International Journal of Electronic Government Research
Volume 14 • Issue 3 • July-September 2018

Automated Evaluation of Open Government Data Portals:
A Case Study

Fabiano Rodrigo Alves Nascimento, UNIRIO - Federal University of State of Rio de Janeiro, Rio de Janeiro, Brazil
Junior Cesar da Rocha, UNIRIO - Federal University of State of Rio de Janeiro, Rio de Janeiro, Brazil
Ana Cristina Bicharra Garcia, UNIRIO - Federal University of the State of Rio de Janeiro, Rio de Janeiro, Brazil
https://orcid.org/0000-0002-3797-5157

ABSTRACT

The quality of public data made available by governments is a crucial factor in achieving greater transparency and also for such data to be used effectively by society. Several models of quality evaluation of open data portals have been proposed in recent years. The vast majority of these models incorporate manual evaluation processes, which makes it time consuming and expensive to maintain a continuous evaluation of open data portals. In order to verify the reliability of the results generated by an automated evaluation model, a comparative analysis was performed between an automated model and a manual model. The results showed the degree of convergence of the quality criteria of the automated model in relation to the manual evaluation model.

KEYWORDS

Automated Quality Evaluation, Data Quality, E-Government, Klein, Neumaier, Open Data Portal, Open Government, Openness

1. INTRODUCTION

Emerging from e-Business ideas in the late 1990s, e-Government was initially seen as a concept focused on exploring advances in information and communication technologies as a means of delivering services. Shortly afterwards, e-Government became a key facilitator of the transformation of the public sector to effective governance, transparency, accountability, and citizen participation in democratic processes and public policymaking (Veljković et al., 2014). The so-called Open Government (OG) represents a modern method of governance that offers a new space for openness, transparency, and ongoing dialogue between government and its citizens (Parycek & Sachs, 2010).

According to Veljković et al. (2014), transparency is a crucial ingredient of Open Government; more transparency means better governance, more efficiency, and legitimacy. Transparency can be further disseminated to the transparency of government operations, procedures, and tasks (government transparency) and transparency of government-maintained data (data transparency). The transparency of a government is a means to achieve a responsible government that measures and monitors the results of its actions and takes responsibility for the results. Allowing people to see the internal flows

DOI: 10.4018/IJEGR.2018070105

of government and investigating whether their representatives meet their expectations is an important step toward achieving responsible government.

Open access to government data has become a hallmark of the global open government movement. A legal framework for access to public information is one of four criteria for a government to participate in the Open Government Partnership (OGP) (Daves et al., 2016) which grew from eight member countries in its foundation in 2011 to more than 90 nowadays. The public availability of government information has been determined as a civil and human right by the United Nations' Millennium Development Goals and acts as a key strategy for assessing government accountability (Linders & Wilson, 2011)

In recent years, a number of open data movements have emerged around the world, with transparency and data reuse as two of the main objectives (Attard et al., 2015). The emergence of open data portals could in principle change the way citizens and researchers look for accountability-related data, since these portals, by definition, function as a centralized point of access to government data (Lourenço, 2015). Ideally, making this data available on the Web would lead to more transparency, participation, and innovation throughout society. However, just publishing the data on the web is not enough. To truly advance the open society, the publication platforms need to fulfill certain legal, administrative as well as technical requirements (Braunschweig et al., 2012).

The poor quality of datasets and metadata affects the discovery and consumption of data by stakeholders (Umbrich et al. 2015). Braunschweig et al. (2012) argue that most open data platforms have a lot of published data that is not readable by a machine or is in a proprietary format, which of course prevents reuse by automated tools. In a similar study, Petychakis et al. (2014) have reached the same conclusion, and further attest that the vast majority of the datasets they reviewed are published without specification of an open license.

According to Máchová et al. (2017), the extraction of valuable information coming from open government data sources requires the quality evaluation of them since the quality of data plays an essential role in the use of open data portals and a certain level of data quality is critical for Open Government Data use. Moreover, there is a need for quality evaluation to better understand quality issues in open data portals.

In order to judge the fulfillment of quality requirements, several authors presented different models and criteria to evaluate and diagnose open data portals (Solar et al., 2014). Conforming to Klein (2017), the different approaches proposed make use of arbitrary quality attributes weights to assess the quality of open government data under evaluation instead of taking account of open data users perception. In this sense, his research presents a model that incorporates a more complete set of quality attributes defined through an extensive Systematic Review of Literature (RSL) and specify the quality attributes weights in a methodical process that incorporates the opinion of users and open data specialists. In spite of its noticeable advantages, this model runs under fully manual process and, consequently, it takes some effort and time to go through.

According to Reiche et al. (2014), manual evaluation by humans is the most significant because it is the human who will make use of the repository. However, open government data portals are not static, on the contrary, the environment changes, laws are passed, or data bound for a given period becomes obsolete, and continuous reevaluation becomes essential. However, continuous manual evaluation is not feasible because it becomes costly and time-consuming. In this same line, Heinrich et al. (2007) state that one of the requirements for obtaining data quality metrics is its economic viability, which requires that the measurement procedure be performed with a high level of automation.

In order to reach a scalable and feasible way to execute continually open data portals evaluation, a few works proposed models with some level of automation, such as Yang et al. (2015), Vétro et al. (2016) and Neumaier et al. (2016). Nevertheless, none of these researches compare its results with others open data portals evaluation models to evidence its performance. In the sense of filling this current gap in the research literature, this paper aims to carry out a case study to investigate,