Societal costs and benefits of high-value open government data: a case study in the Netherlands

Welle Donker, Frederika; van Loenen, Bastiaan

Societal costs and benefits of high-value open government data: a case study in the Netherlands

Frederika Welle Donker
Delft University of
Technology
Julianalaan 134
Delft, The Netherlands
f.m.welledonker@tudelft.nl

Bastiaan van Loenen
Delft University of
Technology
Julianalaan 134
Delft, The Netherlands
b.vanloenen@tudelft.nl

Abstract

Much research has emphasised the benefits of open government data, and especially high-value data. The G8 Open Data Charter defines high-value data as data that improve democracy and encourage the innovative reuse of the particular data. Thus, governments worldwide invest resources to identify potential high-value datasets and to publish these data as open data. However, while the benefits of open data are well researched, the costs of publishing data as open data are less researched. This research examines the relationship between the costs of making data suitable for publication as (linked) open data and the societal benefits thereof. A case study of five high-value datasets was carried out in the Netherlands to provide a societal cost-benefit analysis of open high-value data. Different options were investigated, ranging from not publishing the dataset at all to publishing the dataset as linked open data. In general, it can be concluded that the societal benefits of (linked) open data are higher than the costs. The case studies show that there are differences between the datasets. In many cases, costs for open data are an integral part of general data management costs and hardly lead to additional costs. In certain cases, however, the costs to anonymize /aggregate the data are high compared to the potential value of an open data version of the dataset. Although, for these datasets, this leads to a less favourable relationship between costs and benefits, the societal benefits would still be higher than without an open data version.

Keywords: open government data; high-value data; societal cost-benefit analysis; case study; Netherlands.

1 Introduction

1.1 Dutch National Open Data Agenda

Governments worldwide are in the process of publishing their datasets as open data. There have been ample publications emphasising the benefits of open data, and there are especially high expectations of the economic benefits (see e.g. Manyika et al. 2013, Vickery 2011). For this research, we adhere to the eight principles for of open data formulated by the Dutch government: all data that are public, free of charge, licence-free, available without prior registration, machine-processable, accompanied with metadata, findable, and published 'as-is' (i.e. as complete and unprocessed as possible and with the same level of quality and timeliness as used by the government) (Kamerstukken 2015, p.2). In accordance with the G8 Open Data Charter Action Plan[1] of 2013, the Ministry of the Interior and Kingdom Relations formulated a National Open Data Agenda (NODA) in 2015. The NODA commits government departments to identify core datasets[2] and high-value datasets, and to prioritise the release of these datasets as open data. High-value datasets are defined as data that improve democracy and encourage innovative reuse of the particular data (G8 Open Data Charter 2013).

In the 2016 NODA inventory, 27 datasets were classified as 'high-value', of which 19 were already available as open data, four were in the planning as open data, and four were under investigation.

Not all datasets are suitable to be published as open data, as the dataset may contain sensitive data, such as personal data. Such datasets should be transformed before an open data version can be published. And this poses a dilemma: if the dataset is not sufficiently anonymized/aggregated, the dataset can easily be de-anonymized, especially when combined with other datasets (see e.g. Koot 2012). However, if the dataset is over-aggregated, the dataset may be stripped of too many attributes to be still of value for reusers. It requires expert knowledge to determine a "safe" level of aggregation without losing (too much) value.

1.2 Societal Cost Benefits Analysis

A societal cost-benefit analysis (SCBA) provides an insight into the positive and negative effects of a policy (change) or project on the welfare of society. In a SCBA, direct and indirect effects are identified, quantified and monetised for a number of alternatives or scenarios, and tabulated for comparison. The advantage of a SCBA is that the (expected) effects of various alternatives can be used to monetise the effects of a policy / project in advance, and to evaluate the policy / project afterwards. The disadvantages are that SCBAs deal with uncertainties: effects have to be estimated in the future and not all (serendipitous) effects can be foreseen. Not all costs required to execute a policy / project can be isolated from regular operational costs. Not all effects are easy to quantify, which means that such 'soft' effects may appear to

---

[1] https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

[2] Core data are key datasets on National Statistics, National Maps, National Elections and National Budgets.

be weak in comparison to 'hard' effects that are easy to quantify.

1.3 Research question

In 2011, the Dutch government introduced an 'open by default' policy for all government data suitable to be published as open data. Five years later, the Ministry of the Interior and Kingdom Relations wanted to have an insight into the relationship between the actual costs of publishing open data – and in particular of high-value datasets – and the societal benefits thereof.

2 Methodology

The research consisted of two parts. First, from a literature review of international cost-benefit analyses and impact assessment reports, we analysed the (potential) quantitative effects of open data in general. These effects were used as input for a SCBA for five high-value datasets. The five datasets were selected by a supervisory group that had classified these as high-value datasets. The datasets and their status are presented in Table 1.

To assess the costs and benefits, we followed the eight steps prescribed by the General SCBA Guide (Romijn and Renes 2013), recommended for Dutch public sector projects.

The General SCBA Guide requires the identification of a zero-alternative, i.e. the most likely development without (a change of) policy. Then, other policy alternatives are defined, usually related to measures / investments required to achieve short(er)-term and long(er)-term benefits. For each researched case, four alternatives were identified:

- **0 Alternative**: the current status, no changes
- **-1 Alternative**: reversal of open data policy: only fee-based data or no public data
- **+1 Alternative**: change to open data according to all open data principles
- **+2 Alternative**: change to linked open data

We included the alternative of publishing the data as linked open data according to four stars of the Tim Berners-Lee model.[3] Although this requires extra resources compared to open data, linking the data will be more valuable for reusers. A linked dataset has more and positive network effects can be unlocked (Archer et al. 2013).

To identify, quantify and monetise the direct and indirect effects and the costs, we used desk research, interviews with stakeholders and users, web statistics analyses, and Industry Key Figures. For the BAG, we also carried out an online survey with companies that use BAG data.

---

[3] See http://5stardata.info for examples

Table 1: details of researched datasets and their status.

Name of dataset	Responsible department	Data contributors	Current status
Key Register Addresses and Buildings (BAG)	Min. Infrastructure & the Environment	380 municipalities	– Public sector: open data – Private sector: access for free; download for admin. fee
National real-time traffic information (NDW-data)	Min. Infrastructure & the Environment	Dept. of Roads & Water Management, 12 provinces, 2 regions, 4 largest Dutch municipalities	– Real-time data as open data; – data with additional features: data-for-service
Real estate values (WOZ-values)	Min. Finances	380 municipalities	– Limited access for free; – download for selected parties
Child abuse in NL 2005 dataset	Min. Justice	Research and Documentation Centre (WODC)	– Open data with prior registration
Jobs & Wages of Employees in NL (Polisbus)	Statistics Netherlands (CBS)	Employment Dept. & Taxation Dept.	– Only available under strict conditions

3 Results of the Societal Cost Benefit Analysis

3.1 Outcome literature review

We researched over 70 international studies related to the assessment of costs, benefits and the impact of (open) data, both on macro-economic level and on micro-economic level. The literature review showed that in all cases, the benefits to costs ratio was positive, i.e. the benefits outweighed the costs, ranging from a 1:1.6 to a 1:70 ratio, meaning that for every euro invested, the return was between circa €1.16 and €70.

However, the studies all employed a wide variety of assessment methods, making it hard to compare the outcomes. The studies showed that the additional costs of transforming data to be suitable for open data publication are marginal compared to the total data supply costs the organisation has to make. It is, therefore, difficult to distil the additional costs of publishing open data from normal operational costs. Many of the studies focus on the economic benefits of reuse by the private sector, whereas the public sector and citizens represent a large proportion of users (van Loenen et al. 2017).

Table 2: general overview costs of open data

Governance preparation costs
   Stakeholder consultations
   Data inventory
   Open data strategy development
   Renegotiation of existing contracts / licence agreements
   Staff training / capacity building
Data Transformation Costs
   Quality improvement
   Application of structure
   Transformation to OD-principles
   Digitisation of (historical) data
Infrastructural Costs
   Data platform / web service development
   Capacity for download / upload / invoke
   Security measures
   Apps / tools development
Operational Costs
   Data management
   OD platform maintenance
   Lost revenue due to OD
   Publicity costs
   Helpdesk

Table 3: general overview benefits of open data

Direct effects
   Efficiency gains
   Lower transaction costs
   Higher quality of data (due to feedback)
   More use by a broader group
   More applications based on open data
   More requests for additional data
   More questions related to data
Indirect effects
   Improved decision-making
   Effectiveness gains
   Improved accountability / transparency of government
   Improved corruption / fraud detection by citizens
   Avoided costs
Societal effects (+ = benefit, - = risk)
   + Improved public services
   + More competitive market
   + Environmental benefits
   - Easier to identify individuals when combining datasets

From the literature review, we distilled four types of costs, shown in Table 2, and three types of benefits, shown in Table 3.

3.2 Outcome case studies

From our survey, interviews and previous research (Welle Donker 2016), users indicated that they prefer government datasets because such data is considered authoritative data. Where datasets are already available as (semi) open data, users indicated that a reversal to closed data would pose a problem. Many companies have incorporated the data in their day-to-day business processes, and a return to closed data would mean higher transaction costs or even doing without the authoritative dataset.

3.2.1 Key Register Buildings & Addresses (BAG)

BAG data are collected from municipalities and managed on national level. As the public sector is required by law to use Key Register data for the execution of their tasks, the BAG dataset is available as open data for the public sector since 2013. Private sector users can access the BAG for free; to download the data, an annual administration fee of €168 is required. Thus, we consider the BAG as semi-open data, because in a true open data version, there would be no fees. In 2017, the BAG dataset became available as linked data, to enable more effective use within the public sector.

The direct effects of a semi-open BAG (0 Alternative) are lower transaction costs, less double collected data, better data quality, and more use by the private sector to improve their services and to create new products. The indirect effects are a more efficient and effective government, and new and more reliable value-added services. The direct effects of open data for all (+1 Alternative) are estimated at 2.6M€ due to lower transaction costs, and for linked open data (+2 Alternative) 0,4M€ as the extra costs of linking the BAG-data are circa 2.2M€. The societal benefits for the +1 and +2 Alternatives are circa 0,25M€ as consumer surplus. Societal benefits, such as improved fraud detection, cannot be quantified.

Figure 1: Example of Key Register Addresses and Buildings (BAG)

3.2.2 National real-time traffic information (NDW)

The National Database Traffic Information (NDW) is a cooperation of 19 public sector bodies established to streamline traffic information management. Since 2013, NDW supplies real-time traffic information as open data without service guarantees, and also historic data to partners. Companies can elect to use traffic information with a Service Level Agreement in return for a service (e.g. apps) or data (e.g. floating car data) to NDW, a so-called data-for-service agreement. Data acquired under such agreements are currently not redistributed as open data.

The direct effects of open real-time traffic information (0 Alternative) are lower transaction costs and are monetized at

1.08M€. The indirect effects are less lost time and detour costs due to avoiding traffic jams, and new products / services developed by companies, and monetized at 41.1M€. The societal benefits are more efficient use of roads, less CO_2 emissions and less air pollution. These effects will increase with linked data (+1 Alternative) and linked data including (anonymized) data acquired under data-for-service agreements (+2 Alternative), as the information quality will increase (cf. Deloitte LLP 2017) Although the costs of linking the data are circa 0.9M€ for +1 Alternative and 1.9M€ for +2 Alternative, the benefits will nearly double, leaving a net benefit of 85.18M€ for +2 Alternative.

Figure 2: NDW illustrated (source: www.ndw.nu).
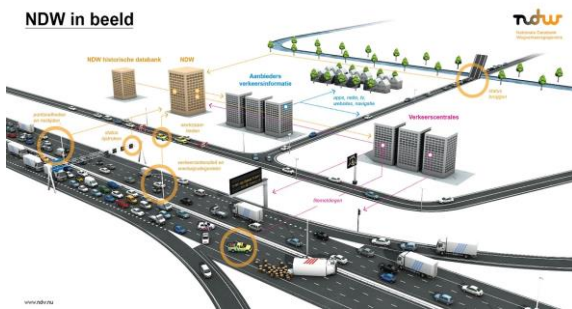
3.2.3 National Real Estate Values (WOZ-values)

Figure 3: screenshot of WOZ-values web service (www.woz-waardenloket.nl)
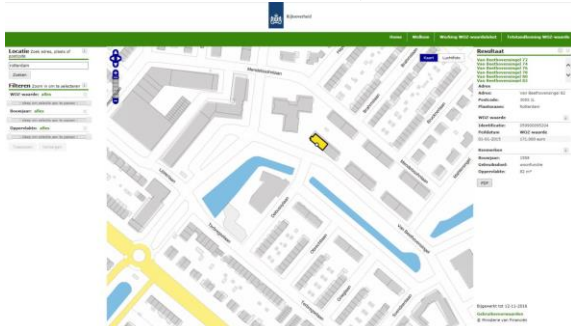
The Key Register Valuation Real Estate (WOZ) contains data related to all real estate in the Netherlands, and is used by government bodies to calculate a number of taxes and local rates, and by authorized parties (banks, insurers and notary solicitors) to avoid real estate fraud. Municipalities value properties annually and supply these values to the national database. As the dataset contain personal data, the Key Register is not available as open data. However, to increase transparency, real estate valuations are accessible for viewing since October 2016. The number of views is limited to a maximum per time period to avoid scraping.

The direct effects of accessible real estate values (0 Alternative) are, to date, lower transaction costs for citizens and higher data quality. Pilot studies had indicated that access to real estate values will lead to fewer formal objections to taxes / rates assessments, and thus, to lower administration costs for the government. These avoided costs are estimated at

48.5M€ for an investment of 3.4M€ for the web service and security measures. The expected indirect effects will be more trust in government and more informed choices for house buyers. These effects cannot be quantified. For +1 Alternative, an aggregated/anonymized open data version, and +2 Alternative, a linked open data version, the costs are estimated at 0.14M€ and 0.16M€ respectively.

3.2.4 Child Abuse in the Netherlands anno 2005 data

The Child Abuse NL 2005 dataset was collected by Research and Documentation Centre WODC ('Wetenschappelijk Onderzoek- en Documentatiecentrum'), a department of the Ministry of Security and Justice tasked with performing research for the Ministry. The Child Abuse NL 2005 dataset was published as open data in 2007 via a national portal for social science data. The dataset does not comply to open data principles (propriety format and prior registration required), and is, therefore, classified as semi-open data. Web statistics show that since 2013, circa 11 unique users downloaded the dataset 65 times. The dataset is very hard to find using generic browsers, and search terms.

The direct effects of semi-open data (0 Alternative) are, to date, limited to reuse by a small number of researchers. The indirect effect of increased scientific knowledge and better decision-making cannot be quantified. If the dataset were available as open data according to open data principles and accessible via the general open data portal (+1 Alternative) or as linked open data (+2 Alternative), the dataset will be findable and reused more frequently. The investment costs are between €500 and €900 respectively.

3.2.5 Jobs and Wages of Employees in Netherlands (Polisbus)

Polisbus dataset contains quantitative and qualitative data related to job and wages of all employees in the Netherlands. The data are supplied by the Employment Department and the Taxation Department to Statistics Netherlands (CBS). The dataset contains personal data and is not publically available. Only authorized users (mainly researchers) may access an anonymized version under very strict conditions in a secured environment. Although CBS adheres to an open data policy, Polisbus is classified as microdata. The Polisbus dataset is one of the most requested microdatasets of CBS.

The direct effects of restricted data (0 Alternative) are 7.32K€, the costs to access the data by ca. 400 users. The indirect effect of increased scientific knowledge and better decision-making cannot be quantified but are significant as the data is mainly used by Ph.D. researchers. If an aggregated version were available as open data and accessible via the general open data portal (+1 Alternative) or as linked open data (+2 Alternative), the dataset will be reused by many more users. The investment costs would be between 140K€ and 165K€ respectively, as the choices and models for aggregation require careful consideration. However, CBS has no plans for an open data version of Polisbus.

3.3 Conclusions of case studies

The first two cases, BAG and Real-Time Traffic Information, show that once open data are established, the data become an integral part of operations and services by the public sector and the private sector. The benefits are mainly found in efficiency and effectiveness gains, which have a flow-on effect for society. For data that are only recently accessible (WOZ-values), the societal benefits have yet to emerge. The first signs are positive though. For open datasets that are not findable (Child Abuse NL 2005) or not open at all (Polisbus), the societal effects cannot be quantified. In the first case, the investment costs to comply to open data principles are negligible. In the latter case, an aggregated version of Polisbus as open data is not under consideration, as the transformation costs are high in relation to expected benefits.

Most cases showed that where high-value datasets are published as open data, use of the dataset, both by the public sector and by the private sector, increased significantly. Although there is a loss of revenue, this loss of often balanced by lower administrative costs, which may be significant (see e.g.de Vries et al. 2011). The societal benefits – although not always obvious – show a positive relation to the costs incurred by the data provider. To demonstrate the positive balance, it may be better to consider the *costs* of no longer providing open data rather than the *economic benefits* of open data (see e.g. Lateral Economics 2016).

It should be mentioned that for all researched organisations, open data supply is not a core task, and only represent a small part of the total operational costs. The additional investment costs for linked data appear high. However, the benefits in the longer term, will outstrip the initial costs. For organisations that have to yet invest in open data, it would be recommended that the datasets should be considered for transforming into linked data at that stage (if data are suitable), as the extra costs of linking the data are relatively low compared to the costs of open data

References

Archer, P., Dekkers, M., Goedertier, S. and Loutas, N. (2013) *Study on business models for Linked Open Government Data (BM4LOGD)*.

de Vries, M., Kapff, L., Negreiro Achiaga, M., Wauters, P., Osimo, D., Foley, P., Szkuta, K., O'Connor, J. and

Whitehouse, D. (2011) *Pricing of Public Sector Information Study. Models of Supply and Charging for Public Sector Information (ABC) Final Report*, Brussels: Deloitte Consulting,.

Deloitte LLP (2017) *Assessing the value of TfL's open data and digital partnerships*, London.

G8 Open Data Charter and Technical Annex, (2013)

Kamerstukken (2015) *Nationale open data agenda 2016 (NODA)*, kst II 2015/16, 32 802, nr. 20, Den Haag: Ministerie van Binnenlandse Zaken en Koninkrijkrelaties.

Koot, M. R. (2012) *Measuring and predicting anonymity*, PhD dissertation, University of Amsterdam.

Lateral Economics (2016) *Permission granted: The economic value of data assets under alternative policy regimes. A Lateral Economics report for the Open Data Institute*

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S. and Almasi Doshi, E. (2013) *Open Data: Unlocking innovation and performance with liquid information*, McKinsey Global Institute.

Romijn, G. and Renes, G. (2013) *Algemene Leidraad voor maatschappelijke kosten-batenanalyse*, Den Haag: Centraal Planbureau / Planbureau voor de Leefomgeving.

van Loenen, B., Ubacht, J., Labots, W. and Zuiderwijk, A. (2017) 'Log File Analytics for Gaining Insight into Actual Use of Open Data', in Vieira Borges, J. and Dias Rouco, J. C., eds., *17th European Conference on Digital Government (ECDG 2017)*, Lisbon, 12-13 June 2017, Reading, UK: Academic Conferences and Publishing International Ltd, 238-246.

Vickery, G. (2011) *Review of recent studies on PSI re-use and related market developments*, Paris: Information Economics.

Welle Donker, F. (2016) *From Access to Re-use: a user's perspective on public sector information availability*, PhD. dissertation, Delft University of Technology.