Government Information Quarterly 31 (2014) 622–630

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

Using open government data to predict war: A case study of data and systems challenges

Andrew Whitmore

Department of Information Systems and Decision Science, University of South Florida, 8350 N. Tamiami Trail, SMC-C263 Sarasota, FL, United States

ARTICLE INFO

Available online 27 September 2014

Keywords:
Open data
Case study
Research methods
Military conflict
Data quality
System design
Government portal

ABSTRACT

The ability to predict future military engagements would be a boon to combatants, contracting companies, investors, and other stakeholders. While governments may seek to conceal plans for impending conflict, they must spend large sums of money mobilizing and equipping soldiers in preparation for deployment. Thus, examining government spending patterns might yield insight into future military conflict. This article reports on an attempt to explore the possibility of using open U.S. Department of Defense (D.O.D.) contracting data to identify patterns of spending activity that can predict future military engagement. The research in this article followed a two-stage method. The first stage involved the exploration of the research question in the context of a specific case, the U.S. invasion of Iraq in 2003. The second stage assessed the open government contracting data used in the research and classified data and systems problems that were encountered according to an established analytical framework for open data barriers. The analysis demonstrated that the use of U.S. D.O.D. contracting data to predict future war has promise; however, a number of problems with the data and online portal prevented the derivation of conclusive, generalizable results. These problems were related to the open data barriers of task complexity and information quality. A detailed description of how these barriers manifested and directions for overcoming them are presented.

1. Introduction

The ability to predict future military engagements would be a boon to combatants, contracting companies, investors and other stakeholders. While governments may seek to conceal plans for impending conflict, they must spend large sums of money mobilizing and equipping soldiers in preparation for deployment (for example, see Harrison, 1988). Thus, examining government spending patterns might yield insight into future military conflict. This article reports on an attempt to explore the possibility of using open U.S. Department of Defense (D.O.D.) contracting data to identify patterns of spending activity that can be used to predict future military engagement.

Over the last decade, the United States has been engaged in a series of military conflicts including the wars in Afghanistan and Iraq. As part of the process of getting ready for and engaging in war, the U.S. Department of Defense (D.O.D.) paid billions of dollars to private defense contracting companies for the procurement of goods and services. For example, private contractors comprised 54% of the U.S. D.O.D.'s workforce in Iraq and Afghanistan (Schwartz, 2010) and military contracts comprised 20% of all Iraq related spending in 2010 (Singer, 2010). In the U.S., the Federal Government uses private contracting companies to provide the goods and services required for the operations and activities of government agencies and offices. The goods and services are provided by government contractors (vendors) who enter into contracts with the government that specify what work needs to be performed or what goods need to be delivered. These contracts cover an enormous domain of goods and services ranging from office supplies to nuclear missile components. Although there are numerous permutations, the basic contracting process begins with a government agency posting an advertisement online that details the nature of the contract, the goods and services to be delivered, the nature of the competitive selection process (for example, lowest bid price) and other related information. The companies interested in competing for the contract submit their bids and then a selection process follows where the agency offering the contracts chooses a specific vendor (for comprehensive contracting procedures, see U.S. Government, 2005). The overwhelming majority of U.S. government agencies rely on contracting, and the U.S. Department of Defense (U.S. D.O.D.) engages in the greatest level of contracting activity by a fairly wide margin.[1]

Given the extent to which the U.S. D.O.D. relies upon military contractors, it seems plausible to explore the notion that the contract granting activity of U.S. D.O.D. agencies could be used as a proxy measure for classified operational activities. If there is a correlation between contracting activity and classified operational activities, then large increases in military contracts might be a leading signal for future

E-mail address: whitmore@ieee.org.

[1] Source: https://www.fpds.gov/downloads/FPR_Reports/Fiscal%20Year%202007/Total%20Federal%20View.pdf.

http://dx.doi.org/10.1016/j.giq.2014.04.003
0740-624X/

military conflict. While other researchers have used historical data to predict military actions and outcomes (Dupuy, 1979), the use of defense contracting data for these purposes has not been previously documented. The defense contracting data used in this research was accessed through the usaspending.gov data portal. Section 4 provides more details about the data portal and its use in this research.

While assessing the feasibility of using open defense contracting data to predict future conflict, it was necessary to assess the value of the open government data itself in terms of its ability to yield unbiased, consistent and generalizable results in accordance with common research standards. This assessment of the open data helps shed light on the benefits of open government data (for example, see Kassen, 2013; Linders, 2013) as well as any barriers to the use of these data (for example, see Janssen, Charalabidis, & Zuiderwijk, 2012; Zuiderwijk, Janssen, Choenni, Meijer, & Sheikh Alibaks, 2012).

Furthermore, this research highlights the complexities involved in multidisciplinary research that involves computer science, information science and digital government. Computer science is focused on the theoretical and technical aspects of computing (Denning, 2005); in this case the efficient design and implementation of the web database system that houses the contracting data. Information science is focused on how users interact with and utilize (Borko, 1968) the government database and portal. Digital government examines the effectiveness of the government-to-citizen communication and information sharing (Marchionini, Samet, & Brandt, 2003) that the database and portal enable. There is a dependency between the individual disciplines in multidisciplinary research. For example, a poorly designed database (computer science) will adversely affect user interaction with the system (information science) which will lead to ineffective government to citizen communication (digital government). This paper helps shed light on these interactions as we progress into the open data era.

The remainder of the paper is structured as follows: Section 2 presents a review of a subset of related literature and identifies the article used as an analytical framework to model the challenges encountered when working with the usaspending.gov data and portal. Section 3 describes the method employed in this research. Specifically, the section addresses how the use of defense contracting data to predict war and open data barriers are related to one another. Section 4 presents the research case, a detailed description of the data source, the analytical approach, the analysis results, a discussion of the emergent issues that inhibited validity and generalizability, and a tie-in with open data barriers. Section 5 maps specific issues that emerged during the use of the usaspending.gov data portal to the chosen classification framework for open data barriers. A set of recommendations for how the data and portal could be improved to offset these challenges is also presented. Section 6 presents the conclusions of the study and describes how the combination of open data classification frameworks and the examination of specific data portals can be used together to provide a roadmap for improved open data implementations.

2. Related literature

Government data portals have the potential to be a great boon to academics and practitioners by allowing easy online access to vast amounts of data without the need for repeated data requests, transcription of data from print to electronic formats, and other tasks that would limit user interest and data usefulness. By opening their data, government agencies have the potential to promote transparency, increase citizen participation and spur innovation (Nam, 2011). Furthermore, open data initiatives can help citizens learn about government activities, improve government accountability, and enable citizens to participate in the political process (Janssen, 2011). Open data initiatives can also provide the data independent parties need to evaluate the quality of government policy-oriented decision making (Napoli & Karaganis, 2010). Despite all these proposed benefits, there is no clear indication that these implementations are actually successful in fulfilling their missions. While benchmarking the success of government information systems has become a common practice in the area of e-government (for example, Jansen, de Vries, & van Schaik, 2010; Whitmore, 2012a), there are currently no metrics by which to evaluate the success of government open data initiatives (Bertot, McDermott, & Smith, 2012). Going forward, government agencies might make use of a benchmarking framework based on an open government data stage model that has been proposed (Kalampokis, Tambouris, & Tarabanis, 2011).

A large number of factors could cause an open data implementation to be unsuccessful. For example, the importance of the quality of the data and the portals that contain the data cannot be overstated. The field of information systems has thoroughly demonstrated the impact of data and system quality on user adoption (*for example*, Venkatesh & Davis, 2000; Venkatesh, Morris, Davis, & Davis, 2004). While government data portals offer many theoretical benefits, they frequently do not receive levels of use that correspond to their potential benefits. Despite attempts by government agencies to stimulate innovative use of open data, response from external stakeholders has been tepid (Yang & Kankanhalli, 2013). While the promise of open data is great, in practice the process of making government data usable is fraught with impediments at every step (Zuiderwijk et al., 2012). These process problems can create a set of barriers that limit the usefulness of open data.

The barriers limiting user adoption of government data portals take many forms. Researchers have shown that data quality continues to be a major issue for open government statistical data (Karr, 2008). Data quality plays a critical role in the level of use of government portals (Detlor, Hupfer, Ruhi, & Zhao, 2013). Some have suggested that the release of quality data is not always a priority for government agencies due to a lack of incentives (Conradie & Choenni, 2013). Users of open government data have complained about a lack of metadata for released documents, useless or inconsistent formats, and other barriers to usability (for example, Kerschberg, 2011). These data quality issues can frustrate user attempts to make sense of the data and therefore limit data interpretability and understandability. These factors have been shown to present a barrier to accessibility and ultimately, use (Strong, Lee, & Wang, 1997).

In addition to discouraging use, quality issues can also impact accessibility. One frequently cited data quality issue, a lack of metadata, can impede retrieval of relevant information in public sector information systems (for example, Christian, 2001; Quam, 2001; Whitmore, 2012b). Also impacting accessibility, researchers have been struggling with the over-classification of information and by the recent use of new document classifications that can make records increasingly difficult to obtain (Feinberg, 2004; Strickland, 2005). Others have pointed out that data are frequently not current and there is typically a lack of opportunity for public participation in the delivery of the open data (Lee & Kwak, 2012). Other barriers impeding the usefulness of open data initiatives include a lack of domain knowledge on the part of users (King, Liakata, Lu, Oliver, & Soldatova, 2011), a lack of interoperability between datasets (Mclaren & Waters, 2011), and a shortage of the statistical knowledge required to successfully work with the datasets (Janssen et al., 2012).

The challenges mentioned earlier only touch the surface of the barriers to open data implementation success. Janssen et al. (2012) proposed a comprehensive classification scheme for these barriers. These researchers conducted a literature review and set of interviews with civil servants in Europe in order to generate a classification of adoption barriers for open government data. Their two-tier classification scheme consisted of six top-level barrier categories (institutional, task complexity, use and participation, legislation, information quality, and technical) and numerous specific barriers associated with each top-level category. This research utilizes this classification framework to illustrate the barriers experienced when working with the usaspending.gov data portal.