Government Information Quarterly xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Government Information Quarterly

journal homepage: www.elsevier.com/locate/govinf

# On the barriers for local government releasing open data

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

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

ARTICLE INFO

Available online xxxx

*Keywords:*
Open data
PSI
Process barriers
Data release

ABSTRACT

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



## 1. Introduction

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

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

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

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

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

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

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

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

http://dx.doi.org/10.1016/j.giq.2014.01.003
0740-624X/

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

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

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

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

## 2. Related work

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

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

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

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

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

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

## 3. Approach

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

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

### 3.1. Participatory action research

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

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

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

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

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

Please cite this article as: Conradie, P., & Choenni, S., On the barriers for local government releasing open data, *Government Information Quarterly* (2014), http://dx.doi.org/10.1016/j.giq.2014.01.003

supply data for project purposes, and second by highlighting the type of the data that is most wanted and the potential re-uses of the data.

While stakeholders such as non-governmental organizations, the media and citizens feature as important actors, our focus had been on incorporating stakeholders from outside the public sector, such as the local creative industry. Several commercial companies took part through their active involvement in our educational program. The cooperation between creative industry as potential user of data, the public sector as data supplier, the education and the research is shown in Fig. 1. By bringing these stakeholders together, a broad view of the issues can be sketched, with education and creative industry acting as data users, the public sector supplying data, and researchers cataloging data release in order to understand barriers to release.

## 4. Implementing PAR

In order to come to a deeper understanding of the underlying processes for the barriers to data release, we have studied the participating services provided by the organizations at the municipality. This happened first through an exploratory workshop with consortium partners where previous data release experiences within each service were highlighted. This workshop was attended by a representative from each of the participating services (n = 6). The exploratory workshop was followed with desk research about the services. A bigger group of participants from the consortium also filled out a questionnaire (n = 14), while an additional external group of civil servants (n = 50) later completed the questionnaire. Questions were based around themes such as judicial, social, cultural, economical and technical barriers to data release. Statements relating to these themes were given, asking the respondents to agree or disagree, on a 5-point Likert scale. Example statements include: *Because my department generates income through licensing data, I think my department should retain the right to license the data; or within my department, too much of the data is too privacy sensitive to release*.

Answers from the questionnaire formed the input for subsequent in-depth interviews with a group of participants involved in data management, policy development and digital data collection (n = 18). The interview questions were related to the type of the data used by the department, and to the processes pertained to data storage and use. A subset of respondents (n = 8) that used data extensively in their tasks also participated in a workshop where some statements regarding data use were introduced by attendees, using input from the previous interviews. The statements focused on how data is produced and the relation of data to the core task of the service. A second workshop with key persons (n = 6) was organized where the focus was specifically on the benefits and barriers of data release. In the left part of Fig. 2, we summarize activities that we carried out for five services.

In the case of the Police department, in addition to interviews mentioned above, a discussion evening was hosted and the thresholds to data release were discussed in a plenary session (n = 21). In a follow-up session, a second round of interviews was conducted with decision makers inside the department (n = 2) to further understand barriers to release data related specifically to the Police, and followed by a review of the experiences gained from data release, attended by a representative from the national open data initiative and a policy officer from the Ministry of the Interior and Kingdom Relations (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties) (n = 4). In the right part of Fig. 2, we summarize the activities for the Police department.

Additionally, we interviewed respondents outside of the project consortium, to get a broader view, most prominently from the Management Service (Bestuursdienst). The goal of these initial interviews was to explore how data release is currently facilitated within the organization and what the relation between data use and task execution is.

In-depth interviews were created based on the themes that emerged from the first open workshop about data release. These results were compared with existing research on data release by national governments, to form a better picture of the subject.

In the remainder of this section, we present the profiles of each of the participating public sector organizations, introducing how data is used internally. Following this, we present five cases from our education program that was used as a driver to understand data release.

### 4.1. Profile of the participating public sector organizations

The first impression gathered from the interviews is the large variety between how the processes inside the selected services impact the use and ownership or attitude to data. City Development executes and develops urban development policies such as new housing complexes, or new traffic routes. For this task, City Development uses data extensively. The organization has a traditional decentralized data structure, where data is used to develop policy, but is often not re-used subsequently.

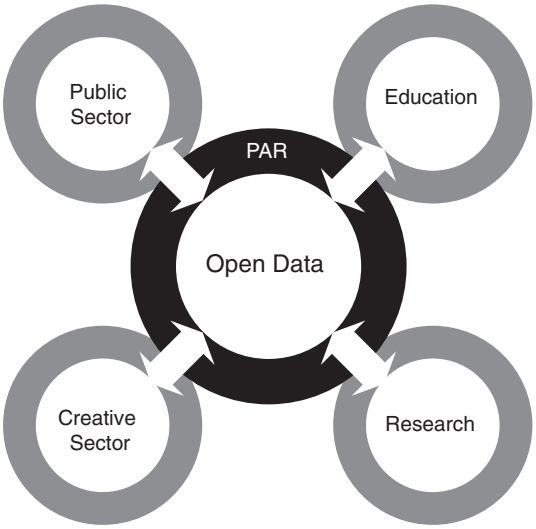


**Fig. 1.** Stakeholders in the participatory action.

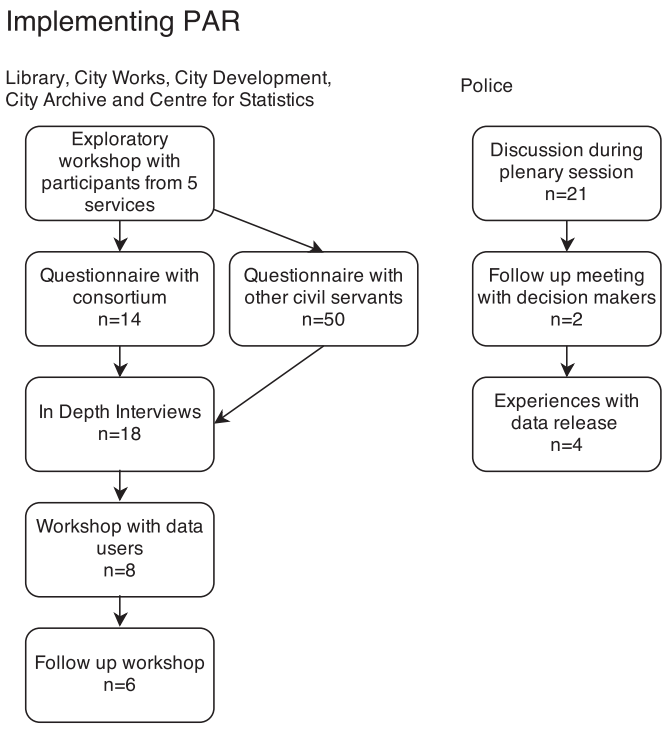


**Fig. 2.** Implementing PAR.

Please cite this article as: Conradie, P., & Choenni, S., On the barriers for local government releasing open data, *Government Information Quarterly* (2014), http://dx.doi.org/10.1016/j.giq.2014.01.003

The sources of the data also vary. In some cases, it is gathered by a certain department, for example through sensors on the road. In other cases, data might be bought externally, or requested from a service of the department controlling the data. This results in opaque data ownership.

While City Development creates policy, the City Works organization has a task to maintain the objects and infrastructure of the city. This includes not only trimming trees, maintaining garbage bins, or removing unwanted growth from bicycle paths, but also large infrastructure project planning and advising. Additionally, City Works also produces a variety of products, such as detailed maps of the city. These tasks require intensive use of data, either because data is needed to identify the location of objects for maintenance, or simply because it is a core aspect of the created products such as maps. City Works also supplies data to organizations inside the local council. For all these tasks City Works has traditionally been very reliant on data management.

The Library organization relates to City Works and City Development in the sense that data is essential for the internal functioning of the service. Data about books is collected to make them findable online or in a local catalog, but the library also keeps data about all their customers, to ensure efficient service delivery. This library data is either personally bound, or might be subjected to copyright law. For example, the metadata about books might contain descriptions supplied by publishers, unavailable for re-use. However, the Library is also in the midst of reorganization and in the future might not be part of the city council, at least not in its current form. This essentially means that the data gathered and used by the library might not be considered as PSI. Subsequently, the legal frameworks, as recommended on European and national levels might not be applicable. This creates uncertainty about the status of the data and to what extent it might be made available.

The City Archive organization is already an active participant in open data, albeit in a largely non-digital form. Currently, the archive consists of almost 20 km of archived material, stretching back 800 years. The legacy of, until recent years, always working with physical artifacts means that digitization is not traditionally part of the role of the archive. However, the archive has already digitized more than 150 TB worth of data. This data has not been imported into any centralized system for easy retrieval. There is thus a significant amount of available information, but a proportionally small amount is currently accessible through the online digital archive. The City Archive also possesses very little geographical information about cataloged objects, complicating its use in maps or other geographic applications.

The Centre for Research and Statistics organization uses data extensively in performing analysis, often as a project commissioned by other departments. As part of their tasks, the Centre for Research and Statistics thus gathers large amounts of data, sometimes requested from other departments, through statistical analysis or via questionnaires. Many datasets are available via a platform owned by the department. The center's data is not presented in a form that allows it to be re-used easily and the information is also published in an enriched format, i.e., the data is not downloadable in raw format.

Finally, the Police department gathers data extensively in order to maintain safety and order. Data from the regional Police is already used by journalists to present information about crime statistics, while the Police also supplies data to city council departments such as the Centre for Research and Statistics. This data is produced internally and is also highly related to persons.

As part of the project, several case studies in the form of application development were used to evaluate and test data release from the participating services. These cases were developed in collaboration with the local council, using the input and results from the interviews. In total, three successive rounds of projects were run, coinciding with the semester durations. Lessons learned from first iterations were accounted for in subsequent project. The aim of the cases and applications is not the application itself, but rather to test and experience the data release in practice and to illustrate the potential re-use of the data.

These applications were all developed as part of the project, with researchers assisting students in ideation and development, while public officials contributed by sharing domain knowledge about the data and informing students about details such as metadata. For certain education programs and projects, public officials also attended final project presentations. Additionally, local industry participated in the education program by offering assistance and guidance to students.

## 5. Observations

The data collected through the development of the case applications results in a number of observations. In the following two subsections we report on these observations. We have classified the observations in two mainstreams: observations that pertain to underlying thresholds of data release, and therefore indirectly determine the release of data, and observations that directly determine the release of data. From the latter observations, we derive a set of indicators that are relevant for the release of data. In Section 6.1, we outline the main underlying thresholds that play a role in data release. Section 6.2 is devoted to the observations and a set of indicators that directly determine the release of data.

### *5.1. Underlying thresholds*

#### *5.1.1. Fear of false conclusions*

A recurring theme during our workshops and discussions was the fear of false conclusions being drawn from the released data. An example of this would be the potential decrease in property value if details about policy plans of new city developments would surface, especially if these were not yet finalized. Interpretations of data by outsiders without prior knowledge about the goal of data gathering could also lead to unwanted situations. For example, research performed on the demographic profile of certain neighborhoods might not appear nuanced enough to an outsider who did not have knowledge about the statistical methods used. This applies especially to the Centre for Research and Statistics, where raw questionnaire data is analyzed before being presented. In the case that there are judicial frameworks in place to mitigate such issues, they might not be known to individuals dealing with a particular data request.

#### *5.1.2. Financial effects of PSI release*

Economically, licensing fees form a barrier to data release. The participating organizations such as City Works earn income through the sale of their data. While research has shown that the data recovery model results in proportionally low amounts of income and that less restrictive data release can in some cases lead to increases in revenue or, in the case of free data, lower transaction costs (de Vries et al., 2011), the services find themselves in a transitional phase. Even discarding a proportionally low amount of income from data licenses needs some sort of compensation, especially if an entire department currently depends on data sales. Decreased government spending aggravates the problem since it is not a given that lost income from releasing previously licensed data for free will be compensated by the national government or the council.

#### *5.1.3. Opaque ownership and unknown data locations*

There are also significant questions regarding ownership of data in our cases. Due to vertical data management in the past, data sharing between departments has not always been the case historically. As a result, services and, in extreme cases, departments within a service lack a complete picture about which data others control. Additionally, even if the location of data is known, its ownership remains opaque, inhibiting the data release. Merely identifying a dataset that would strictly speaking be suitable for release does not warrant its appropriateness as open data. For example, the Centre for Research and Statistics obtains data from other departments to conduct its research, and as such might not necessarily be the owner of the data.

Please cite this article as: Conradie, P., & Choenni, S., On the barriers for local government releasing open data, *Government Information Quarterly* (2014), http://dx.doi.org/10.1016/j.giq.2014.01.003

*5.1.4. Priority*

Finally, from our interviews and workshops, it is clear that open data is not necessarily a priority within the council and does not form part of existing tasks. This is not a consequence of unwillingness, but simply because data release is not part of the regular work for many data professionals. In combination with this, the immediate benefits of data release are not always explicit, making it an extra task, without clear incentives.

While the interviews conducted provide insights into the experience of data release by the data professionals in the city council, it neglects exploring the actual process of data release in an empirical form. In the following section, we elaborate not only on the released data, but also on the data that has easily been released. From these observations, we derive a set of indicators that determines the release of data.

*5.2. Indicators for data release*

For our educational program, the focus has been to release data that can be described as the low hanging fruit. This was for the most part non-personal data related to objects in the city, for reasons mentioned earlier. A majority of data relates to public property — in essence everything controlled by the municipality that can be observed in the public space. Given this, a major supplier has been City Works. The data released includes locations of objects like trees, hospitals, public art, or parks. However, release has been swift not only because of the types of data, but also due to the central data storage at City Works.

Data released by the City Works was not made available for re-use outside of our education program. The reason behind this is the lack of clear judicial rules with regard to data re-use on a wider scale. This effectively means that the data cannot strictly be referred to as open data: it may not be used without restrictions outside of our education program and is not available under any license.

Despite the ease with which data about objects was released, other types of data that relate to the internal processes of the council are much more sensitive. For example, data about the maintenance costs of objects has not been released, even within the education program. This would include data about how often trees are treated for various disease. Anecdotal evidence from our education program suggests that this type can form interesting input for applications. Also from the perspective of transparent government, this data seems valuable. However, in the short term, this is not forthcoming.

The lack of judicial frameworks and the ease of data release stand in stark contrast with the data from City Development. This department has made explicit decisions about data release, by classifying the data under Creative Commons Zero (CC0) license. Applying this license means that, unless otherwise indicated, the contents of data objects such as traffic information are available for any type of re-use. An exception to this might be the presence of copyrighted material inside a document. For example, a policy document could be supplied, unless it contains a photograph that does not fall under the CC0 license. However, it is prohibited to create the impression that the local council agrees with any conclusions that might arise from the data, when interpreted by external parties (Creative Commons, 2011).

Despite having seemingly unambiguous permission to release data to the public, the City Development department is often in the dark as to where data actually resides. Even if the data location is fully known, it does not warrant release. An illustration of this is a request for the locations of public transport data. While City Development controls this data, the data is owned on a metropolitan regional level. As a result, requests for the data's release needs to go to the metropolitan level, even though the data location is known. Due to its ownership being external, it is subsequently unclear whether the data can be classified under the CC0 license mentioned above. This illustrates that defining the judicial aspects of data release does not translate to release. Some practical or technical issues remain.

Due to the product focused work process present at City Development, the extreme decentralization is not conducive to an open data policy. Additionally, departments within City Development are controllers of datasets, but do not necessarily own them. This potentially creates not only uncertainty as to the applicability of the license, but also ambiguity as to who has the final word in releasing the data. Furthermore, while City Development has been very clear in taking care of the judicial side of open data, the data that is currently classified for re-use does not compromise all data. The real-time status of many parking garages in the city is such an example.

While City Development and City Works both deal for a big part with data in a digital format, the City Archive is historically much more focused on physical artifacts. Data release by City Archive has up until now been limited to the opening up of the search Application Programming Interface (API), in order to access the object metadata. There is thus no raw, unedited data about archived objects such as high-resolution images, or historical maps available in digital form, for reasons discussed earlier.

As mentioned before, the Library has more comprehensive digital service provision, due to the reliance on data to manage the collection and serve the public. However, the only set of information that is currently available for use contains book metadata. The use of this data currently falls into a gray zone: bibliographic information gets bought from a commercial organization, which has implications on its release as open data. For education purposes, however, re-use of the data would be allowed. The Library also poses other data that is without copyright restrictions, which comprises information about book lending, or statistics about members. This data could theoretically be personally identifiable, which complicates its use as open data. For both the Police and Library in cases where data is anonymized it might still be possible to extract personally identifiable data, through methods discussed by Kalidien et al. (2010).

A majority of the regional Police data is generated inside the department. As a result, issues such as copyright and licensing fees are not relevant. Permission to release data can also be granted within the department, easing data release. At the time of writing, however, the Dutch Police is merging to form a national organization, which might mean that the ease with which data can be released could change. Of most relevance to the release of Police data is the domain knowledge needed to understand the data. For example, the amount of violence associated with a street robbery might seem high to the larger public, but violence is in almost all cases a pre-condition for a street robbery to occur. This thus gives the impression of increased violence.

Finally, the Centre for Research and Statistics already provides large amounts of data through their own portal, even though the data is not in raw format. Some of this data is currently being provided for re-use, but in some cases, such as the Safety Index, the underlying data is sourced from external organizations, resulting in opaque ownership of data. Raw datasets might also contain individual responses to interviews and questionnaires, resulting in privacy risks. However, data is stored centrally by the service, which eases data release significantly.

From these observations, we note that the judicial issues, which include privacy and copyright concerns, are important when considering data release, but the way data is collected, how it is stored and how it is used are also important roles in data release. Therefore, we distinguish the following indicators that may be important for data release:

- Data storage, i.e., is data stored centrally, or decentralized?
- Use of data, i.e., the way data is used by the department.
- Source of data, i.e., how is a set of data obtained?
- Suitability of data for release, i.e., are there rules and regulations that determine whether a dataset may be released or not, such as privacy or copyright?

In Table 1, we summarize how the several departments score on each of the indicators together with the results of dependent variable “data released”. This overview of the release of data for use within the education program provides an initial view to the barriers of data release as experienced by the various participating partners. In the

Please cite this article as: Conradie, P., & Choenni, S., On the barriers for local government releasing open data, *Government Information Quarterly* (2014), http://dx.doi.org/10.1016/j.giq.2014.01.003

**Table 1**
Indicators for data release and core task.

| | City Development | City Works | Library | City Archive | Centre for Research and Statistics | Police |
|---|---|---|---|---|---|---|
| Core task | Long-term urban planning, execution of development policy | Maintenance of city infrastructure | Book and music lending, archiving of certain historical objects | Archiving documents, maintaining historical artifacts | Gather data for the development of reports | Maintain order and safety |
| Use of data | Execute development policy | Maintain public objects, producing geographical products | Support core services, internally and externally | Enhance services, catalog objects | In primary process, develop reports for other services | Support task of maintaining public safety |
| Source of data | Externally gathered | Internally produced | Internally produced through services processes and externally bought | Metadata internally created, artifacts externally supplied | Statistical analysis produced internally, source data gathered externally | Internally produced |
| Data storage | Decentralized | Central | Central | Central (metadata only) | Central | Mixed |
| Suitability of data for release | Medium: combination of personally identifiable information, live sensor data and object data | High: many sets related to public objects, non-personal | Low: data personally identifiable or subject to copyright | High: data not personally identifiable and owned by the archive | Medium: source data contains personal data, raw data open to interpretation and often complex | Low: data is mostly personally identifiable |
| Data released | Locations of public transport, state of bridges and traffic congestion | 200 data objects pertaining objects in the public space, 3d renderings of the city, web services of city services | Book metadata | Search API of archive metadata | Social index data, questionnaire data | Social index data, questionnaire data |

following section we will elaborate more on how these findings can impact open data policy on a local level.

## 6. Lessons learned

First and foremost, the act of publishing open data is new for local governments. This results in many questions by public servants about how data requests should be handled. Because a framework is lacking to answer these questions, data release stumbles. However, beyond the fact that data professionals lack clear guidelines allowing data release to external parties, there are some more fundamental issues concerned with data release.

In our case, due to the urgency of the education program, the most effort has also been on acquiring easy datasets, without prioritizing data according to demand. As a result, our inquiry reveals not only where data is most easily retrieved, but also, by virtue of not being easily accessible, what types of data need further exploration. This revealed that the easiest data to technically and practically release was non-personal object data such as trees or benches especially due to the centralized data storage of such data.

Below we will elaborate more in depth about the main lessons learned from the research.

### *6.1. Nuanced approach to data release*

While the municipality typically presents itself as one entity to the outside world, there is a much more complex organizational reality. Even though our focus had merely been on six services, the heterogeneous character of these services is immediately apparent. This ranges from the types of tasks conducted, such as servicing the public (Library) to developing policy (City Development). Additionally, while some of the services are involved with interacting with the public as a core task (City Archive), others focus more on maintaining objects in the public spaces. Our attempts at data release further emphasized this view. This complex picture effectively also means that these services deal with data in different ways. City Development, as an organization that is primarily concerned with developing and executing policy, uses data for the creation of a product (such as a policy recommendation). The data itself in this context merely supports a task. As a result, the data requested by the service remains unused, except for the times when it is being applied in policy development or creation of various products such as informational publications about the city.

These processes of data usage thus impact data release. Those organizations where data is used and stored very centrally have shown to be much quicker in being able to release data, as seen in the Library and City Works cases. This does not take into effect the suitability of the data for release. The converse is true for City Development, where data is important and the emphasis is on the products that result from the data. In this case the data storage is decentralized and ownership opaque, thus complicating data release.

The suitability of PSI for release as open data also varies. Based on two factors, privacy sensitivity and copyright restrictions, we identify that the internally produced data about public objects are the most suitable, while at the other end Library data, which is partly bought from third parties and pertains to persons, has a lower suitability.

Different barriers to data re-use apply. While only one service, City Development, has made explicit arrangements with regard to judicial matters, this service, as our interviews and data release show, is not in a position to release a lot of data under a CC0 license for re-use because of the way it treats data.

This too would be the case for data release by the City Archive. In the event that the archive has a legal framework about data release it still might not possess the types of the data that is easily released, due to physical limitations to what the archive can digitize or due to financial restrictions. Conversely, the library can release, and has released, data for re-use. However, this data is not widely reusable due to copyright,

and the data that is not subject to copyright law can be sensitive to privacy.

Given this, it seems that a nuanced view is needed for data release strategies on a local level, with all the implications this has for associated policy. In practice, this would mean that a focused policy would be needed, based on the data processes found inside the service. Prior to formulating policy it is thus important to look at how services deal with data storage by taking into account whether data is easily accessible, stored centrally, or available in digital formats and what the other financial, internal or judicial aspects might be.

### 6.2. Data release for its own sake

The difficulties of data release, either as a result of technical issues, uncertain judicial frameworks, or uncertain economic outcomes or as a result of simply not being able to find the data, mean that data release can still be a costly and labor intensive activity. Our experience up until now has been that simply requesting "all" data yields no results.

This means that choices are necessary to decide which data should be released. In reflecting back on the motivation behind PSI release, it is clear that the aim should not be merely the release of data for its own sake. Rather, the ambition of PSI release policy is to contribute to transparent government, innovation and increased public participation. This, however, is easier said than done. The current state of affairs means that to ensure efficient use of released data, it is important to consider the return on investment, not merely in a financial sense, but in the context of the highest re-use achieved by datasets when compared to effort needed to realize its release. A more focused approach to data release has the potential to mitigate some of the problems identified earlier. For example, if finding the location of data is currently prohibiting the release, a demand-based approach to data release might be applicable. To facilitate the release of data based on demand, it remains necessary to inform potential end users of the types of data currently controlled by a service or department. In essence, this would imply a better cost and benefit analysis of data release.

Additionally a stronger focus can be achieved without big investments into infrastructure: by exploring data release on a small scale, data is released with clearer end-goals, which also explicitly demonstrates the benefits of PSI. This could be facilitated by small pilots, with easily publishable data with a low cost associated with opening data.

### 6.3. Large scale implementation

In relation to the heterogeneity of the separate public sector organizations and in combination with the transaction costs associated with data release we identify that the large-scale implementation of open data infrastructure is likely to encounter significant difficulties. First and foremost, such an implementation would be troublesome to achieve due to the inherent heterogeneity of the services as discussed earlier. For example, significant efforts would first need to take place to ensure efficient data release and due to the differences in place at the various services this process could be lengthy and costly. In our context, merely clearing the judicial hurdles for the participating services in cases where data is deemed usable could slow down development significantly. Even more illustrative of this is the change needed in internal processes, as would be the case in organizations with very fragmented data storage.

Furthermore, a large-scale implementation implicitly requires heterogeneous data sources to be made available. As discussed above, even where there are judicial frameworks in place and data release is technically feasible, there are still transaction costs related to the release of data, such as locating datasets, or digitization. The resources needed to build a large-scale infrastructure that could facilitate data release might not be justified in relation to the uncertainty of the actual re-use of the data that is finally released. This holds especially true in a time when governments on local and national scales in industrialized countries are seeking to reduce expenditures.

This sketches the conclusions of the research. In the final section we will reflect on how these barriers might be addressed in our own work and some suggestions for further research.

## 7. Conclusion and outlook

In this article, we explored data release on a local level. For this purpose, we conducted interviews, questionnaires, and organized workshops and performed desk research. Through this we identified underlying thresholds that inhibit data release, summarized in Section 5.2. Most importantly, we identified indicators for data release, reported in Table 1, based on our observations. These include the use of data within a service, where data is stored, the source of the data, and the suitability of data for release. Based on these observations, we recommend a few practical steps to realize open data on a local level.

Firstly, identify data owners and producers of non-personally identifiable data. These organizations offer data that might not encounter significant barriers to release such as privacy, opaque ownership or judicial issues. These might be information that is readily observable, such as the location of objects in the public space.

Secondly, the costs still associated with data release in terms of locating data and, in some cases, getting permission to publish sets might prohibit the release of large amounts of data. Given this, it could be beneficial to ensure that data requests can be made, giving potential re-users an overview of what might be available in cases where data publication is significantly hindered by costs associated with finding the data.

This aim of this is to contribute to a better understanding of PSI release on a local level. If the benefits of open data, such as increased transparency, smarter government and innovation, are to be realized, an examination of open data release on a local level could be conducive to reach this aim.

Further research should be conducted to study the effects of these efforts and their potential for mitigating data release barriers. While the six examples in our research suggest a dependency between these indicators more studies could show whether this is indeed the case. A better understanding of these processes could result in more focused efforts to release data, by starting with organizations where barriers are the lowest.

## Acknowledgments

We thank our colleagues from the Rotterdam University, both researchers of the research lab Creating 010, the teachers and students of the School of Communication, Media and Information Technology involved in the Rotterdam Open Data projects and the participants from the council of Rotterdam. Special thanks to Judith Lemmens and Nathalie Stembert. The work has been partly funded by the Rotterdam Media Commission (grant number 2011.006.SK) and RAAK Publiek (grant number 2011-13-16P).

## References

Bertot, J. C., & Choi, H. (2013). *Big data and e-government: Issues, policies, and recommendations, 1–10.*

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: e-Government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly, 27*, 264–271.

Bertot, J. C., McDermott, P., & Smith, T. (2012). Measurement of open government: Metrics and process. *2012 45th Hawaii International Conference on System Sciences* (pp. 2491–2499).

Boulton, G., Rawlins, M., Vallance, P., & Walport, M. (2011). Science as a public enterprise: The case for open data. *The Lancet, 377*, 1633–1635.

Choenni, S., van Dijk, J., & Leeuw, F. (2010). Preserving privacy whilst integrating data: Applied to criminal justice. *Information Polity, 15*, 125–138.

Creative Commons (2011). Creative Commons — CC0 1.0 Universal. Retrieved April 10, 2012, from. http://creativecommons.org/publicdomain/zero/1.0/

Please cite this article as: Conradie, P., & Choenni, S., On the barriers for local government releasing open data, *Government Information Quarterly* (2014), http://dx.doi.org/10.1016/j.giq.2014.01.003

Dawes, S. S. (2010a). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly*, *27*, 377–383.
Dawes, S. S. (2010b). Information policy meta-principles: Stewardship and usefulness. *2010 43rd Hawaii International Conference on System Sciences* (pp. 1–10). : IEEE.
de Vries, M., Kapff, I., Achiaga, M. N., Wauters, A., Osimo, D., Foley, P., et al. (2011, October). *POPSIS: Pricing Of Public Sector Information Study*.
Geiger, C. P., & Von Lucke, J. (2012). Open government and (linked) (open) (government) (data). *JeDEM — eJournal of eDemocracy and Open Government*, *4*. (pp. 265–278).
Graves, A. (2011). A case study for integrating public safety data using semantic technologies. *Information Polity*, *16*, 261–275.
Hughes, J. (2003). Commentary: Participatory action research leads to sustainable school and community improvement. *School Psychology Review*, *32*, 38–43.
Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, *27*, 371–376.
Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, *28*, 446–456.
Kalidien, S., Choenni, S., & Meijer, R. (2010). Crime statistics online: Potentials and challenges. *Proceedings of the 11th Annual International Conference on Digital Government Research* (pp. 131–137).
Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*. http://dx.doi.org/10.1016/j.giq.2013.05.012.
Kemmis, S., & McTaggart, R. (2000). Participatory action research. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 567–607). CA: Thousand Oaks.
Landsbergen, D. (2010). Government as part of the revolution: Using social media to achieve public goals. *Electronic Journal of e-Government*, *8*, 135–147.
Matheus, R., & Vaz, J. C. (2012). New perspectives for electronic government in Brazil: The adoption of open government data in national and subnational governments of Brazil. *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance* (pp. 22–29).
McLaren, R., & Waters, R. (2011). Governing location information in the UK. *The Cartographic Journal*, *48*, 7.
Prieto, L. M., & Rodríguez, A.C. (2012). Implementation framework for open data in Colombia. *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance* (pp. 14–17).
Robinson, D., Yu, H., Zeller, W., & Felten, E. (2009). Government data and the invisible hand. *Yale Journal of Law & Technology*, *11*, 160 (160, 160–175).
van Dijk, J., Kalidien, S., & Choenni, S. (2013). Development, implementation and use of a judicial data space system. *ICEGOV 2013, Int. Conference on e-Government*. Seoul: ACM.
Whyte, W. F. (1989). Advancing scientific knowledge through participatory action research. *Sociological Forum*, *4*, 367–385.
Wilson, S.C., & Linders, D. (2011). The open government directive: A preliminary assessment. *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times* (pp. 262–271).
Yang, Z., & Kankanhalli, A. (2013). Innovation in government services: The case of open data. *IFIP International Federation for Information Processing 2013* (pp. 644–651).
Zhang, J., Dawes, S. S., & Sarkis, J. (2005). Exploring stakeholders' expectations of the benefits and barriers of e-government knowledge sharing. *Journal of Enterprise Information Management*, *18*, 548–567.
Zuiderwijk, A., Janssen, M., & Choenni, S. (2012). Open data policies: Impediments and challenges. *12th European Conf. on e-Government, June 14–15* (pp. 794–801). Barcelona, Spain: Academic Publishing Limited, UK (2011).
Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Sheikh Alibaks, R. (2013). Socio-technical impediments of open data. *Electronic Journal of e-Government*, *10*, 156–172.
Zuiderwijk, A., Janssen, M., Meijer, R., Choenni, R., Charalabidis, Y., & Jeffery, K. (2012). Issues and guiding principles for opening governmental judicial research data. *Proc. EGOV 2012, 11th European Conf. on Electronic Government, September 3–6, LNCS*. Kristiansand, Norway: Springer-Verlag, German.
Zuiderwijk, A., Jeffery, K., & Janssen, M. (2012). The potential of metadata for linked open data and its value for users and publishers. *JeDEM - eJournal of eDemocracy and Open Government*, *4*, 222–244.

**Peter Conradie** was a researcher at Creating 010, an interdisciplinary research group of the Rotterdam University of Applied Science. His research interests include value of open data and the co-creation of public services and e-government. He currently works at the Department of Industrial System and Product Design, Ghent University.

**Dr. Sunil Choenni** is the head of the Research & Documentation Centre for the Ministry of Justice in The Hague and a research professor at Creating 010, an interdisciplinary research group of the Rotterdam University of Applied Science. His research interests include privacy, database technology and e-government.

Please cite this article as: Conradie, P., & Choenni, S., On the barriers for local government releasing open data, *Government Information Quarterly* (2014), http://dx.doi.org/10.1016/j.giq.2014.01.003