Government Information Quarterly xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Government Information Quarterly

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

Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities

Ricardo Matheus*, Marijn Janssen, Devender Maheshwari

Delft University of Technology, Faculty of Technology, Policy and Management, Jaffalaan 5, 2628 BX Delft, The Netherlands

ARTICLE INFO

Keywords:
Data science
Dashboards
E-government
Open government
Open data
Big data
Smart City
Design principles
Transparency
Accountability
Trust
Policy-making
Decision-making

ABSTRACT

Dashboards visualize a consolidated set data for a certain purpose which enables users to see what is happening and to initiate actions. Dashboards can be used by governments to support their decision-making and policy processes or to communicate and interact with the public. The objective of this paper is to understand and to support the design of dashboards for creating transparency and accountability. Two smart city cases are investigated showing that dashboards can improve transparency and accountability, however, realizing these benefits was cumbersome and encountered various risks and challenges. Challenges include insufficient data quality, lack of understanding of data, poor analysis, wrong interpretation, confusion about the outcomes, and imposing a pre-defined view. These challenges can easily result in misconceptions, wrong decision-making, creating a blurred picture resulting in less transparency and accountability, and ultimately in even less trust in the government. Principles guiding the design of dashboards are presented. Dashboards need to be complemented by mechanisms supporting citizens' engagement, data interpretation, governance and institutional arrangements.

1. Introduction

Governments are more and more utilizing data in all aspects of their functioning. *Data science* in government deals with the extraction, interpretation and presentation of insights from unstructured and structured data that can be either closed or opened. An important area of data science is to visualize the data in dashboards. Cleveland (2001) argues that data science consists of multidisciplinary investigations, models and methods for data, computing with data, pedagogy, tool evaluation, and theory. Government data scientists need in-depth knowledge of statistics and data analytics for analyzing data, as well as knowledge on the use of techniques and instruments for predictive purposes and to visualize the results. By combining disciplines, new insights and applications can be created and communicated using dashboards. Nevertheless, data scientist also need to have an understanding of other elements like the policy-making, organization, legislation and public values. This knowledge allows them to positioning the data in the context and to understand its use and implications.

Data science is an essential area for governments, as they collect a lot of data in various areas (geographical, traffic, social security, energy, etc.) that can be combined or enriched with data from smart devices and other sources such as discussion forums, social media, and private sector data (Janssen, Matheus, & Zuiderwijk, 2015). The making of sound decisions depends on the use of high-quality data (Chengalur-Smith, Ballou, & Pazer, 1999). Data might be an enabler for creating new innovative applications (Marsh, Pane, & Hamilton, 2006) to improve public values like security, safety, transparency and accountability.

In data science, the sharing, use and interpretation of data are key aspects in bridging the gap between the government and the public. Platforms can be created to share data (Brown, Fishenden, Thompson, & Venters, 2017). The use of data and the accompanying instruments will likely influence government policy-making, resulting in new applications, but can also impact the interaction with the public (Ganapati, 2011a). Dashboards can be used to release information for governmental decision-makers (Maheshwari & Janssen, 2014), but also for the public to scrutinize government actions, to engage in the decision-making processes and to improve decision-making. Dashboards should help to facilitate transparency, governance, trustworthiness and enable citizens' to participate in decision-making in smart cities (Allio, 2012).

Recently, data science and dashboards have gained more and more attention in the public sector. A dashboard is often created by having a webpage which visualized all kinds of data given a certain purpose. For example, in 2009, the US federal government developed dashboards with federal stimulus funding (www.recovery.gov) aiming transparency

* Corresponding author.
E-mail addresses: r.matheus@tudelft.nl (R. Matheus), M.F.W.H.A.Janssen@tudelft.nl (M. Janssen).

https://doi.org/10.1016/j.giq.2018.01.006
Received 22 November 2016; Received in revised form 1 November 2017; Accepted 26 January 2018

and accountability of national economic recovery policy (Ganapati, 2011b). The Dutch government uses dashboards to enable the monitoring of large IT-projects by the public (www.rijksictdashboard.nl). As such, dashboards are becoming an important means of communicating and interacting with the public to create transparency and to achieve accountability. The latter refers to the answerability for one's actions or inactions and to be responsible for their consequences (Roberts, 2002).

In daily life, a dashboard is a control panel of the driver of a car in which the driver can see what happens and has controls to intervene. Few (2006) defined dashboards as "*a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so the information can be monitored at a glance*" (p. 34). This definition might be challenged, as governments might have multiple sometimes conflicting objectives, and it does not acknowledge the ability to intervene. Furthermore, dashboards can enable the zooming in on more detailed information and might not be limited to a single screen. Therefore, in this paper we define dashboards as "*the visualization of a consolidated set data for a certain purpose, which enables to see what is happening and to initiate actions*". The purpose varies depending on the focus on the public or for use for policy-making. In our definition the possibility to intervene is a key aspects, as only viewing data without having the possibilities to take actions based on the result has limited use and might result in the abandon of the dashboards. Dashboards are often part of public organizations' '*open government*' efforts, which aim at creating transparency and stimulating engagement with citizens and business. The efforts of open government can result in a more democracy (Cuadrado-Ballesteros, 2014), efficiency (Navarro-Galera et al., 2016), and transparency, accountability, collaboration and engagement (Bertot, Jaeger, & Grimes, 2010; Dawes & Helbig, 2010) and trust into the government (Lourenço, 2015). Despite the promises, employing dashboards is often a difficult endeavor. Data is often context-specific, and without in-depth knowledge of the context in which the data is collected, interpretation will likely be wrong (Matheus & Janssen, 2013). As such, data science in government requires in-depth skills and knowledge about the inner workings of the governments and its environment (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012).

The objective of this paper is to understand and to support the design of dashboards for creating transparency and accountability. Literature and two case studies are investigated to identify benefits, risks and principles for designing dashboards in the public sector. Following these principles can lead in realizing the benefits and overcoming the risks. A cycle describing the data cycle of open and private data for dashboards is presented to show how value can be created from data.

This research is conducted by reviewing literature about dashboards and complementing the literature by investigating two case studies in detail. The literature review was performed using the top 20 journals of 2015 in the Scimago Journal Rank (SJR) in the Information System (IS), Information Systems and Management, and, Library and Information Science. All the journals were surveyed, using the keyword "dashboard". Around 130 papers were found, considering title, abstract, keywords and citations. The major part of the papers mentioned only dashboard in the text, whereas the research was not focused on dashboards. Only 9 of these papers were related to smart cities. The scant literature shows that dashboards suggests that dashboards are underexplored, whereas they are essential in data science.

This paper is structured as follows. In Section 2, two different cases are presented in which both the government and the public plays a major role. The cases and literature are used to understand the value creation mechanisms, benefits, risks and challenges when designing and using dashboards, which is presented in the Section 3. Section 4 presents design support for developing dashboard in the form of design principles and a data cycle for dashboard. Finally, conclusions are drawn in Section 5.

2. Dashboards in Smart City practice

Due to the economic boom and access to credit in Brazil, there has been an increase in the number of cars (3 million cars for 6.5 million inhabitants) resulting in huge traffic jams, even outside of rush hours. The Smart City dashboards in Rio de Janeiro, Brazil, were created to solve problems related to public transportation and traffic. For this an infrastructure, a dashboard, and a data portal with more than three thousand datasets and seven APIs for real-time data use (www.data.rio) were developed and used by the Center of Operations Rio (COR). COR is situated in a four floor building that reunites almost 30 local secretariats, public and private enterprises to identify and solve in real-time issues on the city. COR collects around 4 Gigabytes (GB) of data on transit every day. This includes data about bus stops, car accidents, constructions works, and accidents like tree falling down on the streets. This excludes real-time GPS (geographical location collected using the Global Positioning System – GPS) data coming from buses driving their routes. Every 2 min, data is collected from the position of over 8000 buses. This result in another dataset which amounts to some 12 GB per day. With data from COR and the work of the data scientist group Pensa dashboards are created with the objective of putting key organizational elements into a consolidated format using several visualization tools, gauges, graphs, charts, and pictograms. The Pensa (www.pensa.rio), a group of big data scientists at Rio de Janeiro City Hall, enabled COR to visualize the data in a structured, integrated, and organized manner at a glance. The process of data analyses used by Pensa in Rio de Janeiro is based on questions made by the political decision-makers (mayor and secretariats) to create dashboards.

2.1. Case 1: traffic dashboards

The first case represents a partnership between the Smart City of Rio de Janeiro and the Social GPS Smartphone application called 'Waze' (www.waze.com) to employ citizen-generated data combined with government data as shown in Fig. 1. This app allows citizens to send real-time information about traffic conditions and accidents to the city. The COR uses the Waze application to send real-time information about route changes, flood routes, traffic jams, and car accidents to its citizens. The result is a combination of open data from the Rio de Janeiro City Hall and user-generated content collected through Waze about Rio de Janeiro's 7 million inhabitants. This big data is presented in real time using electronic panels positioned all over the city.

Two dashboards were created based in the Big Data Analytics. The first dashboard is showed in a big video-wall remembering the North American Space Agency (NASA) Lyndon Johnson Center, in Houston, as an interactive map version at COR called GeoPortal (Dashboard A in Fig. 1). This dashboard shows in real time, 24 h per day, 7 days per week, where the traffic jams and accidents are. Combined with other data sources, including 900 cameras that can turn 360°, spread over the city recording 24 h of the traffic, allows public servants at the operational level to make decisions to solve traffic problems. Accidents street holes and other issues are easily identified with traffic jams and prescanned using video cameras. The most-suitable team to deliver the service and solve the problem will be called. For example, one team might be able to arrive faster, but has no health equipment or technical capacity to solve the issue. This helped also the local government to plan the long-term public service delivery based in frequency and type of events, positioning proper teams to solve long-term issues.

The second dashboard is based on the same data, but it is presented in different format to the public. Ten dashboards were built in the top-10 most congested parts and directions of avenues in Rio de Janeiro (e.g. Dashboard B in Fig. 1). The top-10 most congested places were identified with Big Data analysis from the data scientist group Pensa, taking into consideration data from several internal databases (GPS data from buses, traffic jams reported by civil servants, speed of traffic from speed traps, car accidents, and others) and the Waze application