Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program

Thorhildur Jetzek [1]

Department of IT Management, Copenhagen Business School, Howitzvej 60, 2000 Frederiksberg, Denmark

ARTICLE INFO

ABSTRACT

Current literature on open government data has uncovered a wide range of challenges related to these important initiatives. The problems encountered include: insufficient data quality and interoperability, problems regarding governance and motivation, lack of capabilities, and heterogeneous political and ideological agendas. A common open data infrastructure might resolve some of these problems, however, implementing such an infrastructure is a highly complex task. This longitudinal case study of the Danish Basic Data Program (BDP) is intended to improve our understanding of the challenges related to providing open access to government data through open data infrastructure. The BDP aims to improve the quality of selected government data, make them more coherent, and improve accessibility through the implementation of a common data distribution platform. The program is expected to increase government efficiency and stimulate innovation. This case study describes the evolution of the BDP and identifies the main structural elements of an open data infrastructure. Data analysis uncovered four tensions, which are identified as key challenges of an open data infrastructure implementation. These tensions are presented with four suggested governance strategies that were used in the BDP case. The main contribution of the paper is a process model where the main phases and mechanisms of an open data infrastructure implementation, use and impacts are identified and explained.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

In August 2011, under an increasing need for savings coupled with a continual pressure to deliver high quality public services, the Danish Agency for Digitization published their e-Government strategy for 2011–2015. The two main goals of the strategy were: a) to ensure that citizens experienced a collaborative public sector and b) to reduce the costs of service delivery (Agency for Digitization, 2011). An additional important element to the strategy was an increased focus on re-using data across all levels of government. While Denmark enjoys a strong tradition of maintaining legislative public registers, there was a growing need for improvement in data quality and data management in order for authorities and businesses to harvest the full potential of these data, particularly across organizations (Agency for Digitization, 2012). Previous analysis had uncovered problems with both gaps and redundancy in data sets, as well as the lack of a consistent policy for who could utilize the data and for what purpose. Moreover, both public authorities and private businesses were suffering from cumbersome payment procedures and legislation that stood in the way of reuse. These issues were causing what has been identified by economists as sub-optimal use of the data resource (see for instance Nilsen, 2010 and Pollock, 2008).

The value of accessible, interoperable and contextually appropriate data is becoming apparent in many areas of society. It is estimated that the value of more effective use of data resources in the US health care sector alone could be USD 300 billion annually (McKinsey & Company, 2011). Moreover, the value of improved use of data across Europe´s public sector is estimated to be around EUR 250 billion annually (McKinsey & Company, 2011). The growing technical and semantic ability to access and merge disparate types of data, as well as the availability of sophisticated data analysis techniques and technologies, offers the potential to vastly increase the overall value of data. It has been suggested that one of the key enablers for increased use of data is the liquidity and openness of data, in the sense that the data are coherent, shareable and published in machine-readable formats (McKinsey & Company, 2013). The annual economic value potential of *open* data over seven key sectors is estimated by McKinsey and company as being over 3 trillion USD globally (McKinsey & Company, 2013). However, an increasing body of research has shown that while the *potential* of data-driven value generation is vast, it is still very difficult for most organizations to generate and harness this value in practice. (Conradie

E-mail address: tj.itm@cbs.dk.
[1] Twitter: @HildaJetzek; LinkedIn: dk.linkedin.com/in/thorhildur.

& Choenni, 2014; Janssen, Charalabidis, & Zuiderwijk, 2012; Martin, Foulonneau, Turki, & Ihadjadene, 2014; Zuiderwijk, & Janssen, 2014a). A proposition of this paper is that a lack of a common open data infrastructure might explain some of these practical problems.

This paper addresses the question of *How can the tensions in a multi-stakeholder open data infrastructure implementation be addressed through governance strategies*? This question is approached by examining the Danish Basic Data Program (BDP), which was initiated in 2012. The original aim of the program was to increase public sector efficiency through better internal use of data. However, the program soon took on a leading role in shaping the open data infrastructure in Denmark.

The paper is organized as follows: Firstly, the topic is motivated by discussing previous research and evidentiary claims, which illustrate the need for improved understanding of open data infrastructure implementations. Following this, the three different theoretical frameworks used to analyze the implementation of open data infrastructure in Denmark are reviewed. Thereafter, the research approach is discussed, followed by a description of the case and its context. The main phases of the program are summarized in a process model, which is the main theoretical contribution of the paper. Four governance tensions and four key implementation strategies are presented in the Analysis and discussion section. The paper concludes by discussing the foremost contributions to knowledge and practice, as well as the limitations of the research approach.

2. Motivation and frameworks for analysis

The original aim of the BDP was to implement a common data layer for a limited number of key sets of public data, with the goal of improved efficiency within the public sector (see further discussion in Section 4). The central premise was that by giving all users from public and private sectors access to the same data; double entry of data, as well as various shadow registers, would be eliminated, effectively raising the quality of the data and stimulating its use. However, as the program progressed, it evolved into a more generic *open data infrastructure* program, including a number of additional dimensions necessary for such an infrastructure. Infrastructure in general can be defined as basic physical and organizational structures needed for the operation of a society or enterprise, or the services and facilities necessary for an economy to function (Sullivan & Sheffrin, 2003). Accordingly, I define open data infrastructure as *the digital infrastructure needed to enable effective data sharing and use across boundaries*. Implementing a data infrastructure is a complex initiative that amongst other things requires a change in a number of organizational and societal structures, which are often resistant to change. This important non-technical side of digital infrastructure has recently been conceptualized as *soft infrastructure* (King & Uhlir, 2014).

As open data policies are a recent phenomenon, applicable systematic research that identifies different policies and outlines their impact has not emerged yet (Barry & Bannister, 2014; Hujiboom & Van den Broek, 2011; Zuiderwijk & Janssen, 2014c). Moreover, current research is inadequate in illustrating the necessary actions for open data publishing *after* development of policy. The relationship between the context of open data initiatives (technical, social, organizational and political), various elements of open data and data platforms, and the impacts of such initiatives, needs to be clarified (Zuiderwijk & Janssen, 2014b). A literature review revealed that information science literature offers insights into specific areas such as data quality and use of metadata, and computer science literature offers a sizable corpus of research on linked data principles, data models and platforms (See for instance Bountouri, Christos Soulikias, & Stratis, 2009; Höchtl & Reichstädter, 2011). However, a void still remains in work that presents a holistic view of the implementation and governance challenges faced by organizations in the act of publishing open data. Data infrastructure as a concept is in general not well represented in the literature, with the exception of spatial data infrastructures. Moreover, while it has been

proposed that open data infrastructure initiatives require a central governance organization willing to provide various support activities (Hofman & Rajagopal, 2014), this proposition needs more supporting evidence.

This paper is intended as an input into this emerging body of research. I suggest that by engaging in specific (context dependent) implementation strategies, the governance tensions associated with implementation of multi-stakeholder open data infrastructure can be alleviated. Moreover, I propose that if the data are to be effectively and diversely utilized by multiple stakeholders, as opposed to a particular group of users using data for specific purposes only; then an important phase of an open data infrastructure implementation must be transformation of data. A transformation not solely from *closed to open data*, but also from *siloed to liquid data*. A more detailed discussion of what this entails is offered in Section 2.1.

2.1. Dimensions and affordances of liquid open data

Most of the current open data initiatives are driven by governments. Innovators have been shown to value government data due to factors such as data quality, reliability of source, and comprehensiveness (Lakomaa & Kallberg, 2013; Jetzek, Avital, & Bjørn-Andersen, 2014). In well-resourced countries, few other institutions can provide such consistent and all-encompassing data, making open government data a valuable input to economic activity (Davies, 2013).

From a supplier perspective, the transformation to open data is proposed to be a function of availability, accessibility, format and license (Davies, 2010). From a demand perspective, openness is suggested to combine unrestricted availability of data with accessibility and technical interoperability (Tammisto & Lindman, 2011). In practice oriented literature, the term open data is interpreted in a variety of fashions, as evidenced from the many different working definitions found online. The Open Knowledge Foundation defines open data as "data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike" (Open Knowledge Foundation (OKF), 2015). However, this definition lacks reference to the technical dimensions of open data. Alternatively, Berners-Lee's five stars of linked data specify a number of technical dimensions.[2] However, the five stars are not really an open data definition, but rather a maturity model that focuses on how to gradually transform data into linked data, which indicates a method of publishing structured data so they can be interlinked and discovered through semantic queries.

An overview of multiple working definitions of open data is presented in Verhulst, Noveck, Caplan, Brown, and Paz (2014). This overview shows that the currently used definitions usually highlight 2-4 dimensions that are important for openness of data, in many cases with a specific purpose or ideology in mind. It also shows that no two definitions are based on exactly the same dimensions. The Open Data Barometer defines "truly open" data as data that are available online, in bulk, and under an explicit open license (Davies, 2013). However, in spite of the fact that considerable efforts have been exercised to make diverse government data available to the public, less than one in ten public datasets reviewed in seventy-seven countries in 2013 could be classified as truly open according to the Open Data Barometer definition (Davies, 2013; Höchtl, Davies, Janssen, & Schieferdecker, 2014).

To add to this discussion, I propose a construct I call *Liquid open data*. This construct was synthesized from different working definitions of open data, with an added insight from a review of the open data literature, as well as the analysis of this case study data. The construct consists of five main dimensions that offer seven important affordances to potential users. I propose that an open data infrastructure, which enables cross-boundary use of data by multiple stakeholders for a variety of

---