Nudging the government

How open data can be used to make behavioural governance work both ways

Prof. Dr Ulrich Greveler (Germany) | Rhein-Waal University of Applied Sciences

1 Introduction

Data-driven regulation has become a new type of governance. Identifying number plates and human faces using computer vision, and tracking the position of consumer devices deploying sensors in the public space, are facilitating new master plans for the smart country and its smart cities, where government services are provided only to those who need them (Hodgkinson 2011). Politics are struggling to keep pace and include available data and newly identified threats in the development of new policies. Behavioural governance is helping to utilise data collection in order to influence consumers and citizens without the need to enact strict penalties for unwanted behaviour and without the requirement to enforce regulations with huge bureaucratic overheads. The computer industry's default answer to the how of politics is what could be called 'solutionism': problems are to be dealt with via apps, sensors and feedback loops—all provided by start-ups (Morozov 2014). Google's Eric Schmidt was even more optimistic in 2014: IT start-ups would provide the solution to economic inequality.

In our work we aim to analyse the potential of *open data* and *freed data* as a transparency safeguard employable by citizens and consumers in order to check the necessity and efficiency of government-induced, data-driven regulation and to represent a new level of transparency: data can facilitate government accountability, transparency and citizen participation.

Open data preferably includes non-textual data such as map coordinates, sensor data, figures from public records and any collected data on citizens, as long as it is not related to individuals so that the privacy of citizens is not jeopardised. While open data can come from any source, the data issued by government agencies and the municipal sector has the highest potential for citizens and researchers, as well as application developers, to engage. By encouraging public organisations to make non-sensitive data discoverable via public networks, citizens are empowered to harness it and combine it in new ways so that public decision-making becomes more accountable.

With the aid of concrete example data sets based on current ongoing projects and open data portals, we illustrate how citizens can contribute to a full *read–write society*, observe and influence public decision-making processes and contribute to it in an effective way. This may include very different data sets such as car parking in the inner city, traffic counts, citizen self-service Web portals, as well as e-government online security and teacher resources, just to give an impression of the vast aspects covered by public open data and the chances that computer science, among other disciplines, may offer to nudge public decision-makers.

2 Nudging

The concept of nudging has caught the attention of the media, scholars and public policy-makers as a way of altering citizens' behaviour in a predictable way without prohibiting any choices or implementing economic incentives or disincentives for unwanted behaviour. Nudges are explicitly designed to uphold full freedom of choice (Sunstein 2015).

Prominent examples of nudge concept exploration are the *UK Behavioural Insights Team* and the *US White House Social and Behavioural Sciences Team* (BIT 2012). Regarding public policies, the following nudges are frequently considered: *default rules* (e.g. opting out of default organ donor programmes instead of distributing consent forms), *the use of social norms* (e.g. 'most people do not use plastic bags' campaigns), *an increase in convenience* (e.g. displaying park and ride signs to avoid inner-city traffic) and *disclosure* (e.g. 'How much does our city pay for waste management?').

The advent of digitalisation and connected databases enables the government to help inform citizens' personal decisions: government data mining, for instance, could notify users of the relevant health risks, useful government programmes for which they qualify and neighbourhood crime—such personalised information can be particularly effective in nudging citizens to make socially beneficial choices (Linders 2012).

In this paper we aim to apply nudging to (mainly local) public decision-making, requiring that the potentially unfortunate behavioural process or decision-making pattern is identified as the result of cognitive boundaries or habits and could be nudged towards a better option: a well-informed, accountable way of reaching decisions based on data that could be inspected by anyone having a stake in the decision.

3   Open data and open government

Open data and open government have gained much awareness recently, both from academia and the public sector itself. Since 2009 governments worldwide have been developing *Open Government Data* programmes, with the US and the UK being the two leading countries in this sector (Peled and Nahon, 2015).

The low cost of digital data publication puts governments in a position to create the information obtained while delivering governmental services that are accessible to the general public. The preferred result from a data user's point of view is the publication of open data. Open data is data that anyone can access, analyse, visualise, use for any purpose and share with others. One purpose of a government open data publication is to advance the long-term transparency of government actions, but also to enable the use of open data for business creation, as companies may use open data as a key aspect of their services (Stamova 2016). The data can help citizens to improve their day-to-day productivity and foster informed decision-making as open government initiatives promise to 'open up' government operations to citizens (Clarke and Margetts 2014). The data can also help bring data start-ups to the surface; in other words, small enterprises are created explicitly to take advantage of the value of open data, for example, city planning, education, creative industry, location-based services, neighbourhood risk evaluation, delivering access to public sector contracts or creating transport-related products (Seibel 2016).

However, data publication itself may not be sufficient in many cases, as the usefulness of the data is heavily dependent on discoverability by search engines, formats, granularity and respective metadata. From a citizen's perspective, the major obstacles are access to proper data sets and the adequate use of data sets (Zuiderwijk et al. 2013), while data set fragmentation across websites makes them difficult to find (Boulton, Rawlins, Vallance, and Walport 2011; McLaren and Waters 2011). Sufficient metadata is a key element of successful data exploitation (Zuiderwijk, Jeffery, and Janssen 2012).

4 Local public decision-making processes and data processing

Local autonomy of cities, municipalities and districts is a building block of many constitutional states. It preserves the decision-making processes at the local level and maintains a decentralised separation of powers. The case examples in this paper are located in the state of North Rhine-Westphalia in Germany. In this state, checks and balances to the mayor and local government staff are provided by local parliament structures, for example, city councils (*Stadträte*) being elected by the city's population. The council is the main body of the city's autonomy. The elected council members are not full-time professionals; in other words, they carry out their duties after work in the late hours of the day. In order to be able to perform their major tasks—in particular, passing the local government's proposed revenues and spending for the upcoming financial year—council members need to be informed and educated on local matters by local government staff. They also need access to municipal data sources and adequate data aggregation and visualisation. Another major source of information is the constituents who contact their elected officials and provide background information on pressing issues and current challenges.

There is an inherent conflict of interest regarding the information provided by local staff to council members: exercising the council's rights to check and balance the executive powers can make the staff's jobs more difficult since

they need to publicly justify work results and might receive unpleasant criticism from council members who represent their constituents and relay citizens' complaints. Hence, local government staff can be tempted to withhold information or present information in such a way that makes it difficult for recipients to spot certain details hidden behind the statistics.

With the help of accessible government data, citizens, media, as well as elected officials, are enabled to analyse data on their own and to foster transparency of government actions. The public is then able to review and scrutinise policies without being dependent on the offered rationales and accumulated data by the decision-makers.

Regarding the multidisciplinary nature of the topic of government data analysis and state transparency, we want to highlight technical methods that are naturally provided by the communities of computer scientists or data scientists: opening up raw data sets that are not published but which can be generated from harvesting online services or digital documents; the automated collection of data by scanning or crawling websites, digital interfaces or networked services; parsing machine-readable data; and computer vision and natural language processing in order to generate machine-readable data for further data-driven analysis. Data-driven science itself employs methods and theories drawn from mathematics, statistics and computer science.

In real-world scenarios raw data is often hidden behind online accessible services but can be freed by using robots (sometimes called crawlers and spiders) or scripts (little pieces of software automating the execution of recurring tasks) that automatically traverse the website and database content and output it to a comprehensive file. Note that this automatic extraction of data must not be confused with 'hacking into a database' (e.g. cracking weak passwords to gain unauthorised access), as the latter is considered illegal in most countries, while automatically crawling web pages is the underlying technology of many web search engines. However, it should be noted that the automated extraction of files can lead to the discovery of data that was obviously not intended to be published but is technically accessible as a result of negligent system configuration. In these cases a code of ethics and professional conduct should be applied: confidential data should be excluded and site owners informed.

5 Case examples

We present four cases from our ongoing research of 2016 and 2017. Note that in two cases the data was provided by the local authorities on open data portals, while in the other two we acquired or liberated the data by technical means on our own.

5.1 Case: non-moving traffic notices

Raw data on non-moving traffic notices ('Strafzettel') in several cities of North Rhine-Westphalia, Germany, has been made public as open data or retrieved from the local government by exercising freedom of information laws. The available raw data is stripped of personal information (i.e. number plates, names of vehicle owners) before publication.
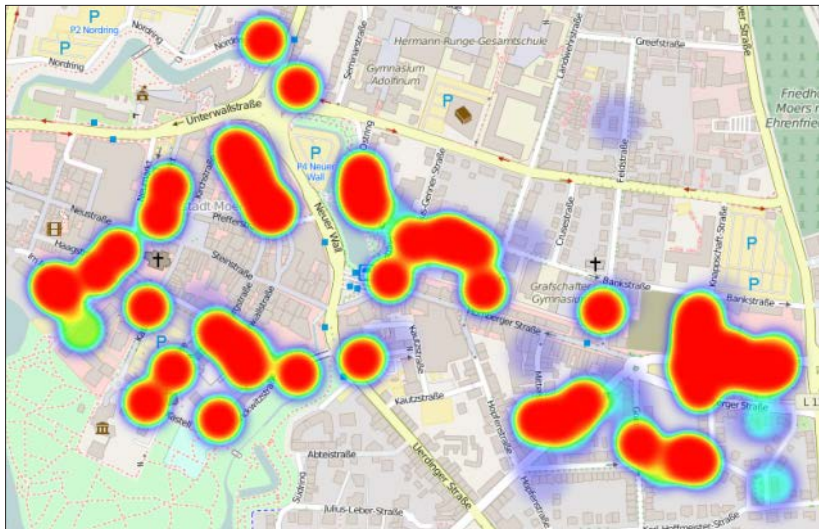

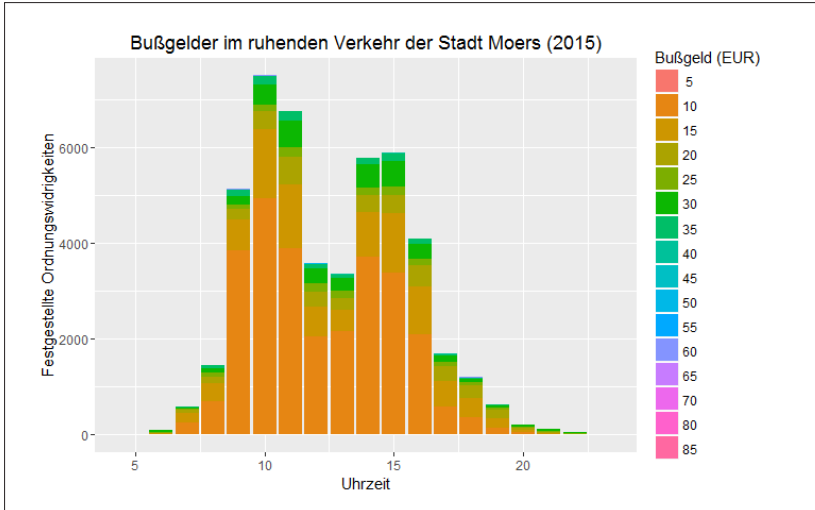**Figure 1a: Heat map visualisation in the city of Moers**

Figure 1b: Daytime charts of non-moving traffic fines in the city of Moers

Some visualisation of the 2015 data is depicted in Figure 1. It becomes visible where the notices were issued and at what time of day fines were issued (between 10 a.m. and 11 a.m. the detected traffic delinquency is most prevalent, which might also reflect the typical working and lunch hours of the assistant police force).

Possible nudge effects: disclosure of inner-city regions where many or few notices are issued. (Fines are issued at certain hotspots that do not necessarily reflect areas with high levels of false parking observation.) Heat maps could be used for informed urban development planning or real estate development. Individual city staff members might work inside hotspots: suspected favouritism can be ruled out by data analysis.

5.2 Case: waiting for administrative appointments

The city of Essen (North Rhine-Westphalia, Germany) offers a service to its citizens to reserve online dates for administrative appointments (e.g. relocation registration, applying for a new ID card, vehicle registration). A citizen selects the desired task in a web interface and is issued the next available date and time (and office location) for the visit.

There has been some public outrage, reported in local newspapers, as the waiting period has been up to fifty days for some citizens. City officials blamed the administrative burden of refugee management as a major factor for the staff work overload (Wandt 2016). The respective data was not made public. In order to make the data available, however, a script was programmed by M. Schurig (Open Knowledge Lab, Ruhrgebiet), which every morning simulates a citizen request to obtain an appointment reservation via the web page. The issued date was logged and the request was discarded afterwards, so no interference with actual administrative business took place. Visualisation of the collected data can be seen in Figure 2. It becomes apparent that at certain regularly occurring intervals the waiting time drops to almost zero, although on the previous day citizens were issued dates that were more than forty days in the future.

Possible nudge effects: disclosure of a gradual positive long-run development. No correlation to events connected with refugee administration is apparent, so the release of wrongful information is discouraged. Informed decisions by citizens lead to shorter waiting periods (e.g. repeat request the next day).
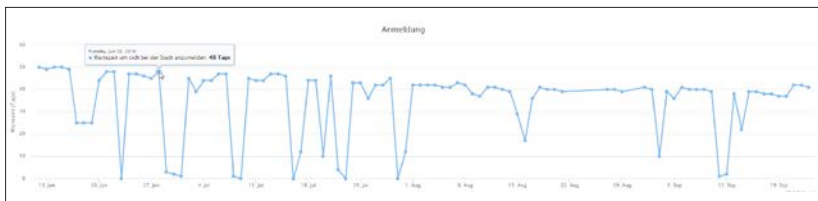

Figure 2: Waiting period in days for administrative appointments, city of Essen

5.3 Case: counting traffic

Traffic counting data provides street-level traffic data for junction-to-junction links. It plays a pivotal role in the decision-making processes of municipal authorities (speeding, congestion, pollution). Raw data sets have become open data for some cities. The data sets include individual occurrences of vehicles, together with a time stamp and vehicle size information (see Figure 3).

```
28.10.14  11:53:03;00;031;01;0
28.10.14  11:54:08;00;037;01;1
28.10.14  11:54:57;00;028;01;1
28.10.14  11:55:28;00;025;01;0
28.10.14  11:57:35;00;021;01;0
28.10.14  12:00:38;00;026;02;0
28.10.14  12:01:33;00;022;01;1
28.10.14  12:02:05;00;028;01;0
28.10.14  12:02:18;00;023;01;1
28.10.14  12:03:25;00;030;01;1
28.10.14  12:03:33;00;038;01;0
```

Figure 3: Traffic count raw data clipping example, city of Moers, Vinner Straße, 2014.

Analysis of traffic count data shows that official traffic count statistics may neglect certain aspects in particular cases:

■ Big timeline gaps in the raw data show that something blocked the traffic or the sensor altogether and rendered the count useless for that particular day.
■ Truck traffic may be a phenomenon at specific times of day, or a constant imposition.

These results cannot be derived from the statistical traffic count data usually provided to the public and to decision-makers.

Possible nudge effects: detection of erroneous data counts can prevent council members from making decisions on inaccurate data or help to find the causes of ineffective street construction measures of the past. Scrutiny regarding data-driven decisions is fostered.

5.4 Case: computer science education

In North-Rhine Westphalia computer science was introduced as an elective subject within secondary schools in the early 1970s. As computer science did not become a core or compulsory subject, and because the majority of German students attend public schools, the lack of achieving competencies in information technology and gaining digital literacy became a matter of political debate. In the recent past the subject was mostly taught by in-service teachers who received their initial professional development on computer science topics but lacked deep knowledge and pedagogical content knowledge (Knobelsdorf et al. 2015).

The Ministry of Education publishes statistics on curriculum development and teaching hours (detailing subjects and types of school) on a yearly basis. However, the bulky 255-page PDF-formatted document (MSW 2016) consisting of printed tables does not meet open data standards, making it a gruelling task to retrieve numbers and metadata from the tables with the aim of analysing and visualising the included data. For this reason, a comparison of taught hours between federal states, or of the volume of computer science education to other subjects (both easily leading to heated political discussions), is difficult.

With the help of semi-automated data-extraction techniques, individual machine-readable data sets were mined from the PDF file.

The bar chart visualisation in Figure 4 shows that taught computer science hours (left bar) represent only a small fraction of the hours taught for other major subjects. Moreover, the analysis showed that a ratio of approximately ten biology teachers per one computer science teacher could be determined.

Possible nudge effects: the disclosure of unbalanced resources in regard to curricula puts pressure on elected officials to strengthen possibly neglected

subjects (such as computer science). Decision-makers might demand data in a form that enables them to conduct their own analysis or call for issuing open data.
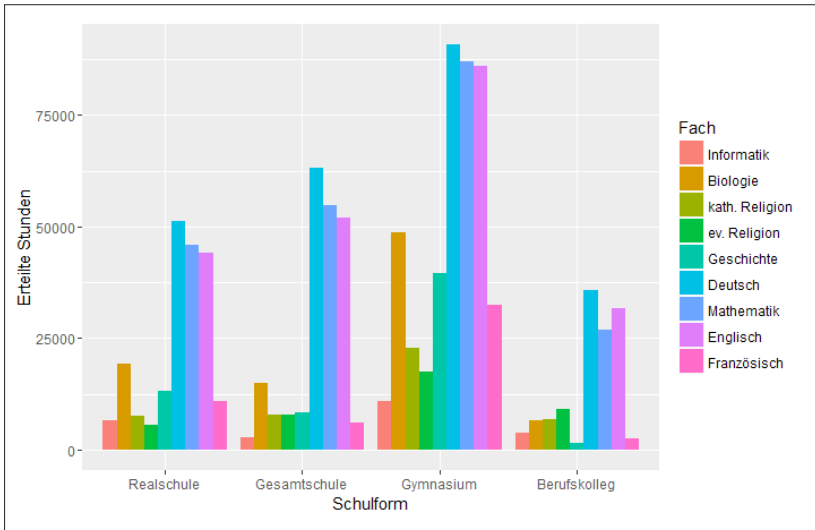

Figure 4: Taught hours per subject and school type in NRW (2015/16)

6 Conclusions

In our work we analyse the potential of open government data as a transparency safeguard that is employable by citizens, as well as decision-makers, in order to check the necessity and efficiency of government-induced, data-driven regulations and to promote public sector accountability. Several cases have illustrated the potential to nudge local decision-makers towards a better-informed, accountable way of reaching decisions based on data.

With the aid of concrete example data sets based on current ongoing projects and open data portals, we have demonstrated how knowledgeable citizens and elected officials can be informed about public decision-making processes and contribute to them in an effective way. Open data can enable the public to understand and reproduce the public decision-making process so that decision-makers are motivated (i.e. *nudged*) to highlight areas of poor decision performance and investigate data-driven methods of improving that performance. With the help of open data, individuals can grow their capacity to participate in society and take responsibility for their local political environment.

The case studies show that computer-scientific methods and techniques can be used to generate, enhance or liberate available data. The multidisciplinary nature of public data analysis for informed decision-making processes and respective controls can be enriched by these methods.

References

BIT (Behavioural Insights Team). 2012. *Applying behavioural insights to reduce fraud, error and debt.* London: Cabinet Office. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60539/BIT_FraudErrorDebt_accessible.pdf

Boulton, Geoffrey, Michael Rawlins, Patrick Vallance, and Mark Walport. 2011. 'Science as a public enterprise: the case for open data.' *The Lancet* 377 (9778): 1633–1635.

Clarke, Amanda, and Helen Margetts. 2014. 'Governments and citizens getting to know each other? Open, closed, and big data in public management reform.' *Policy & Internet* 6 (4): 393–417.

Hodgkinson, Steve. 2011. 'Is your city smart enough? Digitally enabled cities and societies will enhance economic, social, and environmental sustainability in the urban century.' *OVUM report.*

Knobelsdorf, Maria, Johannes Magenheim, Torsten Brinda, Dieter Engbring, Ludger Humbert, Arno Pasternak, and Jan Vahrenhold. 2015. 'Computer science education in North-Rhine Westphalia, Germany – A case study.' *ACM Transactions on Computing Education (TOCE)* 15 (2): 9.

Linders, Dennis. 2012. 'From e-government to we-government: Defining a typology for citizen coproduction in the age of social media.' *Government Information Quarterly* 29 (4): 446–454.

McLaren, Robin, and Robin Waters. 2013. 'Governing location information in the UK.' *The Cartographic Journal* 48 (3): 172–178.

Morozov, Evgeny. 2014. 'The rise of data and the death of politics.' *The Guardian* (20 July 2014).

MSW (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen). 2016. *Das Schulwesen in Nordrhein-Westfalen aus quantitativer Sicht 2015/16.* Statistische Übersicht 391, Düsseldorf.

Peled, Alon, and Karine Nahon. 2015. 'Towards open data for public accountability: Examining the US and the UK Models.' *iConference.* Newport: California.

Seibel, Benjamin. 2016. *Open Data in der Praxis. Bereitsteller und Anwender offener Daten in Berlin.* Berlin: Technologie Stiftung Berlin.

Stamova, Tanya. 2016. *Integrating open data reuse into the business models of German companies. Master's thesis.* University of Twente.

Sunstein, Cass R. 2015. 'Nudging and choice architecture: Ethical considerations.' *Yale Journal on Regulation*.

––– 2015. 'Nudging and choice architecture: Ethical considerations.' Discussion Paper No. 809, Cambridge.

Wandt, Christina. 2016. 'Wartezeit verärgert Bürger – Stadt setzt Security im Amt ein.' *Westdeutsche Allgemeine Zeitung* (13 May 2016).

Zuiderwijk, Anneke, Keith Jeffery, and Marijn Janssen. 2012. 'The potential of metadata for linked open data and its value for users and publishers.' *JeDEM-eJournal of eDemocracy and Open Government* 4 (2): 222–244.

Zuiderwijk, Anneke, Marijn Janssen, Sunil Choenni, Ronald Meijer, and Roexsana Sheikh Alibaks. 2013. 'Socio-technical impediments of open data.' *Electronic Journal of eGovernment* 10 (2): 156–172.