

# How does the U.S. Government use Big Data?

Ethan Moore

September 17, 2021

## 1 Introduction

The name Edward Snowden has become synonymous with internet privacy and controversy. The U.S. government's newly illuminated role in covert data collection has concerned everyday citizens of the consequences. In response the administration at the time pushed for as much open access to government collected data as possible. The use cases in response to the released data are numerous encapsulating departments like the department of agriculture, healthcare, and education. Some uses include analysing education and skill levels, tracking the spread of disease, and examining food production.

## 2 Uses and Results

The department of agriculture uses Big Data to the benefit of farm production. In the book *Big Data in Practice* within the chapter on the U.S government it states that "Milk production of dairy herds across the States has been improved thanks to work to identify bulls most likely to breed high-yielding cows" [Marr, 2016]. The implementation of data collecting on livestock has directly contributed to better milk yields. Another use is within healthcare and the Centres for Disease Control. Their use of Big Data is through "the spread of epidemics and other public health threats." [Marr, 2016]. The department of education also uses Big Data. Their use is for "how the population learns, and assess the level of education and skills among the population in a specific geographical area, again allowing for a more efficient

planning and deployment of resources.” [Marr, 2016]. The results from the governments pursuits in utilizing Big Data culminated in a extensive review. The government review in the book states ”Big Data holds the potential to streamline the provision of public services, increase the efficient use of taxpayer dollars at every level of government and substantially strengthen national security.” and ”a predictive fraud prevention system used by administrators of the Medicare and Medicaid services had prevented \$820 million in fraudulent payments being made since its introduction three years earlier.” [Marr, 2016].

### 3 Methodology

The U.S. governments collection of data was methodically planned. From recognizing the issues to solve, figuring out the technicalities like the type of data, and the challenges needed to overcome issues. These issues that were recognized were from bureaucracy and data between the departments was not freely and openly distributed. The government’s solution of making their Big Data open access can be seen on their website Data.gov. Some of the technical details listed in Bernard Marr’s book details that ”their data available, has grown from 49 datasets at launch to close to 190,000 datasets today.” and ”The biggest individual contributors are NASA (31,000 datasets uploaded), the Department of the Interior (31,000) and the Department of Commerce (63,000).” [Marr, 2016]. The data used and contained within the data sets includes but is not limited to ”climate and meteorological data, food production data from agriculture, statistics on crime and security threats from police departments and federal agencies, population movement data from camera networks and demographic research (i.e. the US Census), economic data from public company records and stock market activity, the movement of people and goods in and out of the country through migration and import/export data, patterns of energy distribution and usage, scientific data collected through federal research and development facilities, epidemiology data from tracking the spread of illness and illness-causing bacteria and information related to the effects of climate change through the Climate Data Initiative.” [Marr, 2016]. The incredible range of data used obviously creates challenges that need to be overcome. The two biggest issues are the public’s trust and understanding of the government data collection. The public’s trust in the book Big Data in Practice see’s it ”simply as “spying”.”

when the subjects are "without the knowledge" [Marr, 2016]. The addition of D. J. Patil the United States first chief data scientist is focused on raising awareness to the Obama administration on being more transparent.

## 4 Conclusion

Private citizens and their private data potentially becoming harvested is a serious topic. The newly created web portal for government data is a great first step to tackling the biggest issues currently for government data collection. Besides the issues and challenges the benefits are great with increased government resource and tax dollar efficiency. There is clearly potential in one of the newest government initiatives.

## References

- [Marr, 2016] Marr, B. (2016). The us government using big data to run a country. In *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*, pages 229–235.