6 Transparency and open data

Tim Davies

Tim is co-director of Practical Participation, and a Web Science and Social Policy PhD student at the University of Southampton. Tim has written extensively on open data and participation. His work in recent years has included facilitating social reporting at the Internet Governance Forum, developing linked-data demonstrators for IKM Emergent and supporting the emergence of an eco-system of applications of data around the International Aid Transparency Initiative (IATI).

6.1 Introduction

Since the Enlightenment, in the 17th and 18th centuries, it has been argued that transparency is a fundamental aspect of democratic politics. Although first enshrined in law in Sweden in 1766, it is only in the later part of the 20th century that freedom of information (FOI) acts spread across modern democracies, reaching the United Kingdom in 2000. These reactive transparency laws, giving citizens the right to request information from Government, have been joined in the last few years by open-data policies, encouraging proactive disclosure of Government information in the form of structured datasets. With open data, citizens asking, for example, for information on how budgets are spent, should no longer have to trawl through hundreds of printed pages. Instead, they should be able to load up a spreadsheet from Government, and filter and search for the information they need. In a complex state, where the scale of information held inside Government grows exponentially with the rise of vast databases and digital-by-default services, access to data may be the only way for citizens to effectively exercise oversight of Government. Non-digital copies of the information are just too cumbersome to work with.

6.2 The emergence of open data

In the UK, a national open-data initiative was announced in 2009 by then Prime Minister Gordon Brown. The announcement came in the context of the expenses scandal and a crisis of confidence in politicians. It was primarily framed in terms of transparency, but also emphasised the importance of opening data to support innovation, public participation and economic growth. The form the initiative took was centred on a data portal at data.gov.uk that provides access to datasets from departments across Government and mirrored the model taken in the United States, where the data.gov site was launched in May 2009. The US open-data policy developed out of a memo on open government issued by Barak Obama on his first day in office and emphasised the triad of transparency, participation and collaboration. In the United Kingdom, open data has remained high on the agenda across a change of Government, with the coalition pushing for further release of Government datasets, strengthening emphasis on the potential contribution that the innovative re-use of Government datasets may make to the economy.

In early 2013, over 9,000 datasets were listed on data.gov.uk, covering issues from food hygiene to school locations, and from prescribing practices of GPs to geological models and oil production statistics. And it's not just Government that is making a move towards open data. By choice, or in response to external mandates, many other sectors are also moving towards open publication of datasets online. Science funders and ‘open-science’ advocates are asking for datasets generated during research to be archived and, in some cases, to be made publicly available. Targeted transparency policies are being used by Government to force certain sectors to disclose data (e.g. on food nutrition, car safety and environmental impacts72). In addition, the Open Data Institute, founded in late 2012, is advocating for private firms to publish datasets online, sharing everything from product specifications to supply chain information. These open data are just a small sub-set of the total amount of data available digitally. Vast datasets of social interactions, captured by services like Facebook and Twitter, are partially accessible for re-use, albeit not as open data. Also, private firms

---

hold vast datasets on their customers, which are used to drive their business decisions. It is no surprise that the Economist referred to this as ‘the data deluge’\(^{73}\).

However, when it comes to public dialogue, where does all this data get us? What difference does it make to have open data? How far does Government transparency alter the relationship between citizen and state? The following sections explore these questions by highlighting a number of ways in which transparency, open data and dialogue might meet. Before this though, it will help to have a clear view of what makes open data different.

### 6.3 Defining open data

A dataset is a collection of ‘facts’: individual atomic descriptions of the world. Taken alone, a single cell or item within a dataset is effectively meaningless. It is only when we add context, and start to represent and analyse data, that it becomes meaningful information that can form the basis of dialogue and decision-making. When you have a printed table of figures or a map showing the location of particular services, then someone has already chosen how data should be contextualised, and has fixed an interpretation. However, if you have the underlying dataset and the tools to work with it, then you might be able to create your own interpretation, focusing on the aspects of the data you feel are important or mixing the datasets together to create new information and understanding. This is the value of open data. To be open data, a dataset must be accessible, machine-readable (i.e. in a form where you can manipulate it with digital tools), and licensed to permit re-use, rather than restricted by copyright or intellectual property rights.

In practice, open datasets vary in how far they are open to different interpretations. Some datasets listed on data.gov.uk are little more than summary statistics pre-computed by departments, while others are ‘raw’ datasets reporting original measurements. Creating a dataset involves being selective about what to record and how to encode it\(^{74}\). Using a dataset often requires considerable tacit knowledge about a policy area and the way the data have been collected. The balancing act between protecting the privacy of individuals and releasing datasets built up from individual records also means that choices over anonymisation and aggregation affect the data that makes it into the open, and what can and can’t be transparent\(^{75}\). It is also worth re-iterating that not all data now available to drive policy are open. Just as ‘commercial sensitivity’ is one of the reasons for non-disclosure under the Freedom of Information Act, many datasets that may inform policy or that are drawn on in consultation responses, remain private, owned by third-parties and not covered by the Government’s transparency and open-data policies. This is important to note when considering the balance of power in public dialogue.

### 6.4 Data informing dialogue

For over 40 years, proposals for a new airport at Notre-Dame-des-Landes, near the city of Nantes in France, have generated significant controversy. Arguments over its economic benefits, environmental impact and forced eviction of those on the proposed site have divided politicians and community members. In November 2012, with thousands of activists occupying the site in opposition to construction, a group of journalists and active citizens, frustrated by constant contradictory statements about the project, started compiling data related to the airport plans. In January 2013, they opened up this collection of data to the public\(^{76}\). Their vision is that this collection of datasets (some fully meeting the open-data definition, others approximating it) will support public conversations over the proposed airport and will allow the creation of dynamic visualisations, animations, web services and graphics that can better inform public debate.

The idea of open data supporting more informed public dialogue is a compelling one. However, as the airport example above suggests, using open data to inform dialogue is not simply a case of placing datasets online. It also requires data to be curated and for intermediaries to help make sense of complex datasets. Even with open data, the information that most citizens receive will still be filtered


\(^{76}\) http://dataaeroportnotredamedeslandes.wordpress.com/
through editorial judgements. Having datasets available will not inevitably drive more evidence-based or rational arguments in general. Yet, in a number of settings, open data does create a space for new actors to enter the debate. In particular, there have been civic-minded, open-data analysts and technologists engaging with open data and learning, in depth, about issues through the process of building platforms that work with the data. For example, creators of the openspending.org platform, visualising public spending from countries across the world, had to learn about the complexities of public finance to interpret and present Government datasets. Also, the ‘Clear Climate Code’ project involved developers rewriting climate prediction software models to better comprehend how vital climate predictions were generated. Whether, as we go beyond the stage of early experimentation with open data, such community-led projects will be sustainable remains to be seen and it is likely that they will not be equally distributed across all policy areas. Nevertheless, they illustrate a new resource for dialogue in the outputs created and in the emergence of new, self-taught, citizen experts in certain policy areas. Such groups may be able to put new issues on the agenda for dialogue, and to play a role as participants and facilitators in existing areas of discussion.

The exact role that open data will play in a formal dialogue depends on the dynamics of the issue at hand and the sorts of data available. In some cases, independent intermediaries may already have created diverse interpretations and presentations of data, which can act as a useful input to a deliberative process. In other cases, few uses will have been made of available data (just because data are out there, doesn’t mean anyone will necessarily have made the investment of time to work with them) and a dialogue project may need to commission or otherwise catalyse the creation of resources that can inform discussions. In cases where expert input is vital, open data offers the opportunity for experts to more transparently ‘show their working’, laying bare the data underlying their conclusions (albeit, requiring a culture change and new ways of working for many experts). Again, in other cases, it may not be appropriate to draw upon data at all and the purpose of dialogue may need to be to provide space for stories, rather than statistics, understanding the experiences and opinions of citizens, rather than their response to selected data.

6.5 Data-driven decision-making

While transparency and open data could be used to ‘lay out the facts’ in front of citizens and support more informed policy discussion, an alternative possibility exists, with open data being but a footnote to larger shifts towards data-driven decision-making. Here, dialogue is displaced by positivist perspectives that view answers as already there in the data, ready to be extracted by data scientists. In the 2013 Annual Letter for his grant-making foundation, Bill Gates wrote about the importance of measurement to the design of social policies. In this, he drew an analogy between the role of data in developing a better steam engine in the early 1800s, and addressing modern challenges in medicine and education. If only we have the right measurements and better data, the argument runs, our problems can be solved. Yet, social problems rarely have simple solutions. The purpose of dialogue is to weigh up not only different evidence, but also to consider different notions of the good and debate the ends that we should pursue.

In the field of open data, just as in debates over evidence-based policy-making, there is a tension between the view that data can provide definitive answers, and the view that they are an input into a process of deliberation. In some cases, rather than bringing new dialogue participants to the table, open data might be used to bring in new ‘solution providers’, drawing on Government data and using statistical models to suggest optimum policies. The rise of smart-cities projects, in which large corporations seek to gain access to data flows from across urban area and to optimise everything from energy policy to transport flows, is just one example of this happening. Here, statistics captured in datasets and the algorithms that process them are what drives policy and practice, rather than citizens’ stories and lived experience. Transparency may let those with the capacity to analyse dataset increase their control over policy, rather than distributing and decentralising control, as many advocates of open data have suggested it would.

The ways in which transparency and open data policy moves forward from here will have a big impact on the outcomes we see: whether providing resources for dialogue, or displacing it. The following

77 http://clearclimatecode.org/
section explores the need to move beyond a transparent and open-data policy that is simply based on disclosure of information to one that is based on supporting dialogue with and around datasets.

6.6 Open data and ongoing dialogue

Meijer and others\textsuperscript{79} divide open government into two components: vision, and voice. Under vision, they place reactive (FOI) and proactive (open data/publication schedules) transparency. Under voice, they place formal and informal arenas for citizens to engage in decision-making. Based on a review of 103 papers, they suggest that ‘vision’ components of open government are rarely connected with initiatives on voice, and argue that “open government is much too important to leave it to the ‘techies’: scientists and practitioners with backgrounds in law, economics, political science and public administration should also get involved to build sound connections between vision and voice”. What might these connections look like?

De Cindio suggests taking datasets and visualisations, and embedding them in online deliberative spaces to create discussions around data\textsuperscript{80}. There has been a growing recognition in e-participation over recent years of the need to take discussions to where people are rather than necessarily expecting everyone to come to some central space. De Cindio points to the need to take open data out into a variety of digital environments and to focus attention on the technical and social features of environments that are needed to support discussion around data. The Five Stars of Open Data Engagement\textsuperscript{81} takes a different tack, highlighting the features that a Government open-data initiative needs to support dialogue. It suggests that open data initiatives need to:

\begin{itemize}
  \item ★ Be demand driven – focusing attention on the data that citizens ask for and prioritising data release based on demand
  \item ★★ Put data in context – with good metadata (i.e. descriptions of where the data came from, guidance on how to analyse them and examples of existing analysis)
  \item ★★★ Support conversations around data – online and offline. These conversations should be able to involve people from inside Government who know the dataset and it should be possible for citizens to communicate with the data owner
  \item ★★★★ Build capacity, skills and networks – don’t stop at just publishing data, seek to build communities around the datasets and make sure all key stakeholders have the capabilities they need to work with them
  \item ★★★★★ Collaborate on data as a common resource – recognising that Government should be open to data coming in from citizens and giving data out; and that tools and services for working with data can be created collaboratively between citizens, state and private enterprise
\end{itemize}

This model envisages open data as a locus for ongoing dialogue between citizens and state, but also points to the significant work and culture change required to make this happen. Right now, data portals tend to act as a firewall between citizens interested in data and the civil servants responsible for those datasets. This makes each invisible to the other, rather than being spaces to connect together producers and users of information.

As well as dialogue around individual datasets, there is also a need for dialogue around open-data policies and the underlying data collection practices of Government as a whole. Sciencewise has already hosted an ‘open-data dialogue’ exploring citizen attitudes to wider sharing of datasets, many of which are built up from data collected from citizens during the course of their interactions with Government in the first place\textsuperscript{82}. Where the line lies between openness and individual privacy or between transparency and protected space for policy deliberation are important questions, as are questions about what data Government should collect in the first place. For example, the coalition

\textsuperscript{81} "www.opendataimpacts.net/engagement/"
Government decommissioned many surveys that previously provided data on policy impact and there have been discussions of whether, in future, the census may be replaced by data bought in from private firms (e.g. credit reference agencies). The consequence of this is that such data may be trickier to provide openly due to their commercial nature\textsuperscript{83}. Jo Bates has argued\textsuperscript{84} that transparency and open data are not neutral and can be used as political tools in service of particular agendas. Maintaining trust in transparency and open data requires good governance regimes to be established around it and these should include public representation. Whether the sector transparency boards that have been established in many UK Government departments will provide this role remains to be seen.

6.7 Open everything?

There are aspects of transparency that remain distinct from open data. In calling governments to account, access to documents through reactive transparency and FOI rights is likely to remain important. Not all the information that is needed for dialogue exists as datasets – much may be in the form of narrative accounts, evidence and opinions. However, the rise of open data does represent a significant shift in how information flows between state and citizens.

In a world of open data, open source and open access, authority is no longer secured through being in a privileged position with respect to some key information. Rather, it has to be produced through public discourse and performance\textsuperscript{85}, and involves appeals to data and to lived experience. This involves a cultural shift.

6.8 Conclusions

Transparency and open data are resources for classic dialogue and the basis for new forms of ongoing dialogue. They bring new actors into the public sphere and allow information to be presented in ways that make issues more accessible. Yet, without careful attention to process, open data can also disempower those whose stories are not captured in the statistics and those who do not have the capacity to conduct arguments through numbers and datasets. Also, as part of a wider data deluge, it can tip the balance of power in favour of those with the computing power and skill to process vast quantities of data.

Although it might be 250 years since Sweden first put government transparency on the statute book, the way in which contemporary transparency and open data will unfold in practice very much remains to be shaped and seen.

7 Complexity, public policy and public dialogue

Greg Fisher

Greg is the Managing Director of Synthesis, a think tank devoted to using the emerging paradigm of complex networks in the social sciences to tackle social and public policy concerns. He is also a Senior Research Associate of the London School of Economics’ Complexity Group. Prior to establishing Synthesis with Paul Ormerod and Bridget Rosewell in 2011, Greg spent 2 years researching the new science of complex systems, and how it relates to economics and finance, at the think tank ResPublica. Greg has also worked for a hedge fund as a global macroeconomic strategist and at the Bank of England in a spectrum of roles that mixed economics and finance.

\textsuperscript{83} See discussion at www.opendataimpacts.net/2012/12/notes-on-a-national-information-infrastructure/
