

Tech Insights: Open Data

Office of Technology Strategies (TS) / Architecture, Strategy, and Design (ASD)

Recap

October 2016



The TS office within OI&T's ASD, interacts not only with the Enterprise Architecture pillar offices, but also with multiple external vendors, stakeholders within OI&T, and with strategic offices across the enterprise. TS works closely with IT and business owners to capture business rules and provide technical guidance as it relates to Data Sharing across the enterprise, specifically for interagency operability.



Introduction

This recap examines three previous Tech Insights: Open Data, Open Data II, and Big Data, and combines their highlights into one concise review of open data. This recap begins with a defining overview of open data, followed by an examination of its benefits, its implementation challenges, and its interconnection to big data. We close with a discussion of open data adoption at VA.

Open Data

Data represents the quantities, characters, or symbols that have been curated and analyzed for patterns in order to answer research questions of a descriptive, explanatory, and predictive nature. Since data must be processed in order to inform, it is not information, but rather the meaningful "signal" isolated from noisy information.

Open data is data designated by its owner as available for free use and distributed by anyone, as long as attribution is given to the original source. For Federal agencies, providing open data is mandated by the White House Office of Management and Budget (OMB)'s [Open Data Memorandum](#), requiring agencies to build and modernize open data information

systems.

According to OMB's Memorandum, open data should be consistent with the following principles:

- Public – Federal agencies must presume data should be open to the public to the extent permitted by law, yet remain subject to privacy, confidentiality, security, or other valid restrictions
- Accessible - Open data must be available in machine-readable, convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched
- Described - Open data must fully describe the [metadata](#), or data about the data, so that consumers of the data can understand its strengths, weaknesses, analytical limitations, security requirements, as well as how to process it
- Reusable - Open data must be available under an open license that places no restrictions on its use
- Complete - Open data must be published in primary forms (i.e., as collected at the

source), with the finest possible level of granularity that is practical and legally permitted

- Timely - Open data must be available at the agency's earliest convenience
- Managed Post-Release - A point of contact must be designated to assist with data use and to respond to complaints about adherence to these open data requirements

Under OMB's Open Data Memorandum, agencies are now releasing valuable datasets that fuel collaboration across the public and private sectors.

Benefits of Open Data

There are many benefits open data gives to agencies, enterprises, and consumers alike. Focusing solely on the government's point of view, the business benefits of open data include, but are not limited to: saving time and money responding to Freedom of Information Act (FOIA) requests; avoiding duplicative internal research – which could save billions in operating costs; sharing complimentary datasets held by other agencies; and empowering employees to make better-informed, data-

Tech Insights: Open Data

Office of Technology Strategies (TS) / Architecture, Strategy, and Design (ASD)

Recap

October 2016

driven decisions.

From an economic standpoint, open data can make an even larger impact. The increased efficiency it affords can result in new products and services, and even a consumer surplus due to cost savings and better products. This transparency improves the operational efficiency of Federal agencies by creating deep analytical insights, measurements for policies, opportunities for the design of products and services, and interagency interoperability.

Implementation Challenges

There are some road blocks, however, to the widespread implementation of open data policies in both the public and private sector. A critical issue hindering implementation in both sectors is the ability to meaningfully analyze the data. For government agencies, supplying the public with open data is not sufficient for its widespread use.

Government agencies not only need to post their data in an accessible format, but they also need to ensure that ordinary people know what data is available and how to become familiar with it.

Another challenge to implementing open data is the navigation of privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), to ensure that the datasets shared do not violate the rights of anyone, especially Veterans. VA has found ways to share plenty of useful information about where, how, and who it is serving without compromising personally identifiable information (PII) or protected health information (PHI).

Big and Open Data

Big data is characterized by the four Vs: volume, velocity, variety, and veracity. Given the dispersion of more data collecting machines (smart phones, apps, and internet-of-things devices, for example), big data sets contain significantly greater volume than traditional data sets. This expanded network of data-collecting computers also increases the velocity, or frequency, of data generation and capture. More data gathered more often likewise increases the variety of data generated and captured. Though big data sets have an expanded variety, velocity, and volume of data, insights analyzed from them are still limited by the veracity, or truth, of the data. Veracity is a measure of the quality of the captured data.

The availability of big data sets and the advancement of the comput-

ing capabilities to store, process, and analyze big data have transformed the way business leaders and everyday citizens make decisions on a daily basis. As a result, decision-making is increasingly [data driven](#), and relies on probability predictions that were not available just a few years ago.

As organizations continue to use big data to support decision-making, open data can facilitate these efforts by expanding the scope of information available to agencies, creating transparency, enabling experimentation with previously inaccessible data sets, and empowering innovators to find new solutions to organizational issues from their analysis. In this way, open data can contribute to greater efficiency and productivity.

Open Data at VA

VA adheres to OMB's Open Data Memorandum by providing the [VA Open Data Website](#), a data repository where anyone can access data, tools, and resources to develop web and mobile applications, design data visualizations, and create stories directly from VA resources. The available data contains information such as VA facility locations, homelessness resources, and family caregiver services. Information posted here is not Veterans' PII or PHI.

Further, VA's [National Center for Veterans Analysis and Statistics](#) makes public the same data which VA uses in its own operations and which is reported to Congress and other stakeholders. Current data sets available from VA include patient satisfaction surveys, Veterans benefits and compensation information (number of veterans receiving disability compensation or pension by county, age, and percent disability rating for each state), and geographic distribution of VA expenditures.

For more information on open data and big data topics, please check out the original Tech Insights: [Open Data](#) (Volume 2, Issue 2), [Open Data II](#) (Volume 3, Issue 2), and [Big Data](#) (Volume 3, Issue 5). You can also view Enterprise Design Patterns (EDPs) on interoperability and data sharing [here](#). Read all our technology topics in the Office of Technology Strategies' [Tech Insights](#). If you have any questions about open data, don't hesitate to [ask TS](#) for assistance or more information.