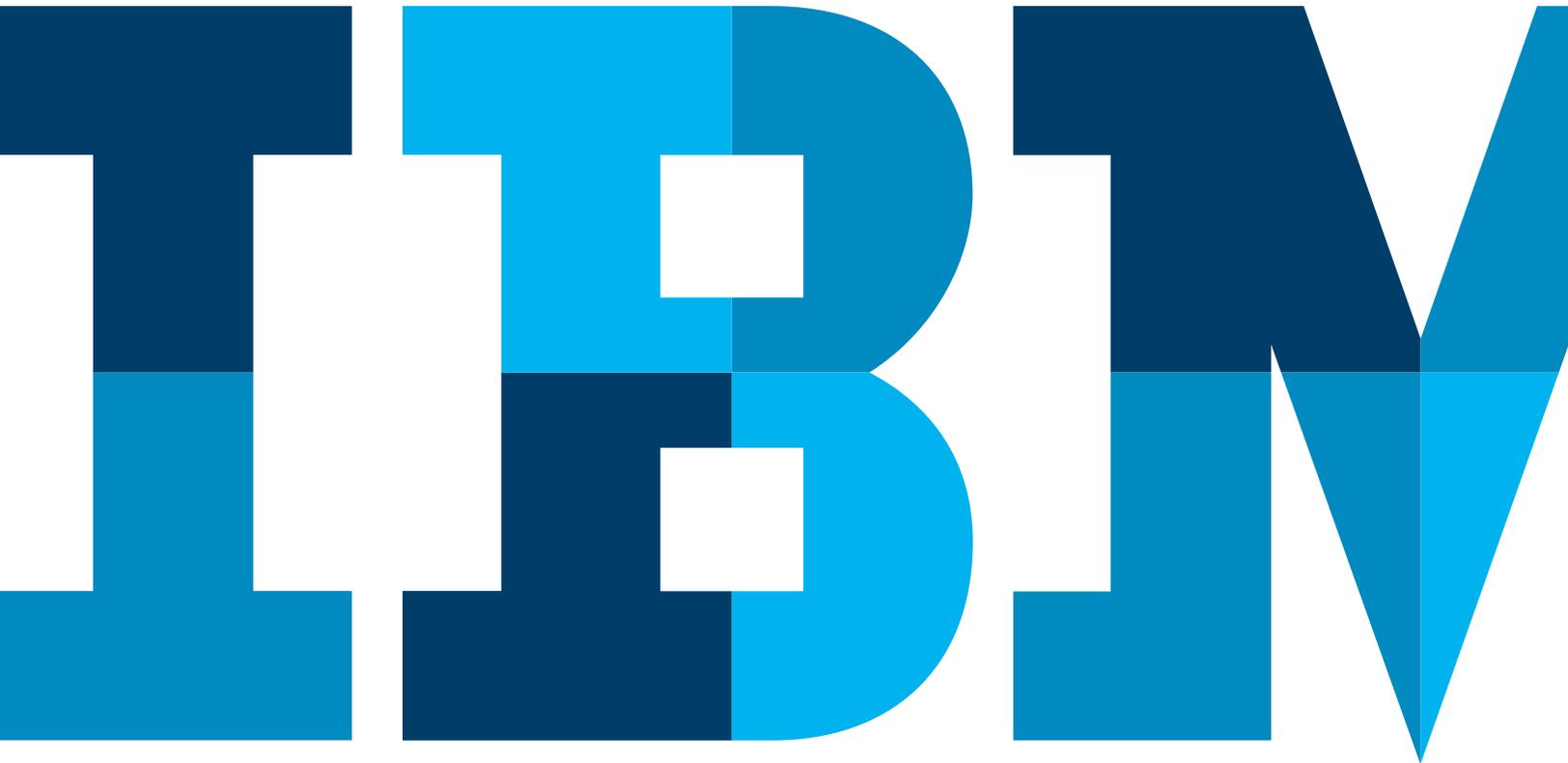


# Data-driven government: Challenges and a path forward



## Contents

- 2 Executive summary
- 2 What is data-driven government and what are the benefits?
- 3 Challenges
  - The privacy responsibility
  - What data?
  - Readiness for use
  - Managing data
  - Using the data
- 7 The path forward
- 8 Conclusion

## Executive summary

Governments have never suffered from a lack of data. In fact, one could argue that, historically, they have suffered from an excess of data, much of which was rarely used. Some countries maintained large warehouses for storing this information for the period specified by law or regulation. This information included personnel records of employees, court records, tax records, benefit eligibility records and more. After it was stored, most of it was probably never accessed again and when it was needed, often it could not be found. Additionally, staff had to be available to manage record retention requirements and destroy the data after the legal limits of its life had been reached.

The potential value of this data was recognized by some many years ago but it was very difficult to extract that value with the tools of that time. In this world of paper, people generally thought in terms of “files,” not “data,” and that the bits of information in the files had no value without the context of the entire file. How the world has changed.

Governments are now using computer technology to capture and harvest this data and extract value from it. However, more can be done to apply it to agency missions.

This paper explains data-driven government and challenges related to it. It then outlines a way forward in the journey to a data-driven government. Although the principles apply to all agencies, the focus of this paper is the government agencies responsible for revenue collection and distribution of social benefits. Generally, these are tax, customs and social services agencies, and they are referred to as revenue management agencies in this paper.

## What is data-driven government and what are the benefits?

Governments were among the early users of computing technology. Initial widespread use of this new technology emerged in the 1950s and by the 1960s, full information technology (IT) divisions were formed with centralized data centers and large data entry organizations appearing on the business side. From a data usage standpoint, two factors distinguished this period from the current state. First, rarely was 100 percent of the available data captured electronically. It was too expensive. Second, the nature of the technology meant that it was stored sequentially in “flat” files and until relational databases emerged, extracting any intelligence from the full set of information was a slow and complex process.

This brief history shows a journey toward data-driven government that has been in the works for a long time. Governments have been quick to see the advantages of computer technology and to adopt it when the business case for doing so was clear. For revenue management agencies, who have always been concerned about “leakage,” the earliest and still most advanced uses have been in the area of fraud and error detection. Because not all filings could effectively be audited each year, agencies were running their flat files against complex formulas as early as the late 1970s to determine which cases should be audited.

As fraud and error continues to be perhaps the greatest challenge to these agencies, technology continues to evolve to better meet these needs. But at the same time, most agencies are only in the early stages of applying analytical approaches to becoming more effective in other parts of their mission responsibilities, such as:

- Measuring the effectiveness and efficiency of service delivery channels
- Evaluating where maximum benefit could be obtained from the next investment (for example, whether having 10 more telephone assistants or 10 more auditors would have the most impact)
- Conducting more effective policy impact assessments
- Increasing the accuracy of projections for revenue, expenditures or services demand
- More complex assessing of social services client needs to provide a more effective solution on an individual basis
- Using social sentiment analysis to determine agency profile in social media and who the “influencers” are
- Detecting internal fraud and errors in procurement and expense reporting
- Adopting more effective management dashboards tailored to the needs of each layer of management

A 2013 study by IBM revealed that only 50 percent of managers made even half of their decisions based upon data and analytics.<sup>1</sup> The reason for so many decisions being made without analysis is that the data and analytical capabilities are just not available. Data-driven government is about solving that problem.

In short, a data-driven government is one where, for all critical decisions, actionable information is available when and where needed. The benefits are almost incalculable. A few examples include:

- Sounder governance and control
- Optimized fraud and error detection, mitigation and prevention
- Improved services based on insights gained from those being served
- Improved efficiency through intelligence networks, which can lead to reduced costs
- Improved public perception of the agency

This is smarter government, enabled to make better decisions.

## Challenges

Based on experiences, intuition, the technical press and public media coverage in general, perhaps the case for data-driven government seems obvious. So why hasn’t everyone done this already? Because even when the decision is made to become a smarter agency, many challenges remain.

### **The privacy responsibility**

Some consider privacy to be the proverbial elephant in the room. But to degrees that can vary, individual privacy is valued and the expectation that government be prudent in protecting it is reasonable. The purpose of this paper is not to debate what proper privacy policies should be. That is up to each government and the assumption in this paper is that each government would proceed with the recommendations in a manner consistent with their own policies.

Of note are those agencies that have been successful in making business cases for policy changes when a clear benefit can be shown while still maintaining prudent privacy protections. Additionally, citizen privacy expectations are very likely to change over time. Detailed information that many carefully protect is made publicly available by many in the younger generation though Facebook and other social media. What will their attitudes be as they move into the workforce?

### **What data?**

The amount of data available both inside and outside government that could be useful is growing at a dramatic rate. Yet, agencies responsible for revenue management could benefit from more data. But do they need and could they use all of it? Obviously not. The first area of emphasis for each agency is to capture all of their own data electronically. To lower the cost of data capture, many are not only offering electronic submission options but are also moving to grow those channels to the point of mandating them. The impact on citizens and businesses might create policy obstacles but, for most agencies, this will gradually go away. For many analytical tasks, such as measuring program effectiveness and constituent segmentation, internal data could be all the agency needs. But for others, such as fraud detection and revenue projection, data from outside the agency is essential for maximum effectiveness.

Next is the valuable data in other agencies of the same government. Many revenue agencies share among themselves or with other entities of their governments. But the extent of the sharing varies widely due to a variety of issues that include privacy and the technical capability of the agency to actually use the data. Examples of data already found to be useful are business license registrations, taxpayer or business compliance profiles and motor vehicle, boat and other personal property transaction records.

Valuable data also exists outside of government. A good example is commercial business information services, along with banking account information if privacy rules allow access.

A trend that has recently emerged is the increased sharing of government data over national borders. Although the case for such sharing in the customs arena is obvious, the case for doing so in taxation and social services agencies is growing rapidly. For many years, data exchange treaties between countries have existed; however, internal privacy policies that restrict what can be shared, along with language and data formatting issues, have severely limited the actual exchanges. Through the leadership of the World Customs Organization, the EU and the OECD, these obstacles are being addressed. And to be clear, cross-border issues do not only relate to being able to capture payment of the right revenue, but they also relate to compliance in payment of benefits to which an individual now residing in another country might be entitled.

*The US Internal Revenue Service has been using data analytics to enhance compliance activities for 40 years. They also use their data to understand taxpayer behavior, educational needs and services needs based on categories such as occupation, age group and postal code.*

Another increasingly important data source is unstructured data. Not many years ago, the only value of technology for unstructured data was that it enabled agencies to store the content and retrieve it for reading while destroying the paper. Internal unstructured data can now be captured and analyzed with modern tools. Additionally, initial work is being done by some agencies to harvest unstructured data from the Internet and social media for the detection of individuals who are encouraging or selling schemes to fraudulently avoid payments or to “maximize” government benefits.

The world is clearly in the age of big data.

### Readiness for use

Having a large quantity of data even at the internal level does not mean that this data is usable in its raw form. Data must be understood by analytical tools. For example, if a business is named Smith and Jones Enterprises, yet various systems label them as “Smith and Jones Enterprises,” “SME” or “Smith & Jones Ent,” unless the data is “cleaned” or analytical systems are trained to recognize that they are the same, they will be analyzed as separate enterprises. Likewise, an internal data dictionary is needed; otherwise different systems might have different definitions for terms such as “enterprise,” “benefit” or “child,” which can make cross-system analysis either inaccurate or extremely complex.

“Stovepipe” systems that do not have the capability to communicate with each other at all are also a challenge. The current trend is to implement integrated systems during major modernization efforts or to build bridges by using master data management tools to get a single view of the citizen. However, observations by IBM indicate that stovepipes are still the rule rather than the exception. Some agencies are even building new stovepipe solutions.

### Managing data

So governments have all this data and many parts want to use it. What is the most effective and efficient way to assure that the data is available when and where needed, fully in compliance with all privacy regulations and at a reasonable cost? This is the data management challenge. Many agencies in many countries have stories about dealing with multiple warehouses full of data. In at least one case, this exceeded 30, yet many of these warehouses contain much of the same data. In the early stages of a new technology and the necessary experimentation to determine its business value, the “I want my own” mentality of many divisions in each agency is to be expected. However, in the age of big data, this is extremely inefficient and only marginally effective.

One could easily argue that from this point forward, the most valuable asset any agency has, aside from its human resources, is its data. Yet very few agencies have a data management function. About 10 years ago, the role of a chief data officer (CDO) began to appear on company organizational charts and in consulting and research writing. But few are the government agencies that have established such a position. In these organizations, data responsibility is scattered between IT, privacy organizations, security and business units, and no one is managing data as if it were a valued enterprise asset.

Typically, the recommendations are that the CDO report to the CEO, but others say that the COO is an option also. A few organizations have put this function under the CIO. This positioning is based on the traditional view that data is an IT issue. Data is not an IT asset, however; it is a business asset. Wherever it is placed, the fact remains that data management is an issue that should be on the agenda of those government agencies that wish to maximize the value of this asset.

### Using the data

Data alone does not translate to data-driven government. An individual data element has little value beyond its applicability to its citizen, business or other subject. The higher value comes, not from the individual data elements themselves, but from using all the data to obtain insightful or actionable information and have it available when and where it is needed. Such value requires deep analytics. For as good as the data might be, only by using and presenting the data through an analytics layer can the deeper insights hidden inside be revealed.

Some agencies have analytical tools custom-built by their own staff or vendors, but the common practice is to use commercial products and solutions. Some provide the tools for building analytical models as standard components. But other analytical tools are needed to support a wider range of analytical needs. Some examples include:

- **Identity analysis** tools that determine who is who, who knows who and who has a family or business relationship with whom. These help in areas such as identify theft and establishing relationships between apparently distinct entities.
- **Visualization tools** that show analysis results in charts and graphs and on maps. Often, this provides more insight than simply seeing summary results in numbers.

- **Unstructured data analysis** tools enable detection of patterns or trends in internal unstructured files, Internet and social media data sources. They are used to identify those of interest to the agency or to dig deeper in the process of working a case on an individual citizen or business.
- **Cognitive analysis** tools are the latest analytical capabilities and provide for advanced analysis of both structured and unstructured data based on machine learning and intelligence that becomes smarter the more it is applied to the problem.
- **Information dashboards** extract and present the information needed for various purposes, such as management. This information generally comes from a range of sources.
- **Analytical tools for auditors and analysts** that, because analytical tools do not totally replace people, provide a workbench for individuals with responsibilities such as auditing or research.
- **Social sentiment analysis tools** that enable assessment of social media to detect patterns of what is being said about agencies so they can better understand their image in the community and the broader perception of their programs and services. The agency can adjust programs accordingly and to effectively manage their own social media presence.

These tools can all play a role in making revenue management agencies smarter. All will likely be considered and agencies will have the challenge of identifying those needed and how to integrate them to support a coherent and efficient analytics program.

**Using analytics in government**

A number of social services agencies, such as Alameda County in California and Medway Council in the United Kingdom, now use analytics to:

- Reduce improper benefit payments.
- Support employees with more complete and timely case information.
- Better target their client population.
- Deliver dramatically improved program results.

**The path forward**

So how does an agency move forward from the current situation? Although some new analytical capabilities could be selected from the options that have been identified in this paper and used to start a project, best practices suggest a more strategic course of action.

The first step is full consideration of the role data and analytics should play in the strategic plan for the agency’s future. Would you like to be an organization that has the right information at the right time for all the decisions you need to make? Who owns the data strategy currently? Do you have a data governance process or are data initiatives driven out of stovepipes with IT having to react accordingly? Should you have a chief data officer? These questions require strategic discussions at the most senior level of the agency and if those discussions have not taken place, they should happen soon.

An enterprise data strategy should then be accompanied by an analytics strategy. In what program areas do you want to apply analytics? What tools do you need? How do you assure efficiency in the acquisition and application of these tools? A best practice is to develop an analytics reference architecture to assist with defining the analytical services needed for the ideal end state. Figure 1 shows an example of such a reference architecture for a tax agency. This could be easily modified for a customs or any social services agency.

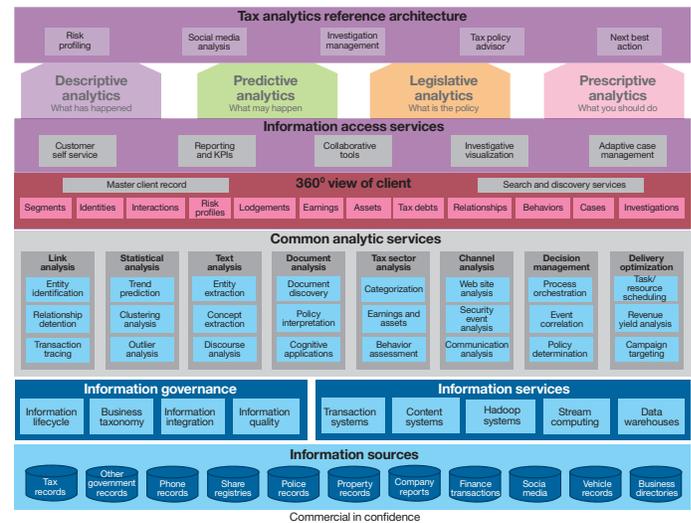


Figure 1: Tax analytics reference architecture.

The architecture starts with a foundation of the data or information sources and incorporates data management and governance along with the levels of services needed to get the desired analytical capabilities. This is the highest level description. Behind this description, each agency can build more detailed use cases, which would ultimately enable development of a technical analytics architecture that defines the tools and necessary integration to create the data-driven agency.

IBM understands that no agency has the budget to build this entire solution set quickly. However, many are spending new money on more projects and solutions each year with no vision for the investment that leads to a true enterprise analytics capability. Also, many, if not all, of the capabilities already developed during previous analytics projects most likely can fit into the framework and be used more broadly across the organization. The key point is that if you don't know where you are going, you don't know where you will end up. The purpose of developing the reference architecture is to make sure that, over time, each investment builds toward this ultimate enterprise capability.

## Conclusion

In a data-driven government, actionable information for all critical decisions is accessible when and where needed. The opportunities for better, smarter government through optimal use of data are clear as are the challenges. The time to start the journey is now.

## Author

**Terence (Terry) Lutes** is the Director, Global Revenue Management at IBM. He is also a member of the Board of Directors of the International Tax and Investment Center. He previously spent 27 years at the US Internal Revenue Service. His international experience includes serving as a US representative to the OECD and serving as a Principal at The M Group.



---

© Copyright IBM Corporation 2015

IBM Corporation  
New Orchard Road  
Armonk, NY 10504

Produced in the United States of America  
April 2015

IBM, the IBM logo, and [ibm.com](http://ibm.com) are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NONINFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

1 "Analytics: A blueprint for value," IBM Institute for Business Value. October 2013. <http://ibm.co/18NKegY>.



Please Recycle