



Delft University of Technology

Towards decision support for disclosing data: Closed or open data?

Zuiderwijk, AMG; Janssen, MFWHA

DOI

[10.3233/IP-150358](https://doi.org/10.3233/IP-150358)

Publication date

2015

Document Version

Final published version

Published in

Information Polity

Citation (APA)

Zuiderwijk, AMG., & Janssen, MFWHA. (2015). Towards decision support for disclosing data: Closed or open data? *Information Polity*, 20, 103-117. <https://doi.org/10.3233/IP-150358>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Towards decision support for disclosing data: Closed or open data?

Anneke Zuiderwijk* and Marijn Janssen
Delft University of Technology, Delft, the Netherlands

Abstract. The disclosure of open government data is a complex activity that may create public value yet might also encounter risks, such as the misinterpretation and misuse of data. While politicians support data release and assume that the positive value of open data will dominate, many governmental organizations are reluctant to open their data, as they are afraid of the dark side. The objective of this paper is to provide a decision-making model that assists in trade-offs between the pros and cons of open data. Data disclosure is dependent on the type of data (e.g. its sensitivity, structure and quality) and the context (e.g. organizational policies, legislation and the political influences). Based on the literature and fifteen in-depth interviews with public sector officials and data archivists, this paper identifies contextual and dataset-related variables which influence a trade-off. A decision-making model is presented capturing trade-offs, and in this way providing guidance for weighing the creation of public value and the risks. The model can be used for decision-making to open or not to open data. It is likely that the decision regarding which data should be opened or closed will shift over time.

Keywords: Open data, decision support, data release, trade-offs, positive effects, risks, variables

1. Introduction

Governments and politicians all over the world have stated to support the disclosure of open government data to accomplish a wide variety of public values (e.g. [1–3]). Public values have been defined as “collectively defined objectives that emerge from a process of collective decision-making” [4, p. 36]. Public value is a multi-dimensional construct which reflects “collectively expressed, politically mediated preferences consumed by the citizenry” [5, p. 358]. Examples of public values are better services, enhanced trust and social capital [6]. Open data-related public values are transparency, accountability and privacy. Similar to most other public goods, both positive and negative value might be created with open data.

The creation of positive public value with open data requires that governmental datasets are opened. Yet, although various datasets have already been released [7–9], more datasets appear to remain closed. Some governments are reluctant to open their data [10], because of the complexities that are associated with this activity e.g. [11–13], for example due to the legislation (e.g. the disclosure of data may conflict with citizens’ right to information privacy [8]), or because of the possible adverse effects [14]. In particular, the fear that data will be misinterpreted by the public contributes to the reluctance to release open data [15,16].

*Corresponding author: Anneke Zuiderwijk, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, the Netherlands.
E-mail: a.m.g.zuiderwijk-vaneijk@tudelft.nl.

Both politicians and scholars appear to implicitly assume that the values and benefits of open data dominate the disadvantages and risks of open data (e.g. [14,17,18]). Only few studies discuss the potential risks of openness in government and open data (e.g. [19,20]). Although the literature provides some insight in positive effects, risks, and variables that influence the positive effects and risks, these are mainly separated and not integrated in a decision-making model for publishing open data. There is a lack of literature which scrutinizes how these separate variables are related, and, more specifically, which trade-offs need to be considered when a civil servant wants to publish a dataset. There is barely any discussion going on in the literature concerning whether certain disadvantages of open data outweigh their advantages and vice versa, which leaves a void in the decision support for the release of data. As a consequence, an unrealistic impression may be provided of the potential of open data, resulting in exaggerated positive expectations of open data disclosure.

There is a clear tension between the ambitions of politicians and the reality that governmental organizations have to deal with [21]. On the one hand, we see that politicians support data release based on their view on the possible benefits. On the other hand, many governments are reluctant to open their data, as they face various complexities and they fear the potential risks. There is a lack of research regarding which trade-offs civil servants need to consider and what are the potential positive open data effects and the risks in specific contexts. The disclosure is dependent on the type of data. For instance, micro data about lawbreakers may be more sensitive than aggregated data about traffic. Moreover, data disclosure depends on the context of the data. For example, if a dataset is owned by various organizations, the contextual variable of ownership may influence the decision to open or not to open the data [32].

The objective of this paper is to provide a context and dataset dependent model that assists in decision making about whether to release governmental data or not. This research builds on Zuiderwijk and Janssen [22] and is organized as follows. In the following section the research approach is presented, followed by a discussion of relevant related literature regarding variables that may influence the decision to offer open government data. Subsequently, such variables are investigated from the perspective of empirical research. Thereafter we discuss trade-offs of disclosing open government data and conclusions on this topic are drawn.

2. Research method

In this paper we develop a model that assists in deciding whether governmental data may be opened. The generation of the model comprised three steps. First, a literature review was conducted regarding 1) potential benefits of opening data, 2) risks of opening data, and 3) relevant contextual and dataset-related variables influencing the decision to open data. We searched through databases such as Science Direct and Google Scholar using various combinations of keywords including "open data", "benefits", "barriers", "impediments", "challenges", "decision" and "model". In addition to examining the databases, we used a snowball sampling method to identify relevant articles that were referred to in the identified articles. The literature search resulted in an initial overview of potential positive effects, potential risks, and contextual and dataset-related variables influencing data disclosure. From this overview we compiled a list of variables for the decision-making model for open data release (see Section 3).

Second, since we found that there is a lack of research on trade-offs between the potential positive effects and risks of open data and on variables for open data decision-making models, we conducted fifteen interviews to obtain more insight in the trade-offs (see Section 4). A grounded approach was taken, which aimed at conceptualizing 'what is going on' by conducting empirical research [23]. The criteria for selecting interviewees were as follows.

1. The interviewees have experience with the publication of governmental data to the public in the Netherlands.
2. The interviewees are involved in publishing data for Dutch ministries that have already developed an open data policy.
3. The interviewees have experience with publishing different types of data.
4. The interviewees are willing and ready to cooperate in the research and to share relevant information.

We focused on the disclosure of governmental data of Dutch ministries because most ministries already publish data. Moreover, we delimit this study to this specific type of governmental organizations, since we believe that the context, including the level of hierarchy in the government, may influence the decision support model.

To obtain a comprehensive overview of the potential positive effects, the risks and the influencing variables, interviews were conducted with two different types of interviewees. First, interviews took place with civil servants from the Dutch Ministry of Security and Justice and Ministry of Health, Welfare and Sport. Civil servants from these ministries were selected as interviewees since they coordinated the data publication process for the department of the ministry that they worked for, and they were already involved in the actual release of datasets and for answering questions from users about (open) data for several years. The selected civil servants were responsible for making different types of governmental data available to the public as open data. Furthermore, interviews were held with open data archivists who work for an archival organization that functions as an intermediary between governmental organizations that desire to publish public data and the public. They facilitated the publication and use of open government research data, and they were involved in decision making processes for the release of governmental datasets. The archivists were involved because they are involved in the publication of many different types of ministerial data, mainly from the domains of social sciences, socio-cultural sciences, behavioral sciences and humanities. The interviewees had experience with publishing numerous datasets and the decisions surrounding data disclosure.

A pre-defined list of topics was discussed during the interviews. Since the literature did not provide insight in the trade-offs for releasing governmental data, the topics discussed during the interviews concerned these trade-offs, including the potential positive effects, the potential risks and the contextual and dataset-related variables that influence data disclosure. More specifically, questions were asked about the organization that the interviewee worked for, data publication, including reasons for publishing data, stakeholders, amounts of data published, steps taken to publish datasets, business processes, metadata, political, economic, social and technical barriers and potential risks of publishing data. Questions about the above-mentioned topics were asked in each of the interviews to ensure consistency. During the interviews notes were taken of what the interviewees said. From the information and notes that we derived from the interviews we identified and categorized variables that appeared to influence the decision on whether or not a governmental organization releases open data. The interviews provided a few new variables that were added to our initial list of variables from the literature (see Section 4). All interviews were conducted in The Netherlands with Dutch interviewees.

Third, building on the variables derived from the literature and the interviews, the decision-making model was constructed (see Section 5). The model assists in deciding whether or not governmental data may be released, taking into account the potential effects of the disclosure of certain types of data within certain contexts. The findings from these three research steps will be described in the following sections.

3. Research on variables which influence the disclosure of government data

As a first step of this study we searched the literature for variables influencing the decision to release governmental data. The results of this literature study are summarized in this section.

3.1. *Potential positive effects*

The literature reveals a large variety of potential positive effects of open government data which may be categorized as 1) political and social, 2) economic and 3) operational and technical [16]. Potential political and social benefits encompass the stimulation of transparency [24,25], the creation of new governmental services for citizens [26,27], increased participation of citizens [28], and growth of knowledge [29]. Economic benefits concern, for example, the stimulation of economic growth and innovation [30–32]. Potential operational and technical open data benefits include the reuse of data [14], the development of public policies [10] and access to external problem-solving capacity [33].

3.2. *Potential risks*

With regard to the risks of open data, the literature shows that offering open data may conflict with an individual's right to information privacy [8]. Furthermore, the conventional wisdom that opening data results in increased transparency has been challenged [19], and in general it is argued that transparency can also have a tyrannous side [34], which may also apply to open data. More information does not necessarily lead to more informed decisions [16]. Large amounts of data may result in an information overload and may not help to increase transparency. Moreover, being too transparent may result in undesirable surveillance and issues with personal privacy [35]. Another risk of open data is that it may empower mainly those who are already 'empowered' [20]. As data use requires certain skills and techniques [36], and open data users have different capacities, skills and knowledge to use open data and participate in open data initiatives [37,38], those individuals who already have access to open data infrastructures, hardware, software, financial and educational resources and skills can make more effective use of open data than others [20]. In addition, open data can be misinterpreted and misused, which might harm the reputation of a data provider [12]. These potential risks complicate the supply of open data.

3.3. *Contextual and dataset related variables influencing data disclosure*

Only few scholars have paid attention to which contextual and dataset-related variables influence the decision to release governmental data and to which trade-offs this leads. Among a few exceptions, Eckartz et al. [39] constructed a data sharing decision model. According to the model, a data provider should identify the goal of an open data initiative, the expected incentives, the constraints, and the processes to open the data. The constraints may be different for each data provider. Based on a literature review and three use cases for data sharing in the private sector, Eckartz et al. [39] identified five main constraints categories for opening data: ownership, privacy, economic, data quality and technical. They state that these constraints influence the level of openness that is appropriate for certain datasets, yet they did not describe trade-offs. Another decision model for the publication of governmental data has been developed by Zuiderwijk et al. [40]. Their model lists a number of general and dataset related issues for the disclosure of open government data. The identification of a specific issue guides the researcher towards different types of data release with various levels of openness, such as open access and levels of more restricted access. Although the authors mention several potential risks of governmental data release, such as misuse, misinterpretation and negative publicity for the data provider, the study did not define which trade-offs each variable in the model may lead to.

Table 1
Variables for a decision model for the disclosure of governmental data derived from the literature

Variables for a decision model for the disclosure of governmental data	Source	
Potential positive effects	Political and social, e.g. transparency, the creation of new governmental services for citizens, increased participation of citizens and knowledge growth	Bertot et al. [24], Neuroni et al. [25], Lindman et al. [27], Yang and Kankanhalli [26], Conradie and Choenni [28], Harrison et al. [29], Janssen et al. [16]
	Economic e.g., stimulation of economic growth and innovation	Borzacchiello and Craglia [30], Janssen [31], Janssen et al. [16], Zeleti et al. [32]
	Operational and technical, e.g., reuse of data, the development of public policies and access to external problem-solving capacity	Huijboom and van den Broek [14], Janssen [10], Janssen et al. [16], O’Riain et al. [33]
Potential risks	Open data may conflict with an individual’s right to privacy	Kulk and van Loenen [8], Parycek and Sachs [35]
	The statement that opening data results in increased transparency can be challenged and transparency may also have a tyrannous side	Bannister and Connolly [19], Strathern [34]
	Too much transparency may lead to undesirable surveillance	Parycek and Sachs [35]
	More information does not necessarily lead to more informed decisions (e.g. risk on information overload)	Janssen et al. [16]
	Open data users have different capacities; open data may empower mainly those who are already ‘empowered’	Gurstein [20], Albano and Reinhard [37], Yannoukakou and Araka [38]
	Risk on misinterpretation and misuse of data, which might harm the reputation of a data provider	Barry and Bannister [12]
Contextual and dataset related variables which influence effects of open data disclosure	Ownership, privacy, economic, data quality, technical General: policy confidentiality, deletion policy, embargo placement, organizational changes (e.g. time-consumption), ownership, privacy-sensitivity and anonymization, lack of metadata, (re)use of data by organization itself, policy-sensitivity, unlawfulness, completeness and exhaustiveness, representation, validity, reliability. Dataset related: clearness and comprehensiveness of column, row, value, variable and other names, provision of additional reports, overall data quality, other	Eckartz et al. [39] Zuiderwijk et al. [40] Zuiderwijk et al. [40]

3.4. Summary: Variables for the decision-making model

Despite not revealing trade-offs, the literature does provide a number of variables for a decision-making model (see Table 1). The identified contextual variables may influence the potential positive effects and the risks of disclosing governmental data. These variables need to be taken into account in trade-offs for opening data. In the following section we supplement the initial list of variables from Table 1 with variables derived from the in-depth interviews.

4. Interviews on variables influencing the disclosure of government data

In this section we discuss the findings from the interviews. We identified several variables influencing data disclosure that we did not find in the literature. These variables are discussed, as well as the variables that confirm the literature.

4.1. *Potential positive effects*

A first potential positive effect mentioned in the interviews concerned increased visibility, which refers to showing what kind of research is being conducted by the organization and which data are collected. This benefit falls within the category of political and social benefits mentioned by Janssen et al. [16]. Interviewees said that opening data may increase the public's familiarity with their organization and this may improve their reputation by showing that they are an open organization. Moreover, by publishing governmental data, the organization intends to give back to the public what citizens have already paid for through the taxes. Second, within the same category of benefits, the interviews showed the potential positive effects of transparency, accountability and creating a culture of openness, in this way confirming Bertot et al. [24] and Neuroni et al. [25]. For one of the public organizations where we interviewed civil servants, the dissemination of knowledge was a key task stated in their mission, and open data was viewed as a means to contribute to this task.

Within the category of operational and technical benefits mentioned in the literature [16], the interviewees also stated that disclosing data might reduce the workload of both of the investigated ministerial agencies, since they received many individual requests for data. Opening the data by default might result in a decreased number of individual data requests. We did not find this benefit through our literature review. In addition, in the category of operational and technical benefits, it was found that data reuse allows for repeating the data analysis that civil servants conducted and supports checking whether the conclusions that were derived from this analysis are valid and which other results can be derived from the data. Citizens can scrutinize the data and reuse them in new and innovative ways. In this way, the interviews confirmed the literature that data can be opened for the benefit of their reuse [14]. The interviewees did not point at the potential economic benefits of releasing data that we identified from the literature.

4.2. *Potential risks*

Based on the interviews, we created a comprehensive overview of potential risks of disclosing governmental data. First, it was found in the interviews that there is a risk on violating the privacy of citizens unintentionally. The interviews showed that much effort is put on removing privacy sensitive variables from datasets, so that the datasets can be opened. Guidelines about privacy-sensitivity partly help to identify which data cannot be published, yet a large extent of own interpretation by the data provider is still required, and combining data with other sources could still lead to the identification of a person.

Second, open data may have undesirable consequences for the government. Although transparency is often mentioned as a significant advantage of open data (e.g. [24,41]), transparency may also result in a more negative image of the government. Trust in the government might decrease by opening datasets which have poor quality or datasets that support decisions contrary to decisions taken by the government. In addition, opened data may show that governments have taken incorrect decisions. This could harm the reputation of the government. Moreover, the interviews showed that many datasets cannot be released for legal or other reasons. Examples of datasets that cannot be made available to the public are datasets which contain privacy identifying variables, (policy) sensitive variables and datasets which have been created by multiple organizations which have different levels of security, different policies and have to comply with different laws. Publishing those kind of data would lead to undesirable situations, as this would violate the law (e.g. the data protection law) and may harm the reputation of the organization that provides these data (as also found by [12]). In one of the interviews it was pointed out that approximately

only eight percent of the agency's data might be appropriate for publication. This means that much valuable data cannot be made available for reuse.

Third, stakeholders may not profit equally from data disclosure. Citizens are often mentioned as important open data stakeholders (e.g. [42]), yet our interviews revealed that university researchers, other researchers and students seem to profit most from the data released by the interviewees. The disclosed data usually do not answer questions of ordinary citizens, and citizens may not have sufficient statistical skills to use the data. As such, open data can be used by certain groups to strengthen their position (as also found in the literature [20]), instead of creating a level playing field.

Fourth, the interviews showed that there is a risk on misuse and misinterpretation of data, as we also found in the literature (e.g. [12]). In several interviews it was mentioned that very complex data were not made available to the public, because the risk of misinterpretation and misuse was too high. Since open data are freely available to anyone, also people with limited knowledge about how to interpret them can use them. This might result in incorrect conclusions about the results of an analysis of the data. Some people may even intend to misuse the data. One interviewee stated that the misinterpretation and misuse of data is mainly feared because this could damage the reputation of the data provider.

Fifth, there is a risk that data users make decisions based on poor information quality. The literature shows that the quality of open data varies widely [43,44], and that data quality refers to various quality dimensions, such as completeness, correctness, accuracy, timeliness, accessibility, relevance and understandability. Some interviewees believe that data with poor quality should not be published. They stress the risks of opening low quality data and suggest to avoid the analysis and drawing of conclusions on low quality data. On the other hand, there are proponents for releasing and opening low quality data. Their argument is that if these data are not released it might remain at low quality, while the publication of data could help in pointing out on which dimensions the quality of the data is poor, so that this can be improved. The crowd can then comment on the data, and this may create an incentive for the data publisher to improve the data. Both arguments can be valid and this requires a trade-off per dataset.

Sixth, published data can be biased and may provide a blurred or incorrect view on reality. The interviewees indicated that the investigated organizations only publish those data which are not sensitive, not very complex or can do no harm. More sensitive data are kept hidden in the organizations to avoid reputation damage of the organization. This means that only certain types of data are selected to be made available to the public, and these data may favor certain arguments, which implies a bias. For example, only data that favor current politicians in place or that favor arguments for past policy decisions might be released. Data that demonstrates the opposite might not be released. We did not find this potential risk in the literature.

Seventh, the interviews revealed that data are often disclosed as an afterthought. Some interviews showed that the opening of data was not institutionalized and, as a consequence, the opening of data was separated from the daily routines and procedures. Transformations are required to develop a sustainable policy for releasing open data. Some interviews showed that the late focus on data publication may result in datasets which are not publishable exactly because of this afterthought, for instance, when insufficient metadata have been collected. If data publication is institutionalized, data can be created in such a way that they become more understandable, for instance, by adding considerable metadata. It is usually very complex and time-consuming to do this after the dataset has already been created.

Finally, another risk mentioned is that resources can be wasted to publish valueless data. There is hardly any support to determine which datasets are (potentially) valuable. Moreover, the interviews revealed that some of the published datasets have only been downloaded a few times. It is often seen that one person downloads various datasets, which shows that only few people profit from these open data.

The interviewees stated that they do not have much more information about how the open data have been used than the number of downloads, and there is a lack of insight in which datasets are valuable to end-users.

4.3. Contextual and dataset related variables influencing effects of open data

A number of contextual and dataset related variables were found in the interviews which influence the decision to disclose open data. First, the interviewees mentioned that ownership influences whether data can be released, which confirmed the study of Eckartz et al. [39]. In several interviews it was stated that the examined governmental organizations also maintain data of other organizations. These data are not possessed by the governmental organization and for this reason they cannot be published. To quote from one of the interviewees, “the impact of ‘borrowed data’ is high. Much research that we conduct is based on borrowed data”. The interviews showed that several organizations only release data that they have collected themselves.

Furthermore, the interviewees stated that the governmental organizations that they work for can publish data only after an embargo period has expired, which was also found in [40]. The embargo period results in various restrictions. Some civil servants first use the collected data to write governmental reports before they could be published. Thereafter these reports sometimes had to be discussed by politicians.

The desired control over data access may influence the decision to disclose open government data. In both examined public agencies it was seen that data were published in an archive which allowed the data provider to keep some control over the data, because the applicable data license does not allow for disseminating data widely, but only for data download and use by the registered user. For some datasets a more restricted form of data release was used, however, even this appeared not to be sufficient sometimes: “Restricted access for scientific data use appeared not to be enough. Even then we feel the need for more control over the data”. This need for more control was felt because they “want to know what is going to happen with the data”. It was also found that it is often unclear if the data owner or the user can be held accountable for the wrongful use and interpretation of data or low data quality. Such contextual information is important for the decision whether a certain dataset can be disclosed. Moreover, in one of the organizations it was found that some datasets were simply too poorly documented and had insufficient metadata to be able to be interpreted correctly. Finally, low data quality, the risk on privacy violations, policy confidentiality and required institutional changes appeared to have considerable influence on whether datasets could be released.

4.4. Summary: Variables for the decision-making model

The identified variables that influence the decision to release governmental data as derived from the interviews have been summarized in Table 2.

5. A decision support model for disclosing government data

In the previous sections we derived variables for a decision support model for opening governmental data. The identified contextual variables may influence the potential positive effects and the risks of disclosing governmental data. All these variables need to be taken into account in trade-offs for opening data. In this section we describe the model, which consists of a list of typical trade-offs. The model points at variables that influence the decision to release data, as well as potential positive effects and

Table 2
Variables influencing the disclosure of data derived from the interviews

	Variables for a decision model for the disclosure of governmental data
Potential positive effects	Visibility Transparency and accountability Knowledge growth Reuse of data, giving data back to the public (created with tax money) Possible decrease of workload
Potential risks	Privacy can be violated unintentionally Open data may have negative consequences for the government Stakeholders do not profit equally from the opening of data Misinterpretation and misuse Data users may make decisions based on poor information quality Published data may be biased and provide a non-realistic view Opening data as an afterthought, no priority given to data publication, resulting in little attention for public value and solving societal problems Wasting resources to publish valueless data
Contextual and dataset related variables which influence effects of open data disclosure	Ownership of the data Embargo period prohibits the publication of recent data (timeliness) Control over data access Unclear responsibility and accountability within governmental organizations Availability of metadata Data quality Privacy-sensitivity Policy confidentiality Required organizational changes

risks (see Table 3). The first column of Table 3 shows the contextual and dataset related variables which appeared to influence the potential positive effects and risks of open data.

The first contextual and dataset related variable in the model concerns ownership of the data. The release of datasets which are owned by various organizations may be interesting for potential data users, since such datasets combine insights from various organizations. However, a potential risk of releasing such data is that it may be unclear who is responsible and accountable for the data release. Civil servants need to consider the trade-off whether they will obtain permission from all data owners to publish the dataset, which can be very time-consuming. Permission may not be given so easily, since civil servants wish to avoid accidentally releasing data which they should not have released. If they do accidentally release such data, their minister may be held accountable for this mistake, which may have important negative implications. To avoid such problems, civil servants may adopt more risk-avoiding behavior.

Second, governmental organizations may decide to adopt an embargo period for disclosing their datasets [40]. On the one hand, adopting a long embargo period reduces the risk on wrongfully publishing data and data may become less sensitive over a longer time period. Moreover, a longer embargo period allows for data reuse by the organization itself. However, a longer embargo period may also reduce the usefulness of datasets. These effects should be part of the trade-off in choosing a specific embargo period.

A third variable of the decision model that needs to be considered is the objective to publish data to become a more transparent organization. Since the literature challenges the conventional wisdom that opening data results in increased transparency [19] and the interviews indicated that opening data may also have negative consequences for the government, data providing organizations need to decide to which extent they aim to become more transparent. In one respect releasing governmental data may provide the public with more insight in what governmental processes encompass and what public agencies

Table 3
The decision-making model for disclosing open government data

Contextual and dataset related variables	Potential positive effects	Potential risks
1. Ownership: Data are owned by various organizations	The datasets are interesting because they combine insights and knowledge from various organizations	Unclear who is responsible and accountable for data release, complex to obtain permission from all data owners to publish the dataset
2. Embargo period: Do not release very current data	Reduced risk on wrongfully opening data, older data may become less sensitive, data can be reused by the organization itself	Data become less useful when they are older, data's timeliness reduces
3. Objective to publish data: Opening data to become a more transparent organization	More insight in governmental processes, crowd sourcing may lead to suggestions on how governmental organizations may be improved	Organizations may become too open, wrongfully releasing certain datasets, negative publicity for the public agency, decrease of trust in the government, requires considerable resources and changes in organizations
4. Data access and licensing: Data access may be restricted to specific groups of people	More control over the use, learn from how data are reused, minimize data misuse and misinterpretation, might lead to better acknowledgement of data provider	Some stakeholder groups are excluded from data reuse, may result in a less creative data reuse and reduced data reuse because users may be afraid that they violate the data license conditions, less benefit realization
5. Data sensitivity: Adopting a restrictive policy for publishing policy, privacy and otherwise sensitive data	Minimizing the risk on negative publicity, minimizing the risk of politicians and civil servants to be held accountable for taking wrongful decisions, when privacy sensitive variable are removed from the dataset, more datasets can be released	Data selection implies bias, only those data are made available to the public that favor certain arguments or decisions of certain politicians, difficult to determine sensitivity of datasets, removing sensitive variables from datasets is time and resource-consuming
6. Data quality: Disclosing data without having insight in their quality	The public may point out which quality problems the data has, data quality can be improved	Decisions and conclusions may be based on low quality data
7. Metadata: Minimum documentation and metadata provided with opened data	Spending little time on data documentation and metadata, data providers can work with relatively simple systems	Increases risk on misinterpretation of data, makes it more difficult to use datasets on the long term when the data experts cannot be consulted for advise anymore (difficulties for preservation)

do. Disclosing governmental data may also lead to suggestions from the public on how governmental organizations and their processes may be improved. In another respect, opening governmental data to the public may result in too much openness. For instance, public agencies may accidentally release sensitive data that should not have been released. This may result in a more negative image of the government and may decrease the public's trust in the government. Becoming a more transparent organization also requires considerable changes in organizations and making data publication part of daily tasks, and it requires financial and other resources from an organization.

Fourth, in deciding whether to open a dataset, a trade-off needs to be considered regarding the type of access that users will have to the data and the related license for the reuse of the data that will be adopted. Data access may be restricted to specific groups of people, such as people who register on an open data portal or people who inform the data provider of what they want to do with the data in advance. Such restricted data access makes it possible for the data provider to have more control over how the data are reused. Wrongful data use may be detected or even avoided, and the data provider may obtain

more information about how data are reused, so that the data provider may learn from new applications and uses of the data. Furthermore, using a more restricting license may prescribe the data user to better acknowledge the data provider in the reuse of the data. On the other hand, restricted data access locks out certain groups of people from data use and there are fewer stakeholders who otherwise might have profited from the data. Restricting licenses and restricting access may result in less creative data reuse and reduced data reuse because users may be afraid that they violate the data license conditions. As a result, the envisioned benefits of opening data may be realized to a more limited extent.

Additional trade-offs relate to the decision to publish sensitive data, such as (policy-)sensitive and privacy-sensitive data. Civil servants may decide not to publish any policy-, privacy- or otherwise sensitive data or process the data before publishing to avoid any problems at a later stage. This reduces the risk on negative publicity based on conclusions from open data. For example, avoiding the disclosure of policy-sensitive data minimizes the risk that a politician is blamed for having taken wrong decisions in the past based on conclusions derived from the reuse of open data. When sensitive variables are removed from datasets, more datasets can be released since their publication is then not prohibited anymore by law. Potential negative effects of not disclosing (policy-)sensitive data is that only those data are made available to the public that favor certain arguments or decisions of politicians. Then the opened data implies some kind of bias and an unrealistic perspective is created with the disclosed datasets. With regard to privacy-sensitive data, data protection laws require a certain extent of interpretation by the civil servant and it can be difficult to determine which data are privacy-sensitive. In addition, the combination of several variables may lead to the identification of persons, yet it cannot be predicted in advance which combination of variables will lead to this identification. Moreover, removing sensitive variables from datasets is time and resource-consuming, especially since this requires interpretation and deliberation.

Another trade-off for considering governmental data publication concerns whether to release low quality data. Civil servants may decide to open datasets without having insight in their quality. As a consequence, they may publish datasets that are incomplete, inaccurate, invalid or unreliable. Data users may help in pointing out on which dimensions the quality of the data is poor, so that this can be improved. The crowd might comment on the data, try to improve the low quality data and this may create an incentive for the data publisher and help the data publisher to improve the data. In contrast, the low quality data may be reused and decisions and conclusions may be based on these data. Data users may not notice that the data is of poor quality.

Finally, a trade-off that needs to be considered concerns the inclusion of contextual metadata which might cause additional costs and effort. Governments may provide only limited documentation and metadata with the data that they release. The advantage of keeping documentation and metadata efforts to a minimum is that data providers spend only little time on documentation and metadata, and they can work with simple easily understandable systems. Yet, the provision of limited metadata increases the risk on the misinterpretation of open data, since users do not have information about the context in which the dataset has been created and how it should be interpreted. Additionally, a lack of metadata causes difficulties for the preservation of datasets for a long period, since data experts cannot be consulted anymore for advice then.

6. Discussion

In this study we developed a decision-making model for the release of governmental data. Various datasets require a trade-off between the value that can be created and the risks of damage that can be made using the data. Practitioners can use the model to identify variables that need to be considered

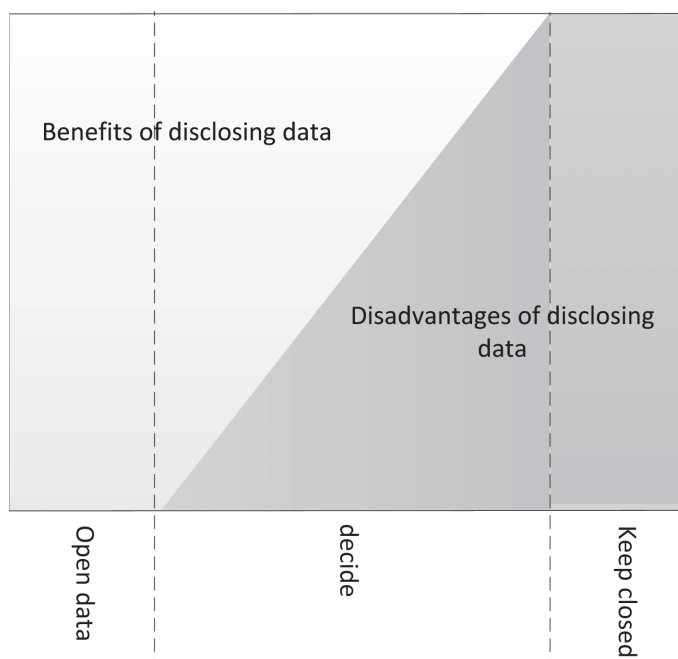


Fig. 1. Decision-making to open or not to open datasets.

in the disclosure of governmental data. The model can help them to identify potential positive effects and potential negative effects that they need to understand to consider the trade-offs of data publication. This may lead to more realistic expectations of the benefits of open data publication. Practitioners can evaluate the model in practice, and scientists can evaluate the relevance of the model in different contexts and further refine it. The model has been focused on the disclosure of Dutch ministerial data, and may need to be adjusted for the publication of other types of data in other countries.

While the model assists in making trade-offs for the publication of governmental data, decision-makers still need to weigh the importance of the variables in the model for each dataset. Decision-makers should analyze each of the variables for their specific situation. For instance, the interviews suggested that the sensitivity of the data depends on the context in which the dataset has been created, by which organization, the political climate and various other variables. Moreover, the interviews suggested that the length of the embargo period depends on the sensitivity of the data, whether the organization desired to reuse the data itself before releasing it to the public, and whether the data will still be valuable after a certain period.

Figure 1 shows the decision making process in which the benefits and disadvantages of opening data are weighed. Some data has many benefits and hardly any disadvantages and can be opened without any discussion. Other data should not be opened without any doubt due to security, privacy or other reasons. There is a huge pile of data requiring a trade-off in which either the benefits or risks may dominate. We do not know how large this part is that organizations need to decide on. Furthermore, it is likely that this changes over time. Since public values represent the needs and preferences of the collective citizenry, public values may change over time, as the needs and preferences of citizens may change [45]. It is likely that the decision regarding which data should be opened or closed will shift over time.

An important issue is to which extent governmental agencies should give up control and accept risks such as the misuse and misinterpretation of open data. This may be more a political decision than one

made by civil servants, and we recommend more research in this field. This raises the question whether data should be opened or whether risk adverse behavior should dominate. Politicians may not have a comprehensive overview of the potential positive effects and risks of disclosing their data (although this may change over time), hence the decision to release data is often left to public servants. Civil servants might have to make the trade-off between potential positive effects and risks, even though considerable societal issues are at stake. In the Weberian system civil servants are expected to have risk averse behavior to warrant the rights and values of citizens [46] and as such the decision not to release data dominates. Hence we argue that trade-offs of the effects and risks of disclosing government data should be central to the discussion on open data by politicians. We plea for creating more awareness for politicians of the trade-offs and issues surrounding open data disclosure to make more informed decisions. Civil servants may feel more supported and confident with data release if politicians would decide on making clear trade-offs for open data publication, which may lead to opening more governmental datasets.

7. Conclusions

In this study a decision-making model was developed for the disclosure of governmental data based on a list of variables which influence the decision to disclose or not disclose datasets, and the potential positive effects and risks of the decision to release the data in a certain form. The identified variables concerned 1) ownership, 2) an embargo period, 3) transparency, 4) data access and licensing, 5) data sensitivity, 6) data quality and 7) metadata. The potential consequences of disclosing data in various forms, such as data of low quality and with a long embargo period, were discussed, and positive effects and risks were identified. Positive effects include the minimization of the risk on negative publicity, obtaining more control over data reuse, learning from how data are reused, obtaining more insight in governmental processes, and allowing the public to point out which quality problems a dataset has. The potential risks that we identified included a lack of clarity about who is responsible and accountable for data release, wrongfully releasing certain datasets, negative publicity for the public agency, a decrease of trust in the government, excluding certain stakeholders from data reuse, and founding a decision on low quality data.

This research provides a number of practical and scientific contributions. It contributes to practice by providing a model that assist governmental data publishers in making decisions surrounding the disclosure of their data. The model can help them to identify potential positive effects and risks that they need to understand to consider the trade-off of data publication. This research contributes scientifically by being the first to provide insight in a decision-making model for open data publication. This model can be evaluated in other contexts, and it can be refined further.

This study is subject to a number of limitations. The interviews have been conducted with a particular group of interviewees, namely with civil servants who work for two ministerial organizations and with archivists who are involved in the release of governmental data on specific topics. Additionally, the interviews have been conducted in the Netherlands only. Different countries with different cultures or different types of governmental organizations may approach open data release in very different ways. Moreover, this study did not incorporate the evaluation of the model in a real-world context. Hence we recommend future research to test the developed model in practice, as well as to examine whether this model holds up in other countries and for other types of governmental organizations. Finally, the developed decision making model did not incorporate the effects of open data in the long run. For example, it neither encompassed the impact of the misinterpretation of open data, nor the incorrect decisions that may be made based on this misinterpretation. We recommend future research to examine which models can be developed for disclosing open data in the long run.

References

- [1] B. Obama. (2013, April 9). Memorandum for the heads of executive departments and agencies. Open Data Policy – Managing Information as an Asset Available: <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.
- [2] B. Obama. (2009, February 21). *Memorandum for the Heads of executive Departments and Agencies: Transparency and Open Government*. Available: http://www.whitehouse.gov/the_press_office/Transparency_and_Open_Government.
- [3] European Commission. (2013, June 13). *Commission welcomes Parliament adoption of new EU Open Data rules*. Available: http://europa.eu/rapid/press-release_MEMO-13-555_en.htm.
- [4] M. H. Moore, *Creating public value: strategic management in government*. USA: Harvard University Press, 1995.
- [5] J. O’Flynn, From new public management to public value: Paradigmatic change and managerial implications, *Australian Journal of Public Administration*, vol. 66(3), pp. 353-366, 2007.
- [6] L. Horner and L. Hazel, *Adding public value*. London: The Work Foundation, 2005.
- [7] A. Meijer and M. Thaens, Public information strategies: Making government information available to citizens, *Information Polity* vol. 14(1-2), pp. 31-45, 2009.
- [8] S. Kulk and B. van Loenen, Brave new open data world? *International Journal of Spatial Data Infrastructures Research*, vol. 7, pp. 196-206, 2012.
- [9] S. G. Grimmelikhuijsen and A. J. Meijer, 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(1), pp. 137-157, 2014.
- [10] K. Janssen, The influence of the PSI directive on open government data: an overview of recent developments, *Government Information Quarterly*, vol. 28(4), pp. 446-456, 2011.
- [11] C. Martin, Barriers to the open government data agenda: Taking a multi-level perspective, *Policy & Internet*, vol. 6(3), pp. 217-240, 2014.
- [12] E. Barry and F. Bannister, Barriers to open data release: A view from the top, *Information Polity*, vol. 19(1), pp. 129-152, 2014.
- [13] A. F. van Veenstra and T. A. van den Broek, Opening moves. Drivers, enablers and barriers of open data in a semi-public organization, presented at the 12th Electronic Government Conference, Koblenz, Germany, 2013.
- [14] N. Huijboom and T. van den Broek, Open data: An international comparison of strategies, *European Journal of ePractice*, vol. 12(1), pp. 4-16, 2011.
- [15] J. C. Molloy, The Open Knowledge Foundation: Open Data Means Better Science, *PLoS Biology*, vol. 9(12), pp. 1-4, 2011.
- [16] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, Benefits, adoption barriers and myths of open data and open government, *Information Systems Management*, vol. 29(4), pp. 258-268, 2012.
- [17] J. Lassnanti, B. Bergvall-Kåreborn, and A. Ståhlbröst, Shaping Local Open Data Initiatives: Politics and Implications, *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9(2), pp. 17-33, 2014.
- [18] M. Kassen, A promising phenomenon of open data: A case study of the Chicago open data project, *Government Information Quarterly*, vol. 30(4), pp. 508-513, 2013.
- [19] F. Bannister and R. Connolly, The trouble with transparency: A critical review of openness in e-government, *Policy & Internet*, vol. 3(1), pp. 1-30, 2011.
- [20] M. Gurstein, Open data: Empowering the empowered or effective data use for everyone? *First Monday*, vol. 16(2), p. n.p., 2011.
- [21] A. Zuiderwijk and M. Janssen, Open data policies, their implementation and impact: A comparison framework, *Government Information Quarterly*, vol. 31(1), pp. 17-29, 2014.
- [22] A. Zuiderwijk and M. Janssen, The negative effects of open government data – investigating the dark side of open data, presented at the Proceedings of the 15th Annual International Conference on Digital Government Research, Aguascalientes, Mexico, 2014.
- [23] B. G. Glaser and A. L. Strauss, *The discovery of Grounded Theory: strategies for qualitative research*. United States of America: Sociology Press, 1967.
- [24] J. C. Bertot, P. T. Jaeger, and J. M. Grimes, Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies, *Government Information Quarterly*, vol. 27(3), pp. 264-271, 2010.
- [25] A. C. Neuroni, R. Riedl, and J. Brugger, Swiss executive authorities on open government data – Policy making beyond transparency and participation, presented at the 46th Hawaii International Conference on System Sciences, Grand Wailea, Hawaii, 2013.
- [26] Z. Yang and A. Kankanalli, Innovation in government services: The case of open data, in *IFIP AICT TDIT 2013*, 2013, pp. 644-651.

- [27] J. Lindman, M. Rossi, and V. K. Tuunainen, Open data services: Research agenda, presented at the 46th Annual Hawaii International Conference on System Sciences, Hawaii, U.S.A., 2013.
- [28] P. Conradie and S. Choenni, Exploring process barriers to release public sector information in local government, presented at the 6th international conference on theory and practice of electronic governance, New York, U.S.A., 2012.
- [29] T. M. Harrison, T. A. Pardo, and M. Cook, Creating open government ecosystems: A research and development agenda, *Future Internet*, vol. 4, pp. 900-928, 2012.
- [30] M. T. Borzacchiello and M. Craglia, The impact on innovation of open access to spatial environmental information: A research strategy, *International Journal of Technology Management*, vol. 60(1-2), pp. 114-129, 2012.
- [31] K. Janssen, Open government data and the right to information: Opportunities and obstacles, *The Journal of Community Informatics*, vol. 8(2), 2012.
- [32] F. A. Zeleti, A. Ojo, and E. Curry, Emerging business models for the open data industry: characterization and analysis, presented at the Proceedings of the 15th Annual International Conference on Digital Government Research, Aguascalientes, Mexico, 2014.
- [33] S. O'Riain, E. Curry, and A. Harth, XBRL and open data for global financial ecosystems: A linked data approach, *International Journal of Accounting Information Systems*, vol. 13, pp. 141-162, 2012.
- [34] M. Strathern, The Tyranny of Transparency, *British Educational Research Journal*, vol. 26(3), pp. 309-321, 2000.
- [35] P. Parycek and M. Sachs, Open government – information flow in web 2.0, *European Journal of ePractice*, vol. 9, pp. 1-12, 2010.
- [36] G. Puron-Cid, J. R. Gil-Garcia, and L. F. Luna-Reyes, IT-enabled policy analysis: New technologies, sophisticated analysis and open data for better government decisions, presented at the 13th Annual International Conference on Digital Government Research, Maryland, U.S.A., 2012.
- [37] C. S. Albano and N. Reinhard, Open government data: facilitating and motivating factors for coping with potential barriers in the Brazilian context, presented at the 13th IFIP WG 8.5 International Electronic Government Conference Dublin, Ireland, 2014.
- [38] A. Yannoukakou and I. Araka, Access to government information: Right to information and open government data synergy, *Procedia – Social and Behavioral Sciences*, vol. 147, pp. 332-340, 2014.
- [39] S. M. Eckartz, W. J. Hofman, and A. F. Van Veenstra, A decision model for data sharing, in *Electronic Government*, vol. 8653, M. Janssen, H. J. Scholl, M. Wimmer, and F. Bannister, Eds., ed: Springer Berlin Heidelberg, 2014, pp. 253-264.
- [40] A. Zuiderwijk, M. Janssen, R. Meijer, S. Choenni, Y. Charalabidis, and K. Jeffery, Issues and guiding principles for opening governmental judicial research data, presented at the 11th Conference on Electronic Government, Kristiansand, Norway, 2012.
- [41] C. Böhm, M. Freitag, A. Heise, C. Lehmann, A. Mascher, F. Naumann, et al., GovWILD: integrating open government data for transparency, presented at the 21st international Conference Companion on World Wide Web, Lyon, France, 2012.
- [42] B. Obama. (2012, September 15). *Digital Government. Building a 21st Century Platform to Better Serve the American People*. Available: <http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html>.
- [43] S. Auer, J. Lehmann, A.-C. N. Ngomo, and A. Zaveri, Introduction to linked data and its lifecycle on the web, in *Reasoning Web. Semantic Technologies for Intelligent Data Access*, S. Rudolph, G. Gottlob, I. Horrocks, and F. van Harmelen, Eds., ed Mannheim: Springer, 2013, pp. 1-90.
- [44] G. Kuk and T. Davies, The roles of agency and artifacts in assembling open data complementarities, presented at the Thirty Second International Conference on Information Systems, Shanghai, China, 2011.
- [45] J. Alford and O. Hughes, Public value pragmatism as the next phase of public management, *The American Review of Public Administration*, 2008.
- [46] F. Sager and C. Rosser, Weber, Wilson, and Hegel: theories of modern bureaucracy, *Public Administration Review*, vol. 69(6), pp. 1136-1147, 2009.