

Researching the emerging impacts of open data

ODDC conceptual framework

July 2013 – ODDC Working Papers #1

Authors

Tim Davies | tim.davies@webfoundation.org | Web Foundation

Fernando Perini | fperini@idrc.ca | International Development Research Centre

José M Alonso | josema@webfoundation.org | Web Foundation

This working paper presents initial literature review and research design work carried out from June 2012 to March 2013 in order to inform the development of the **Exploring the Emerging Impacts of Open Data in Developing Countries** programme. The programme, funded by IDRC and managed by the World Wide Web Foundation, will carry out research over 2013 - 2014. This framework will be refined based on the results and debates emerging from the project.

The funding for this work has been provided by grant 107075 from Canada's International Development Research Centre (web.idrc.ca). Find out more about the ODDC project at www.opendataresearch.org/emergingimpacts

Supported with funding from:



International Development Research Centre
Centre de recherches pour le développement international



Table of contents

1. Introduction	3
2. Open data: definitions, history and dilemmas	5
Operational definitions	6
3. Approaches to open data research	9
4. Open data, governance and emerging impacts: a conceptual framework	12
Open Data	13
Data	13
Licenses	13
Data standards	14
Intermediaries	14
Reconsidering openness	15
Governance	15
The emerging impacts:	16
Transparency and accountability	17
Inclusion and empowerment	19
Innovation and economic development	19
Conceptual framework: review	21
5. Putting the framework into practice	23
Case studies	23
Common assessment methods	26
Cross-cutting research	26
6. Conclusions	29
Bibliography	31

Introduction

1. Introduction

The open data movement holds out the promise of improving transparency, accountability, citizen participation and economic opportunity across developing countries. Citizens in Brazil, Nepal, and Nigeria can use publicly available data on government budgets to track and fight corruption, or to critique public spending policies. Developers and entrepreneurs across Latin America, Africa, and Asia can create web and mobile applications using government data on education, health, and crime, with the potential to promote smarter and more efficient local public services. And donors and advocacy organisations are investing in open data, opening their own datasets, or pushing for open data as part of open government reforms.

Nevertheless, it is not yet clear if open data initiatives are truly delivering on their promises. Worldwide, it is estimated that governments have already posted more than one million datasets on the Internet. Although just a small fraction of these current datasets are from developing countries, this is rapidly changing. Through the Open Government Partnership, governments from more than 50 countries have made concrete commitments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance with a strong emphasis on open data as a means to achieve this. Yet, reliable evidence on the outcomes and impact of open data initiatives remains scarce. Little is understood about how social and political context, open licenses, technical platforms and standards, and the dynamics of data use in different fields affect the outcomes that can be realised from wider sharing of data. It is even possible that well intended initiatives may result in adverse effects by exacerbating inequalities, and negatively impacting existing governance structures.

As open data initiatives spread across the globe, research is needed that can deepen our shared understanding of the potential and practice of open data. Those involved in new programs and initiatives in developing countries need to understand the full value and impact of open data in strikingly different social, economic, and cultural contexts. However, researching open data is a formidable challenge. The publication and use of open data raises many issues, cutting across fields from budget transparency or urban governance, to innovation policy and natural resource management. Open data also connects across many levels of activity, from community-led standard setting, to national government data collection, to grassroots use of datasets. As this paper argues, we can only understand open data in general if we have a detailed understanding of how it operates in specific situations. Open data research also has to engage with the fact that it remains early days in the evolution of open data as a mainstream policy. Of all the government datasets in the world, or all the NGO handled data, only a very small proportion is currently openly licensed and accessible online. Far from removing the need for research, this highlights the need to develop clear evidence and research approaches that can equip diverse stakeholders to engaged in informed dialogue and action to guide the future development of open data.

This working paper seeks to contribute to the conversation on open data research, focussing in particular on open data in developing countries.

In the following sections we offer a brief overview of open data definitions and recent development, before turning to look at different approaches for researching open data. We outline a twin-track approach of looking at macro-level assessments of the context open data operates within, and detailed comparative case studies of open data in use. We then focus in on this second track, exploring the need to connect the study of open data to the study of existing governance processes in transparency and accountability, innovation and economy growth, and inclusion and empowerment. We follow this by outlining a number of open data specific issues that cut across different the different settings where open data may be in use. We end by bringing these elements together in a research framework, and outlining some of the ways in which the IDRC/Web Foundation 'Exploring the Emerging Impacts of Open Data in Developing Countries' research programme will be applying this framework over 2013 – 2015.

Open data: definitions, history and dilemmas

2. Open data: definitions, history and dilemmas

In 2006 a group, inspired by open source software movements published version 1.0 of a statement they called the Open Knowledge Definition (OKF - Open Knowledge Foundation, 2006). This puts forward a definition of what it means to have open content, stating that: “A piece of data or content is open if **anyone** is **free to use, reuse, and redistribute** it — subject only, at most, to the requirement to attribute and/or share-alike” (original emphasis). Applied to data it requires that a dataset be accessible (usually by being online) at no cost, and with no technical restrictions to prevent its re-use. Other widely used principles (see Box 1) have elaborated additional criteria for open *government* data, highlighting the importance of timely publication, and of providing ‘raw’ data, or data at the lowest possible level of disaggregation and granularity (Malmud & O’Reilly, 2007; Transparency International Georgia, 2012).

Sebastopol Open Government Data Principles

Government data shall be considered open if it is made public in a way that complies with the principles below:

1. Complete: All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.
2. Primary: Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.
3. Timely: Data is made available as quickly as necessary to preserve the value of the data.
4. Accessible: Data is available to the widest range of users for the widest range of purposes.
5. Machine processable: Data is reasonably structured to allow automated processing.
6. Non-discriminatory: Data is available to anyone, with no requirement of registration.
7. Non-proprietary: Data is available in a format over which no entity has exclusive control.
8. License-free: Data is not subject to any copyright, patent, trademark or trade secret

Box 1: Open Data Principles, <http://www.opengovdata.org/home/8principles>

The Open Knowledge Definition (OKD), and the related principles, offered a point of consolidation for many distinct groups, all seeking some form of greater access to government datasets. These included commercial firms seeking to build or develop business based on Public Sector Information (PSI)¹, such as state-generated mapping or weather data (Burkert, 2004; Dekkers, Poleman, te Velde, & de Vries, 2006a; Newbery, Bently, & Pollock, 2008; Uhlir, 2009); civil technologists who had developed websites that promoted government accountability by ‘scraping’ data, and who wanted reliable access to the data in future (Tauberer, 2010, 2012); transparency campaigners seeking proactive disclosures from government (OKF - Open Knowledge Foundation & Access Info, 2011); public sector reformers seeking to use technology for more efficient public services (e.g. Steinberg & Mayo, 2007); and open knowledge activists rejecting the assertion of restrictive copyright over publically funded data (Krikorian & Kapczynski, 2010). These groups, with diverse interests, have come to constitute a global movement for open data (Davies & Bawa, 2012).

A pattern for how that movement might evolve globally was set in 2009, when US President Obama’s ‘Open Government’ memo on transparency, participation and collaboration, led to the creation of data.gov, a ‘data portal’ bringing together on one website a range of US government datasets. This was quickly followed by the launch of data.gov.uk in the United Kingdom in 2010, similarly curating government datasets, released by departments in response to a programme of advocacy led by web creator, Tim Berners-Lee. The creation of a data portal provides governments with a concrete action and focal point for opening data. Data portals have often been launched with only a limited number of datasets that would pass muster against the Open Knowledge Definition, being in non-standard formats, or under restrictive licenses. However, data.gov, data.gov.uk and a number of other open government data initiatives (sometimes called ODIs or OGDIs) have sought to increase the quantity and quality of published data over time, including converting selected datasets into linked data formats, and developing new standards for representing certain datasets such as public spending.

¹ Public Sector Information (PSI) can be defined as “information products and services, generated, created, collected, processed, preserved, maintained, disseminated, or funded

Over the course of 2010, as national and city-level open data portals began to spread (some launched by governments, others created by citizen groups independent of government) multilaterals also joined the move towards open data. The World Bank's open data portal was launched in April 2010 providing access to indicator datasets, and starting a World Bank focus on open data that has seen the organisation promote open data to client governments, including supporting the creation of open government data initiatives in Kenya and Moldova (Majeed, 2012; Rahemtulla et al., 2011, 2012). The potential of open data in the developing world was also explored in a study commissioned by the Transparency and Accountability Initiative (Hogge, 2010), which has formed the basis for a number of feasibility studies into whether open government data initiatives might be developed in Ghana, Chile, Uganda, and most recently in Indonesia (Alonso, Boyera, Grewal, Iglesias, & Pawelke, 2013; APC & CIPESA, 2012; Grewal, Iglesias, Alonso, Boyera, & Bratt, 2011; Iglesias, 2011). The globalisation of open data has been further pushed through the Open Government Partnership (OGP), a voluntary association of over 50 nations pledging to "increase access to new technologies for openness and accountability" (Open Government Partnership, 2011). Commitments to open specific datasets, or launch open data portals, have been amongst the most common in country action plans formed during the OGP process (Global Integrity, 2012).

Whilst even in developed countries no rigorous studies yet attribute large-scale impacts to open data, a number of common forms of practice for using open data have emerged. These range from hack-days that bring together technology developers to work with government datasets and 'apps competition' that seek to harness innovation from outside government to generate public services and economic growth (Kuk & Davies, 2011; Shemie et al., 2012), through to 'data journalism' making use of government datasets to generate and add to media stories (Gray, Chambers, & Bounegru, 2012). Together these have come to constitute a 'standard model' of open data, in which open data initiatives are understood to release pre-existing public sector information resources, and technical skilled intermediaries engage in finding and extracting the social and economic value from that data. In Kenya, for example, considerable investment from the World Bank and other donors has gone into building the capacity of the technical community to unlock the potential of the Kenya open data portal.

This growth in the supply of open data has been justified with reference to a wide range of claimed benefits, from the potential of open data to stimulate innovation and economic growth, to support transparency and accountability, and to promote inclusions and empowerment of citizens.

Understanding how far this standard model applies effectively in developing countries, and which elements of it are more or less important for certain desired outcomes from open data is one important area for research. However, with its rapid growth from niche interest less than 10 years ago, to a part of the contemporary policy mainstream, there are many other aspects of open data also to be researched. The next section outlines two complementary approaches to open data research.

Operational definitions

The 'standard model' of open data outlined above, made up of open data initiatives, data portals and technical-mediated use of open datasets for social and economic return, was largely born out of developed country experiences. In looking at open data in developing countries, where the extent of government data holdings, the nature of political relationships, structures of the economy, and levels of Internet access (amongst other factors) may differ sharply from those where the standard model of open data developed, it is important for us to ensure that the definitions that guide our study do not prevent us from understanding both similarities and differences between the trajectories that open data might take in different contexts. For that reason, whilst working with reference to the stringent conditions of the Open Knowledge Definition, and Sebastopol principles, in outlining a research agenda we will operate on the slightly broader understanding that open data is data that is:

- **Generally accessible online** as evidenced by, for example, its inclusion in a national data portal, or the fact that it is being widely accessed by a range of actors operating independently of one another;
- **Machine readable** as evidenced by the use of non-proprietary digital formats, and the data being structured in ways that allow it to be filtered, sorted, reshaped and manipulated without

copying/pasting or re-typing in data;
and

- **Practically / legally re-useable** which may involve the availability of an open license that grants explicit permissions, or may involve the existence of wider legal or cultural frameworks that enables the practical re-use of the data.

It is also important to note that these definitions are not specific to government data, but recognise that the data in use in developing countries (and, for that matter, developed countries) in processes of transparency and accountability, inclusion and empowerment and innovation and economic development, may not only come from government. International agencies, local civil society and the private sector can all form an important part of the data landscape; and in exploring the emerging impacts of open data it is important to adopt definitions that do not exclude this data from focus.

Approaches to open data research

3. Approaches to open data research

Over its short history as a field of action a number of distinct fronts of research into open data have developed, responding to different practice, policy and knowledge needs. Figure 1 suggests these can be usefully classified into three broad groups: open data readiness assessments; open data implementation studies; and impact studies.

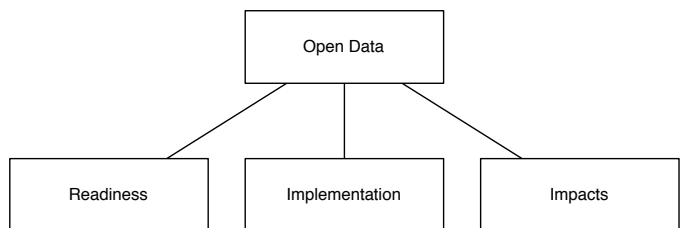


Figure 1: Approaches to open data research

Readiness studies seek to assess whether the conditions in a country, city or sector might be appropriate for an open data initiative to be effective, and may seek to also highlight areas where investment or effort would be needed to get ready for an open data initiative. A readiness study generally needs to have an existing notion of what an open data initiative would consist of, and to decide which factors are important for making an open data initiative a success. So far, readiness studies, such as those from the Web Foundation (e.g. Grewal et al., 2011), the World Bank (Stott & Kaplan, 2013), and CIPESA (APC & CIPESA, 2012), have relied upon the example of open data initiatives in developed countries (in particular, the US and UK examples), as the template for developing a readiness assessment.

Studies of implementation seek to assess whether the conditions for open data, or open data itself, actually exist in a country, city or sector: that is, whether open data policies are being implemented. These studies vary widely in their approach. Some look to count the availability of datasets (OKF - Open Knowledge Foundation, 2012) or the proportion of institutions publishing open data, others focus on the qualities of datasets or open data portals (Braunschweig, Eberius, Thiele, & Lehner, 2012; Craveiro, Santana, & Albuquerque, 2013; Garcia, 2011; Harper, 2012; Murillo, 2012), and others assess whether appropriate legislation and regulations are in place for open data to operate effectively. Again, these studies have to make decisions about what good open data looks like, which datasets are important to survey, and how implementation should be measured. A number of studies, the Web Foundation's first Web Index-derived Open Government Data Index (Alonso, 2012; Farhan, D'Agostino, & Worthington, 2012) included, have sought to reduce implementation measures to a simple index number, offering the possibility of comparing open data implementation in different places as a means to motivate action.

Impact studies ask whether open data has led to change. Generally they focus on whether open data has brought one of the specific benefit that open data advocates suggested would result from opening datasets – such as economic growth (Hammell, Perricos, Branch, & Lewis, 2011; Tong, Irsha, & Revell Ward, 2013) or democratic empowerment (Davies, 2010). As yet, there are no large-scale rigorous studies of open data impacts, and most work remains at the level of ad-hoc and isolated case studies or anecdotes. Many of the existing case studies highlight the complexity of open data adoption and use, and the ways in which local politics and power play a significant role in shaping outcomes. For example, Rath (2012) describes how a project using data to improve district economic planning was derailed when elections distracted politicians from their engagement in the process, and Raman (2012) highlights the potential for open data to have unintended consequences, as when digitised land records led to appropriation of land from the poor by property developers able to use the new records.

In addition to studies that seek to observe existing open data implementation and impacts, action research and applied or design-science approaches can also be employed. Here, research proceeds through pilot projects that look to introduce open data or make use of available open data, and through reflection and evaluation against metrics, to draw out practical lessons on how open data can be published and used effectively.

Future studies of open data impact might take one of two paths. They might seek to measure impacts at a macro-level, looking for statistical correlations between measures of open data implementation, and variables that capture some anticipated impact of open data. For example, trying to measure the relationship between levels of open data publication and economic growth. Alternatively, they may turn

towards the micro-level, seeking to understand the processes through which open data is used in particular circumstances. Whilst, with suitably granular data², some statistical correlations may be explored, we suggest that at present the most important area for focus is on understanding process of open data use, the impacts this can have, and the factors that affect when positive and negative impacts emerge from open data. Here a choice must be made as to whether the research is into the impacts of a particular dataset, the impact of an open data initiative, or the impact of open data in a sector in general.

In the Exploring the Emerging Impacts of Open Data in Developing Countries (ODDC) project our goal is to combine multiple levels of analysis and different disciplinary lenses, bringing together these three aspects of open data research to build up a coherent picture that can inform both local and global policy and practice. We do this by combining a top-down view of the context of open data in different developing countries, and a bottom-up view of the emerging impacts of open data in specific contexts. These are mutually reinforcing strategies, with the top-down view providing a framework for comparative work, and the bottom-up view informing iterative development of the macro-level assessment methods. At present, we are focussed on primarily qualitative work, recognising that the early stage of development of open data means that quantitative studies seeking to ascertain effect sizes are likely to be frustrated by both a lack of data, and a limited understanding of the relevant variables. For this reason we talk in our study of the ‘emerging impacts’ of open data, where our goal is to generate narrative and analytical accounts of how open data is being used, and the kinds of ways it is influencing change, without necessarily making strong claims that establish causation between open data interventions and impacts.

In addition to research that fits the tri-partite classification offered above we also believe it is important to draw upon insights from theoretical, historical and applied research that cuts across readiness, implementation and impact. For example, economists are developing abstract models to estimate potential impacts of open data, whilst design-science and computer science researchers are developing tools and technologies for open data in ways that generate relevant practical and theoretical knowledge. In parallel, historians and political scientists are developing narrative accounts of how politics, personalities and networks have been involved in creating national or transnational open data movements, analysing discourses around open data alongside open data practices themselves. All these lines of research have something to bring to our understanding of open data, and it’s emerging impacts, and across the ODDC project we will adopt a methodological pluralism to draw the best insights into our research. More detail on the ODDC research framework is provided in section 5.

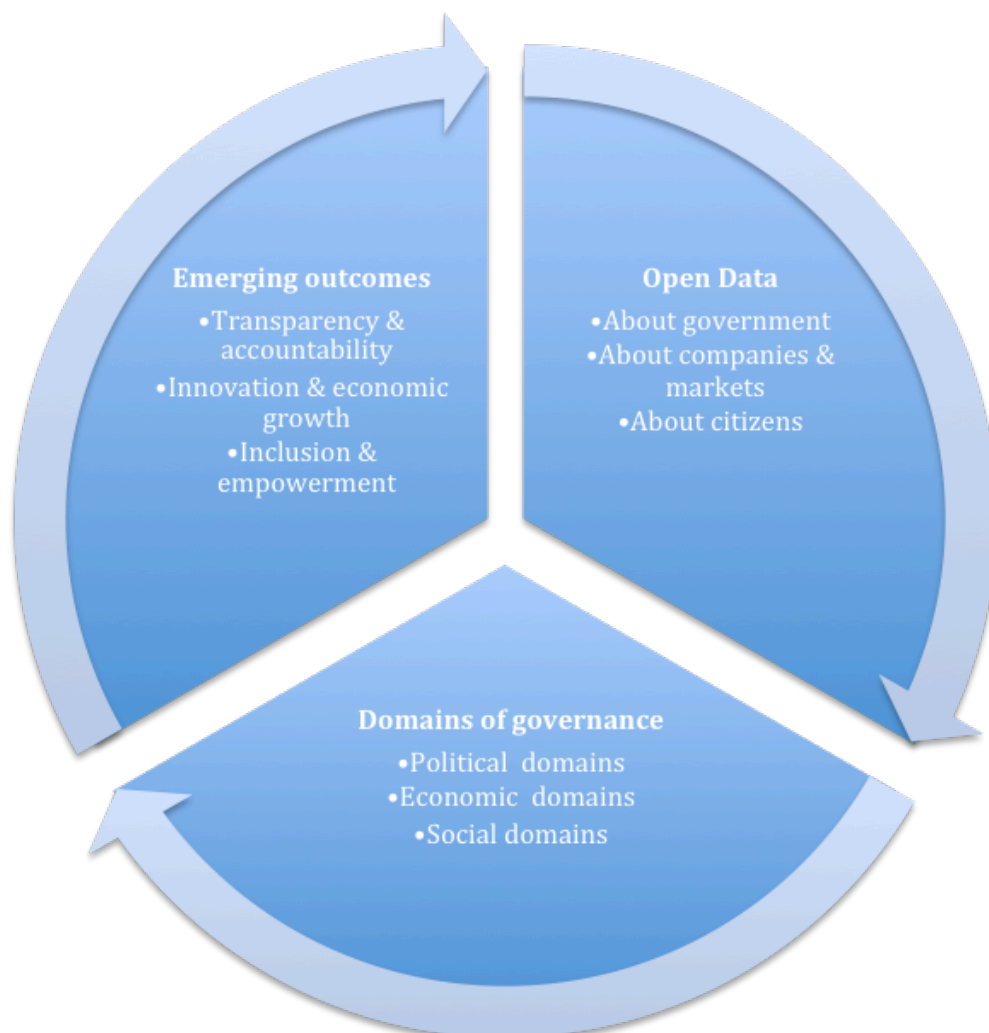
² For example, if reliable data exists on the size of certain market sectors, and there is an increase in the supply of cadastral (geographic) data, then it may be possible to explore the growth or not of the market sectors that draw upon geographical data, and to control for other factors that could have also led to these impacts. However, measurement at the level of national economies (as in, seeking correlations between open data and GDP) or national politics (correlation between open data and character of political debate for example) is unlikely to produce useful results at present due to the presence of many confounding factors, and our lack of detailed understanding of the mechanisms through which open data is operating that would allow us select relevant control variables, or design suitable comparative frameworks.

Open data, governance and
emerging impacts: a conceptual
framework

4. Open data, governance and emerging impacts: a conceptual framework

The emerging impacts of open data will be realised within specific circumstances. The use of open budget data in any country, for example, takes place against a backdrop of political debates about budget, institutional arrangements and power structures, and in the context of existing advocacy, campaigning or scrutiny work. Helbig et. al (2012) have argued that research into open data should investigate the context and dynamics that open data is embedded within, outlining a framework for describing an ‘information polity’ consisting of all the “stakeholders, data sources, data resources information flows, and governance relationships involved in the provision and use of government-held and non- governmental data sources”. Our overall research framework follows this recommendation. We place a particular emphasis on understanding decision-making in different settings, and on unpacking the different mechanisms by which open data may be seen to affect change in distinct governance settings.

However, if differentiation is our only research strategy we can generate only local findings: of limited relevance to other contexts. We also need to understand commonalities between cases of open data publication and use, and to uncover common mechanisms through which open data may be involved in bringing about impacts. To support this we have adopted a conceptual framework that draws attention to three key terms: open data, governance, and emerging outcomes.



Firstly, this recognises that there are many subject areas for open data, and that it is important to understand the subject, structure and status of a dataset in constructing an account of open data impacts. Secondly, it recognises that governance issues exist and are addressed at many levels, involving political, economic and social forces. Lastly, it highlights a range of ‘emerging outcome’ mechanisms through which open data may deliver substantive impacts. In presenting these core

terms as part of a cycle it highlights that there is no linear progression from data, to decision-making, to impact – but rather that these are in on-going interaction. Section 5 will operationalize this conceptual framework by setting out a series of components to include in open data case studies drawing out key data on each of these terms. In this section we further unpack the key terms from the conceptual framework, drawing out in the process critical issues for open data research to engage with.

Open Data

Open data research should be able to develop accounts that are about open data in particular, as opposed to just data, information or transparency in general. Open data research needs to be able to distinguish between the introduction of new *information flows* into a governance setting, and the introduction of *open data flows* into a governance setting. This involves an explicit focus on the role of data (as opposed to information), data licenses, data standards, and the activities of technical and non-technical intermediaries, in affecting the emerging impacts of open data.

Data

A printed table of figures, or a pre-prepared map showing the location of health centres or schools is information: someone has chosen how data should be contextualised, and has fixed an interpretation. The underlying spreadsheet of figures, or a computer file containing structured information on public service locations is a dataset. With the dataset, and the right tools, a user can create their own interpretations of the data: sorting, filtering and representing it in different ways. In most cases, the use of data in governance will involve turning it into information: interpreting and presenting it. However, understanding who controls this process, and where it happens is important.

Data may go through many different processes between being made available as open data and being used. It may be filtered, converted, combined with other data and so on. Tracing the journey data has gone through, and asking whether at each stage it has become more or less valuable for particular users, can aid understanding of the factors that contribute to positive and negative impacts of open data.

Licenses

Many open data advocates emphasise the importance of dataset licensing. That is, attaching an explicit statement to datasets that determines how they can be re-used. As mentioned in section 2, the Open Knowledge Definition (OKD) presents a stringent definition of an open license as one that requires, at most, attribution of the dataset source. According to the OKD an open dataset must be free to use by commercial and non-commercial users alike. This is framed within a principle on non-discrimination against fields of endeavour. The creation of a unified legal framework around open datasets is seen as a particularly important issue as data travels across borders where different intellectual property rights apply to datasets, and as datasets are combined with each other. Incompatible licenses, it is argued, can create significant challenges in determining the legal status of derivative datasets, yet much of the value of open data comes in combining different datasets.

In practice, many datasets that are otherwise accessible online and in open formats do not meet the strong criteria of the Open Definition. A brief survey of development-related datasets in South Africa³ demonstrated that whilst many NGOs, Universities, research projects and government departments published datasets on their websites, few had explicit license statements (Powell, Davies, & Taylor, 2012) although permission for re-use was in some cases implied. The importance of licenses may vary based on the use to which someone intends to put data. If a use involves redistributing the data, then intermediaries may wish to be sure they have the rights to do so; but if a user is taking data direct and analysing it, then they might assume the provision of the data online gives them rights to do this, and so not be overly concerned by unclear licensing.

Understanding how clear licenses facilitate open data use, or how unclear licensing inhibits open data use, in different developing country contexts, in different fields, and for different kinds of open data is important to inform the development of policies and interventions on open data licensing frameworks.

³ See <http://www.opendataimpacts.net/2011/06/a-whole-lot-of-development-datasets/>

Data standards

Data standards can make data sharing and use easier, allowing datasets to be combined, and supporting the creation of tools that work with standardised and compatible datasets from many different sources. Open data relies upon a stack of standards, from data format standards that have resulted from formal standards processes, through the de-facto standards that emerge through ad-hoc collaborations and use. For example, the General Transit Feed Specification (GTFS)⁴ describes a data model for representing information on public transport services. Data published in this format can be made available on Google Maps transport planner, and in other services that support the GTFS standard. Standards exist for many different fields, from the International Aid Transparency Initiative (IATI) standard for describing aid projects and funding flows⁵, to the Open311 standard for exchanging reports of problems with urban authorities⁶. Microformats, Schema.org and linked data vocabularies also provide a range of ‘building blocks’ for describing the contents of datasets in common ways.

Mark Thomson has argued (2013) that the use of common standards enables a more modular approach to development, supporting approaches that embody “adaptive pluralism” (Chambers, 2010) and a greater autonomy for distributed agents to assembly components that meet their local needs. However, whilst standards can simplify data sharing and use, they also determine what it is easy or difficult to express within a dataset. What is, or isn’t included in a standard, and how things are classified and represented, can have significant political and social impacts down the line (Bowker & Star, 2000). For example, in the IATI standard, decisions have had to be made over how to represent the geographical location of aid projects, and how detailed the data is that can be stored in the format. This has an impact on the kinds of geographical visualisations of the data that are possible, or easier, to create. This may have an impact upon how the data is ultimately used. Some standards are developed through open consultative processes, whilst other standards have emerged ‘de facto’ after first-movers or powerful actors decided to use them, and others followed their lead. Access to information on standards may also be unequally distributed, with knowledge on how to combine different standard building blocks transmitted through peer-networks rather than being transparently documented.

In looking at the emerging impacts of open data in developing countries it is important to pay attention to choices made over standards, and whether open data publishers and users are able to work with, or adapt standards to meet their needs, or standards exist as an external force shaping local uses of data.

Intermediaries

Well-structured open data has many uses, yet making use of it directly often requires considerable technical skills and subject-area expertise. The standard narrative around open data suggests that entrepreneurial actors will create ‘apps’ that make data accessible to citizens, but there are many forms of intermediation between data and users. For example, the Code 4 Kenya fellowships programme⁷ has been building the capacity of news media organisations to carry out data journalism and to build accessible products with open data, presenting them both on the Internet and in print (Weinstein & Goldstein, 2012), whilst Development, Research and Training in Uganda have been playing an intermediary role by presenting graphs of government agriculture data on their website⁸. Sharif discusses how Public Sector Information (PSI) can be used by researchers, companies and civil society organisations in Africa – many acting as intermediaries (Sharif, 2009).

Understanding the relationship between data publishers, intermediaries and ‘end users’ of open data is central to understanding where emerging impacts of open data are sustainable and stable, or where they are vulnerable to the loss of intermediary groups. The actions of intermediaries may be oriented towards particular kinds of outcomes from open data, such as transparency and accountability, or innovation and economic growth. States or donors may choose to invest in capacity building for particular intermediary groups, and so it is important for critical research to attend to where investment

⁴ <https://developers.google.com/transit/gtfs/reference>

⁵ <http://www.iatistandard.org/>

⁶ <http://www.open311.org/>

⁷ <http://code4kenya.org/>

⁸ <http://www.drt-ug.org/>

in intermediary activities might bring about the best return for particular groups and for particular desired outcomes from open data policy.

Reconsidering openness

Having open datasets, judged as open under the Open Definition, or some other technical criteria, does not necessarily equate to the openness of the subjects described in the data, nor even to the openness of the data itself to wide re-use. The definitional principles of open data can be considered, in some senses, as a hypothesis about the best practices that will promote wide re-use of data. But, that hypothesis has little to say about the relevance or the quality of the data released, or the technical, knowledge and social resources available to re-users that would enable them to have “effective access” to the data (Gurstein, 2011). Central to the study of open data in developing countries is a questioning of the open definition hypothesis, and an exploration of the role of open data in bringing about other forms of openness: of governments, markets and development processes. That is, an exploration of how open data influences governance.

Governance

In a recent paper titled ‘What is governance?’ Fukuyama defines it as “as a government’s ability to make and enforce rules, and to deliver services, regardless of whether that government is democratic or not” (Fukuyama, 2013). More generally, we can state that governance is centrally concerned with processes of decision-making and implementation (UN ESCAP, n.d.). Although, as in Fukuyama’s definition above, governance is often discussed in terms of the government, and its administration of a nation state (UNDP, 1997), in practice ‘governance systems’ (as processes of making and applying rules) can be found with and without state involvement at all levels of society. Even state-centred systems of governance have become more diffuse over the last sixty years (Kamarck & Nye, 2002), as globalisation and policy complexity has created new sub-national and supra-national governance systems. Open data both enters into existing governance systems, and has the potential to reconfigure and transform them.

Helbig et. al (2012) give the example of a simple governance system seeking to regulate standards food hygiene in restaurants in New York City (which they describe in terms of an ‘information polity’ drawing attention primarily to the informational linkages between actors). They identify the goals of this system (ensuring compliance of the restaurants with the city’s health code), and the actors involved (mayor and council, department of health, restaurant operators, citizens, media and advocacy groups etc.). With these actors identified it is possible to also identify where authority and power are located in the system, both formally (as in, by law or rules), and informally (i.e. as expressed in practice), and to explore how governance decision making and implementation has been handled in the past. For example, in the restaurant hygiene case, governance has taken place through inspections and enforcement activities. Tracking the introduction of new information and data (and open data) into this existing governance system allows the researcher to see how relationships between key actors may change, where new actors enter the system, and how the processes of making and applying rules may be altered. For example, Helbig et. al. find that, following the creation of a consolidated datasets of inspection results, the introduction of a mobile application to provide access to information on restaurant inspections, and a switch to simplified grading of restaurants, citizens became more actively engaged in the governance process. Market forces from citizen choice over which restaurants to patronise were able to influence restaurant behaviours, proving more effective than previous enforcement mechanisms that relied more on enforcement activities carried out through state power (See also Fung, Graham, & Weil, 2007).

This same approach of, of identifying, describing and tracking change within governance systems provides a route to understand the different processes that influencing emerging impacts of open data in a wide range of settings across the developing world. Focussing on changes to governance systems also allows sensitivity to both unintended consequences of open data, and barriers or ‘push back’ seen when open data is introduced. Section 5 introduces a number of case studies being carried out during the ODDC programme, each of which will identify a particular governance system of interest: from city-level budgeting and spending in Brazil, to regulation of extractive industries in India, or monitoring the performance of higher education in South Africa

It is notable in the food hygiene example given above that changes in the information available in the governance system led to a shift from governance through political authority, to an increased role of the market and consumer pressure in securing governance outcomes. In our research framework we draw attention to three ‘domains of governance’ through which decisions may be taken or implemented: the political, the economic and the social. Different disciplinary lenses can be applied to explore how governance is operating in each of these domains:

- **The political domain** focuses attention on the exercise, shaping and control of state power. Political science approaches are particularly valuable here to explore how the opening of data may affect the established balance of power between institutions, and how greater access to information for citizens does or doesn’t lead to political pressure for change. Scholars of media and journalism can also bring insights about how practices of data journalism may affect the political domain. Politics exists at many levels, but generally the reach of political decisions is broad: affecting many different localities.
- **The economic domain** focuses attention on both market mechanisms as a tool of governance (distributing decision making through markets), and on the regulation or promotion of markets, as well as internal economic efficiencies for government from better data use. Where economic theory can help explore how the introduction of data into markets could promote better outcomes, business studies can also contribute to an understanding of the conditions under which open data does or doesn’t result in innovations. In looking at the economic domain, critical attention can also be drawn to whether open data empowers smaller market players, or whether established and wealthy individuals and firms are able to gain the greatest return from open data (Heusser, 2012). Whilst governance of the economic domain (usually driven from the political domain) may be imposed upon all actors in a market, and the rules set through these processes affect market outcomes, governance carried out through the economic domain is generally distributed without central control or ‘designed’ outcomes.
- **The social domain** invites a particular focus on the inclusion of marginalised groups, and on the capacity of individuals and communities at the grassroots to exercise influence over their own lives, without necessarily deploying either political or market power. Social science and community informatics approaches (Gurstein, 2007) may be particularly appropriate here, encouraging the embedded study of how open data affects social relationships and existing processes and practices of governance.

Within each of these domains it is possible to identify different ‘theories of change’, or hypothesis about how open data might affect a governance system. In the following section we summarise some of these theories of change, and survey the existing literature on areas where emerging impacts of open data may be emerged.

The emerging impacts:

In our initial project workshop (Perini, 2012) and an analysis of the literature we identified three broad categories that capture the mechanisms through which commentators suggest open data might bring about change. These are:

- **Transparency and accountability:** open data will bring about greater transparency, which in turn brings about greater accountability of key actors, leading to them making decisions and applying rules in the public interest;
- **Innovation and economic development:** open data will enable non-state innovators to improve public services or build new products and services with social and economic value; open data will shift certain decision making from the state into the market;
- **Inclusion and empowerment:** open data will remove power imbalances that resulted from asymmetric information, and will bring new stakeholders into policy debates, giving marginalised groups a greater say in the creation and application of rules and policy;

As described in the previous section, these outline theories of change each have distinct primary areas of focus

Key theory of change	Key focus	Key disciplinary traditions/Streams
Open data will bring about greater transparency in government, which in turn brings about greater accountability of key actors to make decisions and apply rules in the public interest;	The State (political domain)	Political science, public administration, legal studies
Open data will enable non-state innovators to improve public services or build innovative products and services with social and economic value; open data will shift certain decision making from the state into the market, making it more efficient;	The Market (economic domain)	Economics, business models, regulation
Open data will remove power imbalances that resulted from asymmetric information, and will bring new stakeholders into policy debates, giving marginalised groups a greater say in the creation and application of rules and policy;	The “Excluded” (social domain)	Social science, community informatics

These theories of change are not mutually exclusive. In any governance setting you might find different routes being explored by different actors – as when, for example, one group might choose to use open data on public transport to hold existing service providers to account (transparency and accountability), whilst others may use the data to build commercial mobile applications that help travellers to find the fastest route, or check on the times of buses and trains (innovation and economic development). Nor are these theories unique to open data: literatures exist on each that work on open data should engage with. In the following section we briefly survey some of the existing literature and key issues to be addressed under each of these theories of change.

Transparency and accountability

Although transparency and accountability are frequently discussed together, they are distinct concepts. The pairing have become a mainstay of governance, where deficits of accountability can leave those in power able to practice corruptly, and to serve their own, rather than the public interest. Transparency is an essential ingredient for accountability (Joshi, 2012, p. 4), but is rarely a sufficient condition for it (Kuriyan, Bailur, Gigler, & Park, 2012). Accountability involves the capacity to “elicit justification, render judgment and impose sanctions” on those with power (Joshi, 2012). Whilst accountability relationships may be established internally by key stakeholders in a governance system⁹, such as when an audit institution demands to see the financial records of a government department, or when a public commissioner of road building orders an inspection of build quality and calls in the contractor to explain discrepancies between the specification and delivery, in the context of open data the pairing of transparency and accountability suggests a focus on allowing external actors, citizens in particular, to play a role in holding power to account.

Open data has generally been articulated as a form of proactive transparency, where governments or other actors choose to publish data, in contrast to reactive transparency as invoked in Right to Information (RTI) laws where citizens ask for access to information (Janssen, 2012; OKF - Open Knowledge Foundation & Access Info, 2011). Some moves towards linking open data into reactive transparency have taken place in the United Kingdom, where amendments to the national Freedom of Information Act provide a ‘right to data’ allowing citizens to request structured datasets (HM Government, 2012a). We also note that transparency can be employed by the state as a means of regulating private actors, as in targeted transparency policies, where firms are required to open up information or data on their products or actions – such as safety information, or information on

⁹ These are examples of what Heald, in his paper on ‘Varieties of transparency’, would call ‘upwards’ transparency, highlighting that transparency can be selective and asymmetric, even if in the context of open data and Right to Information (RTI) we would anticipate all actors have (at least formally) equal access to the data. Remembering that even when open data exists, prior closed flows of information (sometimes much richer than the data which has been opened) may persist, is important in describing a governance setting.

pollution from factories (Fung et al., 2007)

Open data is not, however, identical to transparency. Heald, amongst others, argues “Openness might... be thought of as a characteristic of the organization, whereas transparency also requires external receptors capable of processing the information made available” (Heald, 2006 quoting Larsson, 1998). A related point has been made by Gurstein, noting that having data online under open licenses does not mean that everyone has effective access or can make ‘effective use’ of the data. Citizens may face barriers of technology, literacy, education or social capital that prevent them effectively receiving and processing information that might have been made available (Gurstein, 2011). The way in which data is published, the context it is put in, the support on offer to enable access and use, and the presence of intermediaries, all affect how far open data will lead to increased transparency.

Where open data has led to greater transparency, another set of intervening relationships may determine how far it leads to accountability. In a collection of essays on corruption and democracy in Brazil (Power & Taylor, 2011), authors highlight a range of accountability channels, from the ballot box, to audit institutions, media coverage, judicial action and police enforcement. Transparency has the potential to enable new accountability channels, and to affect the operation of existing channels. For example, the use of open data in data journalism has the potential to strengthen the capacity of existing media to hold government to account, and to support the emergence of new media players. Similarly, government ministers in the UK heralded the potential for an ‘army of armchair auditors’ to emerge using public spending data to hold government to account (McClellan, 2011) (individual citizens taking on the role of reviewing government spending from the comfort of their own computers), whereas Speck (2011) notes that in the Brazilian case transparency can help actors outside government place issues on the agenda of the formal audit institutions, affecting how they operate, but neither bypassing nor replacing the need for formally instituted audit. In our workshop exploring theories of change for open data impacts, participants particularly highlighted how public availability of data could empower ‘good’ civil servants to oppose corrupt practices within their institutions without having to turn directly to whistleblowing (Perini, 2012). It is possible also that transparency creates more ‘accountable’ behavior without the need for actual accountability mechanisms to be exercised, as when knowing information on their actions will be made public, and that indiscretion *could* be discovered encourages officials to behave better (Meijer, 2007).

However, the existing transparency literature also highlights the possibility of adverse affects from greater transparency, as more openness can create perverse incentives, limit space for free discussion in politics, lead to ‘gaming’ of the data, can contribute to surveillance of citizens by the state, or can shift power to distant institutions rather than to citizens (Ballingall, 2011; Heald, 2006, 2011; Murray, 2011). Although there has been considerable research into transparency and accountability in the development sector, the locally situated nature of most Transparency and Accountability Initiatives (TAIs) means that cross-cutting research is limited and widely established findings about what makes for effective TAIs, or what conditions lead to positive or negative outcomes, are few (Calland & Bentley, 2012; Joshi, 2012; McGee & Gaventa, 2011). McGee emphasizes the importance of studying transparency and accountability in relation to specific governance issues, rather than in the abstract (McGee, 2011) – a strategy we also recognize and adopt.

As yet, most TAI literature has not explicitly focused on open data. However, over the coming years we anticipate greater connections will be drawn between work on ICT enabled TAIs and open data. As a 2011 report from the Transparency and Accountability Initiative notes, “online and mobile technology tools are beginning to change the transparency and accountability field”, supporting a number of more rapid and responsive accountability projects (Avila, Feigenblatt, Heacock, & Heller, 2011), and increasingly the exchange of well structured machine readable datasets plays an important role in these projects. Fung et al. highlight in particular the possibility that “*Targeted transparency could gain effectiveness through better understanding, design and information technology*”, although noting that “...we are only beginning to grasp the ways in which public policies can harness information to reduce serious risks and improve important services” (Fung et al., 2007, chap. 8).

Inclusion and empowerment

Opening up access to data can help to address asymmetries of information between companies and officials and citizens, NGOs and grassroots groups. With open data there is the possibility for local communities to build up their own understandings and interpretations of key issues, and for intermediaries to contextualize information in ways that make sense to diverse groups, including citizens at the grassroots. Through print-outs, mobile phone-based services, offline access, community radio and participatory workshops (De Boer et al., 2012) data can be taken to local settings – empowering previously marginalised groups, and can provide the basis for feedback loops that enable local communities to shape the knowledge base on which policies are based (e.g. Srinivasan, 2012). However, as Gurstein has noted, open data alone does not necessarily equate to empowerment, and there is a risk that a ‘data divide’ is created, where data only empowers the already empowered (Gurstein, 2011). Whilst evidence from the UK suggests that open data has engaged a number of new actors in thinking about public services and governance (Davies, 2010), little work has been done to map out the users of specific open data, and to explore how far open data is supporting greater inclusion in policy making and governance processes.

There is an important distinction to draw in looking at the impacts of open data on marginalised groups. Open data may support better outcomes for the marginalised by, for example, addressing corruption and empowering regional parliamentarians to secure a better deal for their constituencies. It might similarly support better outcomes by empowering intermediary NGOs and other organizations to secure resources or policy change for a community. Or it may work by supporting the direct engagement of grassroots communities in working with, interpreting and responding to data about their situation. In the first model, marginalised groups may benefit, but are still primarily the objects of development: it is only when data directly empowers marginalised groups that they become subjects of the development process (Perini, 2012; Powell et al., 2012), actively shaping it around their own needs – whether through engagement in policy and political debate, or being able to access knowledge and information that they can use to directly improve their lives. There is also a distinction to be drawn between individual and community empowerment effects of open data. For example, Bates argues that the UK open data agenda has developed to support the marketization of public services, in which citizens are cast as consumers, offered data to help them make individual choices, but in which the potential of collective action to secure social provision of appropriate public services may be side-lined (Bates, 2012).

The conventional articulation of open data initiatives, as involving raw data, technical intermediaries, and only then, end-users, makes tracking the inclusion impacts of open data challenging. However, by widening our focus to include cases of demand-driven open data projects, where grassroots communities have asked for open data, we can explore how far having access to data, as opposed to solely documents, or local knowledge about a situation, impacts upon empowerment at the grassroots.

Innovation and economic development

Markets can be both a focus of, and a tool of, governance. We will address the use of open data as a policy intervention in the economic domain first, before turning to the role of data in governing economic activities.

In markets, decision-making and implementation tasks are distributed widely across semi-autonomous actors, making use of signals such as price to allocate effort and resources. Open data has been described as “digital fuel of the 21st century” (Kundra, 2012), a raw material that can support new economic activity and lead to dramatic breakthrough innovations. Arguments concerning the economic potential of government data were key drivers for open data initiatives, particularly in the EU, where many studies argued that billions of euros in potential economic activity were being lost through the ways governments managed their data, either not providing any at all, or providing it for a fee (Dekkers, Poleman, Te Velde, & De Vries, 2006b; Newbery et al., 2008; Pollock, 2009; Uhlir, 2009). Whilst some of this economic value may come from large scale Public Sector Information (PSI) re-user firms creating products with government data, such as maps or improved weather reports, many advocates of open data have focussed on the potential for open data to be used by Small and Medium Enterprises (SMEs), predominantly in the technology sector, to create new products or find new niche markets (Fioretti, 2010), tapping into the ‘long tail’ of government data and market needs (Anderson, 2006). The release of open government data to stimulate domestic technology industries, and the

creation of new 'start up' firms is a strategy evident both in the UK (HM Government, 2012b), Kenya (World Bank, 2012) and the US (Kundra, 2012) amongst other open data initiatives. Whether or not then open data generates economic returns; who these returns accrue to within a country; and whether standardised data enables cross-border trade in services built on top of data; are all issues important to track in understanding how open data can operate as a tool of economic policy.

Governments may also release open data to stimulate innovation in the delivery of public services. Open government data initiatives have often been linked to Tim O'Reilly's notion of 'Government as a Platform' (O'Reilly, 2010) in which government acts as a provider of data upon which dynamic entrepreneurial actors outside the state can innovate to provide better, more efficient or more customised public services. This taps into an argument about the greater innovative capacity of the private over the public sector, ideas of user and open innovation (Von Hippel & Von Krogh, 2009; Von Hippel, 2005), and the view that both commercial and not-for-profit enterprise can act as intermediaries delivering public service (Mayo & Steinberg, 2007). The widely cited 2008 'Apps for Democracy' contest by the United States District of Columbia has suggested that through awarding just \$50,000 in prizes in an apps contest, developers outside the government put together 47 applications that would have cost \$2.6m if developed internally (UN - United Nations, 2010) although some have raised questions about the sustainability and actual realisation of this value (Nichols, 2010).

In seeking to secure some of the innovation and co-production benefits of open data the Kenya Open Data Initiative has focussed on steps to create an 'eco-system', connecting data providers in government with entrepreneurs and ICT trained young adults (World Bank, 2012). This suggests the hypothesis that enabling open data to drive public and private sector innovation requires more than datasets alone. Understanding the conditions that are conducive to data-enabled innovation, and the kinds of policies that can promote it, is an important area for research. This requires attention to both the micro-level of conditions around particular datasets (for example, whether or not a transport ministry or agency is ready to collaborate with developers from outside government to work on co-producing services), at the mid-level (for example, whether government procurement policies allow it to engage with and appropriate innovative public service ideas from SMEs), and at the macro-level (for example, whether national policies and infrastructure support tech sector innovation). Equally, critical research is needed to assess how far open data enabled innovation serves widespread social needs, or is only able to deal with certain kinds of problems.

We now turn to the role of open data in the governance of markets. All markets require some form of oversight and regulation - yet markets are increasingly complex and hard to monitor. Both citizens and regulators face challenges in exercising effective oversight in many sectors, and it has been suggested that open data can play a role here. For example, the Extractive Industries Transparency Initiative, designed to support monitoring of extractives markets, are exploring the development of a common data standard to assist monitoring of contracts (EITI International Secretariat, 2012); and the OpenCorporates.com project that has collated open data on millions of companies around the world has been actively participating in the EU Financial Stability Board (Taggart, 2012) promoting the use of open data to enhance oversight of the markets implicated in the 2009 financial crash. Data on market activities might be held by government (such that government can choose to release it), or disclosure of information by companies might be mandated by law, in what Fung et al. call 'targeted transparency' (Fung et al., 2007; Fung & Weil, 2010). Although established literatures exist on the role of information in governance of markets, the challenge for open data research is to identify how the provision of this in the form of datasets works in practice. The possibility of connections being made between disparate datasets to identify corruption, or the opportunity for data on contracts to support citizen oversight, for example, depend variously on common standards, identifiers and the availability of tools and platforms that place information in the hands of people with the power to act on it.

In theory, better-governed markets should lead to more sustainable and equitable economic growth. Research may also factor in here questions of environmental sustainability, looking at how open data might impact upon governance of the environmental impacts of economic activity.

Conceptual framework: review

In this section we have shown how studying the emerging impacts of open data might connect with a wide range of existing research agendas, and we have set out an approach to move beyond 'supply side' studies of open data to take a comprehensive look at open data in use. The conceptual framework outlined is a living resource, open to revision as on-going work within the ODDC project digs deeper into the issues around emerging impacts of open data.

Putting the framework into practice

5. Putting the framework into practice

So far we have outlined a brief history of open data and set out different approaches to open data research. We have made the case that research into the emerging impacts of open data should focus on the intersection of open data, governance and emerging outcomes, and we have drawn out a number of cross-cutting themes specifically related to open data. Throughout our discussion we have kept both developed and developing world cases and examples in mind.

In this section we operationalize the research framework to be adopted in the IDRC funded 'Exploring the Emerging Impacts of Open Data in Developing Countries' (ODDC) project taking place between 2013 and 2015. This project, co-ordinated by the Web Foundation¹⁰, and run through the Open Data Research Network¹¹ seeks to reflexively apply principles of openness in its own operation, and whilst working with a core set of funded case studies, will also share methods, research tools and datasets to enable wider collaborative research.

The ODDC programme has three main research components:

- **In-depth case studies of open data in use** in a range of governance settings and country contexts across the developing world, exploring the impact of open data under difference disciplinary lenses'
- **Developing common assessment methods for open data initiatives** to support measurement of the extent of open data adoption and impacts in different countries and sectors;
- **Cross-cutting research** looking at generating practice and policy-relevant learning on open data.

In working with case study partners from a wide range of academic and non-academic backgrounds, and diverse disciplines, and by combining quantitative and qualitative methods, our framework is explicitly interdisciplinary.

Case studies

Our view of open data is broad. We are interested in different kinds of open data, including data from about governments and states, data about companies and markets, and aggregate data about citizens. We are interested in how increased openness of this data impacts a range of different domains of governance: political, social and economic. And we are interested in a range of emerging outcomes: transparency and accountability, innovation and economic growth and inclusion and empowerment. Any case study might involve different kinds of data, governance issues and emerging outcomes, and will be responding to local policy and practice questions as well as cross-cutting research issues. For this reason we highlight six key areas that each case study should address, laying the foundations for a holistic comparative cross-case analysis (Diesing, 1971).

These areas are:

- **The context for open data** – including the political, organisational, legal, technical, social and economic context.
- **The supply of open data** – including data availability, legal frameworks for data, data licenses, and the stakeholders involved in providing data.
- **Technical platforms and standards** – including data formats and data standards use, and any data catalogues, APIs or analysis tools provided by an open data initiative

¹⁰ www.webfoundation.org

¹¹ www.opendataresearch.org

- **The context of the specific governance setting** – including a description and history of the issues in focus, details of key stakeholders, and analysis of how data plays a potential role in this setting
- **Intermediaries and data flow** – documenting the means by which data is made accessible in the governance setting: how, and by who?
- **Actions and impacts** – documenting the experience of those seeking to use data, and providing evidence of intended or unintended consequences.

Figure 2 below illustrates these six core case components, and the necessity of exploring them iteratively, recognising that there is rarely a linear relationship between data supply and impact, but that issues of data supply, data use and governance impact upon each other in various ways.

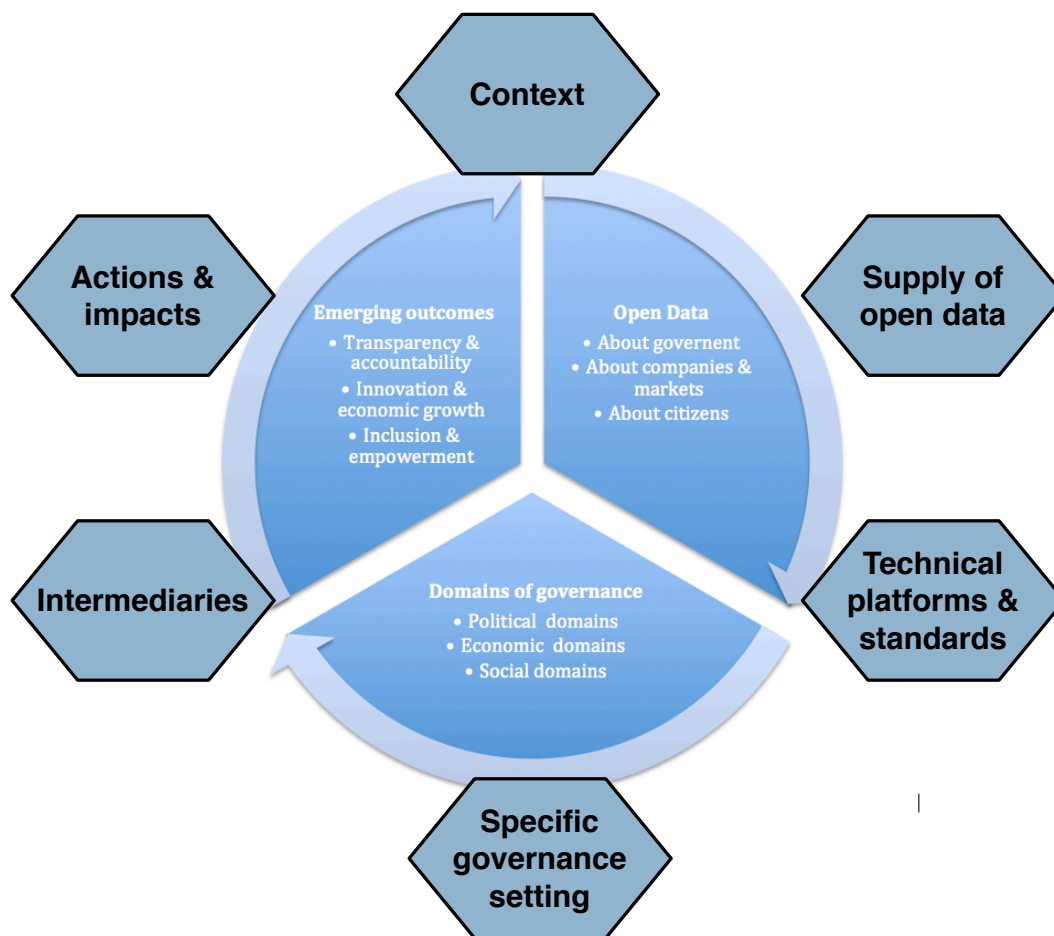


Figure 2: Case study components

The 17 case studies forming the initial network for the ODDC project are shown below, grouped according to the general kinds of data, governance areas, and emerging outcomes that we anticipate these cases may explore.

Budget transparency & governance	Urban Governance
<p><i>Public expenditure data > governance of resources / budgets > transparency and accountability</i></p> <p>An Investigation of the use of the Online National Budget of Nigeria by Relevant Stakeholders (University of Ilorin, Nigeria)</p> <p>Exploring the emerging impacts of open aid data and open budget data in Nepal (Freedom Forum, Nepal)</p> <p>Measuring open data's impact of Brazilian national and sub-national budget transparency websites and its impact on people's rights, especially people living in poverty (Institute for Socioeconomic Studies, Brazil)</p> <p>Case Study on Open Data Initiative of Ministry of Finance on National Budget Transparency in Indonesia (Sinergantara, Indonesia)</p>	<p><i>Public service data > urban governance > more efficient, innovative and inclusive service delivery</i></p> <p>Opening the Cities: Open Government Data in Local Governments of Argentina, Brazil and Uruguay (Instituto Polis, Brazil)</p> <p>Opening Government Data by Mediation: Exploring the Roles, Practices and Strategies of Data Intermediary Organizations in India (HasGeek, India)</p> <p>Quality of civic data in India and the implications on the push for Open Data (Transparent Chennai, Institute for Financial Management and Research, India)</p> <p>Opening the Gates: Will Open Data Initiatives Make Local Governments in the Philippines More Transparent? (Step Up Consulting Services, The Philippines)</p> <p>Open Government in the Philippines: Exploring the role of Open Government Data (OGD) and the use new technologies in the delivery of public services (De La Salle University, Manila, Philippines)</p>
Poverty alleviation	Emerging issues
<p><i>Various data > urban and rural poverty > Inclusion and empowerment</i></p> <p>Exploring the impacts of budgetary information web publishing in the subnational level of Brazil (Research Group on Public Policies for Information Access, Brazil)</p> <p>Investigating the Impact of Kenya's Open Data Initiative on Marginalised Communities: Case Study of Urban Slums and Rural Settlements (Jesuit Hakimani Trust, Kenya)</p> <p>A Monitoring And Evaluation Study on the Deployment of Code4Kenya Applications and Services (iHub, Kenya)</p> <p>How could open data contribute to poverty eradication in Kenya and Uganda through its impacts on resource allocation? (Development, Research and Training, Uganda)</p>	<p><i>Various data > various governance settings > various outcomes</i></p> <p>Open Data in the Judicial Systems: Evaluating Emerging Impact on Policy Design in Paraguay, Chile and Argentina (Center of Implementation of Public Policies for the Equity and the Growth, Argentina)</p> <p>The use of open data in the governance of South African higher education (University of Cape Town, South Africa)</p> <p>Open government data for regulation of resource intensive energy industries in India (The Energy and Resources Institute, India)</p> <p>Taking Stock of the Effectiveness and Efficiency Initiatives in Sierra Leone (Society for Democratic Initiatives, Sierra Leone)</p>

These case studies will be carrying out research over 2013/14, reporting findings in early/mid 2014.

Common assessment methods

Focused case study research will allow us to understand the dynamics of particular uses of open data in influencing processes of decision making and governance, and will allow identification of key contextual factors impacting upon the realization of outcomes from open data, as well as identification of strategies and interventions employed within an open data initiative that can impact upon the quality of openness in governance. However, given the expanding role and importance of open data in society, and the growing uptake of open data initiatives, it is also important to have rigorous methods to assess the scope and value of open data initiatives at the macro level: both for national open data initiatives, and increasingly in particular sectors, such as budgets, aid, urban governance, or natural resource management.

Although various methods have been suggested by organisations and academics (Farhan, D'Agostino, & Worthington, 2012; Murillo, 2012; Yale Information Social Project Open Government Study, forthcoming; The Open Knowledge Foundation Open Data Census), to generate indicators on open data or open government there is no agreement on a robust methodologies for such assessment. None of the proposed methods are comprehensive or rigorous enough to encompass all the key aspects, and potential impacts, of assessing an open data initiative - nor to provide effective coverage of developed and developing nations.

Building on the preliminary work the Web Foundation have undertaken in this field with the multi-dimensional Web Index (Farhan et al., 2012), and pilot Open Data Index (Alonso, 2012), we plan to review existing methods that assess open data readiness and the strengths and weaknesses of open data initiatives, and using these to draft a framework for assessing both. The initial focus will be identifying indicators and variables for assessing the contextual environment of an open data initiative along a number of dimensions (economic, social, political, organizational, legal and technical). These indicators may be used by the local cases to inform their contextual analysis. The wider development of common assessment methods, for exploring strengths and weaknesses of open data initiatives, will then be informed by additional variables, including in particular relevant variables identified by comparative analysis of data emerging from the case studies.

Cross-cutting research

Taking a socio-technical approach to understanding the emerging impacts of open data allows us to identify a range of points that influence the emerging impacts of open data. We anticipate carrying out a range of cross-cutting analysis of evidence from cases and data collected through the common assessment methods in order to respond to key policy and practice debates. In particular, we will look to:

- **Understand the flow of data from open data initiatives, to potential users**, through a range of technical and social intermediaries. This should inform the design of open data initiatives and the design of interventions that can run alongside open data initiatives to realise certain governance outcomes.
- **Understand how the wider context affects the potential of an open data initiative**. This should inform responses to questions of open data readiness, and whether open data initiatives are an appropriate intervention in all kinds of state. It should inform design of open data interventions that are more responsive to local contexts, as well as addressing the relative importance of issues such as legislative frameworks, political support and technical capacity.
- **Understand how global standards, platforms, infrastructure and 'eco-systems' of open data affect local contexts**. As datasets, tools and standards are increasingly developed through cross-border collaborations, it is important to explore what impact the emergence of global open data ecosystems has on the use of open data in developing countries. This should support assessment of the relative importance of bringing more diverse voices into global standard setting processes.
- **Understand the distribution of benefits from open data initiatives**. This will include addressing questions of who is empowered by open data in different settings, and exploring

where potential economic and social value from open data is likely to accrue.

Over the course of 2013 and 2014, through engaging with potential users of the ODDC research findings we will refine and focus questions to ask in each of these areas.

Conclusions

6. Conclusions

Open data has rapidly moved from being a niche interest, to being part of the global policy mainstream. Government-led open data initiatives have spread across the globe, and civil society or technologist experiments using data to improve governance have been spreading organically, from budget monitoring in Nigeria, to court transparency projects in Argentina. Understanding how experience of open data will vary from country to country and context to context, *and*, understanding the common features of open data that are shaping its implementation in these diverse settings, requires broad-based research. It requires research that can engage with both existing realities of governance, with the particular properties of Internet-based open data. In this paper we have set out the foundations of a research agenda that should resist both pure social shaping or technological determinist accounts, and that is intended to generate policy-relevant findings. We have introduced in particular the ODDC project, which will, over 2013 and 2014 be seeking to put this framework into practice, and to address many of the concerns and issues outlined in this paper.

These projects will be sharing learning and insights along the way at www.opendataresearch.org

Bibliography

Bibliography

- Alonso, J. M. (2012). The Open Data Index. *Web Foundation Blog*. Retrieved December 2, 2012, from <http://www.webfoundation.org/2012/09/introducing-the-open-data-index/>
- Alonso, J. M., Boyera, S., Grewal, A., Iglesias, C., & Pawelke, A. (2013). *Open Government Data Readiness Assessment Indonesia* (pp. 1–47).
- Anderson, C. (2006). *The Long Tail: Why the Future of Business is Selling Less of More* (p. 256). Hyperion.
- APC, & CIPESA. (2012). *Uganda Open Government data readiness study*.
- Avila, R., Feigenblatt, H., Heacock, R., & Heller, N. (2011). *Global mapping of technology for transparency and accountability: New technologies*.
- Ballingall, R. (2011). Is There a Case For Limiting Democratic Openness? *Public Policy and Governance Review*, 2(2), 52–62.
- Bates, J. (2012). “This is what modern deregulation looks like” : co-optation and contestation in the shaping of the UK’s Open Government Data Initiative. *The Journal of Community Informatics*, 8(2).
- Bowker, G. C., & Star, S. L. (2000). *Sorting Things Out: Classification and Its Consequences* (p. 389). The MIT Press.
- Braunschweig, K., Eberius, J., Thiele, M., & Lehner, W. (2012). The State of Open Data Limits of Current Open Data Platforms. *WWW2012*.
- Burkert, H. (2004). The Mechanics of Public Sector Information. In G. Aichholzer & H. Burkert (Eds.), . Edward Elgar Publishing.
- Calland, R., & Bentley, K. (2012). The Impact and Effectiveness of Accountability and Transparency Initiatives: Freedom of Information. *tbc*.
- Chambers, R. (2010). *Paradigms, Poverty and Adaptive Pluralism*. Brighton, UK.
- Craveiro, G. da S., Santana, M. T. De, & Albuquerque, J. P. de. (2013). Assessing Open Government Budgetary Data in Brazil. *ICDS 2013*.
- Davies, T. (2010, September 29). *Open data, democracy and public sector reform: A look at open government data use from data. gov. uk*. Practical Participation.
- Davies, T., & Bawa, Z. A. (2012). The Promises and Perils of Open Government Data (OGD). *The Journal of Community Informatics*, 8(2).
- De Boer, V., Leenheer, P. De, Bon, A., Gyan, N. B., Van, C., Gu, C., Tuyp, W., et al. (2012). RadioMarche : Distributed Voice- and Web-interfaced Market Information Systems under Rural Conditions. *Lecture Notes in Computer Science*, 7328.
- Dekkers, M., Poleman, F., Te Velde, R., & De Vries, M. (2006a). *MEPSIR - Measuring European Public Sector Information Resources*.
- Dekkers, M., Poleman, F., Te Velde, R., & De Vries, M. (2006b). *MEPSIR Study - 2006*.
- Diesing, P. (1971). *Patterns of Discovery in the Social Sciences* (Reprint.). AldineTransaction.
- EITI International Secretariat. (2012). Development of an EITI Data Standard.
- Farhan, H., D’Agostino, D., & Worthington, H. (2012). *Web Index 2012*.
- Fioretti, M. (2010). *Open Data, Open Society. Work* (pp. 1–65). Pisa.

- Fukuyama, F. (2013). *What Is Governance?*
- Fung, A., Graham, M., & Weil, D. (2007). *Full Disclosure: The Perils and Promise of Transparency* (1st ed., p. 300). Cambridge University Press.
- Fung, A., & Weil, D. (2010). Open Government and Open Society. In D. Lathrop & L. Ruma (Eds.), (1st ed.). O'Reilly Media.
- Garcia, A. A. (2011). *Methodology for Releasing Free Data (MELODA)* (pp. 1–15).
- Global Integrity. (2012). So What's In Those OGP Action Plans, Anyway? *Global Integrity Blog*. Retrieved from <http://globalintegrity.org/blog/whats-in-OGP-action-plans>
- Gray, J., Chambers, L., & Bounegru, L. (2012). *The Data Journalism Handbook* (p. 242). O'Reilly Media, Inc.
- Grewal, A., Iglesias, C., Alonso, J. M., Boyera, S., & Bratt, S. (2011). *Open Government Data - Feasibility Study in Ghana*.
- Gurstein, M. (2007). *What is community informatics (and Why Does It Matter)?* (Volume 2.). Polimetrica.
- Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2).
- Hammell, R., Perricos, C., Branch, D., & Lewis, H. (2011). *Unlocking growth: How open data creates new opportunities for the UK. Growth (Lakeland)*. London, UK.
- Harper, J. (2012). Grading the Government's Data Publication Practices.
- Heald, D. (2006). Varieties of Transparency. *Proceedings of the British Academy*, 135, 25–43.
- Heald, D. (2011). When Transparency meets Surveillance: External Monitoring of Country Public Finances. *1st Global Conference on Transparency Research* (pp. 19–20). Newark, New Jersey.
- Helbig, N., Cresswell, A. M., Burke, G. B., & Luna-reyes, L. (2012). *The Dynamics of Opening Government Data*. New York.
- Helbig, N., Cresswell, A. M., Burke, G. B., Pardo, T. A., & Luna-Reyes, L. (2012). *Modeling the Informational Relationships between Government and Society: A Pre-Workshop White Paper*. New York.
- Heusser, F. I. (2012). Understanding Open Government Data and addressing its Impact.
- HM Government Protection of Freedoms Act 2012 (2012). Queen's Printer of Acts of Parliament.
- HM Government. (2012b). *Open Data White Paper: Unleashing the Potential*. London, UK: HM Government.
- Hogge, B. (2010). *Open Data Study. Transparency and Accountability Initiative*. Transparency and Accountability Initiative.
- Iglesias, C. (2011). *Open Government Data - Feasibility Study in Chile (Draft)*.
- Janssen, K. (2012). Open Government Data and the Right to Information: Opportunities and Obstacles. *The Journal of Community Informatics*, 8(2).
- Joshi, A. (2012). Do They Work? Assessing the Impact of Transparency and Accountability Initiatives in Service Delivery. *tbc*.
- Kamarck, E. C., & Nye, J. S. (2002). *Governance.Com: Democracy in the Information Age* (Vol. 5, p. 208). Brookings Institution Press.
- Krikorian, G., & Kapczynski, A. (2010). *Access to Knowledge in an Age of Intellectual Property* (p. 646). Zone Books.
- Kuk, G., & Davies, T. (2011). The Roles of Agency and Artifacts in Assembling Open Data Complementarities. *Thirty Second International Conference on Information Systems*.
- Kundra, V. (2012). Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect. *Public Policy*. Cambridge, Mass.

- Kuriyan, R., Bailur, S., Gigler, B.-S., & Park, K. R. (2012). *Technologies for Transparency and Accountability*. Washington DC.
- Majeed, R. (2012). *Disseminating the power of information - Kenya Open Data Initiative, 2011 - 2012*.
- Malmud, C., & O'Reilly, T. (2007, December). 8 Principles of Open Government Data. Retrieved June 1, 2010, from http://resource.org/8_principles.html
- Mayo, E., & Steinberg, T. (2007). *Power of Information Review*.
- McClellan, T. (2011). Not with a Bang but a Whimper The Politics of Accountability and Open Data in the UK.
- McGee, R. (2011). Apparently transparent: Does accountability deliver? *The Broker Online*, 14–18.
- Mcgee, R., & Gaventa, J. (2011). *Shifting Power? Assessing the Impact of Transparency and Accountability Initiatives*.
- Meijer, A. J. (2007). Publishing public performance results on the Internet. *Government Information Quarterly*, 24(1), 165–185. doi:10.1016/j.giq.2006.01.014
- Murillo, M. J. (2012). Including all audiences in the government loop: From transparency to empowerment through open government data.
- Murray, A. (2011). Transparency , Scrutiny and Responsiveness : Fashioning a Private Space within the Information Society. *The Political Quarterly*, 82(4), 509–514. doi:10.1111/j.1467-923X.2011.02245.x
- Newbery, D., Bently, L., & Pollock, R. (2008). Models of public sector information provision via trading funds.
- Nichols, R. (2010). Do apps for democracy and other contests create sustainable applications. *Government Technology*. Retrieved March 6, 2013, from <http://www.govtech.com/e-government/Do-Apps-for-Democracy-and-Other.html?page=1>
- O'Reilly, T. (2010). Government as a platform. In D. Lathrop & L. Ruma (Eds.), (1st ed.). O'Reilly Media.
- OKF - Open Knowledge Foundation. (2006). Open Knowledge Definition. Retrieved March 4, 2010, from <http://www.opendefinition.org/>
- OKF - Open Knowledge Foundation. (2012). Census - Open Government Data Dashboard. *Open Government Data Dashboard*. Retrieved December 2, 2012, from <http://dashboard.opengovernmentdata.org/census/>
- OKF - Open Knowledge Foundation, & Access Info. (2011). *Beyond Access : Open Government Data & the Right to (Re)use Public Information*. Madrid.
- Open Government Partnership. (2011). *Open Government Declaration* (pp. 1–2).
- Perini, F. (2012). *Workshop report: Fostering a Critical Development Perspective on Open Government Data*. Brasilia.
- Pollock, R. (2009, May). The Economics of Public Sector Information. Retrieved March 4, 2010, from <http://econpapers.repec.org/paper/camcamdae/0920.htm>
- Powell, M., Davies, T., & Taylor, K. C. (2012). ICT for or against development? An introduction to the ongoing case of Web3.
- Power, T. J., & Taylor, M. M. (2011). *Corruption and Democracy in Brazil: The struggle for accountability*. University of Notre Dame.
- Rahemtulla, H., Custer, S., Tisacova, I., Jhalla, K., Gigler, B.-S., & Brigham, C. (2012). *The Journey of Open Government and Open Data Moldova*.
- Rahemtulla, H., Kaplan, J., Gigler, B.-S., Cluster, S., Kiess, J., & Brigham, C. (2011). *Open Data Kenya: Case study of the Underlying Drivers, Principle Objectives and Evolution of one of the first Open Data Initiatives in Africa*.
- Raman, B. (2012, April 4). The Rhetoric of Transparency and its Reality: Transparent Territories, Opaque Power and Empowerment. *The Journal of Community Informatics*.

- Rath, S. (2012, February 4). Data Template For District Economic Planning. *The Journal of Community Informatics*.
- Sharif, R. M. (2009). Maximizing the Value of Public Sector Information for Scientific and Socioeconomic Development in Africa | www.KMAfrica.com - Knowledge Management Africa KnowledgeHub.
- Shemie, D., Hirn, M., Raman, V., Paradi-Guildford, C., Sabet, E., Simon, Y., Rop, R., et al. (2012). Water Hackathon: Lessons Learned. Washington DC.
- Speck, B. (2011). Audit Institutions. In T. J. Power & M. M. Taylor (Eds.), *Corruption and Democracy in Brazil: the struggle for accountability*. University of Notre Dame.
- Srinivasan, S. (2012). Mapping the Tso Kar basin in Ladakh. *The Journal of Community Informatics*, 8(2).
- Steinberg, T., & Mayo, E. (2007). Power of Information Taskforce Report (Beta).
- Stott, A., & Kaplan, J. (2013). *Open Data Readiness Assessment Tool (January 2013 version)*.
- Taggart, C. (2012). *The Closed World of Company Data An examination of how*.
- Tauberer, J. (2010). Case Study: Govtrack.us. In D. Lathrop & L. Ruma (Eds.), (1st ed.). O'Reilly Media.
- Tauberer, J. (2012). *Open Government Data*.
- Thomson, M. (2013). *Development 2.0 and beyond: Challenges for ICT4D in 2013*.
- Tong, A., Irsha, H., & Revell Ward, D. (2013). *Market Assessment of Public Sector Information*.
- Transparency International Georgia. (2012). Ten Open Data Guidelines | Transparency International Georgia. Retrieved July 16, 2012, from <http://transparency.ge/en/ten-open-data-guidelines>
- Uhlir, P. (2009). *The Socioeconomic Effects of Public Sector Information on Digital Networks: Toward a Better Understanding of Different Access and Reuse Policies: Workshop Summary* (p. 105). Washington, D.C.: National Academies Press.
- UN - United Nations. (2010). *United Nations E-Government Survey 2010: Leveraging e-governmnet at a time of financial and economic crisis*. United Nations.
- UN ESCAP. (n.d.). What is Good Governance.
- UNDP - United Nations Development Programme. (1997). *Governance for sustainable human development: A UNDP policy document*.
- Von Hippel, E. (2005). *Democratizing innovation*. the MIT Press.
- Von Hippel, E., & Von Krogh, G. (2009). Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science. *SSRN eLibrary*. MIT Sloan Research Papers. doi:10.2139/ssrn.1410789
- Weinstein, J., & Goldstein, J. (2012). The Benefits of a Big Tent: Opening Up Government in Developing Countries. *UCLA Law Review Discourse*, 38(2012), 38–48.
- World Bank. (2012). *Project paper on a proposed additional credit to the Republic of Kenya for a Kenya Transparency and Communications Infrastructure Project*.

