

# A VA Executive's Guide to Big Data

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



## Introduction

Skimming newspapers, blogs, or even highway billboards, tales of big data analytics tantalize the public with promises of insight and opportunity. VA leadership looks to big data analytics to centralize management, synergize shifts to electronic health records and interoperability, and transform healthcare into a proactive, rather than reactive, service. While big data promises revolution, its impact also exists in areas other than technological innovation and business efficacy. This TS Note begins with an introduction to big data analytics and how it differs from the traditional collection and storage of data. It then broaches legal and regulatory factors developing in the United States (US) and the European Union (EU).

## Big Data is the New Data

Data represents the quantities, characters, or symbols that have not been curated and analyzed for patterns. Data is not information because data must be processed in order to inform. Traditionally, data processing focused on the business value derived from relational databases. Big data, however, is characterized by the four Vs: volume, velocity, variety, and veracity. In terms of volume, more [data was created](#) between 2014 and 2015 than in the entire previous history of humanity. In March 2016, the Veterans Health Administration (VHA) was [reported](#) to have over 1 billion data points. In terms of velocity, this data is gathered often. A [recent study](#) tracking less than 15 apps on an Android phone found that the apps tracked the phone's location more than 3,000 times in a week—or once every three minutes. Because the data is refreshed so frequently,

these data sets quickly lose value, driving the demand for broader and ever more current data.

Big data sets come from a variety of sources, drawing connections between structured, semi-structured, and unstructured data, using expensive and complex analytical tools. Thus, the veracity — or “truth”— of the data depends on the quality of data sets — across time, location, and source. The data sets are combined in the search for patterns and meaning. Matching demand to relevant data is one of the challenges of big data analysis. Together, the complexities of the four Vs hint at the challenge, skill, and expense of deriving value from big data.

## Antitrust

Now consider big data analytics in relation to antitrust regulation. Trusts and monopolies are concentrations of economic power controlled by only a few sources. Consumers suffer in monopolized markets, as reduced competition results in the production of lower quality products, reduced industry innovation, and higher prices to consumers. Infamous monopolies like the Bell Telephone Company were disbanded because monopolies reduce market efficiency and competition.

Consolidated information is similar to a monopolized resource or market. If we understand information to be power, then big data offers the opportunity to gather more information—more power—than ever before and creates insurmountable barriers to entry into certain markets. For example, in the US, few entrepreneurs hope to compete with Google, Yahoo, and Bing; and competitive rivalry between the existing search engines is high, keeping new entrants out of the field. Big data players are preparing for and pre-

The TS office within OI&T's Architecture, Strategy & Design (ASD) interacts not only with the ASD pillar offices, but also with multiple 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.

empting regulation. Google's 2015 restructuring, for example, resulted in a new parent company, Alphabet. Alphabet's umbrella now holds a slimmed-down Google, with focus on its core products, such as YouTube and Android; and many new subsidiaries, such as GoogleX. This action neatly coincided with an antitrust legal battle in the EU. Google is alleged to have abused its market dominance to unfairly rank its own products, such as YouTube, ahead of competitors in search results. This preceded significant 2016 EU data protection policy (see below). Strategic business and investment decisions shape the young field of big data analytics.

Yet big data's consolidation of power does not quite fit the anticompetitive precedent set by past monopolies because data consolidation also produces economic benefits. Organizations use data to streamline processes and activities to make data-driven strategic choices and informed

# A VA Executive's Guide to Big Data

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

decision-making. This contributes to greater efficiency and productivity. For example, knowledge of consumer preferences can be used to improve products, or develop new ones with less risk of failure. Such innovation is healthy.

## Data Regulation Increases in Europe and the US

On both sides of the Atlantic, the conversation around big data is maturing away from the thrill of possibilities toward long-term practicalities. The EU Commission and the US Federal Trade Commission (FTC) increasingly address data regulation. Their actions channel big data's development, potentially limiting its business use and function.

In January 2016, the FTC published, "Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues." This report addresses data management and outlines data best practices for the private sector. To avoid harm and illegal activity while wielding big data, the report outlines considerations for data queries:

- How representative is your data set?
- Does your data model account for biases?
- How accurate are your predictions based on big data?
- Does your reliance on big data raise ethical or fairness concerns?

These guidelines are thoughtful, but not yet backed by legislation or significant court rulings. The EU Commission's General Data Protection Regulation (GDPR), which passed into law after four years of negotiation on April 14, 2016, carries more heft. The GDPR offers significant data protection rights to users, including:

- To be permanently erased from company data stores
- To be notified if personal data is processed, and
- To switch service providers through increased ease of personal data portability

The GDPR requires any organization with activities in Europe, including those headquartered in Asia or the Americas, to meet EU standards. This broad legislation could set the global standard for data management practices.

## Conclusion

The Era of Big Data is just beginning and many questions remain unanswered. Should the private and public sectors have different regulations? How will regulation influence VA's data management, information security, and interoperability, while straddling private sector and internal data systems? The US Federal Government and VA play important roles in carving out the business function and application of big data analytics. Choices made now by big data stakeholders lay the foundation for big data in the future as it applies to every industry, including healthcare. The fireworks are just about to start.

Read more on data analytics and related topics in the Office of Technology Strategies' TS Notes: [Open Data](#); [Open Data II](#); and Enterprise Design Patterns regarding [Interoperability and Data Sharing](#). If you have any questions about big data analytics, don't hesitate to [ask TS](#) for assistance or more information.