



Open Data and Democratic Governance: Policies, Platforms and Practices

Ethics Case Study | 6

Open Data and Democratic Governance: Policies, Platforms and Practices

Anonymous

This is an application form to an ethics committee for research on open data and democratic governance.

Describe the rationale, study aims and the relevant research questions of your study.

Governments across the world are adopting open data policies, leading to the creation of national Open Government Data (OGD) portals (such as data.gov.uk) that provide access to machine-readable datasets offered under permissive intellectual property licenses, and driving a range of interventions designed to stimulate and support re-use of government data.

These policies draw on a range of justifications, including the contribution of open data to transparency and accountability, to supporting citizen involvement in government, and to enabling innovation and economic growth through commercial exploitation of government collected data (Tauberer, 2012). However, for all the recent focus on OGD policy in fora such as the Open Government Partnership (Turner, 2012) and G8 (G8, 2013; Global Integrity, 2012), little is understood about how open data is being used in practice, nor about how certain policy and technology choices in the design of OGD initiatives are shaping the outcomes that may be realised from open data.

This study aims to contribute to practical and theoretical understanding of the relationships between open data and processes of democratic governance by focussing on both the policy and technical environment around open data supply, and the practices of open data use undertaken by diverse actors.

This case study was originally published in draft form on the British Sociological Association Digital Sociology [Study Group blog](#) (2016) under the CC BY NC ND licence.

It addresses three core research questions RQ1-3:

RQ1: How can open data contribute to improved governance, and in particular, to more democratic governance?

RQ2: How do choices in the adoption of technical platforms and standards for open data affect data re-use, and in particular, governance-related re-use of open data?

RQ3: How far are the policies, platforms and practices that support democratic governance uses of open data compatible or complementary with those that support alternative use-cases for open data?

The research is motivated by a commitment to realising the democratic potential of open technologies, and a critical interest in development applied Web Science methodologies.

Describe the design of your study.

The study is designed within a transdisciplinary problem-centred research framework (Leavy, 2011) drawing upon mixed methods. It forms a companion study to a wider research programme titled, coordinated by the researcher in his role as a part-time employee of and funded by. Further details on the relationship of these studies is provided below.

The study follows an extensive-intensive design, addressing RQ1 through a broad analysis of open data policies and practices, followed by an intensive and detailed look at a small number of examples of open data in use in specific governance decision making contexts to address RQ2 and RQ3.

The methods adopted will include:

M1) Secondary analysis of OGD policies and initiatives - a close reading of the Open Government Data policies and associated materials of the UK, USA, India, Kenya and Brazil will be undertaken to identify the justifications offered, and goals set out for, their respective OGD initiatives. A complementary analysis of information on open data policies gathered through the Open Data Barometer (see §11) will also be undertaken to ensure that the selected countries (chosen for their geographical spread and existence of open data policies) are broadly representative of wider OGD policy, and to identify further justifications and goals of OGD.

M2) Secondary analysis: case study synthesis - performing an analysis of case studies produced by participants in the research programme to identify distinct 'theories of change' relating open data to governance decision making (see §11 for more details). [Addressing RQ1 and RQ2]

M3) Primary data collection: 'Follow the data' and Key Informant interviews - carrying out a series of purposive sampled semi-structured interviews with suppliers, intermediaries and re-users of OGD inviting them to describe their goals for open data publication and re-use, the steps they have undertaken to publish or use OGD, other resources they have drawn upon, and how their actions have been enabled or constrained by certain technical and policy contexts they have been operating within. The interview approach and sampling design will be informed by the Socio Technical Interaction Networks (STIN) research strategy of Kling et al. (2003), identifying relevant populations of system interactors, and paying attention also to excluded actors, and potential users of OGD whose attempts to use OGD may not have been successful.

The particular uses of data addressed in these interviews will be determined based on a range of factors, including the possibility of finding examples of practices that engage with a number of different theories of change as identified in (2). Candidate governance settings for these interviews include: the use of public spending data in the UK and USA; the use of open data on crime in the UK and USA; the use of open data on aid spending.

These interviews will be complemented by analysis of secondary materials to validate and cross-check the evidence given. This may draw upon analysis of published documents, and a reading of open source code associated with particular cases of data re-use (Berry, 2011). [Addressing RQ2 and RQ3]

M4) Primary data collection: Participant observation - drawing on the researchers role as an active participant in a number of OGD communities, and his role working on the development of open data standards, a reflective research journal will be maintained to identify key themes to explore during analysis of data from (M1 - M3).



Who are the research participants?

Primary data collection (M3) will be undertaken with individuals and organisations involved in the supply, intermediation or re-use of specific open government datasets. These individuals will be identified through examining public online evidence concerning the supply and use of specific datasets, and use of a snowball sampling method, asking previous interviewees to suggest further informants who may be able to illuminate other aspects of open data supply and re-use.

If you are going to analyse secondary data, from where are you obtaining it?

Three main sources of secondary data enter the project. The first two are addressed in detail, as extra care has been taken here to ensure that the dual role of the researcher (as a PhD student at the University of Southampton, and independently as Research Coordinator of a multi-country study with the World Wide Web Foundation) is carefully managed to ensure a high standard of ethical conduct.

D1) The Open Data Barometer (used in M1)

- the Open Data Barometer (ODB) consists of a multi-country expert-survey of open data policies, implementation and impacts carried out between August and September 2013 by the World Wide Web Foundation, and secondary indicator data from a range of public sources. It is a companion study to the Web Index (Farhan et al., 2012).

In his role with the researcher has been involved in compiling the question framework and research methodology for the ODB and in administering the research. The ODB will be published in October 2013 along with a detailed openly licensed dataset that provides: expert survey scores against key questions on open data readiness and impacts; supporting evidence for those scores (e.g. links to national open data policies); checklists concerning the supply of 15 key national datasets as open data. The publication of this data was an explicit condition on the agreements with all the paid country expert-survey researchers who have undertaken data collection for the Barometer and no personally identifying information will be included in this public data. The expert-survey researchers have each been offered an explicit choice of whether their contribution to the study is kept anonymous or is acknowledged, and raw survey data is handled securely to ensure the safety of any researchers operating within oppressive countries or who express a desire for anonymity.

The data drawn upon in this study will be taken only from the publicly available and openly licensed Open Data Barometer dataset, and will be the same data that any third-party researcher could obtain.

D2) Case Study Data (used in M2) - the project consists of a research network of 17 funded case study research projects exploring uses of open data for governance across 12 countries in the global south (Brazil, Argentina, Uruguay, India, Nepal, Indonesia, Philippines, Nigeria, Sierra Leone, Kenya, Uganda, South Africa). In this role, the researcher co-ordinates this network, and has been responsible for the development of the funding bid that created the network and the overarching research framework for the network). The case study research is all carried out within the ethical framework provided by the Canadian Tri-Council Policy on Ethical Conduct of Research Involving Humans 2 (TCPS2). Over autumn 2013 each of the case studies partners will be publishing a report of their case research, and undertaking a number of writing activities with other network members to identify cross-cutting issues in their research.

The analysis of this data carried out in M2 will (a) draw on published case study materials, or materials prepared for publication (i.e. only those materials in, or destined for, the public domain under the existing ethical procedures of the project), and (b) materials shared with the researcher with explicit consent from the case study partners that it may be re-used in this study. The draft form for gaining this approval is attached, and is designed to ensure that the case study researcher have authority to grant the researcher permission to use any secondary materials, and to ensure that the power dynamics of the funder-funded project relationship are minimised. Throughout the research coordination process the researcher has been transparent about his role as a PhD student, that this is distinct from the project, and that participants are under no obligations to provide materials to support the researcher's PhD work.

D3) Online materials (used in M1, M2 and M3) including government policy statements obtained from government websites - websites detailing re-use of open government data; and openly licensed source-code of tools involved in the re-use of open government data. Principles of fair use will be applied to quoting of any copyright material, and where the license of source code is unclear, efforts will be undertaken to clarify that it is intended for public access and review. Materials from online communities, discussion forums or mailing lists will not be involved in this data collection, and additional ethics approval will be sought should such data collection be required in later stages of the study.

If you are collecting primary data, how will you identify and approach the participants to recruit them to your study?

M3 involves primary data collection through interviews with key informants and open data suppliers and re-users in UK and US cities. First, particular datasets of interest will be identified, and web searches used to identify individuals or organisations involved in the supply of this data (generally government officials), and re-users of the data. These potential informants will be contacted by email with details of the study and an invite to participate. Written informed consent will be sought before interviews by phone, skype or in-person are arranged, where consent to participate in the study will be reaffirmed verbally. Interviewees will also be asked to suggest further informants, who will be similarly approached by email, using the same process.

Will participants be taking part in your study without their knowledge and consent at the time (e.g. covert observation of people)? If yes, please explain why this is necessary.

There will be no covert observation involved in the study, and all participants will be fully informed prior to participation.

How will you obtain the consent of participants?

A copy of the consent form used is attached, this will be sent in advance of any interviews, with consent confirmed both in writing and verbally in each interview.

Is there any reason to believe participants may not be able to give full informed consent? If yes, what steps do you propose to take to safeguard their interests?

No.

If participants are under the responsibility or care of others (such as parents/carers, teachers or medical staff) what plans do you have to obtain permission to approach the participants to take part in the study?

N/A.

Describe what participation in your study will involve for study participants. Please attach copies of any questionnaires and/or interview schedules and/or observation topic list to be used.

The research will involve a short telephone or in-person interview based on a 'Follow the data' methodology (Sands et al., 2012). Follow up request for clarifications may be made by email or phone following the interview.

A draft copy of the semi-structured interview schedule is attached. [REMOVED]



How will you make it clear to participants that they may withdraw consent to participate at any point during the research without penalty?

This information will be provided on the participant information sheet and consent form and participants will be verbally reminded of it during interviews.

Detail any possible distress, discomfort, inconvenience or other adverse effects the participants may experience, including after the study, and you will deal with this.

No distress, inconvenience or adverse effects are anticipated from participation in the study.

How will you maintain participant anonymity and confidentiality in collecting, analysing and writing up your data?

All interviewees will be offered the option of anonymity before and after the interviews through the participant information sheet and consent form. In the open data field, many respondents have existing public online presence where they have self-reported they're involved in the projects that are covered by this study, and experience has demonstrated that a significant number of respondents explicitly prefer to be identified in any research outputs about these projects. Therefore, anonymity is only maintained where requested or agreed with the participant. If anonymity is requested then a close reading of the evidence will be conducted to ensure that any identifying information relating to these individuals is removed from the research write-up, and in analysis and preparatory notes pseudonyms will be used.

How will you store your data securely during and after the study?

The affiliated institution has a research data management policy, including for data retention.

Audio recordings and notes of interviews will be kept on an encrypted drive in a password protected laptop for the duration of the research (backed up to an encrypted backup service) and will be archived in encrypted form to external media following the study.

Describe any plans you have for feeding back the findings of the study to participants.

Interim findings from the study will be regularly blogged and the final PhD write-up will be published online with the link circulated to all those who contributed to the research.

What are the main ethical issues raised by your research and how do you intend to manage these?

In the secondary data analysis, the main ethical issue to be addressed is the potential for the dual role of the researcher, as a PhD student undertaking a thesis project, and, independently (as research coordinator a role which involves acting as funder to southern-based research organisations), to overlap and lead to case study partners feeling obliged to share research findings or evidence with the researcher. This potential tension is being managed through a clear informed consent process for the re-use of any materials in secondary analysis.

In primary data collection, the main ethical issue to be addressed is ensuring that the research does no harm to interviewees, their projects and their interests. This can be addressed by clearly offering anonymity and the opportunity to withdraw from the study at any time.

Please outline any other information you feel may be relevant to this submission.

More details on the Exploring the Emerging Impacts of Open Data in Developing Countries project can be found at <http://www.opendataresearch.org>

Please outline any other information you feel may be relevant to this submission.

More details on the Exploring the Emerging Impacts of Open Data in Developing Countries project see Davies, Perini, & Alonso (2013).

References

Berry, D.M. (2011) *The Philosophy of Software Code and Mediation in the Digital Age* (Palgrave Macmillan).

Davies, T. Perini, F. & Alonso, J.M. (2013) *Researching the emerging impacts of open data ODDC conceptual framework*. Available at [http://www.opendataresearch.org/sites/default/files/posts/Researching the emerging impacts of open data.pdf](http://www.opendataresearch.org/sites/default/files/posts/Researching%20the%20emerging%20impacts%20of%20open%20data.pdf)

Farhan, H., D'Agostino, D. & Worthington, H. (2012) *Web Index 2012*. Available at <http://thewebindex.org/2012/09/2012-Web-Index-Key-Findings.pdf>

G8 (2013) *G8 Open Data Charter: Annex*.

Global Integrity (2012) *So What's In Those OGP Action Plans, Anyway?*. Available at <http://globalintegrity.org/blog/whats-in-OGP-action-plans>

Kling, R., McKim, G. & King, A. (2003) A Bit More to It: Scholarly Communication Forums as Socio-technical Interaction Networks, *Journal of the American Society for Information Science and Technology* 54(1): 47–67. Available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.15.9909\&rep=rep1\&type=pdf>

Leavy, P. (2011) *Essentials of Transdisciplinary Research: Using Problem-Centered Methodologies* (Left Coast Press). Available at <http://www.amazon.co.uk/Essentials-Transdisciplinary-Research-Problem-Centered-Methodologies/dp/159874593X>

Sands, A., Borgman, C. L., Wynholds, L. & Traweek, S. (2012) Follow the Data: How Astronomers Use and Reuse Data. In *ASIST 2012*. Available at <http://asis.org/asist2012/proceedings/Submissions/341.pdf>

Tauberer, J. (2012) *Open Government Data*. Available at <http://www.amazon.com/Open-Government-Data-ebook/dp/B007U69DKU>

Turner, L. (2012) 'So What's In Those OGP Action Plans, Anyway?', available at <http://globalintegrity.org/blog/whats-in-OGP-action-plans>