Chapter 98
From Data to Vision:
Big Data in Government

Rhoda Joseph
Pennsylvania State University – Harrisburg, USA

ABSTRACT

This chapter examines the use of big data in the public sector. The public sector pertains to government-related activities. The specific context in this chapter looks at the use of big data at the country level, also described as the federal level. Conceptually, data is processed through a "knowledge pyramid" where data is used to generate information, information generates knowledge, and knowledge begets wisdom. Using this theoretical backdrop, this chapter presents an extension of this model and proposes that the next stage in the pyramid is vision. Vision describes a future plan for the government agency or business, based on the current survey of the organization's environment. To develop these concepts, the use of big data is examined in three different countries. Both opportunities and challenges are outlined, with recommendations for the future. The concepts examined in this chapter are within the constraints of the public sector, but may also be applied to private sector initiatives pertaining to big data.

INTRODUCTION

The purpose of this chapter is to examine big data in the public sector, and the value of moving from raw data and information to knowledge, wisdom and ultimately to vision. The underlying premise is that data is valuable, and there are vast opportunities to utilize the data to improve decision making. Part of this process requires the application of big data analytics, to ask the right questions and receive relevant answers from the data. Recent studies highlight the value of examining big data in the public sector (Joseph & Johnson, 2013; Malykhina, 2013; TechAmerica Foundation: Federal Big Data Commission, 2012). The main context of this chapter is to examine the use of big data at the country level. For example, the country level of examination in the United States will be the use of big data by the federal government, and agencies of the federal government. The country level provides a broad scope for assessing the global impact of big data initiatives. A global examination provides a framework to not only identify best practices and benchmarks but to also identify areas where opportunities exist; particularly

with emerging concepts and areas such as big data and big data analytics. An examination of these concepts from multiple perspectives such private vs. public; domestic vs international; and local vs global can provide useful information on the subject. This chapter focuses on the public side of the equation.

To enhance this chapter, we examine the use of big data in government through a set of theoretical lens. The theoretical view presented in this chapter is based on the development of a model showing the movement of data-to-vision (DV). This theory is informed by the existing DIKW (data-information-knowledge-wisdom) model that presents a hierarchy of these concepts (Ackoff, 1989; Rowley, 2007; Zeleny, 1987). I present the case in this chapter that the next step after wisdom is vision, and big data can push public sector agencies to develop long-term visions that may not have been possible without big data analytics. Arguably, this model will be relevant in both the private and public sector domains. Private sector examples will be presented to support the model, but the focus will be on opportunities in the public sector.

Next, the chapter will examine rewards and challenges associated with the use of big data in government. Some of the rewards include increased transparency, better decision-making, and transformation of internal processes. On the other side, some challenges include concerns about data quality, security concerns, and compromised privacy of citizen data. Developing wisdom and ultimately vision can sometimes be constrained by political, economic and social variables in the government domain. It is thus important to recognize these challenges, and mitigate some of the risks by effectively and efficiently managing the data resource. Following the above described content, the chapter discussion will proceed with future research directions. Lastly, the chapter ends with a conclusion.

BACKGROUND

Over the last few years big data has becoming an increasingly examined topic in a variety of places. The big data phenomenon is currently gathering significant attention in both the public and private sectors. Big data is defined as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze (McKinsey Global Institute, 2011)". The growth of big data is generated from a number of sources including, online transactions, social media, cellular phones, radio frequency identification (RFID), global positioning systems (GPS) and many other sources.

In both the private and public sectors large amounts of data are captured on a daily basis. In the consumer world retail transactions generate high volumes of data such as sales, returns, refunds, shipping details, credit/debit card details to name a few. The online environment also generates a high volume of content created by individuals and businesses. A single online purchase whether for private or professional purposes contains information about the consumer, device used (desktop, laptop, tablet, smartphone), product details, geographical location, internet service provider, shipping company, vendor contact information, and more. In the public domain entities such as law enforcement departments collect data ranging from traffic violations to detailed information about criminal activities such as burglaries and homicides. As the sources of big data continue to increase, and the cost of storage needed to save the data continues to decrease, the technology needed to analyzed the data becomes more prevalent and in demand.

As more and more sources are tapped for the generation of big data, organizations must decide on the best ways to capture and utilize the data. This is a critical issue in the public and private sectors. In the private sector, due to higher levels of competition, innovation associated with big data, may be occurring

Related Content

Sequential Patterns Postprocessing for Structural Relation Patterns Mining
Jing Lu, Weiru Chen, Osei Adjeiand Malcolm Keech (2008). *International Journal of Data Warehousing and Mining (pp. 71-89).*
www.irma-international.org/article/sequential-patterns-postprocessing-structural-relation/1814

Dimensionality Reduction with Unsupervised Feature Selection and Applying Non-Euclidean Norms for Classification Accuracy
Amit Saxenaand John Wang (2010). *International Journal of Data Warehousing and Mining (pp. 22-40).*
www.irma-international.org/article/dimensionality-reduction-unsupervised-feature-selection/42150

Recommender Systems Using Collaborative Tagging
Latha Banda, Karan Singh, Le Hoang Son, Mohamed Abdel-Basset, Pham Huy Thong, Hiep Xuan Huynhand David Taniar (2020). *International Journal of Data Warehousing and Mining (pp. 183-200).*
www.irma-international.org/article/recommender-systems-using-collaborative-tagging/256169

A Solution to the Cross-Selling Problem of PAKDD-2007: Ensemble Model of TreeNet and Logistic Regression
Mingjun Wei, Lei Chai, Renying Weiand Wang Huo (2008). *International Journal of Data Warehousing and Mining (pp. 9-14).*
www.irma-international.org/article/solution-cross-selling-problem-pakdd/1802

Classification of Peer-to-Peer Traffic Using a Two-Stage Window-Based Classifier with Fast Decision Tree and IP Layer Attributes
Bijan Raahemiand Ali Mumtaz (2012). *Exploring Advances in Interdisciplinary Data Mining and Analytics: New Trends (pp. 174-188).*
www.irma-international.org/chapter/classification-peer-peer-traffic-using/61175