Log File Analytics for Gaining Insight into Actual Use of Open Data

van Loenen, Bastiaan; Ubacht, Jolien; Labots, Wouter; Zuiderwijk-van Eijk, Anneke

Publication date
2017

Document Version
Final published version

Published in
Proceedings of the 17th European Conference on Digital Government

Citation (APA)
van Loenen, B., Ubacht, J., Labots, W., & Zuiderwijk-van Eijk, A. (2017). Log File Analytics for Gaining Insight into Actual Use of Open Data. In V. Borges, & J. C. Dias Rouco (Eds.), *Proceedings of the 17th European Conference on Digital Government* (pp. 238-246). Academic Conferences.

ACADEMIAMILITAR
MILITARYACADEMY

Proceedings of the 17th European Conference on Digital Government

Military Academy
Lisbon, Portugal
12-13 June 2017

Edited by
Major-General Vieira Borges
and
Lieutenant-Colonel José Carlos Dias Rouco
Military Academy, Lisbon, Portugal

acpi

A conference managed by ACPI, UK

The Proceedings of
17th European Conference on
Digital Government
ECDG 2017

Hosted By
Military Academy
Lisbon, Portugal

12th-13th June 2017

Edited By
MGen João Vieira Borges Lt
Col José Carlos Dias Rouco

Review Process
Papers submitted to this conference have been double-blind peer reviewed before final acceptance to the conference. Initially, abstracts were reviewed for relevance and accessibility and successful authors were invited to submit full papers. Many thanks to the reviewers who helped ensure the quality of all the submissions.

Ethics and Publication Malpractice Policy
ACPIL adheres to a strict ethics and publication malpractice policy for all publications – details of which can be found here:
http://www.academic-conferences.org/policies/ethics-policy-for-publishing-in-the-conference-proceedings-of-academic-conferences-and-publishing-international-limited/

Conference Proceedings
The Conference Proceedings is a book published with an ISBN and ISSN. The proceedings have been submitted to a number of accreditation, citation and indexing bodies including Thomson ISI Web of Science and Elsevier Scopus.

Author affiliation details in these proceedings have been reproduced as supplied by the authors themselves.

The Electronic version of the Conference Proceedings is available to download from DROPBOX http://tinyurl.com/ecdg2017 Select Download and then Direct Download to access the Pdf file. Free download is available for conference participants for a period of 2 weeks after the conference.
The Conference Proceedings for this year and previous years can be purchased from http://academic-bookshop.com

E-Book ISBN: 978-1-911218-37-1
E-Book ISSN: 2049-1034
Book version ISBN: 978-1-911218-38-8
Book Version ISSN: 2049-1026

Published by Academic Conferences and Publishing International Limited
Reading
UK
44-118-972-4148
www.academic-conferences.org

Contents

Preface

These proceedings represent the work of researchers participating in the 17th European Conference on Digital Government (ECDG 2017) which is being hosted this year by the Military Academy, Lisbon, Portugal on 12-13 June 2017.

ECDG is a recognised event on the international research conferences calendar and provides a valuable platform for individuals to present their research findings, display their work in progress and discuss conceptual and empirical advances in the areas of Digital Government. It provides an important opportunity for researchers and practitioners to come together to share their experiences of researching in this varied and expanding field.

The conference this year will be opened with a keynote presentation by a representative from the Tax and Customs National Authority (AT), Portugal on Digital Government and Tax Challenges. The second day of the conference will be opened by a talk by Colonel Lemos Pires who will address the topic Digital Commanders in the Age of Acceleration.

With an initial submission of 75 abstracts, after the double blind, peer-review process there are 33 academic Research papers, 5 PhD Research, 1 Non-Academic paper and 2 Work in Progress papers published in these Conference Proceedings. These papers represent truly global research in the field, with contributions from Australia, China, Colombia, Denmark, Estonia, Germany, Greece, Ireland, Isreal, Italy, Jamacia, The Netherlands, Nigeria, Pakistan, Poland, Portugal, Russia, Slovakia, South Africa, Spain, Switzerland, Thailand, Turkey, United Arab Emirates, United Kingdom and United States of America.

We wish you a most interesting conference.

MGen João Vieira Borges and Lt Col José Carlos Dias Rouco ECDG Conference and Programme Chairs
Military Academy, Lisbon
Portugal

Log File Analytics for Gaining Insight into Actual Use of Open Data

Bastiaan van Loenen[1], Jolien Ubacht[2], Wouter Labots[2] and Anneke Zuiderwijk[2]
[1]Delft University of Technology, Delft, Faculty of Architecture and the Built Environment, Knowledge Center Open Data, the Netherlands
[2]Delft University of Technology, Delft, Faculty of Technology, Policy and Management, Delft, the Netherlands
b.vanloenen@tudelft.nl
j.ubacht@tudelft.nl
wouter@labots.com
a.m.g.zuiderwijk-vaneijk@tudelft.nl

Abstract: Following open data policies worldwide, an increasing number of public organisations has now published open data that is free to be used by anyone. However, despite the significant increase in use of this open data, the open data providers are mostly not aware of their users and the way in which the data is actually re-used. This is rooted in the principle that open data should be free to use without prior user registration in order to avoid any unnecessary barriers for reuse of the data. However, understanding use and user needs of open data is important to improve the provision of open data and the successful implementation of open data policies. We explored the use of log files of the actual use of open data to identify the users and to explore how the open data is being used. By means of a case study in which we apply log file analytics to the Dutch open geographical data portal we show that this approach is promising for analysing open data use. This approach will yield many new insights for open data providers to improve and fine-tune their open data offer and policy makers will be provided with data on actual use to evaluate their open data policies. Our analysis shows that citizens are a much heavier user of open data than currently assumed: citizens as major users of open government data should be taken much more seriously in the demand driven open data policies. We recommend transferring the pilot project into a permanent monitoring instrument for open data use and exploring additional analytics for using the rich data that log files provide for.

Keywords: Open data – open data use – data analytics – geographical data portal – log file analytics – evaluation of open data

1. Introduction

An increasing number of public organisations worldwide publish open data (Chatfield & Reddick, 2017; Sieber & Johnson, 2015). *Open data* is data available for reuse without any costs and without any restrictions (Open Knowledge Foundation, 2015; Gurin, 2014; Geiger and Von Lucke, 2012). Open data efforts may be supply-driven, which means that there is a unidirectional provisioning of data from governments to end users through a data portal or platform (the so-called *data over the wall model*), or there may be a more active, participatory or responsive level of government involvement with open data end-users (Sieber & Johnson, 2015). In practice, open data efforts are often supply-driven (Sieber & Johnson, 2015; Susha, 2015; Zuiderwijk, 2015). As a consequence, open data providers are mostly not aware of the use and the user of their data (Susha, 2015; Zuiderwijk, 2015).

Having insight in actual open data use allows public organizations and data infrastructure providers to support what Gurstein (2011) refers to as 'effective use'. This means that data providing agencies can take into account the requirements of users, including technical and professional requirements for data interpretation, the language in which data is presented, and the availability of training in data use and visualization (Gurstein, 2011). Moreover, understanding user needs of open data is important for improving the provision of open data and for the successful implementation of open data strategies (Welle Donker and Van Loenen, 2016). When lacking contact with their users, open data providers remain uncertain about the need for a certain dataset, about possible ways to improve the provision of their open data, and how to get into contact with the users to gather feedback (Zuiderwijk, 2015). Therefore, open data providers need to find out who their open data users are (Olausson, 2016).

In order to generate societal value with open data, policy-makers are increasingly aware that the publication process of open data should be demand-driven or even problem-driven rather than supply-driven (Susha et al., 2015). Public agencies should take on a more proactive role, beyond the 'data over the wall-model' (Sieber &

Johnson, 2015). This is important since demand-driven open data provision is expected to result in more benefit and value creation (Susha et al., 2015; Jetzek, 2015; Janssen et al., 2012).

Some research on open data users has already been conducted. For instance, Dawes, Vidiasova and Parkhimovich (2016) state that open data users are usually not ordinary citizens, but they are technologically skilled data analysts or application developers. Open data users are mainly those people who already have access and are already 'empowered' (Gurstein, 2011). Nevertheless, beyond those few studies on open data use, there is very limited insight in how open data is actually used in practice (cf. Van Beuningen et al. 2016; Bregt et al., 2016). In addition, research on how governmental agencies can obtain insight in how their open data is actually used is scant (see Ruijer et al., 2017). Whereas several approaches have been proposed (e.g. Zuiderwijk, 2015), these approaches require relatively much effort and time from the data user and are quite demanding.

This study aims to gain insight in how open data is actually used by analysing log files. The study contributes to existing research by showing how a log file analytics approach can be used to obtain more insight in open data use. It is a means to assess the effects of open data with minimum pressure on the re-user and data provider, but with maximum output.

2. Research approach

In this study we examine open data use by using a single case study approach (see Yin, 2003). The case study selection criteria were as follows:

- The case provides log data that allows for analyzing open data use.
- The case employs open geographical data. This focus is justified by the, compared to other domains, relative maturity of the open geographical data domain (see Van Loenen and Grothe, 2014). Further, geographic data such as topographical maps and the underlying earth observation data, are top-listed by the European Commission and the G8 for release as open government data due to the high demand from re-users (Cabinet Office, 2013; European Commission, 2014).
- The case represents open data utilization in the Netherlands.
- Case study information should be available and accessible.

We selected the case of the PDOK portal operated by the Netherlands Land Registry and Mapping Agency (the Cadastre). The reason for performing a single case study concerns the unique circumstances that the case represents, namely the availability of log data that allowed for analysing open data use. This data can be difficult to obtain by researchers, since the log files are personal data (CJEU, 2011; CJEU, 2016). The literature, reports as well as discussions and open interviews with civil servants working at the Cadastre were used as information sources for our case study.

3. Understanding open data use(rs)

3.1 Case description: Open Geographical Data in the Netherlands

Much Dutch open data is geographic data provided through the generic national open data portal and through the so called PDOK-portal, which is a dedicated portal for geographic data (www.pdok.nl; PDOK stands for *Publieke Dienstverlening op de Kaart*: Public Services on the Map) (Van Loenen & Grothe, 2014). Although there is a growing interest in the PDOK open data sets (with 280 open datasets in 2016 and 4.4 billion hits on open data services) (see Table 1), the Cadastre lacks insight into the actual users behind these numbers. Therefore, the Cadastre is looking for means to collect information about how their open data services are being used and how these may be improved to better accommodate user needs.

Table 1: Key PDOK performance indicators 2012-2016 (PDOK 2012; 2014; 2015; 2016)

Year	2012	2013	2014	2015	2016
#datasets	41	64	78	91	280
#hits on services	Not available	580 million	1.1 billion	1.7 billion	4.4 billion

3.2 Case study findings: Traditional ways of gaining insight in open data use

The literature, reports as well as discussions and open interviews with civil servants working at the Cadastre showed that there are several traditional ways of gaining insight in open data use:

- *Mandatory user registration.* To bridge the gap between provider and user of open data, some organisations require users to register for access to the open data (see for example the Danish Mapping Agency KMS and the UK Consumer Data Research Centre data portal). It provides them with some information on the use(r) and potentially enables them to ask for feedback and input. However, open data advocates that would prefer to stick to the open data principles argue that registration of users is not compliant with the core principles of open data, which state that access to the data must be non-discriminatory (see for example Stott, 2014).

- *Voluntary registration of the user.* The second approach is identical to the first approach, with the difference that the registration is voluntary. In the Netherlands this approach was used for the topographic dataset of the Cadastre (see Bregt et al., 2013). When the user clicked on the topographical data for download, a screen pops up asking the user to fill out a survey about the user and to register as reuser. This approach, especially with the use of the survey instrument, is time consuming for users and not giving comprehensive and complete, and therefore potentially biased, information to the data provider.

- *Social media channels.* Approach three uses social media as a way to link data providers and users. Users and sometimes also data providers, can start social media groups around a specific dataset. For example, the topographic dataset in the Netherlands has a Linkedin group. The data provider is following the events in the group and responds to specific issues that are raised The disadvantage of is that new user groups are not identified and that only specific datasets are addressed, so an overall view on other datasets is missing.

- *Establishing a user group.* The fourth approach gains insights from a frequently meeting user group Establishing a user group attracts well-informed users but leaves out on new user groups (e.g. start-ups), since these are unknown, not visible or not organised in a formal manner such as through associations and the like.

- *Additional service provision.* A data provider may be in contact with data users through the provision of additional open data services (e.g. a service notifying users of new dataset updates, a data quality feedback service or a newsfeed service). The disadvantage is that only a selection of all data users is reached.

- *Organising (ad hoc) events.* Data providers may organise hackatons, data rally's and the like to explore the opportunities of open data and to become acquainted with the users of the data (e.g. https://www.opendata-award.nl/). The impact of these ad hoc events on the improved communication between data provider and user is limited, because of the ad hoc character and the limitation in the kinds of users they attract.

In summary, the approaches discussed towards identifying users and involving them in a strategy to improve datasets have their advantages for specific datasets, drawing on the expertise of serious user groups. But they are limited by the tension between required user registration and the principle of free access to the data and by the user group pre-selection that ignores (potential) new users. We therefore explored an alternative approach that both requires a minimum effort of users and is sufficiently informative for open data providers to evaluate or review their open data operations and strategies. In the next section we present our research steps towards developing a new approach linked to user log data.

3.3 An alternative approach: log file analytics

Our case study provided user log data, containing information about the actual use of the web services on the open data platform. We used a five-step approach to analyse the log data and analyse its potential for gaining insight in data use and users.

First, the Cadastre provided us with the log data of the open datasets provided through PDOK. The log data concerns PDOK use over the period of one year from October 1st 2012 to October 7th 2013, containing more than 3 million (3,042,895) log lines, which equal 515,325,523 hits. Each log line consists of 8 attributes: an IP address, a referrer, the service requested, the method used for requesting the service, the date of the request,

the week number of the request, the total number of requests and the number of errors identified. **Table 2** shows an example of a log line, which can be read as follows. On July 27th 2013 (Column 5) in week 31 (Column 6) user 213.10.x.y (Column 1) requested the service "brtachtergrondkaart" (Column 3) with the wmts method (Column 4). The user did this through the website www.zwemwater.nl. And he used this service 16 times (Column 7) (e.g., zooming in and out on a map requires several requests for the same service). No errors were reported (Column 8).

Table 2: Example of a log line

1	2	3	4	5	6	7	8
213.10.x.y	http://www.zwemwater.nl/	brtachtergrondkaart	wmts	2013-07-29	31	16	0

Second, we adapted the log files to fit the database management system in which the use and user data were structured and logically stored. Bringing log data into a database enabled us to analyse the log data. We used open source MySQL as the Database Management System. Before we started the analysis, we had to clean the data from errors. The data had, for example, several particularities. One was the lack of a public IP-address. Every user on the internet has a public IP address. In the database, however, we found multiple IP addresses in one log line, IP addresses without a public IP address, and other errors. We cleaned the database by selecting only log lines with (at least) one public IP-address. 8% (253,213) of the log lines did not meet this criterion and was removed from the database. The remaining 2,789,682 log lines were processed. They were used by 310,348 unique IP-addresses.

Third, we analysed the data, including the use of the data and the user based on the IP-addresses and the combined data to find the relation between the user categories and the types of web services and open datasets they used. The following section presents the findings from this analysis.

4. Analysis results

4.1 Frequency of open data use: regular and incidental users

Our first analysis focused on the frequency of the use. Remarkably, 77% of the IP-addresses (representing 238,968 IP-addresses) used PDOK for only one day in the entire year of our assessment period. 11% visited PDOK two days and 4% used PDOK three days in total. 8% Of the IP-addresses used PDOK four days or longer and only 4.3% used PDOK seven days or more over the one year period.

Table 3: Use of PDOK in days per IP-address (October 1st 2012-October 7th 2013)

Use	Unique IP-addresses	
1 day	239363	**77%**
2 days	35533	**11%**
3 days	11381	**4%**
4 days	5364	**2%**
5 days	3170	**1%**
6 days	2123	**0.7%**
≥ 7 days	13414	**4.3%**
	310348	**100%**

Based on these numbers we classified the users that used PDOK for seven days or more as regular users (13,414 unique IP-addresses, 4,3% of the total number of unique users, see **Table 3**). The remaining 96% were classified as incidental users. We continued our analytics with the class of the regular users as these can be seen as the most interesting users for feedback on the use of the open data.

4.2 Regular data users: from IP-address to the user

After identifying the use of PDOK, we proceeded to establish a link between the use and the regular users of PDOK by means of assigning the IP-address to a user category. We used three ways to find the user behind the IP-address:

- Using the IP-address of users known by the Cadastre, this revealed 349 users that already had accounts before the datasets were opened;
- Inquiring the host name linked to an IP-address. The domain name system (DNS) was used to link host names and IP addresses. For example, the IP address 145.94.165.135 has the host name 145-94-165-

135.wlan.tudelft.nl. This indicates that the user is connected to Delft University of Technology (tudelft.nl);

- Searching the network name of the IP-address. This information is public. Every IP-address is part of a group of IP-addresses, a network. The European registry, RIPE NCC (Réseaux IP Européens Network Coordination Centre) provides a publicly available database which allowed us to link IP-addresses to a network.

Based on the knowledge to which organisation the IP-addresses belong, we categorised the user groups. In **Table 4** a full overview of the main and sub categories of users is presented, as well as the percentages they each represent in the total of regular users.

Table 4: Overview of number of regular users per category and the percentages of representation in the total number of regular users.

Main category	Sub category	Number	Percentage
Government	municipality	383	2.9%
	environmental service	10	0.1%
	ministry	52	0.4%
	province	28	0.2%
	Water board	33	0.2%
Public	agentschap.nl	1	0.0%
	taxservice	8	0.1%
	duo	1	0.0%
	geonovum	1	0.0%
	cadastre	9	0.1%
	logius	1	0.0%
	Area service	2	0.0%
	Raad van state	2	0.0%
	rijkswaterstaat	11	0.1%
	rvob	1	0.0%
	staatsbosbeheer	1	0.0%
	uwv	7	0.1%
Public order and safety	Fire department	10	0.1%
	police	4	0.0%
	Safety service	6	0.0%
Education	university	65	0.4%
Knowledge	research institution	128	1.0%
Private	advocacy notarial	27	0.2%
	architect	18	0.1%
	consultancy	72	0.5%
	geo	9	0.1%
	grondwegwater	23	0.2%
	engineering	29	0.2%
	it-services	38	0.3%
	other	11	0.1%
	real estate	177	1.3%
	lessor	11	0.1%
	insurance	45	0.3%
Utility service	utility service	2	0.0%
	water company	8	0.1%
ISP	isp-foreign	22	0.2%
	isp-consumer	6425	47.9%
	isp-mobile	33	0.2%
	isp-business	886	6.6%
Search engine		222	1.7%
Total categorised		8822	65.8%
Not categorised		4592	34.2%
Total		**13414**	**100%**

As presented in **Table 4**, 34% of the IP addresses, representing 4,592 users were not categorised because of the manual attribution process that had to go into determine the main and subcategory required significant resources in effort and time. If it were possible to link the name of the network to the database of the Chamber of Commerce, then this categorisation can be established too. However, the Dutch Chamber of Commerce does not provide their data as open data, which made the full categorisation of our user data too laborious (see Van Loenen et al. 2016). Table 4 shows that citizens take a major portion of the total amount of users (isp-consumers: 6,425 of the 13,414, app. 48%). This percentage deviates from the percentages of user groups found in other studies. For example, comparing our research results to the results of a survey among users presented in Bregt et al. (2013) and Bregt et al. (2014) over a similar period (2012-2013) for *one* dataset (topography) provided through PDOK, shows significant differences in the distribution of the use and user groups (see Table 5).

Table 5: Comparing the results of two different research strategies

User category	% of unique IP-addresses of regular users (1 Oct. 2012 -8 Oct 2013) of PDOK	Bregt et al. (2013) survey with 56 responses	Bregt et al. (2014): survey with 148 responses
Public sector users	4%	34%	37%
Businesses	10%	41%	41%
Citizens	48%	5%	22%
Research	1%		
Unknown	34%		

4.3 Linking users to use: types of data use

The attribution of unique IP-addresses to a user group enabled us to link the categories to the actual use. As a last analytical step, we explored the extent to which the user categories used the PDOK data sets. In **Table 6** the percentages per use category are presented. The three top users of the PDOK data sets can be found in the private sector (24,57%), the individual users/citizens (18,57%) and governmental organisations (13,91%).

We see significant differences between the percentage unique users and their part in the usage of the PDOK services. For example, citizens are 48% of the unique users, but only are responsible for 19% of the actual use of PDOK.

Table 6: Overview of users in percentage per use category

Main category	Use
Private	24.57%
ISP	22.65%
of which isp-consumer	*18.57%*
of which isp-business	*4.04%*
Government	13.91%
Public	5.08%
of which Cadastre	*2.70%*
Public order and safety	1.53%
Utility service	1.24%
Education	0.68%
Knowledge	0.52%
Search engine	0.01%
Unknown	29.81%
	100%

5. Discussion

Our log data analytics approach shows the possibility of retrieving detailed user categories and the distribution of use over these categories. As the data were extracted from the actual log files, we consider them as a reliable source for data analysis. The findings are input to the Cadastre for evaluating their open data policy, based on actual use of their data. The analysis offers the open data provider the opportunity to address new user (groups), e.g. by setting up communication channels with these new users.

Through the IP-address not only more detailed information about the use category may be obtained, it also allows the provider to explore the websites of all identified users to find information on the applications that

are developed on the basis of the open datasets. An exploration of the users' websites in our research identified several new applications that the Cadastre was not familiar with.

In this paper we only showed a fraction of the information that is embedded in the log data. Many other overviews can be developed such as for most popular services, use of individual datasets per user category, frequency of use per month/ period of the year, typical steps of new users (starting with service A for dataset AA, then service A for dataset B etc.), development of services throughout the years, use of combination of services, and use of combination of datasets, among many other applications. This data will not, or much more limitedly, be available through other monitoring mechanisms.

In comparison to traditional approaches, log data analytics is an unobtrusive way of gaining more insight into actual use, does not require user registration that is at odds with the core principle that access to the data must be non-discriminatory and it takes all possible user groups into account. As such the log data analytics approach has added value to the other approaches for discovering new user groups, for completeness over all datasets and for being a continuous source of data.

The research presented should be regarded as research in progress resulting in several limitations. First, the log file analysis does not reveal so-called downstream users as the IP-address data relates to viewing not to downloading data. Users downloading the PDOK data may be typical infomediaries distributing the open datasets directly to their users. However, the log file analysis of the open data source does not reveal these downstream users, neither does it show the downstream value chain that builds upon the PDOK data.

Secondly, 75% of the log files included a referrer that can shed a light on users that integrate PDOK-services into their website or application for their end-users. Analysis of the referrers can provide insight into the services that have been developed based on the PDOK services.

Finally, privacy or data protection issues may arise. Although the new method described is appealing as a means to monitor the use of open data, it may not be utilised everywhere. For example, the European Court of Justice has ruled that dynamic and static IP-addresses should be considered as personal data, which may limit the use of these data for monitoring purposes (CJEUE, 2011 & 2016). In our research we took appropriate measures to avoid any data protection issues. For example, we used the PDOK IP-addresses only at a general, not highly detailed level: it was not possible to identify individuals. However, the value of the log files for supporting open data decision-making would increase if it were possible to link the log files to individual users. This may allow us, for example, to categorise the 34% of the users that we currently had to categorise as 'unknown' (see Table 4).

6. Conclusion

We presented and applied a new approach to gain insight into the actual use of open data by performing log data analytics. We showed that this approach offers the possibility to identify users of open data and the use frequency amongst the user groups. Application to the open geographical data portal of the Netherlands shows that this approach yields new types of information without the drawbacks of other approaches that aim at gaining insight into open data use and users. We conclude that log data analytics is not only a feasible approach, but also a cost effective one, especially when the process is automated (cf. Atz et al., 2015), offering many additional analysis that can be employed to monitor the use(r) of public open data services.

We found that citizens take a major portion of the total amount of use(rs). These findings are contrary to the findings of Dawes, Vidiasova and Parkhimovich (2016) stating that open data users are usually not ordinary citizens, but technologically skilled data analysts or application developers. Although some part of the citizens in our research may be 'empowered' (Gurstein, 2011), it might very well be that a fair amount of the frequent citizen users are ordinary citizens with no special technical skills. This may imply that the open data providers using demand driven approaches directed to professional users such as data analyst and application developers ignore the needs of an entire user group using open data. Accommodating these needs is novel direction that is barely addressed yet. It calls for research on the citizen as user of open data and the value citizens generate by using open data.

In addition to a call for further research on the citizens as an open data user, there is also a future research challenge in the possibility of automated log data analytics, for example an automatic link of the IP-addresses to data from the Chamber of Commerce (company name and web address) to improve the categorisation of private companies. Such an automated way of analysing the use of open data, may allow for the real-time monitoring of the use of open data, with direct input to decision-makers on the effect of new policy measures, potential changes in popular datasets and user preferences, all specified by user type and user groups. This would be one of the major prerequisites of successfully implementing truly demand driven open data policies: real-time and ubiquitous policies incorporating 24/7 the direct needs of all reusers.

Acknowledgements

The authors would like to thank the Netherlands Land Registry and Mapping Agency for kindly providing the log data of the Publieke Dienstverlening op de Kaart (PDOK). More information concerning the data processing and analysis can be obtained from the authors upon request.

References

Atz., U., T. Heath, and J. Fawcett (2015) Benchmarking open data automatically. Open data institute.

Bregt, A.K., Ł. Grus, and D. Eertink (2014) Wat zijn de effecten van een open basisregistratie topografie na twee jaar? Report commenced by the Dutch Kadaster.

Bregt, A.K., W. Castelein, Ł. Grus, and D. Eertink (2013) De effecten van een open basisregistratie topografie (BRT). Report commenced by the Dutch Kadaster.

Chatfield, A.T., and Reddick, C.G. (2017) 'A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments', *Government Information Quarterly*, http://dx.doi.org/10.1016/j.giq.2017.02.004

Bregt, A.K., Grus, L., van Beuningen, T., and H. van Meijeren (2016) Wat zijn de effecten van een open Actueel Hoogtebestand Nederland (AHN)? Rapport Wageningen University & Research

Cabinet Office (2013) Policy paper: G8 open data charter and technical annex. 18 June 2013.

CJEU (Court of Justice of the European Union (2016). Patrick Breyer V Bundesrepublik Deutschland, C‑582/14, 19 October 2016, ECLI:EU:C:2016:779.

CJEU (Court of Justice of the European Union) (2011) Scarlet Extended SA v Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM), C-70/10, 25 November 2011, ECLI:EU:C:2011:771

Dawes, S.S., Vidiasova, L., and Parkhimovich, O. (2016) 'Planning and designing open government data programs: An ecosystem approach', *Government Information Quarterly*, vol. 33, pp. 15-27.

European Commission (2011). Communication to the European Parliament, the Council, the European Economic and Social Committee, and the Committee for the Regions. Open data: an engine for innovation, growth and transparent governance, COM(2011) 882 final.

European Commission (2014) Guidelines on recommended standard licences, datasets and charging for the reuse of documents. Official Journal of the European Union: 2014; Vol. C240/01, p 10.

INSPIRE (2007) Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), OJ L 108/1.

Geiger, C. P. and Von Lucke, J. (2012) 'Open government and (linked) (open) (government) (data)', *Journal of e-Democracy and Open Government*, vol. 4, pp. 265-278.

Gurin, J. (2014) *Open data now. The secret to hot startups, Smart investing, savvy marketing, and fast innovation*, New York: Mc Graw Hill Education.

Gurstein, M. (2011) 'Open data: empowering the empowered or effective data use for everyone?', *First Monday,* vol 16, no. 2, n.p.

Janssen, M., Charalabidis, Y. and Zuiderwijk, A. (2012) 'Benefits, adoption barriers and myths of open data and open government', *Information Systems Management*, vol. 29**,** pp. 258–268.

Jetzek, T. (2015) The sustainable value of open government data. Uncovering the generative mechanisms of open data through a mixed methods approach, Copenhagen, Copenhagen Business School.

Kadaster (2017) Website Kadaster at: www.kadaster.com/about-kadaster, last accessed January 25th 2017.

Labots, W. (2016) Gebruiker in kaart; Analyse van loggegevens van de PDOK-webservices om groepen gebruikers van open data te identificeren. BSc. Thesis TU Delft.

Olausson, K. (2016) A step towards aligning supply and demand? - User involvement in supply of open data among ten Dutch public sector bodies. MSc. thesis Utrecht University.

Open Knowledge Foundation (2015) Open Definition version 2.0 [Online]. Available: http://opendefinition.org/od/ [Accessed March 12 2017].

PDOK (2012) Rapportage 2011. [Online].Available at: https://www.pdok.nl/nl/actueel/rapportages. [Accessed March 12 2017]

PDOK (2014) Rapportage Q4 2014. [Online].Available at: https://www.pdok.nl/nl/actueel/rapportages. [Accessed March 12 2017]

PDOK (2015) Rapportage Q4 2015. [Online].Available at: https://www.pdok.nl/nl/actueel/rapportages. [Accessed March 12 2017]

PDOK (2016) Rapportage Q4 2016. [Online].Available at: https://www.pdok.nl/nl/actueel/rapportages. [Accessed March 12 2017]

Ruijer, E., S. Grimmelikhuijsen, and A. Meijer (2017) 'Open data for democracy: Developing a theoretical framework for open data use', *Government Information Quarterly*, http://dx.doi.org/10.1016/j.giq.2017.01.001

Sieber, R.E., and Johnson, P.A. (2015) 'Civic open data at a crossroads: Dominant models and current challenges', *Government Information Quarterly, vol.* 32, no. 3, pp. 308–315

Stott, A. (2014) [od-discuss] Registration for accessing open datasets. Available at: https://lists.okfn.org/pipermail/od-discuss/2014-October/001083.html

Susha, I. (2015) Participation in Open Government. Orebro Studies in Informatics 8. Doctoral Dissertation.

Susha, I., Grönlund, Å. and Janssen, M. (2015) 'Organizational measures to stimulate user engagement with open data', *Transforming Government: People, Process and Policy*, vol. 9, pp. 181-206.

Van Beuningen, T., A. Bregt, L. Grus (2016) Professionals in de wolken met open AHN. *Geo-Info* 2016-5, 42-45.

Van Loenen, B., and M. Grothe (2014) 'INSPIRE Empowers Re-Use of Public Sector Information', *International Journal of Spatial Data Infrastructures Research*, vol. 9, pp. 86-106

Van Loenen, B., W.K. Korthals Altes, D. Groetelaers & F. Welle Donker (2016) Ontsluiten handelsregister met open data nader belicht. Delft: TU Delft: 66.

Welle Donker, F., B. van Loenen, and R. Braggaar (2016) Stand in open dataland. Research report for Dutch Ministry of Internal Affairs and Kingdom Relations.

Yin, R. K. (2003) *Case study research. Design and methods,* Thoasand Oaks, SAGE publications.

Zuiderwijk, A., Helbig, N., Gil-García, J. R., and M. Janssen (2014) 'Guest Editors' Introduction. Innovation Through Open Data: A Review of the State-of-the-Art and an Emerging Research Agenda', *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9 no. 2, pp. I-XIII.

Zuiderwijk, A. (2015) Open data infrastructures: The design of an infrastructure to enhance the coordination of open data use, 's-Hertogenbosch, Uitgeverij BOXPress.