Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 32-44
© 2014 Universidad de Talca - Chile

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

Reconciling Contradictions of Open Data Regarding Transparency, Privacy, Security and Trust

Ronald Meijer[1], Peter Conradie[2, 3] and Sunil Choenni[1, 2]

[1] Ministry of Security and Justice, Research and Documentation Centre, The Hague, The Netherlands, r.f.meijer@minvenj.nl, r.choenni@minvenj.nl
[2] Rotterdam University of Applied Sciences, Creating 010, Rotterdam, The Netherlands, p.d.conradie@hr.nl, r.choenni@hr.nl
[3] Ghent University, Department of Industrial System and Product Design, Kortrijk, Belgium, peter.conradie@ugent.be

Received 7 August 2013; received in revised form 31 January 2014; accepted 14 March 2014

Abstract

While Open Data initiatives are diverse, they aim to create and contribute to public value. Yet several potential contradictions exist between public values, such as trust, transparency, privacy, and security, and Open Data policies. To bridge these contradictions, we present the notion of precommitment as a restriction of one's choices. Conceptualized as a policy instrument, precommitment can be applied by an organization to restrict the extent to which an Open Data policy might conflict with public values. To illustrate the use of precommitment, we present two case studies at two public sector organizations, where precommitment is applied during a data request procedure to reconcile conflicting values. In this procedure, precommitment is operationalized in three phases. In the first phase, restrictions are defined on the type and the content of the data that might be requested. The second phase involves the preparation of the data to be delivered according to legal requirements and the decisions taken in phase 1. Data preparation includes amongst others the deletion of privacy sensitive or other problematic attributes. Finally, phase 3 pertains to the establishment of the conditions of reuse of the data, limiting the use to restricted user groups or opening the data for everyone.

Keywords: Open data, Precommirment, Transparency, Privacy, Security, Trust

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 32-44
© 2014 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

1 Introduction

Open Data consists of the data that is not traceable to a person, with the aim to be reused and redistributed by everyone, without restrictions from copyright, patents or other mechanisms of control [31], [38], [45]. In recent years, the publication of Open Data is gaining importance. This is taking place in the context of a growing demand for openness in the public sector [26]. For example, within the European Union, the Obama administration, and national/European levels, various calls are made for increased access to government data [13], [28], [36], [40]. Also within the scientific community an increased access to research data is advocated [41], [44]. Motivations for the publishing of Open Data include the commercial value of data reuse [4], [11], [48], or increased transparency [10], [26], [39]-[40]. Transparency is viewed as a necessary condition for a well-functioning democratic state of law. It serves the legitimacy of public administrations and the trust of civilians in governments [40]. European Commissioner for the Digital agenda, Neelie Kroes, especially emphasizes Open Data's potential for fueling innovation, while also stressing the positive benefits on transparency of the public administration in Europe [28]. Innovations might include new applications based on the published data, or other indirect financial benefits in the form of enhanced efficiency [11], [48].

While Open Data is thus seen as a driver for innovation and transparency, increased openness might also lead to privacy breaches and security violations [27], [29], [47]. Despite these potential issues associated with data publishing, several Open Data initiatives have been started on various government levels. These include both municipal [7] and national level initiatives [42] as well as the efforts made by the European Commission to publish data [16].

What all these Open Data initiatives have in common is that they are, to a certain extent, operating with some more or less defined goals and objectives. Some of these objectives can be categorized as *Public Value*. The *open* side of Open Data provides access for the public eye. This clearly underlines the idea that Open Data is about creating *public* value, as is also illustrated by the strong emphasis on increased transparency and innovations as a result of Open Data [10]-[11], [39], [48]. Open Data, as is demonstrated in [33] is not only about creating Public Value, but also, to an important extent, about conserving and maintaining Public Value. In fact, an Open Data policy has to reconcile multiple seemingly conflicting values.

The concept of Public Value is increasingly popular within both academic and practice settings [51]. Some believe that this concept will be the next *Big Thing* in public management [46]. For a successful Public Value strategy the three elements of Moore's *strategic triangle*, i.e.: *public values/strategic goals*, *authorizing environment* and *operational capability* must be brought into coherent alignment, [34], [51]. Moore describes the elements as follows. The first element, *public values / strategic goals*, comprises the aims of the particular public program. Secondly, the *authorizing environment* refers to the environment in which individuals operate. Moore emphasizes the supporting roles of external stakeholders (i.e., researchers, non-profit organizations or donors). Finally, the *operational capability* includes the resources needed to "achieve the declared objectives" [34]. p. 71.

An Open Data policy presents a problem in the preservation of Public Values as defined by Moore, in the form of several contradictions between values. These contradictions might include privacy, transparency, security or trust. Our objective is the use of precommitment as a policy instrument to overcome these contradictions. For this purpose, we built on [33]. We discuss how we have aligned the Public Value elements for an Open Data policy in the context of the judicial research and registration data at a government research institute, and in the context of the data used for the making and execution of a policy in a local government. We focus especially on the relationships between Open Data and Public Value. Several contradictions exist between these values as will be demonstrated below. In our case, these values form the first of the three strategic triangle elements. The authorizing environment is mainly formed by general policy instructions, and privacy laws and regulations. The *operational capability* in our case consists of the data infrastructure and staff.

The remainder of this paper is organized as follows. In the next section we start with a description of the research approach, followed by a condensed description of Public Value in section 3. In section 4 the central *public values* which are encountered by Open Data policy are described and analyzed while the concept of precommitment is clarified. In section 5 the case of Open Data policy for judicial research and registration data in a government research institute, and the publication of public sector data from local governments is presented. These cases illustrate some solutions for reconciling the conflictions of an Open Data policy with public values. Finally, section 6 concludes the paper.

2 Research Approach

We argue that to be successful, Open Data policy has to reconcile conflicting values. For this purpose, we exploit the notion of precommitment, which is in essence a restriction of one's choices [15], [30]. We conceptualize precommitment in this paper as a policy instrument whereby an organization imposes some restraints on its policy in

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

order to restrict the extent to which values may conflict and the degree to which stakeholders should be concerned about the trustworthiness of that policy.

The identification of the conflicting public values and of precommitment as a useful policy instrument is based on a literature review. By means of two case studies we investigate how precommitment can be implemented [14]. The cases belong to two public sector organizations. The first case study is conducted at a government research institute, the Research and Documentation Centre of the Ministry of Security and Justice - in Dutch *Wetenschappelijk Onderzoek- en Documentatiecentrum* (WODC), and the second at the City Works department of the Municipality of Rotterdam – in Dutch *Stadsbeheer* (City Works).

The data for the first case study is based on analysis of research datasets, literature study, interviews and discussions with WODC researchers and employees as part of a project aimed at improving the existing Open Data policy of WODC. Guidance for opening data was developed by providing a list of issues that play a role in deciding whether to open data, and by presenting an alternative to fully publishing data (i.e. restricted access). Solutions for overcoming some of the issues were given, such as dealing with privacy-sensitive data, deletion policies, publishing after embargo periods instead of not publishing at all, adding related documents and adding information about the quality and completeness of datasets. Design principles for improving the open data publishing process were derived, such as start thinking about the opening of data at the beginning of the open data publishing process, develop guidelines, especially about privacy and policy sensitivity of data, and monitor how the published data are reused [52]-[53]. The data for the second case study, is based on interviews, workshops and observations conducted as part of a research project involving Open Data release by local government (see [7]-[8]). These practice oriented cases were performed with student projects at the Rotterdam University of Applied Science, with the focus on active data release to understand thresholds to public sector data release. Within this project, 25 pilot cases were developed, of which 14 relied on the data from the City Works department of the Municipality of Rotterdam, forming the basis of this case study. Typical examples of these practice oriented cases were mobile or web applications that use geographic information. These included a tool for gathering and comparing data about the quality of the public space [43], an application to calculate how sunny a terrace might be, based on a 3d model of the city, or a website that shows how wheelchair accessible public buildings are.

The goal of both case studies is to highlight how precommitment functions as a policy instrument to align the Public Value elements. Our cases demonstrate how by means of a precommitment instrument - implemented as a data request procedure and imposing certain restrictions on data re-use – combined with a proper data infrastructure, Open Data policy may reconcile potentially conflicting values. Given this, we define our contribution as the use of precommitment as policy instrument for a successful public value strategy in the context of Open Data, aligning the elements of *public values/strategic goals*, *authorizing environment* and *operational capability*.

3 Public Value Framework

The Public Value framework was originally formulated by Moore [34], [51]. Public Value can best be understood and achieved within the notion of the *public sphere*, a democratic space which includes, but is not coterminous with, the state in which citizens address their collective concerns and where individual liberties have to be protected [3]. The government is seen as a creator of Public Value and a proactive shaper of the public sphere politically, economically, socially and culturally [3]. There is consensus in the literature that Public Value can be interpreted as combining (and reconciling), safeguarding, and enrichment of the public sphere with the delivery of the values that are desired by the public [51].

Moore proposes that the use of public resources should result in an increased value, similar to how a value is created in private enterprises, also including the general benefits that are not necessarily financial [51]. Public Value thus includes the value attached not only to relatively concrete outcomes but also to more intangible benefits [23]. The value *trust* is also especially present in definitions of Public Value [23], [35], [51]. p. 7. A strategic triangle is central in Moore's Public Value framework, see [51] p. 5. and [35]. It contains three elements "public values / strategic goals", "authorizing environment" and "operational capability" [51]. For a successful organizational Public Value strategy, these elements should be coherently aligned. This is attained by complying the strategy to three corresponding broad tests, namely that it must be "substantially valuable", "legitimate and politically sustainable" and "operationally and administratively feasible" [34]. p. 71. Successfully, implementing a policy and creating Public Values thus rely on each of these elements in being present and properly aligned.

For Information Systems (IS) we find a clear parallel relation between the Public Value literature and the literature about the approach of embedding human values in IS. For instance, [6] stresses the importance of embedding human values, such as privacy and trust, in the development of information systems. They plead for an explicit agreement with regard to the values that should be included in a design, as a result, extending the view about the IS beyond the original more narrowly defined requirements. We argue that *human values* embraced by government become *public values*, as they are, from then on, part of the conditions, goals and objectives of organization strategies. Thus, public managers, adopting a Public Value approach who aim to create value in IS for e-Government can profit from IS *human value*-design approaches. We also argue that interpreting Open Data

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

initiatives in the Public Value paradigm may help to clarify the policy problems, which Open Data may encounter and, in doing so, may help to raise and increase Public Values.

4 Public Values in Open Data

Open Data consists of data that is not identifiable to a person with the aim to be reused and redistributed by everyone, and is without any restrictions from copyright, patents or other mechanisms of control [31], [45]. The idea behind opening public data is to make the information that is generated or collected by organizations in the public sector re-usable for third parties. This idea is founded on the acknowledgement that citizens are taxpayers and therefore have access rights to this data. They have this right wherever financially feasible and, when published, it will not violate any laws or rights relating to privacy, either for citizens or government staff [19], [31], [45].

4.1 Transparency, Trust, Privacy, Security as Public Values

Transparency is a central value that drives Open Data. Transparency is viewed as a necessary condition for a well-functioning democratic state of law. It serves the legitimacy of a Public Administration and the trust of civilians in the government [40]. The scientific community is also calling for more transparency with its own research data. We have observed that in The Netherlands, the trust in scientific research is an object of discussion. Regularly, messages that mention cases of questionable research practices appear in the media. A growing distrust against science seems to appear, a distrust which is fed by a series of incidents fully described in the media [41]. Several cases of fraud have been discovered in recent years [25]. The proposed measures point to more openness and transparency. Besides good data management, peer pressure, archiving and sharing are advocated. These elements support the replication of research results. As a consequence the chances of fraud decrease, while the chances of discovering fraud increase [4], [41]. Therefore we argue that openness contributes to transparency and via transparency, it is expected to contribute to trust of civilians and other stakeholders, in the government and science [29], [39]-[40], [54].

While the relationship between transparency and trust is controversial (see [21]-[22], [32]), there is an expectation of increased trust as a result of increased transparency of data. For example, [39] argues that transparency can indeed lead to higher trust, especially noting the importance of data transparency in law enforcement as a means to create trust. Similarly, [44] argues that transparency of research data can result in increased trust, and [50] notes that transparency is positively associated with trust in government.

The IS evaluation framework based on Public Value of [23] focuses upon citizens' and clients' experiences of service provision and service outcomes as contributors to the formation of public trust. They show that trust is related to the extent to which people feel that an e-Government service enhances their sense of being well informed, gives them greater personal control and provides them with a sense of influence or contingency [23]. In the context of law enforcement the authors of [39] note that people feel they have lost control over their own data and they do not know who handles personal data, when and for what purpose. This concern can be answered by increasing transparency of these operations [39]. The principle of transparency is that information should be shared while data is collected. Possibilities for control must be created and people must be assured that there is no abuse [39]. As will be argued in section 3.3, constraints imposed on the access to data therefore are important for trust in deciding how far one party needs to trust the other and vice versa. Transparency is also important for trust, as by transparency the unilateral restraints imposed can be verified by the stakeholders.

As described above, Open Data refers to data that does not reveal personal identity. We argue that privacy is thus another central Public Value in Open Data. By means of the Data Protection Directive the European Union requires that if personal data is processed, this should be done fairly, lawfully and for specified, explicit and legitimate purposes, see article 6 in [17]. The purposes for which the data is processed must be explicit and legitimate and must be determined at the time of collection of the data [17], [29]. In The Netherlands the important principles of privacy protection are anchored in the Dutch Privacy Protection Act (DPPA), such as finality, legitimacy, proportionality and subsidiarity, transparency and data subject's rights. Finality refers to the purpose for which personal data is collected. This purpose should be explicit and the processing of collected data must be compatible with the purpose for which the data was collected. Legitimacy refers to the process of data collection and also to the context of the data, according to which data must be processed in a proper, careful, and legal manner. Moreover data must be relevant, sufficient, not excessive, and correct [49]. Proportionality demands that the means used to collect data is proportional to the intended purpose. Subsidiarity demands the use of the alternative means which minimizes the use of privacy sensitive data. Transparency refers to the right that the data subject is entitled to know if someone is processing data about him. The data processing party has the obligation to identify itself to the data subject and has to inform her/him about what data it processes and the purposes of processing [49].

The fourth and final Open Data value we discern in this paper is security. Security is a comprehensive notion. However, we derive this value from the privacy value. To prevent accidental or malicious disclosure, modification, or destruction of records and data sets, data security is indispensable [12]. Research and registration databases may contain privacy sensitive data. The opening of this data therefore should be done in strict compliance with the privacy value to prevent privacy breaches. In other words, the security of the data has to be protected. In Section 3.2 below, we will examine this value in more detail.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 32-44
© 2014 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

4.2 Contradictions between Open Data and Public Values

Having identified the values playing a crucial role in Open Data from the literature, we analyze here how they are related to each other. Figure 1 illustrates the Open Data values as described above. It depicts the public Open Data values of transparency, trust, privacy and security and the way they assumingly relate to each other and to three selected intermediary elements of replicability, information overload and reliability. The way that public values and intermediary elements relate to each other is depicted by arrows. These elements may reinforce each other, indicated by a + sign, or may be contradictory, indicated by - sign. The intention of this chart is not to be complete. Its intention is to show the apparent contradictions between the public values on which we focus in this paper. The intermediary elements will only be described briefly in relation to the values. An extensive discussion of these elements is beyond the scope of this paper.

As discussed before, Open Data - making data accessible for (re-)use to the public – is assumed to contribute to transparency. By giving access to research and (semi)government data, civilians, policy makers, journalist, audits and scientists get opportunities to control, verify the data, replicate research findings or create new findings. It is assumed that this results in maintaining or increasing trust. However there is a *dark side* to opening data without constraints or restrictions imposed on the access to data. This *dark side* may lead to several contradictions in the Open Data policy values.

In the first place Open Data may conflict with privacy. The opening of data is seriously impeded when privacy sensitive data is at stake. Open Data may not seem to be personal data at first glance, especially when it is anonymized or aggregated. However, it may become personal data by combining it with other publicly available data or when it is deanonymized [12], [29]. Anonymizing data cannot be *100% privacy proof*. Even when data with a high aggregation level is shared, the risk that one is able to deduce or abduce privacy-sensitive information remains [37], [47]. Opening up data without taking into account the privacy risks attached, may lead to privacy breaches with possibly very negative consequences for the trust of the respondents who participated in research. Furthermore, it influences the trust of civilians in research or government, and reduces the trust of other individuals whose data is recorded. We have found a possible negative relation between the privacy value and trust. To prevent privacy breaches, it is necessary to eliminate privacy sensitive attributes. However, this may have a negative impact on the possibility of using the Open Data for replicability as some of the attributes needed to ensure the replication cannot be accessed any longer. Thus as privacy is protected, not all results may be replicable as a consequence. This may have negative impact on trust.

Next, Open Data may conflict with security. Identity disclosure from surveys or administrative data might be used by private or public groups to target or harm individuals, population subgroups, or business enterprises [24]. Privacy of civilians thus needs to be thoroughly protected. Civilians expect that public organizations follow the rules and procedures carefully in order to protect them and their privacy. Civilians that are harmed due to incorrect following of the rules and procedures may cause social unrest. Therefore, measures to enforce that the rules and procedures are followed correctly should be taken into account while developing infrastructures for data sharing in the public domain [47].

Thirdly, Open Data may conflict with transparency, via the intermediary element of information overload. In the literature we find that information overload occurs when the information received becomes a hindrance rather than becoming potentially useful [2]. Information overload is related to the quantity and diversity of information available [1]. We therefore argue that as governmental organizations possess large volumes of data about many subjects the opening up of this data may cause information overload.

Finally, Open Data may have negative effects on trust via the intermediary element of validity and reliability of the results in cases where the data is reused. This may concern reuse on the basis of the data provided but also on the basis of extension of data with other (open) data sources. We argue that as data are opened, the governmental control on reliability and validity decreases due to a possible lack of a proper interpretation. Third parties may use the opened data in ways that weaken these elements. Data from administrative databases might for instance be misinterpreted and misused with the stigmatization of groups as a consequence [27]. [19] also mentions misinterpretation of data as a general challenge of Open Data, raising questions about how wrong such interpretations might be handled.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 32-44
© 2014 Universidad de Talca - Chile

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

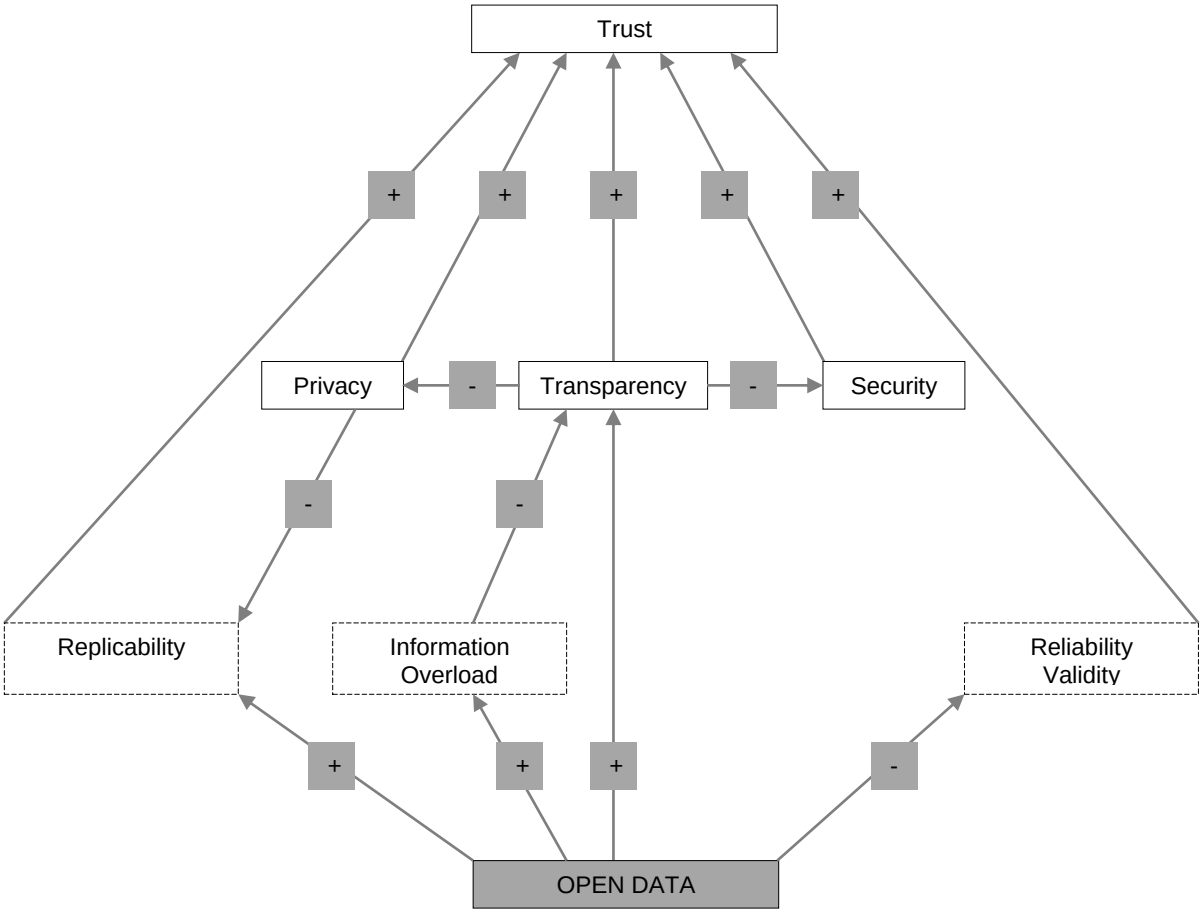

Figure 1: Illustration of the contradictions between open data and public values

4.3 Precommitment as Policy Instrument to Bridge Contradictions

Based on the literature review, we have found several contradictions between the Open Data values. Consequently, to succeed, Open Data policy has to reconcile these values. We argue that the constraint imposed on the access to data is important for creating trust, in deciding how far one party needs to trust the other and vice versa. We argue that precommitment is necessary to bridge the contradictions in Open Data.

Precommitment is a restriction of one's choices [15], [30]. It implies imposing some constraints. According to [15] p2, individuals might benefit from having specific options unavailable, available only with a delay, or at a greater cost. For example, precommitment may be aimed at overcoming impulsivity (e.g., in gambling machines to require the gambler to preset a limit on his or her expenditure, after which the machine deactivates [30]). While precommitment can be viewed as a strategy to restrict personal choice, it can also be seen as a way to foster trustworthiness [9], [18]. By preemptively placing restrictions, the degree with which others have to be concerned about trustworthiness decreases. To apply precommitment in the context of Open Data policy, we conceptualize precommitment as a policy instrument where an organization – in this case responsible for Open Data - imposes some restraints on its policy in order to restrict the extent to which the values may conflict. As a result, other stakeholders have to worry less about the trustworthiness of that policy.

To limit possible conflicts between Public Values and Open Data values several restraining options for opening data are possible. In the first place privacy sensitive data might be irrevocably deleted. Furthermore, datasets may be completely anonymized before they are archived, thus limiting the future possibilities to replicate research or to link the datasets to other data to create new datasets that might reveal personal identities. Secondly, the data including sensitive information, such as personal information, might be opened for specific goals or target groups only. Some data (e.g. registry databases of police and prosecution) might only be distributed for specific purposes and made available to specific organizations – in compliance with privacy laws and regulations in vigor. These might include research institutes or educative institutes (using the data for educative purposes). Thirdly, before the opening of research data to the public, all privacy sensitive data needs to be thoroughly removed. As a result, the opened datasets will contain only limited information and can be analyzed only in a (very) restricted way. Data access might also be restricted to certain groups, or the results or interpretations of data might be verified for accuracy. Fourth, data can be made accessible in an indirect way by the provision of exclusively highly aggregated data only. Such

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 32-44
© 2014 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

data is generated by data experts. Lastly, access to data might be restricted to a certain timeframe, or only accessible through restricted channels.

To illustrate how these restraining options are applied in a public sector organization, we introduce two case studies. The first concerns some data published at a government research institute, while the second focuses on the data published by a local government. The case studies present contradictions between Open Data policy and public values, including privacy, transparency or security, and illustrate how precommitment can be used as a policy instrument to bridge these contradictions. Some of the data in these case studies could not initially be defined as Open Data prior to release, but the process followed in these cases, allows the exploration of the boundaries to data release, while preserving public values.

5 Case Studies

The goal of the case studies is to illustrate how precommitment may be applied as a policy instrument to align the Public Value elements. Our cases demonstrate how by means of a precommitment instrument - implemented as a data request procedure and imposing certain restrictions on data re-use – combined with a proper data infrastructure, Open Data policy may reconcile potentially conflicting values. The data request procedure is aimed at the opening of data. The data request procedure explores the boundaries of which data may or may not be released, exploiting pre-commitment as instrument. Within the process for data requests, we discern three phases. Figure 2 illustrates these three phases.

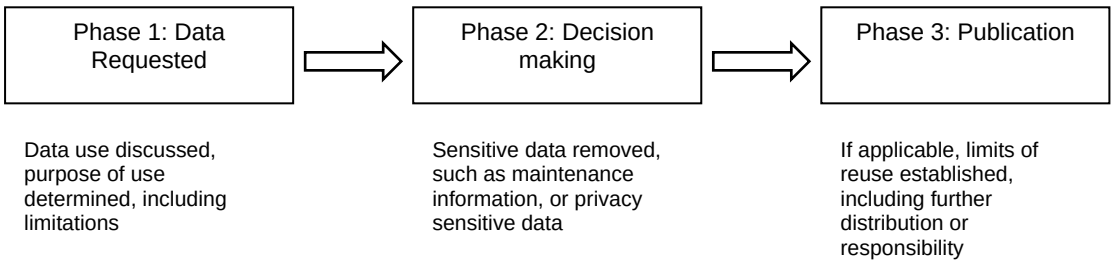

Figure 2: Process of data release and precommitment steps

First, a data request is made to the appropriate authority or individual within the organization. This request might include descriptions of the type of data, the purpose for which and the context within which the data might be re-used. Secondly, a phase of decision-making occurs, where the suitability of the data request is evaluated. During this phase data might be filtered, aggregated or anonymized, based on the needs of data reuse presented in phase 1, and a decision is made about publication of the data. The final phase concerns the publishing of the data, either online or via other channels. It is accompanied by agreements about the reuse of the data, such as restricting certain types of uses or taking measures to ensure that the data is not falsely interpreted.

In the following two cases we illustrate how precommitment can be implemented as a policy instrument within each of these phases. These cases belong to two public sector organizations. A summary of this process for the two case studies is shown in Table 1.

5.1 Case Study 1: Open Data Policy of a Government Research Institute

At the Research and Documentation Centre of the Ministry of Security and Justice - in Dutch *Wetenschappelijk Onderzoek- en Documentatiecentrum* (WODC) - data is gathered to define the current and future research agenda of the Dutch Ministry of Security and Justice, to answer policy-related questions and to indicate the possible implications of research findings for standing policy [52]. For this purpose the WODC systematically collects, stores, enhances and provides criminal justice information produced by the WODC staff or by the external organizations commissioned by the WODC.

The WODC strives towards transparency, thus investing in trust, while giving priority to protecting privacy. The WODC aims to facilitate the reuse of research data, as this may provide the organization with benefits such as the possibility to scrutinize and validate the data and to decrease the workload of the WODC. The WODC works with confidential judicial research and registration data, so that issues such as confidentiality and privacy-sensitivity should be thoroughly taken into account [27], [54]. The WODC therefore has developed a procedure to share the collected data as much as possible with other parties, while protecting privacy and being in compliance with the restrictions of the privacy protection principles and laws. This procedure is combined with a data infrastructure to manage the contradictions of different values. In two subsequent sections we discuss the data infrastructure and the procedure for data release, focusing on the three phases of request, decision, and publication.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 32-44
© 2014 Universidad de Talca - Chile
This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

5.1.1 Data Infrastructure

Data from concluded research projects is collected. Data from those projects that is qualified for public opening is centrally stored in compliance with the Dutch Privacy Protection Act (DPPA). Privacy sensitive data is deleted unless explicitly needed for further research (e.g., for longitudinal research and for monitoring projects). Public safety registration data is stored in a Data Warehouse (DW), containing police and justice data (e.g., for policy research purposes). A DW ensures a uniform approach to data for interpretation purposes and ensures maximum accessibility. Privacy is protected as the DW is anonymized, i.e., has been stripped off from the directly identifying attributes like names and addresses. In the DW problems around inconsistencies, reliability, and validity are tackled [5]. Protection of the archived research data and the DW data forms the basis of the WODC data access policy.

The WODC may decide to open its research data to the public. The research data of the WODC is open for everyone once - contented to the high criteria of DPPA. Here confidentiality matters and therefore the data is uploaded on the server of the Dutch Archiving and Networked Services (DANS). Before opening the research data to the public, all privacy sensitive data, which may lead to disclosure of identity directly and indirectly is removed. Moreover, the WODC may permit scientific organizations to have a restricted access to privacy sensitive research data for scientific purposes on the basis of a data request. The DPPA allows the (re)use of personal or privacy sensitive judicial data, under certain conditions, for scientific purposes. Public safety registration data might be released only for scientific research to scientific organizations. The WODC regularly receives individual data requests from scientists for permission to reuse research data or for an extract of public safety data from the DW. Extracts from the DW may in principle be opened, but for scientific research only. Finally, the WODC gives access to the data by providing highly aggregated and non privacy sensitive data on demand. This data is generated by data experts and concerns DW data requests mostly.

Every request is thoroughly audited by the WODC data request procedure. The decision to the public opening of research datasets, which is uploaded to DANS, is processed in an equally thorough way, by a separate procedure. This procedure is the focus of a major redesign at the moment of the production of this paper.

5.1.2 Procedure for Data Requests

The procedure for public opening of datasets and the data request procedure are rigorous procedures which are aimed at sharing data with other parties as much as possible while thoroughly protecting privacy. With the aid of these procedures the WODC manages to protect privacy sensitive attributes in datasets in compliance with the security policies.

In phase 1, the request phase of the current public opening procedure, several times per year the statistics department of the WODC requests the provision of the datasets of completed research projects of the research departments. This is executed via sending mails. Within phase 2, the decision making phase, the decision to deliver research datasets is taken by the division managers, who consult the employees of each division to see which datasets are appropriate for publication. The decision is made on the basis of a limited number of guidelines aimed, amongst others, at privacy protection and the compliance to confidentiality rules. In the final phase, the data publication phase, the appraisal of the division manager leads to the uploading of the data to DANS. Before uploading the data, the statistics department checks the data thoroughly to prevent possible privacy breaches. The datasets are uploaded together with the metadata that comprises a summary and a link to the research reports based on the datasets.

In phase 1, the requests phase, the WODC discerns two subtypes of data requests in order to contribute to Open Data (i.e., give access to citizens). These subtypes of data requests are: requests for Statistical Information and requests for Data Supply. Statistical Information is aggregated data on which people do not aim to edit the information. This information can be based on registration as well as research data of the WODC. The output provides a minimal opportunity to be edited. Requests for Data Supply can be subdivided in requests for reuse of research data from published research or requests for an extract of registration data from the data warehouse (DW).

Within phase 2, the decision making phase, the WODC distinguishes three steps, which in the context of the WODC data focuses on the protection of privacy. At first an experienced data manager carefully studies a data request to see which variables are necessary for an applicant and whether the required variables could be delivered from the centrally archived research data or the DW. Requests for statistical information concern mainly highly aggregated and not privacy sensitive data. The decision to deliver this data is made by the data request coordinator. In case of requests for data supply, the data manager prepares a preliminary decision to the request by making a report with considerations such as legal requirements, policy sensitivity and amount of workload for the WODC. This document is sent to the workgroup of DPPA asking them to examine the legal conditions of the request. The data request is tested on the criteria of the DPPA by the workgroup. In this step every kind of convertible (personal) facts that are not in agreement with privacy laws and rules are removed. When necessary a Trusted Third Party is included in the data request project in order to prevent the unnecessary transfer of privacy sensitive data [47]. The subsequent and final step is the judgment of the board of directors. Board members discuss the request looking at the advice written by the data manager. Based on their experience and comments of the DPPA workgroup, the board members decide whether or not the data should be delivered to the requesting party and if so, based on which conditions.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

In the final phase, the data publication phase, an appraisal of the board of directors leads to delivering data. This is accompanied by the signing of a standard agreement by the applicant. This agreement relates to data reuse, in addition to the specific conditions related to a particular dataset. Statistical information is delivered without any further procedure. There is no need to sign any agreement whatsoever.

5.2 Case Study 2: Local Government Data Release

The second case study focuses on the City Works department of the Municipality of Rotterdam – in Dutch *Stadsbeheer*. This department is tasked with maintaining infrastructure in Rotterdam. They also create products such as maps while providing engineering consulting services [20]. This department also effectively supplies geographical information to other departments within the council. In order to perform these tasks, City Works collects, stores and enhances data about the public space. This data might refer to objects in the public space such as benches, public art, trees, or to complete and detailed maps of the city. Data is sometimes collected by the department directly, for example, by counting the occurrence of a physical object in the public space or by hiring external service providers to perform periodic measurements.

Especially within a two year research project to examine data release (see [7]-[8]), several requests for data have been made. These datasets were used as part of the education program of the Rotterdam University of Applied Science, where students used the data to experiment and create applications. Within this case we will examine two examples of data release, focusing on how the principle of precommitment is applied to the publication of the data.

5.2.1 Data Infrastructure

While City Works does not yet have an explicit Open Data policy, it does publish certain amounts of data proactively. This might include information about the locations of council facilities, 3D models of the city, or the locations of other public sector organizations such at the police. This data might be published on City Works website or on an Open Data catalogue. Once available in the catalogue, the data is available for reuse without restriction. For data that is not yet published online, data requests might occur. This also forms part of a policy to publish data on demand.

Internally, the data is accessible through Geographic Information Services (GIS) applications, and the object data in particular can be exported as standalone files. Certain types of data is currently available online as Open Data, in the form of web services. Besides maintaining the public space, a part of City Works is also in charge of garbage collection. For this task and in order to perform maintenance in the city, City Works has access to a fleet of vehicles that include garbage collection trucks or other cars used for maintenance or transport of civil servants, including Aldermen, or other council members. Especially in the case of garbage collection and city maintenance, the vehicles are tracked to optimize routes, additionally, data is recorded about vehicle usage to track and improve vehicle maintenance. This information is hosted on a DW within the garbage collection department of City Works.

The release of the first dataset over trees might conflict with the intermediary element validity and reliability, as a result of misinterpretation of data, resulting in false conclusions. While the tree itself is an object that can be observed and does not require privacy protection, the department also records the health status of the trees. According to City Works, publishing all the data about the trees, including the current health status, could lead to undesirable actions by citizens [7]. These might include misinterpreting tree illness as also being hazardous to personal health, or taking personal measures to treat the trees.

The second example within this case is the movement data about municipality vehicles, including garbage collection, maintenance and transport vehicles. The data recorded include the status (stopped or moving), location, speed or direction of the vehicle. The release of this dataset might conflict with the privacy and security. While the data can, and has been, anonymized, combining the routes driven by the trucks with the driver's schedule, can reveal the identity of an individual driver. Additionally, the routes of council members can be traced with the data, presenting a potential security or privacy risk (e.g., revealing the home addresses and other sensitive information about civil servants).

5.2.2 Procedure for Data Requests

Below we describe the procedure for data release and focus on the precommitment measures taken to avoid conflicts with public values as previously described.

In phase 1, a data request is processed. This request is placed with the program manager for Open Data within the council. Included in this request are the reuse purposes of the data, i.e., the purposes for which the data is requested or the period during which the data might be used. The limits of reuse are discussed with the program manager, who subsequently relays the data request to the data owner within the council, in this case City Works. Data requests might also be placed with the data controller, in cases where the owner of the data does not directly control the data.

Within phase 2, decisions are made about whether the data can be published, and if so, under which circumstances. For the data about trees, effort is undertaken to remove all maintenance information from the data. This includes the state of the trees, whether they are ill or not, but also the costs associated with maintenance. Efforts are also made

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

to ensure that the dataset only includes objects that are maintained by City Works. The remaining data items are still rich enough to allow experimentation with the data. For example, the data still includes information such as the exact location of the trees, the date when a tree was planted, or type of tree (including the sex of the tree). For the data on vehicle history, all personal data is removed from the data. This is done through irretrievable removal of driver information and replacement of vehicle type information. Data thus retains enough details to be reusable within the goals stipulated in phase 1, but critical information about persons is removed.

In phase 3, the data is published. The tree data, due to lack of security or privacy risks, is made available for reuse by a larger audience, while the vehicle data is provided under more strict conditions. A non-disclosure agreement is signed by all users of the data, where restrictions such as copy restrictions are agreed upon. A further potential conflict in the form of misinterpretation and misuse is contained through stipulating that any publication of data analysis would first be disclosed to City Works. Finally, data is available only through logging in on an internal database within the university, limiting the possibility of data being released online without restriction.

5.2.3 Summary of Precommitment Procedure

With these two case studies, we presented two public sector organizations that, through the use of precommitment as a policy instrument, explore the boundaries of Open Data release, while preserving the public values of privacy, transparency and security. This procedure occurs within 3 phases, where phase 1 concerns the data request, phase 2 focuses on decisions about the data itself, such as anonymization or legal requirements. Finally, in phase 3, data is published, based on the reuse agreement or restrictions to dissemination, when applicable. Table 1 summarizes the results.

Table 1: Summary of the precommitment procedure

Procedure for data release	Case 1: WODC	Case 2: City Works	Precommitment mechanism
Phase 1: Data Requested	Discerning data types: Statistical Data, or Data Supply	Limits of the reuse agreed to via a request, evaluated by the program manager and data owner	Determining the type and content of the data request.
Phase 2: Decision making	The data manager studies the request, defining necessary attributes. Data suppliers of requests take into consideration the legal requirements, removing privacy sensitive or otherwise problematic attributes.	Restrictions on data contents are applied (e.g.: removal of personal information or maintenance data)	Adapting the data based on the data request and in compliance to legal requirements. Removing privacy sensitive or otherwise problematic attributes.
Phase 3: Publication	Based on the data request and the data, a reuse restriction agreement is signed or data is published without restriction.	Publications are restricted, based on the data reuse request and data contents.	Establishing the conditions of reuse of the data, limiting use to restricted user groups or opening data for everyone. Based on the data request and published data.

6 Conclusions

Our contribution in this paper was the use of precommitment as a policy instrument that allows a successful public value strategy in the context of Open Data, aligning the elements *public values/strategic goals*, authorizing *environment and operational capability*. To do this, we have highlighted a relationship between Open Data and public values. This relationship is described by the values trust, transparency, privacy, and security. As we have argued, several contradictions between these values exist. To solve these contradictions we have introduced the notion of precommitment: a policy instrument whereby an organization imposes some restraint on its policy in order to restrict the extent to which values may conflict and stakeholders have to be concerned about the trustworthiness of that policy. We have elaborated this notion in a rigorous data request procedure. To manage the contradictions between values we combine this procedure with a data infrastructure. Through two case studies, we have illustrated how these contradictions are bridged at a governmental research institute and within a local government.

The data request procedure used in these cases can be summarized as follows. In this procedure, precommitment is operationalized in three phases. In the first phase, restrictions are defined on the type and the content of the data that might be requested. The second phase involves the preparation of the data to be delivered according to legal

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

requirements and the decisions taken in phase 1. Through this data preparation, privacy sensitive or otherwise problematic attributes may be removed. Finally, phase 3 pertains to the establishment of the conditions of reuse of the data, limiting use to restricted user groups or opening data for everyone. This is based on the data request and published data.

The cases illustrate how the priority to protect privacy of citizens and public servants ensures public safety and maintains the integrity of data. This priority, on the one hand, leads to the limitation of the opening of data to the public, but, on the other hand, it gives the opportunity to Open Data in a restricted mode for both scientific and innovation goals. In doing so, we presented precommitment as a policy instrument to bridge the contradictions of public values and Open Data policy.

A DW assures reliability of answers to statistical information requests. Hereby privacy is protected by presenting only very highly aggregated data to the general public. Requests for data supply mostly concern data on the level of unique identifying records in research datasets or registration data. Therefore in these types of requests most of the time, privacy sensitive data is involved. In these cases the data is supplied only for scientific goals in compliance with the privacy laws and regulations. Moreover, by verifying the data on legal and policy confidentiality points of views on several moments in different phases, the chance of failure is reduced to a minimum, thus maintaining trust. Opening data to third parties such as scientists, developers or students creates possibilities for reuse. Especially in the case of scientific data, results may be replicated, contributing to trust.

Acknowledgments

The authors wish to thank Mortaza Shoae Bargh and Rochelle Choenni for their valuable contribution to this paper, in particular for verification of English.

References

[1] D. Bawden, C. Holtham and N. Courtney, Perspectives on information overload, Aslib Proceedings, vol. 51, no. 8, pp. 249-255, 1999.
[2] D. Bawden and L. Robinson, The dark side of information: Overload anxiety and other paradoxes and pathologies, Journal of Information Science, vol. 35, no. 2, pp. 180-191, 2009.
[3] J. Benington and M. H. Moore, Public Value: Theory and Practice. Great Britain: Palgrave Macmillan, 2011.
[4] G. Boulton, M. Rawlins, P. Vallance, and M. Walport, Science as a public enterprise: the case for open data, The Lancet, vol. 377, no. 9778, pp. 1633-1635, 2011.
[5] S. Choenni and R. Meijer, From police and judicial databases to an offender-oriented data warehouse, in Proceedings of the IADIS International Conference e-Society, Avila, Spain, 2011, pp. 98-105.
[6] S. Choenni, P. van Waart and G. De Haan, Embedding human values into information system engineering methodologies, in Proceedings ECIME 2011, 5th European Conference on Information Management and Evaluation, Italy, 2011, pp. 101-108.
[7] P. Conradie and S. Choenni, Exploring process barriers to release public sector information in local government, in Proceedings 6th International Conference on Theory and Practice of Electronic Governance, Albany, NY, 2012, pp. 5-13.
[8] P. Conradie, I. Mulder and S. Choenni, Rotterdam open data: Exploring the release of public sector information through co-creation, in Proceedings ICE 2012: International Conference on Engineering, Technology And Innovation, Munich, 2012, pp. 187-196.
[9] P. Dasgupta, Trust as a commodity, in Trust: Making and Breaking Cooperative Relations, Electronic. (D. Gambetta, Ed.). Department of Sociology, University of Oxford, New York: Blackwell Publishing, 2000, pp. 49-72.
[10] S. S. Dawes, Stewardship and usefulness: Policy principles for information-based transparency, Government Information Quarterly, vol. 27, no. 4, pp. 377-383, 2010.
[11] M. Dekkers, F. Polman, R. te Velde, and M. de Vries. (2006, June) MEPSIR: Measuring European public sector information resources. European Commission. [Online]. Available: http://ec.europa.eu/information_society/policy/psi/docs/pdfs/ mepsir/final_report.pdf.
[12] D. E. Denning and P. J. Denning, Data security, ACM Computing Surveys, vol. 11, no. 3, pp. 227-249, 1979.
[13] Donner, Betreft Hergebruik en Open Data: naar betere vindbaarheid en herbruikbaarheid van overheidsinformatie. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, Den Haag, pp. 1–9, 2011.
[14] L. Dubé and G. Paré, Rigor in information systems positivist case research: Current practices, trends, and recommendations, Mis Quarterly, vol. 27, no. 4, pp. 597-635, 2003.
[15] J. Elster, Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints. Cambridge: University Press, 2000.
[16] European Commission. (2013, April) Data | European Union Open Data Portal, European Union Open Data Portal. [Online]. Available: http://open-data.europa.eu/.
[17] European Parliament of the European Union, Directive 95/46/EC, Official Journal L 281, pp. 0031-0050, 1995.
[18] D. Gambetta, Can we trust trust?, in Trust: Making and Breaking Cooperative Relations (B. Blackwell Ed.). Oxford: Blackwell Publishing Ltd., 2000, pp. 213-237.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

[19] C. P. Geiger and J. Von Lucke, Open government and (Linked) (Open) (Government) (Data), JeDEM - eJournal of eDemocracy and Open Government, vol. 4, no. 2, pp. 265-278, 2012.
[20] Gemeente Rotterdam. (2013, April) Stadsbeheer. Rotterdam Worldportworldcity. [Online]. Available: http://www.rotterdam.nl/gw.
[21] S. G. Grimmelikhuijsen and A. J. Meijer, The effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment, Journal of Public Administration Research and Theory, vol. 24, no. 1, pp. 137-157, 2014.
[22] S. Grimmelikhuijsen, G. Porumbescu, B. Hong, and T. Im, The effect of transparency on trust in government: A cross-national comparative experiment, Public Administration Review, vol. 73, no. 4, pp. 575-586, 2013.
[23] M. Grimsley and A. Meehan, e-Government information systems: Evaluation-led design for public value and client trust, European Journal of Information Systems, vol. 16, no. 2, pp. 134-148, 2007.
[24] M. Gutmann and K. Witkowski, Providing spatial data for secondary analysis: Issues and current practices relating to confidentiality, Population Research and Policy Review, vol. 27, no. 6, pp. 639-665, 2008.
[25] J. Heilbron, Wetenschappelijk onderzoek: dilemma's en verleidingen, 2005.
[26] K. Janssen, The influence of the PSI directive on open government data: An overview of recent developments, Government Information Quarterly, vol. 28, no. 4, pp. 446-456, Oct. 2011.
[27] S. Kalidien, S. Choenni and R. Meijer, Crime statistics online: potentials and challenges, in Proceedings of the 11th Annual International Digital Government Research Conference on Public Administration Online: Challenges and Opportunities, Pueblo, Mexico, 2010, pp. 131-137.
[28] N. Kroes. (2011, October) Public data for all – opening up Europe's public sector. European Commission. [Online]. Available: http://ec.europa.eu/commission_2010-2014/kroes/en/blog/public-data-for-all-%E2%80%93-opening-up-europes-public-sector
[29] S. Kulk and B. Van Loenen, Brave new open data world, International Journal of Spatial Data Infrastructures Research, vol. 7, pp. 196-206, 2012.
[30] Z. Kurth-Nelson and A. D. Redish, Don't let me do that! - models of precommitment., Frontiers in Neuroscience, vol. 6, p. 138, 2012.
[31] LinkedGov. (2011, January) What is open data?. LinkedGov. [Online]. Available: http://linkedgov.org/what-is-open-data/.
[32] A. Meijer, Understanding modern transparency, International Review of Administrative Sciences, vol. 75, no. 2, pp. 255-269, Jun. 2009.
[33] R. Meijer, S. Choenni, R. S. Alibaks, and P. Conradie, Bridging the contradictions of open data, in Proceedings 13th European Conference on eGovernment, Como, Italy, 2013, pp. 329-336.
[34] M. H. Moore, Creating Public Value: Strategic Management in Government. Cambridge, MA: Harvard University Press, 1995.
[35] J. O'Flynn, From new public management to public value: Paradigmatic change and managerial implications, Australian Journal of Public Administration, vol. 66, no. 3, pp. 353-366, 2007.
[36] B. Obama. (2009) Transparency and open government. The White House. [Online]. Available: http://www.whitehouse.gov/the_press_office/Transparency_and_Open_Government
[37] P. Ohm, Broken promises of privacy: Responding to the surprising failure of anonymization, Social Science Research Network, vol. 57, no. 6, pp. 1-64, 2009.
[38] Open Knowledge Foundation. (2011, September) What is open data?. Open Data. [Online]. Available: http://okfn.org/opendata/.
[39] J. Rajamäki, J. Tervahartiala, S. Tervola, S. Johansson, L. Ovaska, and P. Rathod, How transparency improves the control of law enforcement authorities' activities?, in Proceedings 2012 European Intelligence and Security Informatics Conference, Odense, Denmark, 2012, pp. 14-21.
[40] ROB, Gij zult openbaar maken: Naar een volwassen omgang met overheidsinformatie, Den Haag, 2012.
[41] C. J. . Schuyt, Zorgvuldig en integer omgaan met wetenschappelijke onderzoeksgegevens, Advies van de KNAW-commissie onderzoeksgegevens, Amsterdam, 2012.
[42] N. Shadbolt, K. O'Hara, T. Berners-Lee, N. Gibbins, H. Glaser, W. Hall, and M. C. Schraefel, Linked open government data: lessons from Data.gov.uk, IEEE Intelligent Systems, vol. 27, no. 3, pp. 16-24, 2012.
[43] N. Stembert, P. Conradie, I. Mulder, and S. Choenni, Participatory Data Gathering for Public sector reuse: Lessons learned from traditional initiatives, in IFIP EGOV (2013): IFIP Electronic Government, vol. 8074, pp. 87-98, 2013.
[44] V. Stodden, The scientific method in practice: Reproducibility in the computational sciences, Cambridge, Massachusetts, MIT Sloan Research Paper, No. 4773-10, 2010.
[45] K. Sweeny, Open Data: Meaning, context and implication, 2009.
[46] C. Talbot, Public value-the next big thing in public management?, International Journal of Public Administration, vol. 32, no. 3-4, pp. 167-170, 2009.
[47] S. W. van den Braak, S. Choenni, R. Meijer, and A. Zuiderwijk, Trusted third parties for secure and privacy-preserving data integration and sharing in the public sector, in Proceedings of the 13th Annual International Conference on Digital Government Research - dg.o '12, New York, 2012, p. 135.
[48] R. Velde, Public Sector Information: Why Bother?, The Socioeconomic Effects of Public Sector Information on Digital Networks: Toward a Better Understanding of Different Access and Reuse Policies, 2009.
[49] J. A. G. Versmissen, Achtergrondstudies en Verkenningen: Sleutels van vertrouwen TTP's, digitale certificaten en privacy, Den Haag, 2001.
[50] E. Welch and C. Hinnant, Internet use, transparency, and interactivity effects on trust in government, in Proceedings of the 36th Hawaii International Conference on System Sciences, Hawaii, 2003, pp. 1-7.

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762014000300004

[51] I. Williams and H. Shearer, Appraising public value: Past, present and futures, Public Administration, vol. 89, no. 4, pp. 1367-1384, 2011.
[52] WODC. (2011, June) Organization: Aim and role. English WODC. [Online]. Available: https://english.wodc.nl/organisatie/
[53] A. Zuiderwijk, M. Janssen, S. Choenni, and R. Meijer, Design principles for improving the process of publishing open data, Transforming Government: People, Process and Policy, vol. 8, no. 2, 2014.
[54] A. Zuiderwijk, M. Janssen, R. Meijer, S. Choenni, Y. Charalabidis, and K. Jeffery, Issues and guiding principles for opening governmental judicial research data, in Proceedings 11th IFIP WG 8.5 International Conference, EGOV 2012, Kristiansand, Norway, 2012, pp. 90-101.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

www.manaraa.com