

# Exploring the Emerging Impacts of Open Data in Developing Countries network meeting report

**April 2013**

**Dates:**

24th - 26th April 2013

**Location:**

The Open Data Institute, London



The project is funded by the Canadian International Development Research Centre (Grant 107075) and is managed by the World Wide Web Foundation.

### Contents:

[Introduction](#)

[Setting the scene: project launch and introductory presentations](#)

[Project launch](#)

[Case studies: discussion and questions](#)

[Case study presentations](#)

[Discussion issues](#)

[Research framework and methods](#)

[Approaches to open data research](#)

[Developing the research framework](#)

[Linking the framework to methods](#)

[Next steps on methods:](#)

[Thematic and regional clusters](#)

[Thematic clusters](#)

[Regional clusters](#)

### Introduction

In April 2013 project leaders from the Exploring the Emerging Impacts of Open Data in Developing Countries (ODDC) research network gathered in London for a two-day network meeting. The meeting was an opportunity for project leaders to get to know each other, and to explore shared research issues in understanding emerging impacts of open data. This report captures key discussions and learning from the event and is a record for participants.

### Setting the scene: project launch and introductory presentations

#### Project launch

The evening before the network meeting, we attended an evening reception at Lancaster House co-hosted by the Open Government Partnership steering committee. At this reception, Web Foundation founder, and Web inventor, Sir Tim Berners-Lee formally launched the ODDC project, noting the importance of the project to the work of the Open Government Partnership (OGP) in exploring open government data issues across the world. Early findings from the ODDC project will be presented at the Open Government Partnership plenary meeting in London on October 31st/November 1st.

The ODDC project is the result of two-years of preparation by the World Wide Web Foundation (WF) and Canada's International Development Research Centre (IDRC). The open call for proposals that resulted in the selection of the case studies that form the core of the research network was developed following a fringe workshop alongside the first OGP plenary meeting in Brasilia, Brazil in 2012.

#### A brief history of open data

The full network meeting opened with a presentation from Professor Nigel Shadbolt, Chair of the Open Data Institute. Key themes from the presentation and discussion included:

- **The history of open data**, going back to project like Florence Nightingales 'rose diagram'<sup>1</sup> and John Snow's 'cholera map'<sup>2</sup> which made use of data collection and visualisation to influence public policy in the 1850s;
- **The history of the UK open data initiative** which Nigel described as resulting from a combination of 'politics and luck'. The initiative started in 2009 and led to a data portal being launched in 2010. The initiative was continued following a change of government in 2010. Creating and showing evidence of the impact of open data has been key in the development of the initiative.

---

<sup>1</sup> See <http://understandinguncertainty.org/coxcombs>

<sup>2</sup> See [http://en.wikipedia.org/wiki/1854\\_Broad\\_Street\\_cholera\\_outbreak](http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak)

- **The importance of data standards** and the considerable work that it has taken in the UK to show the benefits of standardisation, through projects like AKTive PSI in 2005-2006. Pre-existing interoperability across government significantly facilitates open data publication and use.
- **The role of the web** in supporting the emergence of open data, including by driving a minimal marginal cost of open data. Nigel also discussed the potential of semantic web technologies to link up datasets from across government.
- **The importance of certain datasets** such as location data, to provide linkage points for other datasets. Identifying the importance of certain 'core reference datasets' in different countries will be an important cross-cutting question for ODDC.
- **Personal data**, both in terms of how open data can raise privacy concerns when data derived from personal information is shared, and how ideas of giving individuals access to their own personal data are being explored alongside open data initiatives through programmes such as MiData<sup>3</sup>.
- **Building open data ecosystems** that require change in behaviour of data owners and users, businesses and others. The Open Data Institute has been established to build capacity and increased the demand for open data, incubating innovators, companies and skills for working with open data.

Whilst focussed on learning from the UK context, the presentation and discussion also focussed on the spread of open data ideas across the world. Through comparing across case studies in the ODDC project we hope to understand the common elements that open data initiatives all share, and how they are localised in different countries and contexts.

## Case studies: discussion and questions

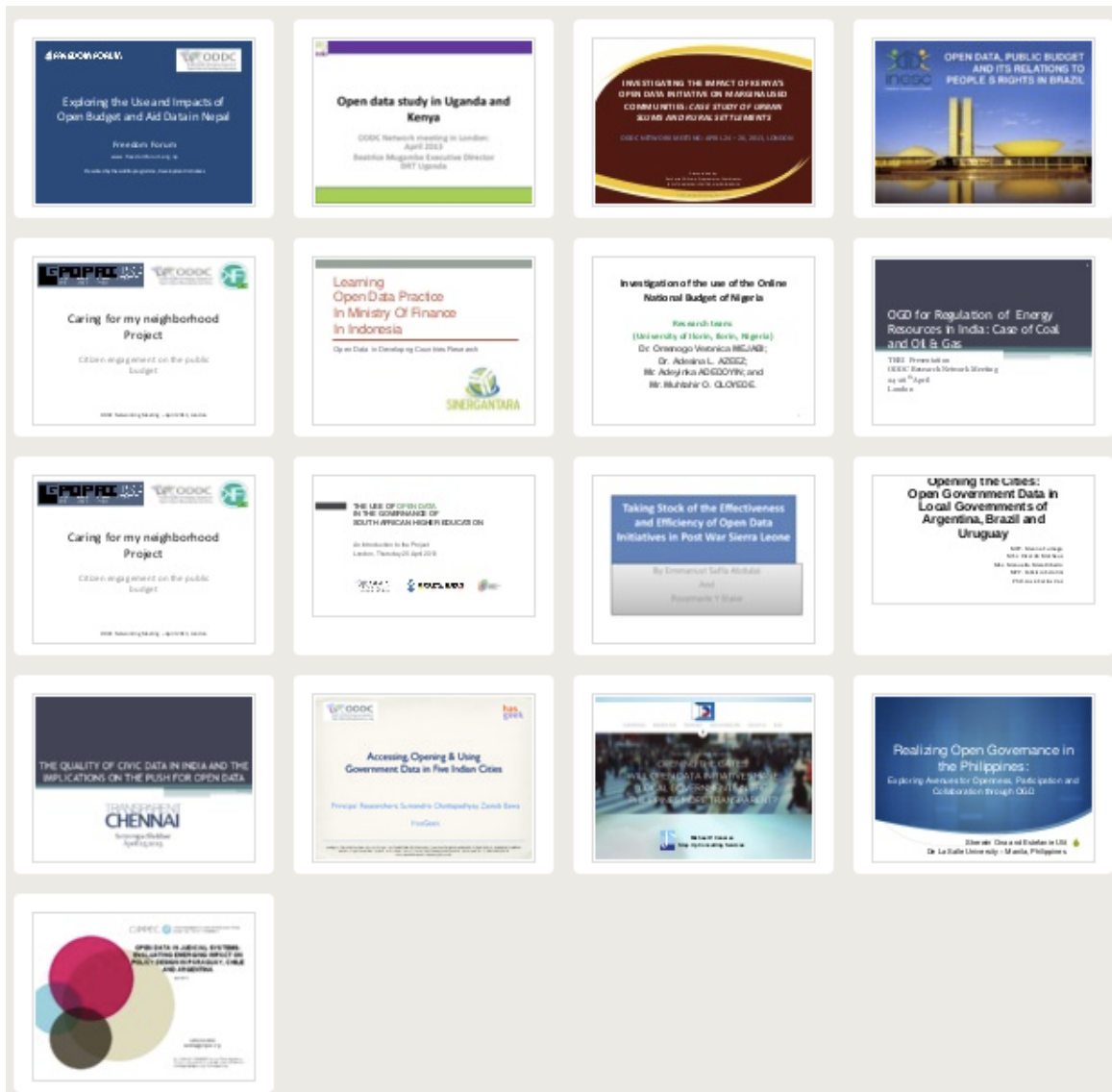
### Case study presentations

Most of the rest of the first day of the meeting consisted of presentations on each of the case studies that will be taking place over the coming year. These presentations provided an overview of the context in which each open data initiative was talking place, the data in focus, and the governance and decision making issues being explored. Slides from each presentation can be found at: <http://www.slideshare.net/odrnetwork/tag/april2013>

---

<sup>3</sup> See

<https://www.gov.uk/government/policies/providing-better-information-and-protection-for-consumers/supporting-pages/personal-data>



## Discussion issues

During the presentations there was time for discussion. The list below outlines a number of the questions raised during discussion.

## Drivers for open data

- Is the development of open data responding to local need or global agendas?
- What is the incentive for policy makers to put data up there? What are the policy alternatives?
- What are the businesses cases for open data in different settings. For example: education, health and extractives industries?
- Who are the commercial actors with an interest in open data?
- Should open data be seen as a tool to address specific issues? Or a public good that can support innovative problem solving?

- By linking open data to particular governance issues we risk treating it as a narrow functional tool, only to be used once a problem is identified. But in many cases, such as the example of GP prescription data being released in the UK<sup>4</sup>, data was published, and then outsider found innovative ways to re-use it to address governance challenges.
- What is the role of national government in driving open data? Or can effective initiatives emerge solely at the municipal level?

### Data availability and supply

- Sometimes data doesn't exist at all, or isn't in a digital form. The data that is available may be politicised in various ways.
- What about grey areas: some people (often the marginalised) can benefit from being in the grey area? Officially 'illegal' practices can become legitimised over time (e.g. squatting) - how does data interact with this? There can be about situations where particular (illegal/informal) citizen groups consciously avoid enumeration by government surveyors (spatial or social surveys)? Should missing information about such groups be necessarily read as political decisions on part of the state, or should be seen as coming from survival strategies of these citizen groups?
- How do government records and data relate?
- The condition of machine readability is only met by very few of the 'datasets' published by governments across the world: a lot of the data in our cases is PDF or otherwise non machine-readable right now.

### Data quality

- What criteria can be used to assess good quality of data? Whether data is good quality or not may be relative to the use it is being put to.
  - The Five Stars of Linked Open Data (<http://5stardata.info/>) outline a way of assessing the quality of data against the linked data paradigm;
  - The Open Government Data principles at <http://www.opengovdata.org/home/8principles> outline a normative framework for good quality government data for transparency. But these raise the question of what 'complete' and 'primary' data means in different thematic domains?
  - There is a framework for assessing the specific qualities of Brazilian open budget data at <http://bit.ly/12GNjLz>
- What are unintended consequences of releasing poor quality data?
- Governments are often unwilling to release data, because of the threat to prestige of releasing bad data and concerns about data being taken as 'word of god' when it might not be accurate. How far does this issue differ in importance depending on government, civil service and national cultures?
- Do open data ecosystems accommodate the dynamic nature of data? Can citizens demand to have data corrected? If it is corrected, will these corrections be reflected

---

<sup>4</sup> See <http://blog.revolutionanalytics.com/2012/12/nhs-prescription-analytics.html>

where data has been re-used, or do outdated copies of the data perpetuate and hang around?

### Data use

- Who is using this data? What for? What kinds of data are they asking for?
- What role do boundary partners<sup>5</sup> play in shaping the use of open data for governance?
- What's the connection between NGOs and hackers in using open data?
- What role do coding cultures play in shaping the nature of open data<sup>6</sup>?
- What role are big firms playing in driving practice around open data?  
e.g. Google; IBM; Microsoft; Amazon
- What methodologies do companies and developers use to measure the impact of applications built with open data? What metrics are needed?
- What about literacy levels in different countries? Or within countries - e.g. slums?
  - Who gains access and who are the intermediaries?
  - Is work on open data too centred on the web, to the exclusion of other means of accessing information?
- How does accessing open data on a mobile phone affect what someone can do with it?
- What is relationship between open data and Right to Information (RTI) / Freedom of Information (FOI)? Can open data operate effectively in countries where there is no FOI law?
- What role are data standards playing in support data re-use?
  - We don't imagine that every city will have capacity to create data for everything and have all the communities of developers needed to serve all needs. So - with so much experimentation, issues and areas - what are the specific issues with water, transport, other things that all cities might face, and how are ideas and standards and technologies being shared between cities.

## Research framework and methods

### Approaches to open data research

On the first day of the meeting we discussed a range of different approaches to open data research, and existing research we knew of.

The Web Foundation **Open Data Feasibility Studies**<sup>7</sup> look at open data readiness, assessing the political, economic, social, legal, technical and organisational resources in a country available to support implementation of an open data initiative.

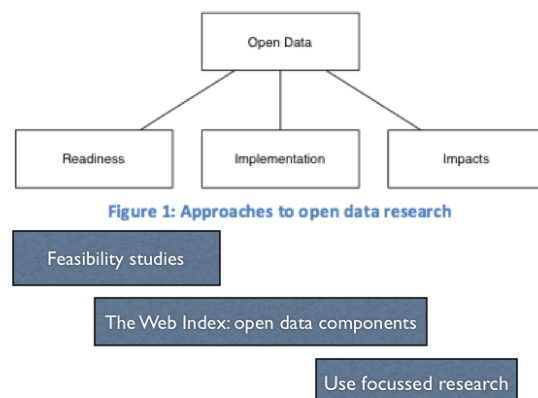


Figure 1: Approaches to open data research

<sup>5</sup> See [http://en.wikipedia.org/wiki/Outcome\\_mapping#Identifying\\_boundary\\_partners](http://en.wikipedia.org/wiki/Outcome_mapping#Identifying_boundary_partners)

<sup>6</sup> See <http://www.swansea.ac.uk/staff/academic/artshumanities/pcs/berryd/>

<sup>7</sup> See <http://www.webfoundation.org/projects/open-government-data-feasibility-studies/>

**The Web Index** is a multi-country survey combining expert-survey evidence with secondary indicator data to provide an assessment of how far different countries have the foundations in place for effective use of the web, and how far the web is having social, political and economic impact in those countries<sup>8</sup>. The 2013 Index will include additional questions on open data to measure how far countries have implemented open data policy, by looking at which datasets are available as open data. This is a similar approach to that taken by the Open Data Census<sup>9</sup>. Less research exists that looks at impacts of open data. The majority of evidence remains at the level of anecdotes. The ODDC project is designed to address this gap, using rigorous case study methods to identify **emerging impacts** of open data.

We talk about **emerging impacts** because we are not trying to establish full causality at this stage of the research. In other words, we are *not* looking to say “If this dataset is opened, then that impact will definitely result”. Instead, we are looking to be establish detailed well researched accounts of how open data is being used, and the impacts that those involved in open data use believe are occurring; and we are looking to use triangulation and a range of different methods to check these reports of impacts from open data. In accounting for how open data is being used we are also looking to identify unintended consequences, and where expected impacts have not resulted. We will be using conceptual tools from readiness and implementation assessments to describe the context of our detailed cases of open data use.

### Developing the research framework

On the second day of the meeting we discussed the project research framework. The framework highlights that open data comes from a range of sources, is used in a range of different governance settings, and can have a range of different impacts. The development of open data is not linear, but involves the dynamic relationship of these: where introduction of open data can create new possibilities for governance and decision making, feeding into changes in the availability of data over time.

During discussion of the framework we identified some modifications to be made, including renaming the **‘Actions and impacts’** component to ‘Use and impact’ to emphasise that many of the emerging impacts of open data may not result directly from data use, but from individuals or organisations making use of other platforms and tools that have been built with open data. For example, when a parent uses a website to locate a school for their child, it is hard to describe this directly as ‘data use’, but their choice of school may be influenced by open data which was used to create the website by an intermediary.



---

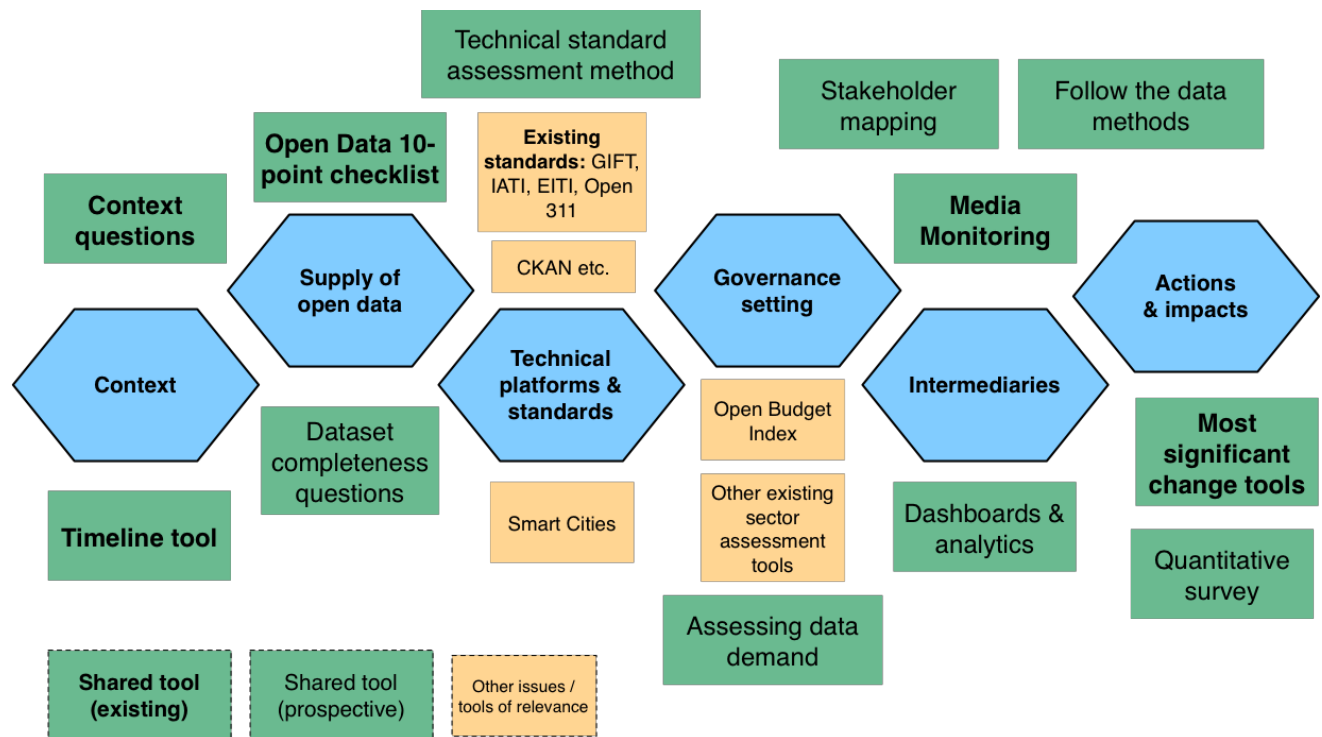
<sup>8</sup> See <http://www.thewebindex.org>

<sup>9</sup> See <http://census.okfn.org/>



### Linking the framework to methods

To enable comparisons and cross-cutting research between ODDC case studies we need to make use of shared research tools wherever possible. Using post-it notes we identified a range of existing and prospective shared research tools, and other issues or existing initiatives we should have reference to in different parts of our research. We then broke into groups to explore a number of these methods in more detail. The results of the exercise are shown below.



This highlights the existing tools the ODDC project has developed which include

- [Context questions](#) which projects can use in writing a contextual assessment of the open data initiative they are studying<sup>10</sup>.
- **Timeline tool** (detailed in the context questions tool) which can be used to record and visualise key events in the development of an open data initiative;
- [10 point dataset supply checklist](#) for assessing the technical features of a dataset<sup>11</sup>.

In addition to the 10-point checklist we discussed the need for methods that can measure whether a dataset is **complete, primary and timely** (Dataset completeness questions). These methods will need to be customised for different contexts - as the criteria and measurement methods for a complete, primary and timely dataset may be different from criteria and measurement methods for education, health or judicial data.

<sup>10</sup> See <http://bit.ly/162qOWb>

<sup>11</sup> See <http://bit.ly/11hazxy>

### **Discussion notes: Principles of open data / data completeness**

- In particular cases and countries, machine readable data might not be universally accessible. We may need to consider also including human readable forms of data; and data which may not be available electronically;
- Does access to underlying public records feature as part of the definition of quality open data?
- We need to look at the definition of data
- It is Important to have data as disaggregated as possible: different groups have different needs.
- Historical data and comparison are important
- Not just open government data, but also data made available by companies and research groups.

Development of a framework could start from the eight principles of OGD. There is applied work for budget, and the group are open for new cases and considerations. There is a need for discussion of what 'non-discrimination' means in the principles?

We also identified the need to methods that can explore the impact of open data standards (**Technical standard assessment methods**). These should be able to: introduce a particular standard (such as Open 311, the International Aid Transparency Initiative standard, or the Open Contracting standard), and support an assessment of how this standard affects what can and can't be done with data. These methods may draw upon evidence gathered from assessments of data demand.

A tool for **assessing open data demand** is needed to understand the kinds of data that people in a governance setting want. Simply asking 'what data do you want' doesn't always get the best results, so improved interview questions and approaches are needed for this.

The need for **stakeholder mapping** methods that can be used to understand a governance setting, and to work out the important intermediaries to be interviewed in more depth during research, was identified.

### **Discussion notes: Stakeholder mapping**

- Needs to take place at demand and supply side
- It is easier to map the supply side, but harder to map the demand side
- Stakeholder mapping can be used to identify the hypothetical end; in other cases use to identify key informants and experts; and to identify the data flows;

Process:

- (1) Look at context; settings, service providers; get list of NGOs and Media
- (2) Categorisation: incentives, motivations, use of data
- (3) Workshops to do the mapping and expand list: identify new stakeholders and

- intermediaries
- (4) Draw how the stakeholders interact with each other / process mapping: through the supply chain of data; creation of platforms and apps; users of end info/knowledge; advocacy.

Alongside case-specific interview approaches we discussed **media monitoring** as a means to assess the role of the media as open data intermediaries, including monitoring mainstream and social media.

### Discussion notes: media monitoring

There are number of key variables for monitoring: Name of Newspaper (To be selected); Name of Author; Author's contact; Length of article (no. of words); Date of Publication; News Items; Themes; Prominence; Direction ; Tone; Quality of the Report; Assumed or Established Sources; Implication for impacting policy; Implication for impacting public opinion; Prose commentary/summary/comments

#### Considerations:

- Are newspapers available online as well as in print? (time saving?)
- Are there tools to monitor online newspapers for content?
- Ensuring that criteria enable objective research
- What time frames should we be using to monitor print media? - Forward looking? Backward looking? What is an acceptable duration to make the research legitimate?

#### Social Media Monitoring

Possible tools for extracting details from social media include: Data sift - <http://datasift.com/> ; Swift river - <http://www.ushahidi.com/products/swiftriver-platform>; Weka - <http://www.cs.waikato.ac.nz/ml/weka/>.

The use of **web analytics from app-builders and technology intermediaries** is also an area to explore.

For identifying emerging impacts a number of methods may be used. Firstly, **quantitative and qualitative survey and focus groups methods** will be appropriate for a number of cases seeking to understand the breadth of awareness and use of open data.

### Discussion notes: survey methods

- A quantitative survey may draw upon baseline research with pre- and post- interventional assessments of some indicator
- In both quantitative and qualitative survey we need to think about sampling: who to

survey & talk to;

- Consider issues of age and who is using the Internet;
- Understanding the digital divide and questions of gender;
- Questions of awareness: sometimes the awareness is subtle;

Focus group discussions will help us to capture some of the awareness levels

Various cases are developing survey and interview schedules which it will be valuable to share with the network.

To better establish the connection between impacts and data, **follow the data** methods are being developed which seek to use interview, sampling and triangulation techniques to understand the role that open data has played in a particular governance setting.

### Discussion notes: follow the data

- Start from a **particular use of data** - and interview the user about how they made use of the data; or start from a dataset, and interview the data owner about how the data has been made open, and who might be using it;
- **Plot visually** the steps involved in creating or using the data, using open question to solicit information about why the data, or product of the data, is the way it is;
- Where there are gaps in the narrative, **identify extra informants** to interview, or sources of information that can fill these gaps;
- At each stage of the journey from data to impact, identify the format the data is in, and other key features of the data;

The method can also be used where there is no open data available by asking someone who could use data to describe a current task they perform, and then asking how they would make use of open data in this setting.

Using logos and pictures of datasets, and drawing a picture of the way data has been used with an interviewee can be good ways to solicit information. Asked if they have used particular datasets people might not recognise the dataset name - but showing the dataset on a screen or a screenshot or logo on a card - can help remind people about their own data use.

Lastly, **most significant change (MSC) methods**<sup>12</sup> might be used, both to assess the impact of open data, and the impact of our own research into it. MSC is an evaluation method often used in development projects, which asks people to identify the most significant change brought about by an intervention. It may provide a means to rigorously explore emerging impacts of open data.

### Discussion notes: most significant change

---

<sup>12</sup> See <http://www.kstoolkit.org/Most+Significant+Change>

### Two possible approaches

- Evaluate the network and the project impact: reflect on the possible impact that your project as resulted in?
  - Were you able to convince other groups to take on your open data advocacy?
  - Were you part of public or academic groups on OGD?
  - Need a self-monitoring and evaluation
- Evaluate interventions
  - Using most significant change tool to interview people who have used data or seen it used.

### Next steps on methods

An area for sharing methods tools will be developed on the Open Data Research website: [www.opendataresearch.org](http://www.opendataresearch.org)

## Thematic and regional clusters

### Thematic clusters

During the selection of proposals for the ODDC project we originally organised cases into a number of thematic clusters: budget transparency; urban governance; poverty alleviation; and emerging issues. During the project presentations it became clear that these clusters did not work well for all projects. We explored alternative clusters, and held discussions in small groups.

### Regional clusters

On the Friday we met in regional clusters to identify **key issues** for open data research in each region, and opportunities for further collaboration.

### Africa

<b>Key issues</b>	<ul style="list-style-type: none"><li>● To see how we can influence legislation around open data;</li><li>● How to advocate to push for freedom of information acts in respective countries;</li><li>● Increasing supply of open data;</li><li>● Teaching / increasing the level of awareness of open data - sensitisation with different stakeholders</li><li>● Using MSC tools to showcase the stories of change that have come from using open data (also creative commons approach)</li><li>● Improving the discoverability of data</li></ul>
<b>Key stakeholders</b>	<ul style="list-style-type: none"><li>● Government</li></ul>

	<ul style="list-style-type: none"> <li>● Leveraging existing partnerships - e.g. Kenya ICT board</li> <li>● Raise awareness amongst NGOs and CSOs - people who can pressure the government work on open data and the legislation around it</li> <li>● Sensitise academia</li> <li>● Media</li> </ul>
<b>Opportunities</b>	<ul style="list-style-type: none"> <li>● ICT4D conference in South Africa in December</li> <li>● Open Government Meeting in Nairobi - in May</li> <li>● Planning regional meeting / conference on Open Data and looking towards future</li> </ul>
<b>Needs</b>	<ul style="list-style-type: none"> <li>● Support around organising an event - and getting out of Nairobi Cape Town or Ghana to new locations</li> </ul>

### Latin America

<b>Key issues</b>	Focus on influencing governance at different levels: government, CSOs, Universities etc.
<b>Opportunities</b>	<ul style="list-style-type: none"> <li>● Conference and unconference in June in Montevideo</li> <li>● Keep momentum - identified some OGP related events that could help to carry out agenda <ul style="list-style-type: none"> <li>○ Projects can contribute to this - at least in October by sharing 'low-hanging fruits' of what is being discovered.</li> </ul> </li> <li>● Thinking about Open Knowledge Conference in September</li> </ul>

### Asia

<b>Key issues</b>	<ul style="list-style-type: none"> <li>● Integrating open data and RTI discussions</li> <li>● Actual data usage</li> <li>● Understanding restrictions to data (e.g. security, privacy, confidentiality etc.)</li> <li>● Hindrances to open data (e.g. capacity - supply and demand side / technical limitations in the data / lack of nation interoperability frameworks)</li> </ul>
<b>Key stakeholders</b>	<ul style="list-style-type: none"> <li>● Local government (city, municipality, bureaucrats + politicians)</li> <li>● Commercial associations/guilds, professional organisations</li> <li>● Media organisations and journalists (from more right/freedom to information journalism to data journalism)</li> <li>● Specific govt depts (eg, finance, local govt, energy and natural resources etc.)</li> </ul>

	<ul style="list-style-type: none"><li>● IT and technical arms of the govt.</li><li>● International forces (world bank etc)</li><li>● Regional pressures (regional cross country influences?)</li></ul>
<b>Opportunities</b>	<ul style="list-style-type: none"><li>● A regional meeting in july</li><li>● Joint panel proposal for IGF 2013 in Bali (october): presentation from all south and south-east asian ODDC projects.</li></ul>
<b>Needs</b>	Travel support for participants.

### Organisational issues: network communication

During the meeting we discussed the different communication tools available in the network and agreed the tools we would use. The results of this discussion can be found in [the updated Communication Guide for the project](#)<sup>13</sup>.

We agreed that the network website at [www.opendataresearch.org](http://www.opendataresearch.org) would be updated in the coming months to reflect themes in open data research, rather than being centred around the individual projects.

### Next steps:

- The network will continue to communicate over e-mail and regular web meetings. Regional clusters are encouraged to develop their own activities.
- Case studies are expected to start reporting findings by early 2014.
- Future meeting plans will be developed regional in clusters, and as a network, later on in 2013.

### Participants & thanks:

#### Participants

Alexandre Ciconello; Beatrice Mugambe; Emmanuel Saffa Abdulai; Estafanie Ulit; Fabrizio Scrollini; Fernando Perini; Francois Van Schalkwyk; Gisele Da Silva Craveiro; Ilham Cendekia; Isabel Bucknal; Jose Alonso; Leonida Mutuku; Michael P. Cañares; Nidhi Srivastava; Omenogo Mejabi; Ricardo Matheus; Sandra Elena; Satyarupa Shekhar; Sherwin Ona; Sumandro Chattapadhyay; Tim Davies; Veena Aggarwal; Zacharia Chiliswa; Zainab Bawa

#### Thanks:

With thanks to: Professor Nigel Shadbolt; Sir Tim Berners Lee and Hania Farhan for input into the meeting. And to: Andrea Cox, Jade Goodman and the team at the The Open Data Institute for hosting us, and Betty Purwandari for logistical support throughout the event.

---

<sup>13</sup> See <http://bit.ly/YUTyKV>