

OPEN DATA

ISSUE 2, VOLUME 2

OFFICE OF TECHNOLOGY STRATEGIES (TS)

INTRODUCTION

In a past Tech Insight, we examined open source development, how VA is adopting, and how VA should use collaboration software and development tools to embrace open source in the future. This Tech Insight will cover a similar topic, open data, which is a trend in data management policy that can help both the public and private sectors operate more efficiently and stimulate innovation. This overview will discuss open data's principles, benefits, challenges, and how VA is sharing its own data via the Open Government Initiative.

OVERVIEW

Open data is data that can be freely used, reused, and redistributed by anyone, as long as attribution is given to the original source. If you use sites like Kayak, LinkedIn, or Weather.com, then you are a contributor to the estimated \$3 trillion a year in additional value that open data provides to innovators. Allot these sites pull in public information from countless sources and aggregate them into a user-friendly application program interface (API) that allows you to easily book a flight, find a job, or check the weather.

On May 9, 2013, the White House Office of Management and Budget (OMB) released an open data policy requiring agencies to build and modernize information systems in a way that maximizes interoperability and information accessibility, maintains internal and external data asset inventories, enhances information safeguards, and clarifies information management responsibilities. The White House also developed Project Open Data to help agencies adopt the new policy by providing a collection of code, tools, and case studies—a community resource to which anyone (government staff, developers, or the general public) can contribute. These initiatives reflect the President's recognition that data is a valuable national resource and a strategic asset to the government, its partners, and the public.

According to the government, open data should be consistent with the following principles:

- *Public* - Agencies must adopt a presumption in favor of openness to the extent permitted by law and subject to privacy, confidentiality, security, or other valid restrictions.

- *Accessible* - Open data are made available in machine-readable, convenient, modifiable, and open for-mats that can be retrieved, downloaded, indexed, and searched.
- *Described* - Open data are de-scribed fully (through metadata)so that consumers of the data can understand their strengths, weaknesses, analytical limitations, security requirements, as well as how to process them.
- *Reusable* - Open data are made available under an open license that places no restrictions on their use.
- *Complete* - Open data are published in primary forms (i.e., as collected at the source), with the finest possible level of granularity that is practicable and permitted by law and other requirements.
- *Timely* - Open data are made available as quickly as necessary to preserve the value of the data.
- *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.

BENEFITS OF OPEN DATA

There are many benefits to open data beyond its ability to let you book a flight from multiple airlines' ticketing data. From 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 – saving billions
- Discovering complimentary datasets held by other agencies
- Empowering employees to make better-informed, data-driven decisions

From an economic standpoint, open data can have even larger impacts. The increased efficiency it affords can result in new products and services and even a consumer surplus due to cost savings and better products. Open data can make big data easier to handle by creating transparency, enabling experimentation with data, and empowering innovators to find ways to have big data support human decision making.

CURRENT CHALLENGES

There are some road blocks, however, to the widespread implementation of open data polices in both the public and private sector. Even though the US has over half a million datasets available on data.gov, this widespread availability of data means nothing if the average person doesn't know how to analyze it. Therefore, Federal 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 be conversant in 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 being shared don't violate anyone's (especially Veterans') rights. 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).

OPEN DATA AT TS

President Obama's Open Government Initiative calls on Federal agencies to break down barriers to transparency, participation, and collaboration between the government and the people it serves. One of VA's answers to the Open Government Initiative (www.va.gov/open/) is Blue Button, which allows Veterans to download their personal health information from their MyHealthVet account. The availability of Blue Button's source code (not Veteran health data) has allowed developers to create such things as The Blue Button Connector, which makes health data that is increasingly available to the public more transparent. VA believes this application will help fuel the creation of new products and services. The usefulness of the Blue Button Connector resource will grow as the community contributes to it and The Department of Health and Human Services is even soliciting feedback on how to make it better.

OPEN DATA AT VA

VA is implementing the open data policy through its website, www.va.gov/data/, where anybody can access data, APIs, tools and resources that can be used to develop web and mobile applications, design data visualizations, and create stories directly from VA resources. The available data is a by-product of the work VA does for Veterans (information such as VA facility locations, homelessness resources, and family caregiver services) and VA is careful to note that this is not personal data. VA is encouraging visitors to the site to use the data to create products like applications, infographics, or stories of any kind. VA's National Center for Veterans Analysis and Statistics makes freely available to the 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 per-cent disability rating for each state), and geographic distribution of VA expenditures.

If you have any questions about open data, don't hesitate to ask TS (askTS@va.gov) for assistance or more information. Check out earlier [Tech Insight editions](#) for more information.

TS TECH INSIGHT SERIES

The monthly Tech Insight series aims to help readers make better decisions and be more informed customers (of Office of Information & Technology's products and services) by providing them with high-level overviews of technology issues that impact or will impact VA's Information Technology (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 TS. View all TS Tech Insights [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.