

Socio-technical Impediments of Open Data

Anneke Zuiderwijk^{1,2}, Marijn Janssen¹, Sunil Choenni^{2,3}, Ronald Meijer², and Roeksana Sheikh Alibaks²

¹ Delft University of Technology, Delft, The Netherlands

² Research and Documentation Centre – Ministry of Security and Justice, Den Haag, The Netherlands

³ Rotterdam University of Applied Sciences – Creating 010, Rotterdam, The Netherlands

a.m.g.zuiderwijk-vaneijk@tudelft.nl

m.f.w.h.a.janssen@tudelft.nl

r.choenni@minvenj.nl; r.choenni@hr.nl

r.f.meijer@minvenj.nl

r.sheikh.alibaks@minvenj.nl

Abstract: There is an increasing demand for opening data provided by public and private organisations. Various organisations have already started to publish their data and potentially there are many benefits to gain. However, realising the intended positive effects and creating value from using open data on a large scale is easier said than done. Opening and using data encounters numerous impediments which can have both a socio and a technical nature. Yet, no overview of impediments is available from the perspective of the open data user.

Socio-technical impediments for the use of open data were identified based on a literature overview, four workshops and six interviews. An analysis of these 118 impediments shows that open data policies provide scant attention to the user perspective, whereas users are the ones generating value from open data. The impediments that the open data process currently encounters were analysed and categorized in ten categories: 1) availability and access, 2) find ability, 3) usability, 4) understand ability, 5) quality, 6) linking and combining data, 7) comparability and compatibility, 8) metadata, 9) interaction with the data provider, and 10) opening and uploading. The impediments found in literature differ from impediments that were found in empirical research. Our overview of impediments derived from both literature and empirical research is therefore more comprehensive than what was already available. The comprehensive overview of impediments can be used as a basis for improving the open data process, and can be extended in further research. This will result in the solving of some impediments and new impediments might rise over time.

Keywords: open data, open government data, impediments, barriers, challenges, problems, user perspective.

1 Introduction

Recently, open data is gaining importance in the context of a growing demand for openness of public and private organizations. For instance, the Obama Administration and the European Commission increasingly plea for openness (Obama, 2009, European_Commission, 2003). Public and private organisations are under increasing pressure to release their data to a variety of users, including researchers, citizens, businesses and civil servants. Organisations in countries from all over the world have already started to publish their data and several examples of its use are already known. For instance, in the Netherlands an application was developed showing historical pictures, films and other information during a city tour, based on the geographical location of the user (<http://www.appsvoornederland.nl/apps/vistory-the-interactive-historical-video-app>). Another example comes from the Care Quality Commission in the United Kingdom, which publishes rates of survival after heart surgery, aiming to help people who need heart surgery to make informed choices about their care and treatment and to improve the quality of healthcare (<http://heartsurgery.cqc.org.uk/survival.aspx>).

The potential value of open data is said to be enormous and is expected to result in applications and practices we are currently not aware of (Dekkers et al., 2006, European_Commission, 2011b). However, the process in which data are published, found, used, linked, reused and discussed, which is here referred to as the open data process, seems to encounter many socio-technical impediments. *Impediments* are those factors which hinder or block the use of open data. Socio-technical impediments concern both humans and technology. According to Bostrom and Heinen, “the technical system is concerned with the processes, tasks, and technology needed to transform inputs to outputs”

whereas “the social system is concerned with the attributes of people (e.g., attitudes, skills, values), the relationships among people, reward systems, and authority structures”. The technical and social systems are assumed to interact (Bostrom and Heinen, 1977a, Bostrom and Heinen, 1977b).

The open data process implies that open data should not just be seen as a product, but as an on-going process (Janssen and Zuiderwijk, 2012), as new applications and use of open data might result in new insights which might result in new ways of using open data. In the open data process, public organizations open their data which can be used by citizens, businesses, researchers, civil servants and other types of users. From a high-level, the open data process can be divided in five basic steps. First, government organizations produce, collect and integrate large amounts of data to be able to fulfil their tasks (see box 1 in Figure 1). The production of these data is funded by public money. Second, public bodies decide whether they will open, i.e. publish, their data (see box 2). This can be on their own website, in a national portal or any other platform, such as the Engage platform (www.engage-data.eu). According to the European Commission, opening data should be realized by making all kinds of non privacy-sensitive data available on the internet (European Commission, 2011a). Government data that are published are then referred to as open data. Open data that are found by potential open data users (see box 3) “can be used, reused and redistributed by everyone, without restrictions from copyright, patents or other mechanisms of control” (LinkedGov, 2011, Open_Knowledge_Foundation, 2011, Sweeney, 2009) (see box 4).

A possibility of reusing open government data is by linking them to other data, so that relationships with other data can be shown (Berners-Lee, 2009). The Linked Data that are the outcome of this linking are defined as “a collection of interrelated datasets on the Web” (World_Wide_Web_Consortium, 2011, <http://www.w3.org/standards/semanticweb/data>). Data that are both open and linked, referred to as Linked Open Data (LOD), are data that meet the requirements of open data and that show relationships among the open data as well. Open government data obtain more value when they are linked compared to open government data that are not linked (Berners-Lee, 2009). Additionally, in the ideal open data process, feedback information on the use of the data is provided to the public bodies that produced the data (see box 5), so that these data may be used to improve work processes, such as policy making processes of public bodies (Janssen and Zuiderwijk, 2012). The open data process will be used to identify impediments.

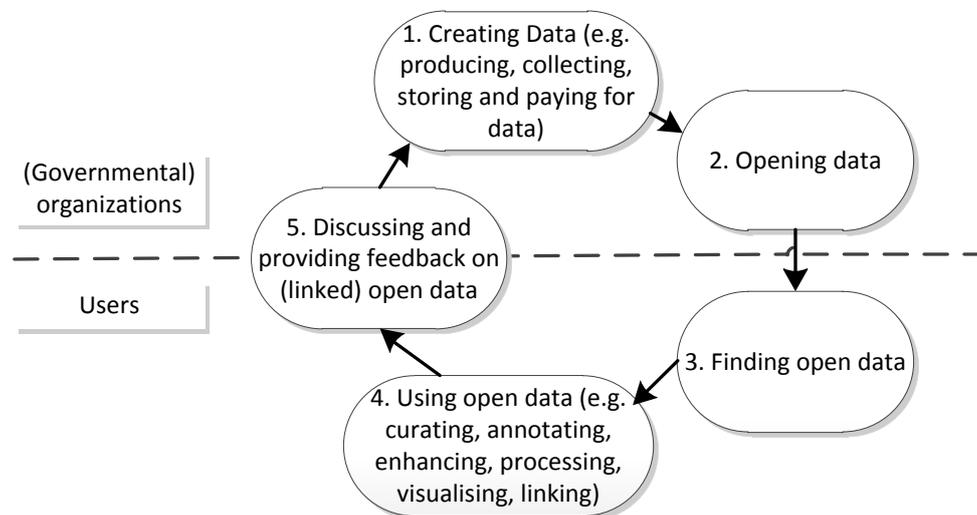


Figure 1: High level representation of the open data process including both (governmental organizations) and users

In spite of its significance, no extensive overview of the socio-technical impediments of the open data process is available from the perspective of the open data user. Many studies have investigated socio-technical impediments that are related to the perspective of the provider of the data (for instance, Huijboom and Broek, 2011, Zhang et al., 2005, Conradie and Choenni, 2012, Meijer and Thaens, 2009, Janssen et al., 2012, Zuiderwijk et al., 2012b), whereas the impediments that users of open data experience are often neglected (Blakemore and Craglia, 2006, Janssen et al., 2012). Creating an overview of the socio-technical impediments is relevant, as an extensive overview of these impediments of the open data process could provide practical knowledge about which

impediments currently exist and which challenges should be addressed by open data policies and strategies.

This paper aims to provide an extensive overview of socio-technical impediments of the open data process from the user's perspective by analysing the results of a literature overview, interviews and workshops. In the following section, the research approach will be presented. Subsequently, we describe a background for the identification of impediments based on open data policies and the impediments that we found in literature. This is followed by an overview and discussion of socio-technical impediments, including tables with categorized impediments and discussions of source differences, possible explanations of these differences and limitations of this research. We conclude the paper with the main findings of this impediment research.

Parts of this paper were published in the proceedings of the 12th European Conference on eGovernment (ECEG) (Zuiderwijk et al., 2012a).

2 Research approach

The aim of this paper is to create a comprehensive overview of the various types of socio-technical impediments. Such a list can be used to improve open data infrastructures, policies and strategies. The open data process could be improved and positive effects can be realised. We opted for using various sources, as this is expected to provide a more comprehensive overview of impediments than a single source. We started by creating a literature overview, which resulted in the identification of initial impediment categories and our initial list of impediments. Thereafter, empirical research was conducted by organizing four workshops and six interviews, as these sources could be used to reach different types of open data users, including researchers, civil servants, developers, journalists and companies.

2.1 Literature overview

The literature overview is created by searching for journal papers, conference papers, books, governmental and non-governmental reports and other information in various databases, including Science Direct, Scopus, TU Delft Repository, Google Scholar and Google. Keywords that were used during this search were combinations of the terms open data, open government data, linked open data, Public Sector Information (PSI), open data use, open data impediment, open data barrier, open data problem, open data restriction and open data challenge. In total, approximately 1.555.000 documents were found in the Science Direct database, 552.000 in Scopus, 7.000 in the TU Delft Repository, 2.000.000 in Google Scholar and 211.000.000 in Google. There might be an overlap between those documents. The documents were filtered by searching for impediments, barriers, problems and challenges for the use of open data. Most of the obtained documents appeared not to be useful, because they did not describe any impediments for the open data process. In total, 37 publications were selected that were relevant and from these an overview of the impediments for the use of open data was created. The relevance of the hits was determined by the search machines and by scanning the titles and abstracts of the documents.

2.2 Interviews

To obtain a better understanding of the impediments that were described in literature and obtain more in-depth information, semi-structured expert interviews were conducted with six key persons who worked with open data and/or worked often with people that worked with open data. The interviewees mainly worked in social sciences and humanities disciplines, as no sufficient open data infrastructures exist in these disciplines and for this reason impediments are expected to be experienced by these interviewees. The decision to interview those six persons was based on their experience with the use of open data, their recognized and respectable reputation and position, their track record of publication production, their track record of participation in conferences and their reach beyond strict and specialized communities. Finally, the decision to interview these six persons was based on their willingness to participate in this research. All interviews were conducted in December 2011 and January 2012. During the interviews, the interviewees were asked which types of open data and metadata they used, in which way, which impediments they noticed during their use of open data and metadata, which challenges existed for the use of open data and metadata and what their requirements and needs were for open data and metadata.

All interviews were transcribed. After conducting the interviews, the interviewees were asked whether a transcript of the interview could be sent to them for approval. All interviewees agreed with this and all transcripts were sent to the interviewees within a few days. The interviewees were asked to provide comments on the transcript within two weeks in case that they did not (totally) agree with it, so that the transcript could be assimilated.

2.3 Workshops

To broaden our knowledge about which socio-technical impediments currently influence the open data process, we conducted four workshops at international events. The workshops aimed at engaging a diverse composition of open data users, as different users are expected to mention different impediments. Furthermore, multiple workshops were conducted in various countries so that a large number of people with different nationalities would be reached, which decreases the risk on invalid or country specific conclusions. Table 1 shows the workshops that were organized.

Although several ideas about the impediments of the use of open data were discussed by the organizers of these workshops, only those impediments that were confirmed by the participants are described in this paper.

Table 1: Organized workshops to obtain information about socio-technical impediments for the open data process.

Conference and workshop title	Location and date	Number & type of participants	Aim(s)
1) International Conference for E-Democracy and Open Government (CeDEM12) , "Open Linked governmental data for citizen engagement - A workshop about the benefits and restrictions of open linked governmental data and the role of metadata in citizen engagement" (90 minutes).	Danube University of Krems, Krems an der Donau, Austria. May 4, 2012	17. Mainly civil servants (federal government and municipalities), academic researchers, students.	The aim was to gain an insight in the way that open data are used and reused at present, including the benefits and impediments that are typical for currently available open government data. Regarding the benefits and impediments, in particular the role of metadata were discussed during this workshop.
2) Annual International Conference on Digital Government Research (DG.O2012) , "Linking open data - Challenges and Solutions" (half day)	Robert H. Smith School of Business, University of Maryland, USA. June 4, 2012	26 (after the break 22). Mainly researchers (universities, government, other research organizations).	The organizers presented an overview of recent developments in the field of open governmental data and examples of the use of these data. Furthermore, the main aims and the current status of the ENGAGE project were presented, which aims to develop an infrastructure of the provision and use of open government data. Subsequently, participants were asked to fill out the ENGAGE questionnaire, which aimed to find out the needs of (potential) users of open public sector data.
3) Samos 2012 Summit on Open Data for Governance, Industry and Society (Samos Summit) , "Open Data Requirements" (90 minutes)	University of the Aegean in Samos, Greece. July 3, 2012	16. Mainly students, academic researchers.	The aim was to exchange knowledge and ideas about the status of open data, its benefits and challenges and user requirements. Participants were asked to fill out a questionnaire about requirements of open data use, to write down on post-its the three requirements that were most important according to them and to discuss underlying impediments, ideas and experiences. Finally, the organizers of the workshop presented the interim results of the questionnaire and put forward several propositions, which provided the basis for an interactive discussion among all participants.
4) International Federation for Information Processing - Electronic Government Conference (IFIP EGOV 2012) , "A workshop about using open public sector data: The ENGAGE project" (half day)	University of Agder, Kristiansand, Norway. September 3, 2012	12 (after the break 10). Mainly researchers (universities), civil servants, companies.	The aim was to exchange information about the status of open data, its benefits and challenges and user requirements. Participants were asked to fill out a questionnaire about requirements of open data use. The interim questionnaire results were presented to and discussed with the participants.

3 Literature background

In this section we describe several open data policies, which have resulted in impediments for the open data process. Finally, the impediments that were found in literature are presented.

3.1 Open data policies

Impediments of the open data process can be dealt with by developing good open data policies. An important open data policy is the EU Public Sector Information (PSI) Directive, in which a common legislative framework was presented which regulates making data of public sector bodies available for re-use (European Commission, 2003). In this report, the European Commission (EC) argued that a general framework "is needed in order to ensure fair, proportionate and non-discriminatory conditions

for the re-use of [PSI]” (p. 1) and that “PSI is an important primary material for digital content products and services” (p. 1). After the launch of the PSI-directive, many directives and implementation guidelines followed. In 2006, the EC developed a policy for the reuse of her own information sources which includes the statement that all general accessible data of the EC should become available for everyone, usually for free (European_Commission, 2011a). In 2009, the Obama Administration stated that a primary goal was the establishment of an unprecedented level of openness of the Government (Obama, 2009). The Obama Administration published an Open Government Directive some months afterwards (The_White_House, 2009). Building on former policies, the EC has presented an Open Data Strategy for Europe, in which more evident rules on making the best use of government-held information are presented (European_Commission, 2011b).

An important change of the Open Data Strategy of 2011 compared to directives and guidelines that were released by the EC before, is that it was made “a general rule that all documents that are made accessible by public sector bodies can be re-used for any purpose, commercial or non-commercial, unless protected by third party copyright” (p. 1) and that “public bodies should not be allowed to charge more than costs triggered by the individual request for data (marginal costs)” (p. 1). The EC will lead by example; it will open its data for free through a new data portal (European_Commission, 2011b).

In May 2012, the Obama Administration published the Digital Government Strategy, which aims to 1) enable the American people to access high-quality digital government information and services anywhere, anytime, on any device; 2) ensure that as the government adjusts to this new digital world, we seize the opportunity to procure and manage devices, applications, and data in smart, secure and affordable ways; and 3) unlock the power of government data to spur innovation and improve the quality of services for the American people (Obama, 2012).

Despite the considerable attention that has been given to open data policies, the policies barely pay attention to overcoming the impediments of the use of open data. Most open data policies are internally focused and pay hardly any attention to the possible users who are needed to generate the intended impact of the policies (Zuiderwijk and Janssen, Forthcoming).

3.2 Impediments found in literature

Although the open data movement is guided by PSI-directives, strategies and national policies, open data policies are accompanied by many impediments. Previous research has shown that there are at least three main categories of impediments of current open data policies from the perspective of the user. First, data access impediments exist, which are related to creating, opening, finding and obtaining the data (box 1, 2 and 3 in Figure 1). Second, data use impediments directly restrict the use of open data (box 4 in Figure 1). Third, data deposit impediments make it difficult to store, discuss and provide feedback on datasets (box 5 in Figure 1). The main impediments of current open data policies are shown in Table 1.

Table 1: Impediments derived from literature.

1) Data Access Impediment

Description	Literature
Gaining access to data is difficult.	(Boulton et al., 2011, McLaren and Waters, 2011, European_Commission, 2011b)
PSI is published at several fragmented sources, which makes it hard to find them.	(Vickery and Wunsch-Vincent, 2006)
PSI information is deposited at several fragmented sources, so that users have difficulties with finding the data.	(FP7-ENGAGE, 2011)
A major barrier is the access to proper and appropriate datasets.	(Conradie and Choenni, 2012, Janssen et al., 2012)
The data are temporarily not available on the website	(Veljković et al., 2011)
The data are only partly available; incomplete information	(DataCite, 2011, Janssen et al., 2012)
No access to the original data, only processed data. Apps hide the complexity but also potential of the use of open data	(Janssen et al., 2012)

Description	Literature
Many PSI-websites require taking action from the view of the user, for example registration, membership, filling a form, obtaining written permission or a fee	(Blakemore and Craglia, 2006, Meijer and Thaens, 2009, Murray-Rust, 2008, Janssen et al., 2012).
Access to the data is restricted to a group of users	(DataCite, 2011)
Users may have to accept a variety of use agreements before they can get access to the original data. There is no uniform set of licensing terms for its re-use	(DataCite, 2011, Judge, 2010, Vickery and Wunsch-Vincent, 2006, Janssen et al., 2012)
Significant differences in national rules and practices persist. This leads to fragmentation of the internal information market and hinders the creation of cross-border information services	(European_Commission, 2011a)
Datasets are fragmented and offered at several websites, which are in some case hard to find access to datasets are in some cases restricted to specific user groups. There is no central portal or architecture	(Conradie and Choenni, 2012, Janssen et al., 2012)
Difficulty in searching and browsing due to no index or other means to ensure easy search for finding the right data	(Janssen et al., 2012)
The income of several government organizations is based on the selling of data, which makes them reluctant to publish the data	(Huijboom and Broek, 2011)
Scientists generally see published data as belonging to the scientific community, but many publishers claim copyright over data and will not allow its re-use without permission.	(Murray-Rust, 2008)
Many location-enabled publicly owned datasets have not been easy to access and assemble across local and central government bodies	(McLaren and Waters, 2011)
The notion of access is contested and is the outcome of a political process. There is a set of characteristic tensions that make it very difficult for the parties involved (producers, intermediary brokers, information users, and citizens) to find effective common ground in a debate that should focus on building the “information commons.”	(Blakemore and Craglia, 2006)
Administrations typically express reluctance to make their data and metadata (information about the data) available, for various cultural, political, legal, institutional and technical reasons. They keep data and metadata within their legacy systems, fenced and isolated	(European_Commission, 2011c)
The potential user cannot get access to the data, for example because of administrative overhead, different user registrations, terms of access and other barriers	(FP7-ENGAGE, 2011)
Data are kept behind closed doors or pay walls which require that people purchase what information is available	(Vogel, 2011)
There is a lack of information that certain data actually exists and is available	(European_Commission, 2011b)
There is a lack of clarity of which public authority holds the data	(European_Commission, 2011b)
There is a danger of public sector bodies abusing their monopoly power as the only producer of a particular type of data in order to charge excessive prices	(Janssen, 2011)
Policymakers may have made strict access limitation contracts with various commercial providers, so that data cannot be accessed by external stakeholders.	(Napoli and Karaganis, 2010)
There may be practical barriers such as difficulties with getting access to the data (e.g. undue delays).	(Viglione et al., 2010)
Exclusive re-use agreements with one commercial actor or re-use restricted to a government-owned company	(European_Commission, 2011b)
Not all countries world-wide have adopted national open data policies.	(Schellong and Stepanets, 2011, Zuiderwijk and Janssen, 2012)
No research on the needs of data users.	(Zuiderwijk and Janssen, 2012)

Description	Literature
Too much information to process and not sure what to look at and too many data initiatives	(Janssen et al., 2012)

2) Data Use Impediments

Description	Literature
There is no commonly agreed metadata.	(European_Commission, 2011a)
There is insufficient metadata available (especially contextual metadata), so that no decisions can be made about the quality of the data and the way it was gathered and measured	(Xiong et al., 2011, Hernández-Pérez et al., 2009, Schuurman et al., 2008, Zuiderwijk et al., 2012c)
Adequate use of datasets is hindered since metadata with regard to datasets are poorly documented, and therefore the semantics of the data may be ambiguous	(Conradie and Choenni, 2012)
There are concerns with source trustworthiness, data provenance and legal aspects of data consumption	(O'Riain et al., 2012)
No explanation of the meaning of data	(Janssen et al., 2012)
It is difficult to measure the data quality from combined outputs	(Smith, 2011)
How to determine the quality of a dataset is an open question	(Conradie and Choenni, 2012, Janssen et al., 2012)
Limited quality of data. Several countries suggested that the quality of some government data is too limited to permit its publication	(Huijboom and Broek, 2011)
The quality of information is not automatically guaranteed, and insight is needed in this before the information can be used for certain purposes.	(Janssen et al., 2012)
There is a need to identify resources that will maintain the quality of PSI	(Blakemore and Craglia, 2006)
The quality of the ontologies may not be very high	(Klerk, 2011)
Lack of accuracy of the information	(Janssen et al., 2012)
Users may lack domain knowledge, which makes it difficult to compare, link and re-use data. This also gives problems in case that users are forced to employ various arbitrary data transformations to make data usable and comparable	(FP7-ENGAGE, 2011, King et al., 2011, Janssen et al., 2012).
An underestimated subject seems to be the availability of all kinds of capabilities and knowledge levels of users for using complex and more sophisticated data (lack of the necessary capability to use the information)	(Janssen et al., 2012)
Much of the current effort is focused on how to easily use data embedded in software applications, whereas linking and combining data by users requires sophisticated knowledge	(Janssen et al., 2012)
Statistical techniques are often used for the collection, analysis, interpretation, and presentation of data. Yet statistical knowledge is scarce.	(Janssen et al., 2012)
Conflicting data definitions are major obstacles	(Zhang et al., 2005)
Information is not available in a machine-readable format	(European_Commission, 2011a)
Users often have to convert PSI into a format that makes it possible to reuse the data and to compare it with other data	(FP7-ENGAGE, 2011, Janssen et al., 2012)
A lack of open data standards between (levels of) government organizations has been identified as a barrier to open data usage by citizens and businesses and subsequently new open data policy	(Huijboom and Broek, 2011, Janssen et al., 2012)
No uniform policy for publishing data	(Janssen et al., 2012)
Little attention has been paid to systematic and structured research on differences between open data policies.	(Zuiderwijk and Janssen, 2012)
Language barriers and interoperability aspects need to be tackled so that information resources from different organizations and countries can be combined	(European_Commission, 2011a)
Data cannot easily be analysed across different organizations.	(McLaren and Waters, 2011).

Description	Literature
Many location-enabled publicly owned datasets have not been easy to analyse across local and central government bodies.	(McLaren and Waters 2011)
The focus of the PSI debate is focused on an agenda articulated primarily through the power of the PSI producers, and less on the citizen/consumer needs	(Blakemore and Craglia, 2006)
Data users find it difficult to interpret the data, because they are unfamiliar with definitions and categories that are used to present the data. It is therefore difficult to draw conclusions from the data	(FP7-ENGAGE, 2011)
Difficult to involve domain experts in the transformation and use process of a dataset	(Klerk, 2011)
There is much duplication, too little re-use and too little interoperability between datasets	(McLaren and Waters, 2011)
Datasets are expensive	(McLaren and Waters, 2011)
Datasets have restrictive licenses	(McLaren and Waters, 2011)
Privacy legislation is a serious barrier for open data initiatives	(Kulk and Loenen, 2012)
Threat of lawsuits or other violations. For instance, privacy might be violated by using data	(Janssen et al., 2012)
Fragmentation of software and applications	(Janssen et al., 2012)
No standard software for processing open data	(Janssen et al., 2012)
There is a lack of agreement on URIs for resources, which can result in missing associations between resources during integration and entity consolidation difficulty	(O'Riain et al., 2012)
The reader may not be free to do what they wish with the text or data because of licenses	(Molloy, 2011)
Users must comply with standard conditions when they want to use the data	(Judge, 2010)
Database design issues may make it difficult to transform datasets	(Klerk, 2011)
Data is made available only in formats that are difficult or expensive to use	(European_Commission, 2011b)
The conditions for different types of re-use may differ	(Janssen, 2011)
The development of cross-border information products and services is hindered by the many different rules and practices with regard to the procedures and conditions for re-use. Potential re-users do not know where they can find the data they need, or they are confronted with unclear or restrictive conditions under which they can re-use the data	(Janssen, 2011)
There is a lack of transparency about the availability of documents and the conditions under which they can be re-used	(Janssen, 2011)
Data may not be updated timely or they are obsolete and non-valid	(Lee and Kwak, 2012, Janssen et al., 2012)
Data might be simply incorrect, but also essential information about the data sets might be missing, such as the time period in which the data was collected.	(Janssen et al., 2012)
Lack of flexibility in data format	(Lee and Kwak, 2012)
Lack of accuracy and timeliness of data.	(Lee and Kwak, 2012, Janssen et al., 2012)
Data are typically available as is (i.e., in heterogeneous structures and formats), requiring substantial human workload to clean them up for machine processing and to make them comprehensible	(Ding et al., 2011)
Making public only non-value-adding data	(Janssen et al., 2012)
Similar data stored in different systems yields different results	(Janssen et al., 2012)
Lack of effective infrastructure with sufficient computing power to process large data and better social mechanisms to distribute the necessary human workload to stakeholder communities	(Ding et al., 2011)

Description	Literature
No tooling support or helpdesk	(Janssen et al., 2012)
No time to delve into the details, or no time at all	(Janssen et al., 2012)
There may be problems of usability of the data made available (e.g. when data are provided on paper instead of electronically)	(Viglione et al., 2010)
Unavailability of a supporting infrastructure and lack of standards, fragmentation, and legacy. These depict the need for ensuring a good infrastructure before the concept of open data will be widely accepted.	(Janssen et al., 2012)
Use would be stimulated if more information about the way open data were collected and processed was provided by including metadata.	(Zuiderwijk and Janssen, 2012)
Use and participation might be blocked, because there might be no incentives or no added value for users to make use of open data.	(Janssen et al., 2012)
Unclear value: information may appear to be irrelevant or benign when viewed in isolation, but when linked and analysed collectively it can result in new insights	(Janssen et al., 2012)
Focus is on making use of single datasets, whereas the real value might come from combining various datasets	(Janssen et al., 2012)

3) Data Deposition Impediments

Description	Literature
Opening data by data producers is a complex and ill-understood activity.	(Zuiderwijk et al., 2012b)
Threat of privacy violation by opening data and of being legally liable when opened data are misused.	(Kalidien et al., 2010, Zuiderwijk et al., 2012b).
Although many governmental organizations might be willing to open their data, they lack guiding principles derived from practical case studies that help them in doing this.	(Kalampokis et al., 2011, Zuiderwijk et al., 2012b).
No support for making data available	(Janssen et al., 2012)
Legacy systems that complicate the publication of data	(Janssen et al., 2012)
Most of the guiding initiatives are very general (as they have to cover many types of organizations and data) and barely provide practical support for opening up data.	(Zuiderwijk and Janssen, 2012)
Websites raise restrictions on data formats for depositions	(DataCite, 2011, Zuiderwijk et al., 2012a).
Users are asked to register or to become a member of the website	(Graaf and Waaijers, 2011) (DataCite, 2011)
The practice of depositing data is currently still limited to a minority of researchers and data sharing is confined to a limited number of datasets	(Graaf and Waaijers, 2011)
There are complicated licensing procedures or prohibitive fees	(European Commission, 2011b)
Lack of public participation mechanisms	(Lee and Kwak, 2012)
Public organizations do not react on user input	(Janssen et al., 2012)
There are no mechanisms for ensuring that results of open data are used by the government.	(Janssen and Zuiderwijk, 2012)
Debatable quality of user input.	(Janssen et al., 2012)
No process for dealing with user input	(Janssen et al., 2012)
The open data process is usually not viewed as an interaction process between the government and the public.	(Janssen and Zuiderwijk, 2012)
The use of open data might require considerable transformations of public sector organizations.	(Janssen and Zuiderwijk, 2012)
Public servants seem to be largely dominated by the focus on the publishing process and the associated challenges. In addition, they are satisfied if some data are published without thoroughly looking at the actual use.	(Janssen and Zuiderwijk, 2012)

The overview shows that most literature only mentions a limited number of impediments and barriers and there was no overview work, with the exception of (Janssen et al., 2012).

As the before-mentioned impediments are derived from literature, it is interesting to compare them with impediments that were derived from other sources. An overview of the impediments derived from the interviews and workshops will be given in the following section and compared with all the impediments that are mentioned in Table 1.

4 Overview and discussion of socio-technical impediments

In this section, an overview is provided of socio-technical impediments that influence the open data process from the viewpoint of an open data user. In addition, the source from which these impediments were derived is shown, so that a comparison can be made of which types of impediments are provided by which sources. The impediment overview is presented in three different tables. In line with section 3.2, Table 2 provides an overview of access impediments.

Table 2: Overview of socio-technical data access impediments that influence the open data process from the perspective of open data users.

Impediment	L	I	W
1) Availability and access			
Data are not collected, especially detailed data.		X	X
Unclear which data are collected and available.	X		
Unclear which organization collects which data.	X		
Data are not published, as organizations keep these data for themselves.	X	X	X
Data have a license.	X	X	
There is no uniform set of licensing terms for data use.	X		
It is unclear how to get a license for the use of data.		X	
Data are only available on request, with permission.	X	X	
Data are not available for free, a (large) fee must be paid.	X	X	X
Data are not provided continuously; they are not updated, so that monitoring is not possible.	X	X	
Data are only partly available.	X		
Data are only available for a certain group of users (e.g. commercial users, researchers or governmental organizations).	X		
Data are temporarily not available on the website.	X		
No access to the original data, only processed data.	X		
Difficult to obtain recent data of the last few years.		X	
Not all desired formats of the data are available.	X	X	
No good preservation of the data on the long term.	X		X
Governmental organizations sometimes use restrictions that are prohibited according to the law.		X	
No long term commitment to data updates by data providing organisations.			X
Not all countries world-wide have adopted national open data policies.	X		
There is much duplication of datasets.	X		
Lack of regularly updated data.	X		
No research on the needs of data users.	X		
Too much data to process.	X		
2) Find ability			
Data cannot be found.	X	X	
No advanced search possibilities.			X
Metadata cannot be found.	X	X	
Data are fragmented. There is no central portal for the data that you need for studies.	X	X	
Difficulty in searching and browsing due to lack of index or other means to ensure easy search for finding the right data	X		
Data are hidden in reports, cannot be found in a machine-readable format.	X	X	

Table 3 shows use impediments and Table 4 shows impediments for the deposition of reused open data. Within those three main categories, several subcategories are provided. The columns at the

right side of the table show whether the impediment was derived from the literature (L), the interviews (I) and/or the workshops (W). Some impediments were derived from multiple sources.

Table 3: Overview of socio-technical data use impediments that influence the open data process from the perspective of open data users.

Impediment	L	I	W
3) Usability			
Concerns with source trustworthiness.	X		
Data might be incorrect or essential information is missing.	X		
Lack of (domain) knowledge about how to treat the data.	X	X	
Expert advice is needed to use the data.	X	X	
Unclear what new business models should be thought of for the use of open data.		X	
Lack of services given by the data provider to use raw data.		X	
No interoperability of open data infrastructures with other systems.	X		X
The data are not relevant/interesting.	X	X	
No explanation of the applied licences for open data.	X		X
Threat of lawsuits or other violations.	X		
Users are forced to employ various arbitrary data transformations to make data usable and comparable.	X		
Data require substantial human workload to clean them up for machine processing and to make them comprehensible.	X		
Fragmentation of software and applications.	X		
No standard software for processing open data	X		
No time to delve into the details, or no time at all.	X		
No incentives or no added value for users to make use of open data.	X		
4) Understand ability			
Data are not understandable for the general public (e.g. related to jargon).		X	
No explanation of the meaning of data.	X		
Data are not visualized.		X	
Metadata are not visualized.			X
Lack of knowledge about how to interpret the data.	X	X	
Unavailability of a supporting infrastructure.	X		
Lack of a good API.		X	
Lack of skills and capabilities to use the data.	X		X
Lack of statistical knowledge.	X		
Data are provided in a language that the user does not understand.	X		
Metadata are provided in a language that the user does not understand.			X
Datasets are not complete.			X
No information about the provenance (context) of data.	X		X
No information about the time span (validity) of data.			X
No information about the validity of the data.			X
No version management, describing who has done what and when to datasets is available, so that it is difficult to understand the data.			X
No support and/or help and/or training for the use of the data is provided.	X		X
5) Quality			
Data are not reliable.		X	
No knowledge about whether the rating of data is reliable.			X
Difficult to determine the quality of data.	X		
Data are of limited or bad quality.	X	X	
Low quality of ontologies.	X		
Lack of accuracy of the data.	X		X
Data are often based on political approaches and intuition instead of an adequate level of intelligence.		X	
Difficult to measure the data quality from combined outputs.	X		

5) Linking and combining data			
Difficult to link data.	X	X	
Difficult to link data by linking metadata.			
Data cannot be linked to other data.		X	
It is unclear how downloaded data can be related/linked to other datasets.	X		
Linking and combining data requires sophisticated knowledge.	X		
Difficult to perform interdisciplinary research, as one may have enough knowledge about one discipline, but not about the other.		X	
Linking data and tools for linking data are lacking.		X	
No unique identifiers are available. So when data are linked to other data and the original dataset is updated, the other one is not.	X	X	
Focus is on making use of single datasets, whereas the real value might come from combining various datasets.	X		
7) Comparability and compatibility			
Many differences in definitions of data.	X	X	
Different types of open data policies.	X	X	
Different sources give different information about the same topic.	X		X
Differences between countries. The use of names of companies or postal codes could be very different in countries, which could lead to problems with linking data.	X	X	
Differences between organizations, for example, differences in terminology. This makes it very difficult to link and combine datasets.	X	X	
Different data standards are available and used.			X
No harmonised legal framework across countries.			X
No uniform policy for publishing data.	X		
Lack of research on differences between open data policies.	X		
A lack of standards.	X		
8) Metadata			
There is no commonly agreed metadata.	X		
The current metadata provision is insufficient. Especially contextual metadata is lacking.	X	X	X
No metadata about the quality of the data.	X	X	
No metadata about the way the data were gathered and measured.	X		
Metadata can be interpreted differently by different users.		X	
More information about metadata parameters is needed in particular policy sectors, such as knowledge management.		X	
Not enough domain information is provided.		X	
The metadata do not have a recognizable structure (metadata are now usually descriptive).		X	
When metadata contain assumptions for the use of open data, they could point at certain choices and interpretations. This may unconsciously exclude certain ways of reusing data.		X	

After presenting the main results, we discuss the differences between the impediments that are mentioned by different sources and provide possible explanations for these differences. Finally, the limitations of this impediment research are discussed.

Several data access impediments were mentioned by the different sources. The access impediments were divided into the categories 1) availability and access and 2) find ability. Most impediments belonged to the first category. Impediments that were mentioned by all three sources are 1) data are not published, as organizations keep these data for themselves and 2) data are not available for free, a (large) fee must be paid. The fact that all sources show these impediments might be a reason to prioritize them.

Most impediments for the open data process concern the actual use of open data. The analysis shows that open data policies provide scant attention to the user perspective, whereas the user needs to generate value from the open data. The use impediments were divided into the categories usability, understand ability, quality, linking and combining data, comparability and compatibility and metadata. One impediment was mentioned by all three sources, namely the impediment that the current metadata provision is insufficient. The fact that all sources show this impediment might be a reason to give priority to this impediment.

Table 4: Overview of socio-technical data deposition impediments that influence the open data process from the perspective of open data users.

Impediment	L	I	W
9) Interaction with data provider			
No discussion between the data provider and the data user possible.		X	X
No platform or mechanism for participation available.	X		
10) Opening and uploading			
Difficulties with uploading (reused) datasets.	X		
Threat of privacy violation by publishing data.	X		
Threat of being legally liable when opened data are misused.	X		
Lack of guiding principles and support for opening data.	X		
Websites raise restrictions on data formats for depositions.	X		
Users are asked to register or to become a member of the website.	X		
The practice of depositing data is currently still limited to a minority of researchers and data sharing is confined to a limited number of datasets.	X		
No mechanisms for ensuring that results of open data are used by the government.	X		
There is a lack of clarity about the terms of re-use.	X		
There are complicated licensing procedures or prohibitive fees.	X		
Legacy systems complicate the publication of data.	X		
Debatable quality of user input.	X		
Public organizations do not react on user input.	X		
No process for dealing with user input.	X		
The open data process is not viewed as an interaction process between the government and the public.	X		
The use of open data might require considerable transformations of public sector organizations.	X		
Public servants seem to be largely dominated by the focus on the publishing process and the associated challenges.	X		

The deposition impediments were divided into the categories interaction with the data provider and opening and uploading. All deposition impediments except one were derived from literature. None of the impediments found in literature was confirmed by the interviews or workshops.

4.1.1 Source differences

Analysing Table 2, 3 and 4, we see that many different types of impediments for the open data process are mentioned in literature, the interviews and the workshops. Literature often provided high-level impediments, whereas in the interviews and workshops more detailed impediments were revealed. This shows the complexity of categorizing impediments, as some of them are partly overlapping or could be viewed as subsets of others. For instance, we created a separate category for impediments that are related to the quality of the data, but quality related impediments could also be categorized under the category of usability or metadata.

Striking is that the sources do not always confirm each other. Most of the presented impediments are only mentioned by one or two sources, but not by all of them. The literature provided many different types of impediments for the categories of data availability and access, usability and opening and uploading. Impediments in the categories of understand ability and metadata were mentioned less by literature. During the analysis, we noticed that the same impediments about availability and access, metadata and data quality were mentioned by a large number of literature sources. For instance, impediments such as the data are not published, the quality is limited or unclear and insufficient metadata are provided, were mentioned in many publications.

The interviewees mentioned various impediments for the use of open data in each impediment category. They did not focus much on one category, although impediments in the categories availability and access, find ability, linking and combining data and metadata were mentioned slightly more. Few impediments were mentioned for the categories understand ability and opening and uploading.

The conducted workshops provided the smallest number of impediments. In comparison with the literature review and the interviews, the workshops provided few impediments in the categories availability and access, find ability, usability, quality and metadata. During the workshops, relatively many impediments were mentioned in the category understand ability. No impediments were mentioned about linking and combining and opening and uploading data.

Although some similarities were found of the impediments that were mentioned during the four workshops, most workshop impediments of Table 2, 3 and 4 were derived from different workshops. The differences between the workshop impediments cannot be attributed to differences in the type of participants, as most workshops in general had the same participant groups.

4.2 Possible explanation of source differences

We did not expect to find different impediments from different sources. A first possible explanation for the source differences could be the different aims and set-up of the data collection. Most literature communicates a certain idea or provides argumentation for a certain statement and does not aim to provide an extensive overview of all impediments for the open data process. Most literature only mentions a few high-level impediments, but does not focus much on detailed impediments. Impediments mentioned by literature were often used to introduce a topic and to provide support for writing an article. The interviews concentrated on obtaining more in-depth information about impediments for the use of open data and metadata and related background information. The workshops aimed at thoroughly discussing and obtaining feedback about certain ideas, such as open data infrastructures. In all workshops at least one presentation was given about the use of open data, which stimulated participants to ask questions about this use in the context of the presentation and discuss it afterwards, sometimes on the basis of several statements. This might have resulted in a certain bias towards the issues presented or touched upon by the presenter. The context of the various workshops may have been different. The workshop provided a specific context in which the impediments were mentioned. They might only be applicable in that context.

Furthermore, the interviewees and the people attending the workshops were mainly academia. The impediments that are provided in literature do not come from academic persons only, but also from practitioners and policy-makers. People with different backgrounds may have different perspectives on open data impediments.

The differences between the impediments that were found in current literature and the impediments that were found in empirical research may partly be due to the fact that literature does not pay attention to providing a structured overview of impediments. Furthermore open data impediments are likely to change over time, as in the past an impediment could be having no single shop to publish open data, whereas nowadays most countries have introduced an open data portal for publishing their data. Furthermore, in the past the availability of datasets might be an issue, whereas nowadays more and more data are available. However, a more thorough literature review is expected to have resulted in more repetition of the same impediments instead of an extension of the impediments that were found, as much repetition of impediments was already found in our current literature overview. Therefore, we may conclude that the impediments found in literature were incomplete and that our table which includes both literature and empirical research is more comprehensive.

Another possible explanation for the differences in impediments that were mentioned by different sources could be the fact that the field of open data is relatively new. As not much systematic research has been performed in the field of open data, no all-embracing impediment overview might be derived from literature.

4.3 Limitations

In the introduction we argued that creating an overview of the socio-technical impediments is relevant, as an extensive overview of these impediments of the open data process could provide practical knowledge about which impediments currently exist. The previous sections have shown which impediments we derived from literature, interviews and workshop. Nevertheless, identifying all impediments is difficult for several reasons. First, searching for open data literature provides considerable hits in online searching machines and not all those hits could be examined. In addition, during the interviews and workshops impediments may have been mentioned that were interpreted in a different way than that they were meant by the participants or they may not have been noticed at all. Moreover, other sources, such as questionnaires and focus groups, and other groups of open data

users, such as developers and journalists, could provide even more or other impediment information. Moreover, impediments may overlap. Therefore, it is difficult to assess whether the impediment overview that was presented in this paper is complete. Furthermore, impediments have different levels of abstraction, sometimes they are high level, whereas in other situation they are very detailed. This is one reason why comparing literature and our empirical research is difficult.

In the introduction it was stated that an overview of the socio-technical impediments could help us identifying which challenges should be addressed by open data policies and strategies. The impediment overview shows which impediments were mentioned by all sources and could get priority in open data policies and strategies. However, the impediment overview does not show how many times a certain impediment was mentioned by a source. For example, one impediment may have been mentioned in ten literature publications, but never during the interviews or workshops. In the impediment overview this impediment may seem to be not very important, although literature has shown that it is important. These limitations should be taken into account in interpreting the impediment overview.

5 Conclusions

The potential value of open data is said to be enormous. However, realising positive effects and creating public value from using open data on a large scale seems to be too ambitious at this moment, as the process in which data are published, found, used, linked, reused and discussed, which is here referred to as the open data process, encounters many impediments which hinder or block the publication or use of open data. In spite of its significance, no extensive overview of the socio-technical impediments of the open data process is available from the perspective of the open data user.

Based on a literature overview (37 documents), four workshops and six interviews, 118 socio-technical impediments for the use of open data were identified. The analysis showed that scant attention is paid to the user perspective, whereas the user needs to generate value from the open data. The impediments that the open data process currently encounters were categorized in ten categories: 1) availability and access, 2) find ability, 3) usability, 4) understand ability, 5) quality, 6) linking and combining data, 7) comparability and compatibility, 8) metadata, 9) interaction with the data provider, and 10) opening and uploading.

The impediments will likely change over time. New policies and infrastructures will likely solve many of the current impediments, and due to new and higher expectations new impediments might arise over time. For instance, a single shop for open data was an impediment a few years ago, but it is hardly nowadays, as most countries have established an open data portal. Different types of impediments were derived from the different sources. Differences were also found within the same source type. In addition, we conclude that the impediments that were found in literature differed from impediments that were found in empirical research. The impediments found in literature may have been incomplete and our overview of impediments which was derived from both literature and empirical research is more comprehensive. The comprehensive overview of impediments can be used as a basis for improving the open data process so that positive effects can be realised. More research should be performed to assess whether other sources and other groups of users provide more or other impediment information, to assess the importance of the individual impediments and to examine whether certain impediments hinder the open data process more than others.

6 Acknowledgements

This paper is related to the ENGAGE FP7 Infrastructure Project (An Infrastructure for Open, Linked Governmental Data Provision Towards Research Communities and Citizens). The authors would like to thank their colleagues of the ENGAGE project for their input for this paper although the views expressed are the views of the authors and not necessarily of the project. Furthermore, the authors would like to thank all persons that provided information about impediments of the open data process during the workshops and interviews.

References

- berners-Lee, T. 2009. *Linked Data*. [Online]. Available: <http://www.W3.Org/Designissues/Linkeddata.Html> [Accessed October 11 2012].
- Blakemore, M. & Craglia, M. 2006. Access To Public-Sector Information In Europe: Policy, Rights And Obligations. *The Information Society*, 22, 13-24.
- Bostrom, R. P. & Heinen, J. S. 1977a. Mis Problems And Failures: A Socio-Technical Perspective. Part I: The Causes. *Mis Quarterly*, 1, 17-32.
- Bostrom, R. P. & Heinen, J. S. 1977b. Mis Problems And Failures: A Socio-Technical Perspective. Part II: The Application Of Socio-Technical Theory. *Mis Quarterly*, 1, 11-28.
- Boulton, G., Rawlins, M., Vallance, P. & Walport, M. 2011. Science As A Public Enterprise: The Case For Open Data. *The Lancet*, 377, 1633-1635.
- Conradie, P. & Choenni, S. Year. Exploring Process Barriers To Release Public Sector Information In Local Government. In: 6th International Conference On Theory And Practice Of Electronic Governance (Icegov), October 22-25 2012 Albany, New York, United States Of America.
- Datacite. 2011. *Helping You To Find Access And Reuse Data* [Online]. Available: <http://Datacite.Org/Repolist>. [Accessed October 11 2012].
- Dekkers, M., Polman, F., Velde, R. T. & Vries, M. D. 2006. *Mepsir. Measuring European Public Sector Information Resources* [Online]. Available: http://Ec.Europa.Eu/Information_Society/Policy/Psi/Actions_Eu/Policy_Actions/Mepsir/Index_En.Htm. [Accessed October 9 2011].
- Ding, L., Lebo, T., Erickson, J. S., Difranzo, D., Williams, G. T., Li, X., Michaelis, J., Graves, A., Zheng, J. G., Shanguan, Z., Flores, J., Mcguinness, D. L. & Hendler, J. A. 2011. Twc Logd: A Portal For Linked Open Government Data Ecosystems. *Web Semantics: Science, Services And Agents On The World Wide Web*, 9, 325-333.
- European_Commission. 2003. *Directive 2003/98/Ec Of The European Parliament And Of The Council Of 17 November 2003 On The Re-Use Of Public Sector Information* [Online]. Available: http://Ec.Europa.Eu/Information_Society/Policy/Psi/Rules/Eu/Index_En.Htm. [Accessed].
- European_Commission 2011a. Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions. Open Data. An Engine For Innovation, Growth And Transparent Governance. Brussels.
- European_Commission 2011b. Digital Agenda: Turning Government Data Into Gold.
- European_Commission 2011c. *Towards Open Government Metadata*. Brussels.
- Fp7-Engage. 2011. *Deliverable 4.4.1: Engage Services Scenario's*. [Online]. Available: http://www.Engage-Project.Eu/Engage/Wp/?Page_Id=212 [Accessed October 11 2012].
- Graaf, M. V. D. & Waaijers, L. 2011. *A Surfboard For Riding The Wave. Towards A Four Country Action Programme On Research Data*. [Online]. Knowledge Exchange. Available: http://www.Knowledge-Exchange.Info/Files/Files/Downloads/Primary%20research%20data/Surfboard%20for%20riding%20the%20wave/Ke_Surfboard_Riding_The_Wave_Screen.Pdf. [Accessed October 11 2012].
- Hernández-Pérez, T., Rodríguez-Mateos, D., Martín-Galán, B. & García-Moreno, M. A. 2009. Use Of Metadata In Spanish Electronic E-Government: The Challenges Of Interoperability. *Revista Espanola De Documentacion Cientifica*, 32, 67-91.
- Huijboom, N. & Broek, T. V. D. 2011. Open Data: An International Comparison Of Strategies. *European Journal Of Epractice*, 4-16.
- Janssen, K. 2011. The Influence Of The Psi Directive On Open Government Data: An Overview Of Recent Developments. *Government Information Quarterly*, 28, 446-456.
- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. 2012. Benefits, Adoption Barriers And Myths Of Open Data And Open Government. *Information Systems Management*, 29, 258-268.
- Janssen, M. & Zuiderwijk, A. 2012. Open Data And Transformational Government. *Transforming Government Workshop*. Brunel University, United Kingdom.
- Judge, E. F. 2010. Enabling Access And Reuse Of Public Sector Information In Canada: Crown Commons Licenses, Copyright And Public Sector Information. In: Geist, M. (Ed.) *From "Radical Extremism" To "Balanced Copyright", Canadian Copyright And The Digital Agenda*. Irwin Law.
- Kalampokis, E., Tambouris, E. & Tarabanis, K. 2011. Open Government Data: A Stage Model. In: Marijn_Janssen_Et_Al (Ed.) *Egov 2011*. Delft, The Netherlands: Lncs.
- Kalidien, S., Choenni, R. & Meijer, R. F. 2010. Crime Statistics Online: Potentials And Challenges. In: Chun, S. A., Sandoval, R. & Philpot, A. (Eds.) *Proceedings Of The 11th Annual International Digital Government Research Conference On Public Administration Online: Challenges And Opportunities*. Puebla, Mexico: Digital Government Society Of North America.
- King, R. D., Liakata, M., Lu, C., Oliver, S. G. & Soldatova, L. N. 2011. On The Formalization And Reuse Of Scientific Research. *Journal Of The Royal Society Interface*, 8, 1440-1448.
- Klerk, P. D. 2011. *Linked Open Government Data*. Master Of Science, Tu Delft.
- Kulk, S. & Loenen, B. V. 2012. Brave New Open Data World? *International Journal Of Spatial Data Infrastructures Research*, 7, 196-206.
- Lee, G. & Kwak, Y. H. 2012. An Open Government Maturity Model For Social Media-Based Public Engagement. *Government Information Quarterly*, 29, 492-503.
- Linkedgov. 2011. *What Is Open Data?* [Online]. Available: <http://Linkedgov.Org/What-Is-Open-Data/> [Accessed October 11 2012].

- Mclaren, R. & Waters, R. 2011. Governing Location Information In The Uk. *The Cartographic Journal*, 48, 172–178.
- Meijer, A. & Thaens, M. 2009. Public Information Strategies: Making Government Information Available To Citizens. *Information Polity* 14, 31–45.
- Molloy, J. C. 2011. The Open Knowledge Foundation: Open Data Means Better Science. *Plos Biology*, 9, 1–4.
- Murray-Rust, P. 2008. Open Data In Science. *Serials Review* 34, 52–64.
- Napoli, P. M. & Karaganis, J. 2010. On Making Public Policy With Publicly Available Data: The Case Of U.S. Communications Policy Making. *Government Information Quarterly*, 27, 384–391.
- O'riain, S., Curry, E. & Harth, A. 2012. Xbrl And Open Data For Global Financial Ecosystems: A Linked Data Approach. *International Journal Of Accounting Information Systems*, 13, 141–162.
- Obama, B. 2009. *Memorandum For The Heads Of Executive Departments And Agencies: Transparency And Open Government*. [Online]. Available: http://www.whitehouse.gov/the_press_office/transparency_and_open_government [Accessed September 15 2012].
- Obama, B. 2012. *Digital Government. Building A 21st Century Platform To Better Serve The American People* [Online]. Available: <http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html> [Accessed September 15 2012].
- Open_Knowledge_Foundation. 2011. *What Is Open?* [Online]. Available: <http://opendatamanual.org/what-is-open-data/what-is-open-data.html> [Accessed December 8 2011].
- Schellong, A. & Stepanets, E. 2011. Uncharted Waters: The State Of Open Data In Europe. In: Csc (Ed.) *Public Sector Study Series*.
- Schuurman, N., Deshpande, A. & Allen, D. M. 2008. Data Integration Across Borders: A Case Study Of The Abbotsford-Sumas Aquifer. *Journal Of The American Water Resources Association*, 44, 921- 934.
- Smith, P. 2011. *Methodological Challenges In Integrating Data Collections In Business Statistics*. Office For National Statistics [Online]. Available: http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/S5p4_methodological_challenges_paul_smith.pdf. [Accessed October 11 2012].
- Sweeney, K. 2009. *Open Data: Meaning, Context And Implications* [Online]. Available: <http://www.mfe.govt.nz/environmental-reporting/about/partnerships/forum-2010-05-11/kevin-sweeney.pdf> [Accessed October 17 2012].
- The_White_House 2009. *Memorandum For The Heads Of Executive Departments And Agencies: Open Government Directive*. Washington, Dc.
- Veljković, N., Bogdanović-Dinić, S. & Stoimenov, L. Year. Municipal Open Data Catalogues. In: Parycek, P., Kripp, M. J. & Edelmann, N., Eds. *Conference For E-Democracy And Open Government (Cedem11)*, 2011 Danube University Krems, Austria. 195-207.
- Vickery, G. & Wunsch-Vincent, S. 2006. *Digital Broadband Content: Public Sector Information And Content*. [Online]. Available: <http://www.oecd.org/dataoecd/10/22/36481524.pdf>. [Accessed October 11 2012].
- Viglione, A., Borga, M., Balabanis, P. & Blöschl, G. 2010. Barriers To The Exchange Of Hydrometeorological Data In Europe: Results From A Survey And Implications For Data Policy. *Journal Of Hydrology*, 394, 63–77.
- Vogel, L. 2011. The Secret's In: Open Data Is A Foreign Concept In Canada. *Canadian Medical Association Journal*, 183, E375-E376.
- World_Wide_Web_Consortium. 2011. *W3c*. [Online]. Available: www.w3.org [Accessed December 2 2011].
- Xiong, J., Hu, Y., Li, G., Tang, R. & Fan, Z. 2011. Metadata Distribution And Consistency Techniques For Large-Scale Cluster File Systems. *Ieee Transaction On Parallel And Distributed Systems*, 22, 803-816.
- Zhang, J., Dawes, S. S. & Sarkis, J. 2005. Exploring Stakeholders' Expectations Of The Benefits And Barriers Of E-Government Knowledge Sharing. *Journal Of Enterprise Information Management*, 18 548-567.
- Zuiderwijk, A. & Janssen, M. 2012. A Comparison Of Open Data Policies And Their Implementation In Two Dutch Ministries. In: Luna-Reyes, L. F. & Mellouli, S. (Eds.) *13th Annual International Conference On Digital Government Research (Dg.O 2012)*. University Of Maryland, Usa: Acm.
- Zuiderwijk, A. & Janssen, M. Forthcoming. A Framework For Comparing Open Data Policies, Their Implementation And Impact At Various Levels Of Government. *Government Information Quarterly*.
- Zuiderwijk, A., Janssen, M. & Choenni, S. 2012a. Open Data Policies: Impediments And Challenges. In: Gascó, M. (Ed.) *12th European Conference On E-government*. Barcelona, Spain: Esade Ramon Llull University.
- Zuiderwijk, A., Janssen, M., Meijer, R., Choenni, S., Charalabidis, Y. & Jeffery, K. 2012b. Issues And Guiding Principles For Opening Governmental Judicial Research Data. *11th Conference On Electronic Government* Kristiansand, Norway.
- Zuiderwijk, A., Jeffery, K. & Janssen, M. 2012c. The Necessity Of Metadata For Open Linked Data And Its Contribution To Policy Analyses. In: Edelmann, P. P. N. (Ed.) *Conference On E-Democracy And Open Government (Cedem12)*. Krems, Austria: Danube-University Krems.