On the barriers for local government releasing open data

Peter Conradie [a,c,*], Sunil Choenni [a,b]

[a] Creating 010, Rotterdam University of Applied Sciences, P.O. Box 3001 HA, Rotterdam, The Netherlands
[b] Research & Documentation Centre, Ministry of Security and Justice, P.O. Box 20301, 2500 EH The Hague, The Netherlands
[c] Industrial System & Product Design, Faculty of Engineering and Architecture, Ghent University, Graaf K de Goedelaan 5, 8500 Kortrijk, Belgium

ARTICLE INFO

ABSTRACT

Due to expected benefits such as citizen participation and innovation, the release of Public Sector Information as open data is getting increased attention on various levels of government. However, currently data release by governments is still novel and there is little experience and knowledge thus far about its benefits, costs and barriers. This is compounded by a lack of understanding about how internal processes influence data release. Our aim in this paper is to get a better understanding of these processes and how they influence data release, i.e., to find determinants for the release of public sector information. For this purpose, we conducted workshops, interviews, questionnaires, desk research and practice based cases in the education program of our university, involving six local public sector organizations. We find that the way data is stored, the way data is obtained and the way data is used by a department are crucial indicators for open data release. We conclude with the lessons learned based on our research findings. These findings are: we should take a nuanced approach towards data release, avoid releasing data for its own sake, and take small incremental steps to explore data release.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The release of Public Sector Information (PSI) by governmental organizations is getting increasing attention from local, national and international government levels.

On the one side, the motivation for the release of data ranges from the hope of increased involvement of citizens in government, increased transparency and improved decision-making (Bertot, Jaeger, & Grimes, 2010; Dawes, 2010a, 2010b), and aligns with high level ambitions to realize smarter government through data provision (Bertot & Choi, 2013). On the other side, the release of data for a broader use may give a boost to the creative industry, which in turn leads to innovative applications and techniques in the public as well as in the private sector (Graves, 2011; van Dijk, Kalidien, & Choenni, 2013). These developments contribute to the creation of sustainable economies.

However, this increased call for data release by governmental agencies presents data professionals with tasks that are still unknown and with costs that are unclear. Benefits that open data can have also need to be further explored. Having said that, the directives from higher levels of government are putting pressure onto local governments to find answers to such pressing questions in relation to open data. These questions might include what the challenges, potentials, barriers and impact of open data are.

The assumed benefits of open data have resulted in several governments' initiatives facilitating the release of PSI. These governments, among others, include Colombia, (Prieto & Rodríguez, 2012), Brazil (Matheus & Vaz, 2012) and the U.S.A. (Wilson & Linders, 2011). An overview of these initiatives is provided by Yang and Kankanhalli (2013) and Janssen (2011).

While possible barriers and potentials of open data are reported by several researchers (see Section 2), a deep understanding of the underlying processes for these barriers is missing. In this paper, we study the underlying processes that pertain to data release in a local context, i.e., we search for crucial determinants for the release of open data. We have studied the processes involved in releasing data initially within six public sector organizations in Rotterdam, The Netherlands.

We have implemented a participatory action research approach (Whyte, 1989) to identify the indicators that play a significant role in releasing data for the public at the participating organizations. Important elements in this approach are desktop research, workshops, questionnaires, and in-depth interviews with key persons at different levels of the organizations. We have found that important indicators for data release are how the data is stored (distributed/decentralized versus centralized), how the data is obtained, and the way data is used by the organization. Additionally, we examined the suitability of the data itself, based on potential privacy and judicial issues. We discuss how these indicators may contribute to shaping an open data policy at a local level.

So far, studies that pertain to open data lack the focus on the issues experienced on a local level by public sector information professionals. There is an absence of understanding on local government levels on the impact, barriers and opportunities of open data release. This view

* Corresponding author at: Creating 010, Rotterdam University of Applied Sciences, P.O. Box 3001 HA, Rotterdam, The Netherlands.
E-mail addresses: peter.conradie@ugent.be (P. Conradie), r.choenni@hr.nl (S. Choenni).

is shared by Kassen (2013) who states that analysis of open data projects on a local level is needed to understand how the potential of open data can be realized.

Our study focuses on the understanding of the underlying processes entailed by open data at a municipal level. The rationale to choose this level instead of a national level is that data is mainly gathered at local levels, and therefore the support of local governments is of crucial importance for the success of open data. Additionally, while many works exist about open data as emergent theme from a national perspective, a view of the issues from a local perspective is not yet as present.

While certain datasets, released as part of our research, did indeed encounter barriers related to re-use, finance or privacy, we focus here primarily on the underlying processes of data release in order to prevent the dominance of these issues, and also to understand the thresholds to release data from an additional perspective.

The remainder of this paper is organized as follows. In Section 2, we embed our work in the field of open data in more detail. In Section 3, we motivate the use of a participatory action research approach for our purpose, and we describe the implementation of this approach in Section 4. In Section 5 we present our findings from the participatory action research sessions. Section 6 is devoted to the lessons that we have learned so far. Section 7 reports about our future activities and final conclusions.

2. Related work

Open data has gained a lot of attention in recent years, both from academia and the public sector. For example, Jaeger and Bertot (2010) discuss the issues associated with citizens' increased access to PSI, while Bertot et al. (2010) elaborate on the role of ICT as a tool for openness. Dawes (2010a, 2010b) also examines the goal of greater openness through accessibility of information, while Robinson, Yu, Zeller, and Felten (2009) plea for an extended government technical infrastructure for information provision to improve transparency. Wilson and Linders (2011) reflect on the Open Government Directive of the U.S.A., developing a framework for future work, which relates to a measurement framework for U.S. federal transparency introduced by Bertot, McDermott, and Smith (2012). Geiger and Von Lucke (2012) also look at the publication of open data, and the general challenges of data publication.

Also when considering the practical barriers of open data release, various efforts are made to understand and bridge the current thresholds. From a user perspective, major barriers are the access to proper datasets and the adequate use of datasets (Zuiderwijk, Janssen, Choenni, Meijer, & Sheikh Alibaks, 2013). Datasets are also offered in a fragmented way, and offered at several websites, which are in some case hard to find (Boulton, Rawlins, Vallance, & Walport, 2011; McLaren & Waters, 2011). Moreover, the access to datasets is in some cases restricted to specific user groups. Adequate use of datasets is hindered since metadata are poorly documented and, therefore, the semantics of the data may be ambiguous. Adequate metadata are also important to assist improve data re-use (Zuiderwijk, Jeffery, & Janssen, 2012). Furthermore, determining the quality of a dataset is an open question. Zhang, Dawes, and Sarkis (2005) also report on various thresholds to the release of data, including the lack of tools for sharing, and conflicting data definitions.

Zuiderwijk, Janssen, and Choenni (2012) provide an overview of the barriers that users may encounter in using public sector information, such as lack of knowledge to deposit data, or no knowledge about its existence. Additionally, domain knowledge might be needed to understand the data. While these issues are relevant in the context of open data, they are not limited to this domain and also occur for any data re-use generally.

Furthermore, organizations need practical frameworks that support them to decide whether a dataset is eligible to release or not (Zuiderwijk et al., 2012). In these frameworks, special attention should be given to privacy issues. For example, organizations are struggling with the questions whether a dataset is or may become privacy sensitive (Choenni, van Dijk, & Leeuw, 2010). An issue might be the disclosure of personal identities when combining different datasets (Kalidien, Choenni, & Meijer, 2010).

These examples highlight the importance and emergence of open data as a theme. However, they refrain from examining open data in a local context. Furthermore, the literature presented emphasizes some of the issues associated with its release, such as privacy. However, PSI is not always personally bound. Moreover, the barriers that may be encountered depend on the type of data at hand to be released.

Given this, we agree with Kassen (2013) that open data as a concept is best understood through investigation of open data on a local level. As such, we argue that the discussion about open data stands to benefit from the research on open data projects that focus on a local level, not only focusing on the thresholds of release, but also focusing on how the processes present at public sector organizations can benefit or hinder open data.

3. Approach

In order to explore data release, the University of Applied Sciences in Rotterdam explored data release with six local public sector services. They are the City Development (Stadsontwikkeling), City Works (Gemeentewerken), Library (Bibliohteek), City Archive (Gemeentearchief), regional Police and Centre for Research and Statistics (Centrum voor Onderzoek en Statistiek).

The exact profiles of the organization will be further elaborated on in Section 4. First, we motivate in the following the choice for a participatory action research approach and describe how this approach was applied.

3.1. Participatory action research

To understand the barriers of data release, we employed a participatory approach. Formally referred to as participatory action research (PAR), this approach focuses on co-research with participants (Kemmis & McTaggart, 2000; Whyte, 1989).

From the previous sections, it is clear that the barriers – e.g., technical issues, privacy or law – associated with data release require expertise from different branches of knowledge. Seen from this perspective, PAR can be an appropriate methodology, due to its applicability in multidisciplinary research (Whyte, 1989). According to Hughes (2003), where the knowledge is *embedded in a local context*, a PAR approach can be beneficial. Landsbergen (2010) reports on a similar approach in the City of Columbus, where the focus was on social media use within several city departments.

PAR means a more involved approach by all participants. This has meant not only conducting interviews with our participants, but also organizing workshops to discuss emerging themes about data release on a local level. Researchers also aimed to be involved with discussions relating to open data release by attending internal events at the council about publishing data. Furthermore, events were organized to involve external stakeholders in the process, by giving a platform from which issues related to open data could be discussed. Most importantly was the involvement of stakeholders in the development of case studies that explore, guide and illustrate a data release. This was achieved by incorporating projects where open data is used to develop cases within the bachelor program of our university.

As advocated by Kemmis and McTaggart (2000), a PAR approach is iterative with researchers reflecting on new data. This iterative process occurred throughout the duration of the project, since lessons learned in initial stages and initial education projects were incorporated in subsequent data requests, project descriptions and objectives.

The goal was thus to let the projects act as driver for data release in the local council, first by encouraging the participating partners to