

Authoritative Data and Data Governance

Volume 5, Issue 11

Office of Information and Technology (OIT)

Introduction

According to [Forbes](#), 90 percent of the data in the world has been generated over the last two years. The amount of data being generated every day is astounding. Consider these facts: There are now 1.5 billion Facebook users daily; [Google now processes more than 40,000 searches every second](#) equaling 3.5 billion searches per day; and every 60 seconds, 72 hours of footage is uploaded to YouTube, while 216 thousand Instagram posts are being made, and 204 million emails are sent, equaling [2.5 quintillion bytes of data](#). To get a sense of how much data the Federal Government produces, the Department of Commerce puts [out as much as 20 terabytes of data](#) every day. In this Tech Insight, we present an overview of authoritative data and the increased emphasis on data management and governance by examining its key definitions, and assessing the business value impact of trusted data for the [Department of Veterans Affairs \(VA\)](#).

Overview of Authoritative Data

An [authoritative data source](#), as defined by [VA Directive 6518](#), is a source of data or information designated and recognized as “official” that is **trusted, timely, secure** and **used within VA’s information environment** in support of VA. Codified in the same directive is an entry on [data governance](#) describing it as “the exercise of authority, control, and shared decision-making planning, monitoring, and enforcement over the management of data assets.” [Gartner](#) states that data governance is an “effective tool to manage and control the ever-growing amount of data to improve business outcomes.”

The [Data Governance Institute \(DGI\)](#) is the industry’s oldest and best-known source of in-depth, vendor-neutral data governance best practices and guidance. In 2004, DGI introduced the DGI Data Governance Framework in response to an emerging need for a way to classify, organize, and communicate complex activities involved in making decisions about enterprise data.

According to DGI, information technology (IT) does not exist for its own sake—IT exists to meet the information needs of the business. Data governance includes the processes and framework involved in managing data assets. For instance, organizations need to know how to define their data by understanding what “customer” means within different departments and how it should be managed overall.

Suzette Kent, Federal Chief Information Officer (CIO) [said](#) that by 2025, the government will produce at least 163 [zettabytes](#) of data — that’s more storage than 250 billion DVDs! “Agencies

need better automation tools, because they're producing way too much data for humans to process. If one of us tried to process a terabyte of data, we would have to watch the equivalent of 400 90-minute videos. But using technology, and with the right discipline around data, we can process that in seconds," said Kent.

International Business Machines (IBM) has been quantifying data for years and recently set a new world record by capturing 330 terabytes of uncompressed data — or the equivalent of 330 million books — into a cartridge that can fit into the palm of your hand. Several IBM Data Scientists say the convergence of four dimensions ([4 V's](#)) helps both to define and distinguish data as follows:

Volume: The mass quantities of data that organizations are trying to harness to improve decision-making across the enterprise.

Variety: Managing the complexity of multiple data types, including [structured, semi-structured and unstructured data](#).

Velocity: Speaks to data in motion; the speed at which data is created, processed, and analyzed continues to accelerate.

Veracity: The level of reliability associated with certain types of data. One in three business leaders say veracity troubles them the most as they [do not trust](#) the information they use to make decisions.

Help is on the way through the [President's Management Agenda \(PMA\)](#) regarding data, accountability, and transparency for transparent insights. Each Federal agency compiled an inventory of high-value assets as a deliverable as codified in the PMA. Other initiatives to focus agency resources in improving data management and governance include the following:

- The [Federal Information Technology Acquisition Reform Act \(FITARA\)](#) made it clear that Federal Chief Information Officers (CIOs) are responsible and accountable for any IT product or service consumed by an agency.
- The [Digital Accountability and Transparency Act of 2014, or DATA Act](#) is the nation's first open data law. The act requires the US Federal Government to transform its spending information into open data. The DATA Act took two basic steps. First, it required the Department of Treasury and the Office of Management and Budget (OMB) to establish government-wide data standards for the spending information that agencies report to Treasury, OMB, and the General Services Administration (GSA). Second, Treasury and OMB must publish this standardized spending data for free access and download.
- The [DATA Act Information Model Schema \(DAIMS\)](#) is the authoritative source for the terms, definitions, formats, and structures for hundreds of distinct data elements, which tell the story of how federal dollars are spent.

- On January 27, 2014, Congress shed its spotlight on the importance of data when the 113th U.S. Congress adopted [S. Res. 337](#), a non-binding resolution expressing support for the designation of January 28 as “[National Data Privacy Day](#).”
- The [General Services Administration GSA](#) launched "Data to Decisions" ([D2D](#)), a new-generation data and information management platform.

Benefits of Authoritative Data

Today, the Federal Government spends at least \$90 billion annually on data and is developing a [Federal Data Strategy](#) to leverage data as a strategic asset. The fruits of introducing greater data accountability and transparency has already been seen throughout several sectors.

Education and workforce outcomes: Most states are operating or in the process of establishing longitudinal data systems to inform policymakers on workforce outcomes of publicly funded education programs.

Combating opioid and addiction crises: Virtually, every state is in the process of formulating a strategy to monitor and reduce overdose-related deaths and improve addiction treatment and intervention strategies. Data at the local, state, and federal levels can support research and evidence-based policies.

Recidivism: Most levels of the government want to understand the causes of [recidivism](#) and identify effective strategies to reduce it. State and local governments can offer relevant data on housing, substance abuse, Medicaid, employment, education, and transportation.

Fraud, waste, and abuse: Efforts are currently underway in several states to leverage administrative data to ensure Federal and state funds are used appropriately, and those initiatives can be scaled through collaboration with the Federal Government.

Challenges of Authoritative Data

An organization’s data and analytics program faces four key challenges: establishing trust, promoting a culture of diversity, mastering the complexity of running a digital business, and building data literacy for the workforce.

In **establishing trust**, the heart of building confidence in your data is to improve data quality and to verify the information in customers’ systems. Gartner posits that crowdsourcing and automating metadata creation is a way to establish trusted data. For **promoting a culture of diversity**, Gartner states that more diverse data source teams perform better than homogenous ones. In **mastering the complexity of running a digital business**, Gartner says that organizations need to empower multiple small teams to leverage more precise data and analytics platforms that provide more content, more understanding, and more timely responses. When **building data literacy for the workforce**, as business strive to become more data-driven, they will need to establish a common language and culture around data. Gartner

says this can be achieved by providing training in context, creating a certification system for data professionals, and by leveraging augmented analytics.

Authoritative Data Sources At VA

The VA Office of Information Technology (OIT) has produced several products around data. There are seven Enterprise Design Patterns (EDPs) on [Interoperability and Data Sharing](#). Additionally, three Tech Insights highlighting authoritative data sources are published as well: [Data Transformation](#) (examples of how databases are transformed at VA), [Big Data Analytics](#) (how it differs from the traditional collection and storage of data), and [OIT's Open Data II](#) initiative.

In February 2018, VA [announced](#) a research partnership with Alphabet subsidiary, [DeepMind](#), to tackle issues concerning patient deterioration during hospital care. As the healthcare industry generates more data, VA sees an opportunity to share its own information with other partners for research purposes. VA envisions a secure, scalable, and reliable data management infrastructure for moving information across the entire VA network. To get there, VA has created a concept of “[data turnpike](#)” that enforces rules for data flow and usage. Managing data is one of VA OIT's [five modernization priorities](#).

Open [application programming interfaces \(APIs\)](#) accessed by the public via applications and software programs have resulted in an exponential growth of data. The growth of open APIs has given birth to the ubiquitous term “[API economy.](#)” VA is embracing APIs and has implemented a new project, [API Management Platform](#), a next-generation, open, digital platform that enables rapid innovation in core VA functions.

As VA launches the [Electronic Health Record Modernization \(EHRM\) project](#) – which will allow for seamless data sharing between the agency and community providers – the need for preserving data integrity will be paramount.

Conclusion

[Data is the lifeblood of the Federal Government](#). The decisions that flow from data will ultimately shape the future of agencies, increase efficiency in services, and ensure citizen trust in the government's ability to use new technologies like artificial intelligence (AI). In fact, the enterprises with the best people and data available to train an AI application how to do its job will create the most capable AI systems. Therefore, ensuring the integrity of data must be as much of a priority as implementing the technology itself.

The recent [Gartner](#) Data and Analytics Summit 2018 states that data and analytics leaders must champion workforce data literacy as an enabler of digital business and treat information as a second language. [Gartner predicts](#) that by 2019, 90 percent of large companies will have a Chief Data Officer (CDO) role in place. CIO Kent [says](#) that data hygiene also compliments the latest need for increased cybersecurity, as organization and security of data go hand in hand.

Tech Insight Series

The monthly Tech Insight series aims to help readers make better decisions and be more informed customers of OIT products and services by providing them with high-level overviews of technologies that impact or will impact VA's IT environment. Tech Insights introduce topics in an easily digestible fashion by presenting background information on the topic, clearly explaining its importance within VA, and providing recommendations for success from OIT. All Tech Insights are available [here](#).

Disclaimer: This document includes links to websites outside VA control and jurisdiction. VA is not responsible for the privacy practices or the content of non-VA websites. We encourage you to review the privacy policy or terms and conditions of those sites to fully understand what information is collected and how it is used.