Supporting open data use through active engagement

Position paper: W3C Using Open Data Workshop – June 2012

Creating, accessing and using government data (OGD) are all socially situated processes. Whilst technology platforms, common standards and open licenses all play a key role in making data easier to access and simpler to process, they operate against a wider backdrop or organisation and social arrangements, power dynamics, and market conditions which may or may not be favourable to the use of data by different agents to influence policy, generate economic value, or improve the delivery of public services. The gap between the promise and reality of OGD re-use cannot be addressed by technological solutions alone. This paper reviews and collates a number of theoretical and practical perspectives on the development of open data use that point to organisational, economic and social interventions needed to support the greater realisation of value from OGD.

The human element: creating a culture of open data engagement

Whilst data portals can act support citizen access to data, conceived of as a purely technical intervention they can also create a barrier between citizens, journalists or companies with an interest in certain data, and the officers inside government who hold responsibility for that data or for the policy area it relates to. A sense that the open data task is complete once data is published online misses out on the greater potential of open data to act as the start of conversations and collaboration between those inside, and those operating outside of government. Opening up state data for wider use is not new (though the scale at which it is happening with OGD initiatives is perhaps unprecedented), and there is learning to be gained from past work of Local Information Systems (LIS) and regional observatories that have sought to increase use of data by both effective curation of data, and by investing in outreach and capacity building with potential users of data. Introducing the human element to open data initiatives involves recognising the key role that existing data stewards within public organisations can play in making their datasets easier to understand, interpret and work with, and supporting the emergence of new or expanded roles focussing on open data engagement.

At the 2012 UKGovCamp event in London one workshop sought to develop these ideas into a 'Charter of Open Data Engagement', which led to the creation of a draft '5 stars of open data engagement model', seeking to find a non-technical parallel to the '5 star linked data' model (Berners-Lee, 2010). The draft engagement model articulating overarching principles of OGD, five broadly cumulative steps to greater engagement around OGD, and a series of practical questions that could be asked to identify the extent to which an organisation is addressing each 'star'.

Five stars of open data engagement

Government information and data are common resources, managed in trust by government. They provide a platform for public service provision, democratic engagement and accountability, and economic development and innovation. A commitment to open data involves making information and data resources accessible to all without discrimination; and actively engaging to ensure that information and data can be used in a wide range of ways.

Engaging open data should:

★ *Be demand driven*

- Are your choices about the data you release, how it is structured, and the tools and support provided around it based on community needs and demands?
- Have you got ways of listening to people's requests for data, and responding with open data?

★★ *Put data in context*

- Do you provide clear information to describe that data you provide, including information about frequency of updates, data formats and data quality?
- Do you include qualitative information alongside datasets such as details of how the data was created, or manuals for working with the data?
- Do you link from data catalogue pages to analysis of the data that your organisation, or third-parties, have already carried out with it, or to third-party tools for working with the data?

★★★ *Support conversation around data*

- Can people comment on datasets, or create a structured conversation around data to network with other data users?
- Do you join the conversations?
- Are there easy ways to contact the individual 'data owner' in your organisation to ask them questions about the data, or to get them to join the conversation?
- Are there offline opportunities to have conversations that involve your data?

★★★★ *Build capacity, skills and networks*

- Do you provide or link to tools for people to work with your datasets?
- Do you provide or link to How To guidance on using open data analysis tools, so people can build their capacity and skills to interpret and use data in the ways they want to?
- Do you go out into the community to run skill-building sessions on using data in particular ways, or using particular datasets?
- Do you sponsor or engage with capacity building to help the community work with open data?

★★★★★ *Collaborate on data as a common resource*

- Do you have feedback loops so people can help you improve your datasets?
- Do you collaborate with the community to create new data resources (e.g. derived datasets)?
- Do you broker or provide support to people to build and sustain useful tools and services that work with your data?
- Do you work with other organisations to connect up your data sources.

Developed with input at UKGovCamp or subsequently online from: @timdavies; @hokulele; @anthonyzach; @jonfoster; @jagusti and @exmosis

Engaging all users: escaping the 'application fallacy'

Many narratives of open government data implicitly or explicitly assume that the products of OGD use will be 'apps'. Data portals focus on providing application catalogues, and soliciting details of applications built with OGD, and presentations about OGD focus on data visualisations. Yet, many of the important patterns of OGD re-use, and areas where OGD generates impact, involve neither application development nor visualisation (aside perhaps from spreadsheet graphs).

In an analysis of 55 instances of open data use from the data.gov.uk portal, the 'Open Data, Democracy and Public Sector Reform Project' (Davies, 2010) found that many open data uses involved directly identifying facts within datasets (where simple tabular data was most useful to re-users), turning data into information (writing discursive or analytical reports that could draw on the available data), and generating derivative datasets for others to use (see table).

Process (n=instances)	**Summary** (and example)
Data → Fact *Search* *Browse* ***Extract*** (n=8)	**A dataset is used directly to identify a specific fact of interest.** E.g. Finding out the voting history of a local constituency.
Data → Information *Manipulate* *Statistically analyse* *Visualise* *Contextualise* ***Report*** (n=19)	**Content from a dataset is given a single representation or interpretation that is reported in text or graphics.** E.g. Composing a report that "profile [s] communities of interest within [the local area] as part of the Council's equality & diversity agenda".
Data → Interface *Clean, Combine, Subset Data* *Configure interface tools* *Write custom code* ***Provide interface*** (n=26)	**An interface is provided allowing interactive representation of a dataset – providing information customized to the user's input.** E.g. Creating a searchable interactive online map of stations and former British rail assets.
Data → Data *Convert format* *Filter data* *Augment/combine data* ***Provide API*** ***Dataset for download*** (n=17)	**A derivative dataset is provided for download, or access via an API** E.g. I "took Westminster Constituency data, combined it with scraped [General Election] 2005 data from exposed it as RDF."
Data → Service *?* ***Integrate into existing product/service*** ***Create new service*** (n=4)	**A service is provided that relies on open data, whilst not necessarily exposing it to the end-user.** E.g. Using boundary data from the Census to run an application that forwards reports of Potholes to the correct Highways authority.

www.practicalparticipation.co.uk/odi/report

For many uses of OGD, re-users will turn to the tools they already have available, often desktop software and spreadsheets. Whilst some OGD enabled big-data applications, and flexible OGD based consumer services will require the creation of custom applications, in understanding how OGD is being used, and supporting a greater range of re-users, we need to keep in mind the everyday practices that will shape how many users adopt and work with data.

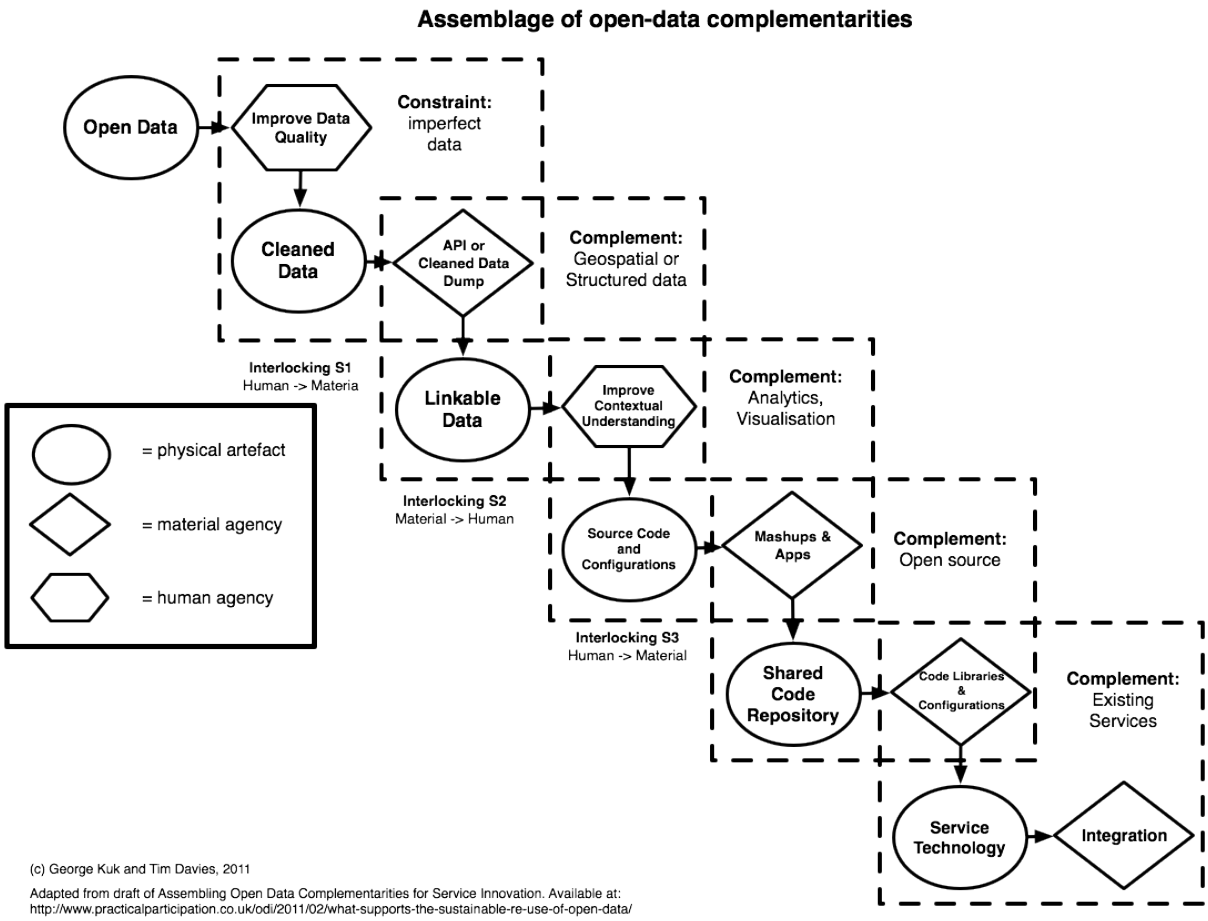
New institutional designs: articulating open 'governance' data

The Crossroad Project Roadmap highlights the need to explore new institutional designs for collaborative governance (RC2.7) in light of the growing availability of open data. Although an open data engagement framework goes some way to articulate the idea of collaboration and co-creation of policy, analysis and products based on government datasets, we can go further to explore how collaborative governance requires a rethinking of the category of 'government data'.

The particular datasets government holds are the product of past political choices, resource constraints and practical considerations (Bowker and Star, 2000; Scott, 1998). The monopoly role that government plays in collecting and curating certain datasets is often an artefact of earlier technologies of data collection and processing. By contrast, with digital technologies, there are significant possibilities for a greater range of datasets to be brought into the policy making process, from private actors, from international institutions, from civil society organisations, and crowd-sourced and co-created by citizens. We might term the total domain of data relevant to policy making as 'open governance data', and when we consider increasing the use of open data, we should question whether the division of data into that which comes from government, and that which comes from other sources is the most relevant one, or whether other distinctions of data provenance are more relevant.

Supporting re-use to move from prototype to product

When it is used to generate applications or visualisations open data can support the rapid creation of prototypes. However, even simply prototype creation often involves a number of steps, from cleaning up data for re-use, linking datasets on common keys, developing source code or configurations to work with a particular dataset, and so-on. Observing participants at a hack-day, or watching established open data communities responding to a new dataset release, we can see technical artefacts of derived datasets or snippets of source code being shared amongst peers, but to outsiders, these artefacts that can make it easier to build off a particular open dataset often remain invisible. The figure below highlights the interaction of human actions and material artefacts in gaining value from open data. Note that often it is only the start and end of this chain that are shared widely, or that portals support the discovery of.



For many open data applications, there is a significant challenge in moving from the prototype stage, to providing an established and generally accessible product. If relying on services (e.g. APIs or derived datasets) provided by third parties, applications developers need a degree of guarantee that the service will continue to exist. The experience of the International Aid Transparency Initiative (IATI) shows that many third-party APIs may be built around raw data, but that maintaining these APIs (dealing with quirks in incoming data, or just keeping them refreshed regularly with new data) involves significant effort for their respective owners, which can lead quickly to stale data and broken applications. Finding models to fund the maintenance of 'public good' layers between open data and simple re-use of that data is a key challenge. Equally, funding and procurement models are needed to support small-scale investment in promising prototype data-driven applications.

Missing pieces in research and technology

Recognising the diverse uses to which data is put, the need for engagement around datasets, the fact that many uses of OGD will draw also upon non-governmental datasets, and that sustainable open data re-use involves a number of layers between data and application or use, we can (a) suggest some areas in need of further research; and (b) highlight some areas for technical innovation.

In the research domain: Firstly, we need to record and understand in more detail all the steps between the release of open data, and a particular application or use: anecdotal accounts tend to focus only on the dataset and the end product, restricting opportunities for us to identify important barriers and enablers relating to the use of open data. Secondly, we need to explore the existing organisational arrangements around particular datasets and their subject domains: for example, there may be little point for an individual to invest significant time in visualising, analysing or exploring a dataset of policy relevance if there is little political chance of change in that policy domain. Thirdly, we need to understand data use from the citizen perspective, identifying the full range of data sources that might meet an information need, and looking at how open government data does, or could, fit into this. And fourthly, we need to explore the micro-economics of taking an open data use from being a prototype, to being an established accessible product.

In the technical domain, whilst tools to support data analysis and visualisation are undoubtedly needed, there is a pressing demand to also explore tools to support social communication and interaction around data and the artefacts that result from open data use. At its simplest this might involve modifying open data portals to better invite users to feedback on how they have used given datasets (going beyond the narrow invitation to 'Tell us about the apps you have built') and republishing this information. Further developments might include identifying effective ways to aggregate conversations and knowledge sharing around particular datasets, addressing the current fragmented sharing of data analysis and re-use that means someone looking at a dataset will find it hard to discover all the different places on the web where it has been used, or where value-added versions of the data are available. Such interventions could have the added value of increasing our understanding of open data re-use, and supporting the development of improved strategies of open data engagement.

Bibliography & links

Open Data Engagement: http://www.opendataimpacts.net/engagement/

IATI / AidInfoLabs: http://www.aidinfolabs.org

Berners-Lee, T. (2006, July). Linked Data - Design Issues. Retrieved from http://www.w3.org/DesignIssues/LinkedData.html

Bowker, G. C., & Star, S. L. (2000). *Sorting Things Out: Classification and Its Consequences (Inside Technology Series)* (p. 390). MIT Press. Retrieved from http://www.amazon.co.uk/Sorting-Things-Out-Classification-Consequences/dp/0262522950

Davies, T. (2010). *Open data, democracy and public sector reform: A look at open government data use from data.gov uk*. Practical Participation. Available from http://www.opendataimpacts.net/report/

Davies, T. G., & Bawa, Z. A. (2012, December 4). The Promises and Perils of Open Government Data (OGD). *The Journal of Community Informatics*. Retrieved from http://ci-journal.net/index.php/ciej/article/view/929/926

Kuk, G., & Davies, T. (2011). *The Roles of Agency and Artifacts in Assembling Open Data Complementarities*. *Thirty Second International Conference on Information Systems* (pp. 1-16).

Scott, J. C. (1998). *Seeing like a state*. Yale University Press New Haven, CT.