What's New in Open Data Assessment Tools

ODRA 3.1, Open Data for Business, Transport and Energy

Webinar
Thursday, 31 March 2016
10:00 AM-12:00 PM EST

Presented by Transport & ICT Global Practice and Development Economics Data Group
10:00 – 10:10 Introduction and main changes in ODRA 3.1 – **Amparo Ballivian**, Lead Economist

10:10 – 10:20 Legal Review section updates and guidance – **David Satola**, Lead Counsel


10:45 – 11:05 Introduction of new Open Data Assessment tools:
   - Open Data for Business (OD4B) – **Alla Morrison**, Program Officer
   - Transport Data Assessment – **Alla Morrison**, Program Officer

11:05 – 11:55 Q&A

11:55 – 12:00 Closing remarks and way forward – **Amparo Ballivian**, Lead Economist
ODRA Legal Issues

Substance

Enabling Environment

- Anonymization
- FoIA
- Licensing
- Reuse
- Fees

Safeguards

- Anonymization
- Data Protection/Privacy
  - limits on the collection of data
  - Specifies purpose for which data is collected at the time of collection
  - Use of the data specified
  - security to protect the data
  - Rights of individuals to know whether the government has data about him/her
  - Use of information must be relevant, necessary, accurate and complete
ODRA Legal Issues
Process

• ? Bank-Executed vs Recipient-executed ?
• ? Level of Review?
• ? “Recommendations”? 

If B-E...

• All usual articles-based restrictions
  • no “design”
  • no “implementation”

• Disclaimer
  • not “due diligence”
  • not “legal advice”
Topics

- Mission organisation
- Assessment points
- Engaging Civil Society
- Scoring
- Datasets
- Action Plans
- Review & Publication
- Non-Assessment issues
Mission Organisation

- Key policy interviews best done early
- Briefing for all Ministries/Public events best done early
- If resources allow, consider partial twin-track approach (eg Supply/Demand)
- Consider using Saturday for civil society/developers
- Test whether right people are being seen
- May be easiest to set timetable yourself
Assessment Points

- Test extent to which policies/laws are implemented and used
- Test extent to which “Presidential orders” are effective
- Question how charges are set, costs of collecting them, % of operating budgets covered, what actually happens to money …
- Data management should include NSDI
- Dimension 8 could now be more standardised with ITU data, WDR and other World Bank data
Engaging Civil Society

- Some client counterparts not good at engaging civil society
  - Very poor turnout for public events
  - Unrepresentative civil society contacts
  - Only “formal” civil society

- Need to consider alternatives
  - Social Media
  - Established networks (eg OGP, OKFN)
  - Public events (perhaps evenings/Saturdays)
  - Government’s own “Civil Society” list
Scoring (1)

- Version 3 dropped colour-scoring of sub-dimensions
  - Rationale: insufficient data points for some dimensions
  - Yet some ODRAs still do it!
  - No harm in doing so, but consistency?

- Version 3 introduced “rubics”
  - Intent: Consistency, Objectivity, Transparency
  - Designed to be used to explain rating
  - Not being consistently used in reports
Scoring (2)

- Consider “Mixed” scores (YELLOW/RED) if needed to highlight partial weakness or avoid “soft Greens”
- Don’t allow client to claim “no Reds” until report is final (!)
Datasets

- Intent was that for each of “key datasets” ODRA should report:
  - How/If data is currently held
  - Feasibility of release: Policy, Technical, Institutional
  - Benefits of release
  - Risks of release (and how they could be managed)
  - Recommended priority for release

- “Supply” evidence from questionnaire and interviewing relevant Ministries

- “Demand” evidence from Civil Society and Business interviews
ODRAs are not doing this consistently

- Some are assessing current status of key datasets – for instance by reference to Open Data Index criteria
- Some are assessing feasibility, benefits and risks of release and making recommendations for priority
- … and in Burkina Faso the ODRA actually released data!
Datasets: Open Data Index Scoring

The key datasets of the open data index are available. The overview of key datasets is illustrated below.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Score</th>
<th>Breakdown</th>
<th>Location (URL)</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport Timetables</td>
<td>5%</td>
<td><img src="image" alt="5%" /></td>
<td><img src="image" alt="5%" /></td>
<td><img src="image" alt="5%" /></td>
</tr>
<tr>
<td>Government Budget</td>
<td>45%</td>
<td><img src="image" alt="45%" /></td>
<td><img src="image" alt="45%" /></td>
<td><img src="image" alt="45%" /></td>
</tr>
<tr>
<td>Government Spending</td>
<td>45%</td>
<td><img src="image" alt="45%" /></td>
<td><img src="image" alt="45%" /></td>
<td><img src="image" alt="45%" /></td>
</tr>
<tr>
<td>Election Results</td>
<td>0%</td>
<td><img src="image" alt="0%" /></td>
<td><img src="image" alt="0%" /></td>
<td><img src="image" alt="0%" /></td>
</tr>
<tr>
<td>Company Register</td>
<td>75%</td>
<td><img src="image" alt="75%" /></td>
<td><img src="image" alt="75%" /></td>
<td><img src="image" alt="75%" /></td>
</tr>
<tr>
<td>National Map</td>
<td>30%</td>
<td><img src="image" alt="30%" /></td>
<td><img src="image" alt="30%" /></td>
<td><img src="image" alt="30%" /></td>
</tr>
<tr>
<td>National Statistics</td>
<td>100%</td>
<td><img src="image" alt="100%" /></td>
<td><img src="image" alt="100%" /></td>
<td><img src="image" alt="100%" /></td>
</tr>
<tr>
<td>Legislation</td>
<td>45%</td>
<td><img src="image" alt="45%" /></td>
<td><img src="image" alt="45%" /></td>
<td><img src="image" alt="45%" /></td>
</tr>
<tr>
<td>Postcodes / Zipcodes</td>
<td>0%</td>
<td><img src="image" alt="0%" /></td>
<td><img src="image" alt="0%" /></td>
<td><img src="image" alt="0%" /></td>
</tr>
<tr>
<td>Pollutant Emissions</td>
<td>75%</td>
<td><img src="image" alt="75%" /></td>
<td><img src="image" alt="75%" /></td>
<td><img src="image" alt="75%" /></td>
</tr>
</tbody>
</table>

Key: Yes | No | Unsure | No data
Datasets: Fully structured analysis

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Environmental monitoring: levels of pollution, energy consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Availability</strong></td>
<td>The Ministry of Environment and Sustainable Development manages</td>
</tr>
<tr>
<td></td>
<td>the Environmental Information System of Colombia (SIAG PPI), it</td>
</tr>
<tr>
<td></td>
<td>is an informative portal for consultation of environmental</td>
</tr>
<tr>
<td></td>
<td>indicators redirected at other services such as the IDEAM.</td>
</tr>
<tr>
<td></td>
<td>Nonetheless, an advanced search of the SIAG portal to see</td>
</tr>
<tr>
<td></td>
<td>what documents are being published in Excel, shows 249 results</td>
</tr>
<tr>
<td></td>
<td>corresponding to data from the system of indicators. The</td>
</tr>
<tr>
<td></td>
<td>search carried out on Google is site: <a href="http://www.siac.gov.co">www.siac.gov.co</a> filetype:</td>
</tr>
<tr>
<td></td>
<td>These documents cannot be directly located on the web portal.</td>
</tr>
</tbody>
</table>

**Feasibility of openness: Political**

- The sphere of competence of the SIAG is the generation and exchange of information to support environmental management and research in Colombia. IDEAM currently publishes most of the environmental indicators.

**Feasibility of openness: Technical**

- Technical capacity for opening environmental data currently corresponds to IDEAM.

**Feasibility of openness: Institutional**

- Institutional, commitments towards opening data are those acquired by IDEAM.

**Benefits and Risks of openness**

- Those indicated by the data of the meteorological and environmental dataset.

**Recommendations**

- Group together environmental indicators in a category called “Environment.” **High Priority**

**Current contribution to open data portal**

- IDEAM publishes the dataset containing ozone levels under the “Mines and Energy” category.
## Datasets: Narrative structured analysis

<table>
<thead>
<tr>
<th>Ref</th>
<th>Dataset</th>
<th>Feasibility</th>
<th>Key Benefits &amp; Risks</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>D32</td>
<td>Road Accidents</td>
<td>Information on type, location and frequency of road accidents is recorded by the police and maintained by the Traffic Management and Road Safety Department (TMRSD). The police are starting to use a mapping device to get coordinates of the location of the accident which is then recorded with the other data. Data is not available publicly but is aggregated in some statistics. This data can be easily anonymized to exclude details of the victims and of driver behavior, and also details of police and other emergency services that attended the accident. Basic road accident data is easily extracted and published in other jurisdictions, including date and time, location, type(s) of vehicles involved, number and level of casualties.</td>
<td>This data is typically used to allow vulnerable road users to avoid accident hot-spots - such as finding safe cycle routes or safe walking routes for children going to school. It can also be used by community group to lobby for improvements in road design or for speed and other restrictions. Personal details of the people involved are not necessary for the data to be useful, and a safely anonymized dataset can be easily produced. Since road accidents inherently happen in public places there is no other sensitivity involved.</td>
<td>QUICK WIN</td>
</tr>
</tbody>
</table>
Summarising Dataset Findings
Action Plans

- Early clients wanted a detailed action plan
- If many actions, indicate priorities/lead actions
- In some countries early action may need to be just preparatory
- All countries seem challenged in getting program started
- Need to be realistic about capacity to deliver
- Portal should not run ahead of policies, even if easier
Review and Publication

- Consider Legal Review & Peer Review before client review
- Check Client review has involved stakeholders
- Consider fact-checking individual datasets with owners
- Some Bank-funded ODRAs have still not been published
Non-Assessment issues

- “Case for Open Data”
- Which license?
- Which software?
- Example solutions
- What do other countries do?

The ODRA is part of a Toolkit! – other items in the Toolkit can help
OPEN ENERGY DATA ASSESSMENT

Pierre Chrzanowski
ICT Consultant, World Bank
31 March 2016
Why an Open Energy Data Assessment?

• Generic ODRA does not address sector-specific issues
• High potential for open energy data for private sector
• ODRA consultants are not necessarily energy experts
• World Bank NEGAWATT Challenge provided an opportunity to pilot energy ODRAs on a municipal level
What does the methodology contain?

- Step-by-step guide to analyze availability of energy sector data, identify energy issues and priorities, and provide recommendations on how open data can help line ministry, regulator, private sector, and civil society

- Based on 8 dimensions of standard ODRA, with specific questions and topics relevant to the sourcing of data in the energy sector

- Examples of positive and negative evidence, successful projects, and key datasets for sector development

- Can be used as a standalone tool or complement the ODRA
Examples of energy ODRA questions

- [Institutions] What is the dynamic of the energy market, in particular regarding development of IPPs* and SMEs; and to what extent does it rely on the energy broker system and broker companies?

- [Supply side] To what extent and how does the energy sector communicate to the population about the status of the energy system (capacity level, power outages, construction plan, etc.)?

* independent power producers
## Example of key datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailed electricity consumption</td>
<td>Average electricity consumption at the most detailed level (building, block, district)</td>
</tr>
<tr>
<td>Energy audit (Certificate)</td>
<td>Results of energy audits for private, public buildings and other infrastructures</td>
</tr>
<tr>
<td>Land Register (cadastral) parcels</td>
<td>Geospatial data of boundaries of lands and real estates.</td>
</tr>
<tr>
<td>Wind Capacity</td>
<td>Geospatial data of wind capacity on the territory</td>
</tr>
<tr>
<td>Biomass</td>
<td>Geospatial data about the biomass resources generated (quantity and composition)</td>
</tr>
</tbody>
</table>
Lessons learned in Kenya and Ghana

1. Work with or seek advice from local energy experts

2. Energy sector is complex. Narrow your scope and define priorities with client in advance

3. Find right incentives / potential projects that could sensitize energy stakeholders to open energy data

4. Leverage energy ODRA to connect energy sector with the local/international open data community

5. Target simple, achievable actions, and ensure follow-up

6. Donors collect and use primary energy data and have incentives to publish it as open data

7. Seek to leverage existing initiatives
Links and contacts

- Methodology and assessments available on http://opendatatoolkit.worldbank.org/
- Version open for comments: http://bit.ly/1UsFUfM
- Feature story “Exploring Open Energy Data”: http://bit.ly/1RInCUk
- Contacts: Anna Lerner alerner@worldbank.org and Pierre Chrzanowski pierre.chrzanowski@gmail.com
OPEN DATA FOR BUSINESS (OD4B)

Alla Morrison

AMORRISON@WORLDBANK.ORG

@ALLAMORRISON
Offers countries an assessment and engagement tool to help catalyze the use of open data by private sector

Is a new open source tool, was added to the Open Data Toolkit in February 2016

Builds on the ODRA work (Chapters 5 and 6) to assess demand for open data from the private sector and on the Open Data Roundtables in the US

Consists of: (1) an introductory briefing on open data for business, (2) a questionnaire for businesses, incl. a scoring sheet, and (3) a roundtable engagement guide

Leads to an action plan to help build awareness of open data and address country-specific impediments to the use of open data by businesses.

The tool can be used as a stand-alone diagnostic, or in combination with an ODRA or a technical assistance project. Elements of the tool have been tested in Kazakhstan, Serbia and Sierra Leone, with many more countries interested.
OPEN TRANSPORT DATA ASSESSMENT

Alla Morrison

AMORRISON@WORLDBANK.ORG

@ALLAMORRISON
Exhibit 3
Three levers can lead to $720 billion to $920 billion per year of value in transportation

$ billion

- Improved infrastructure planning and management: 270–280
- Optimized fleet investment and management: 230–370
- Better-informed customer decision making: 220–280
- Total: 720–920

NOTE: Numbers may not sum due to rounding.
SOURCE: McKinsey Global Institute analysis
Offers a comprehensive assessment of a city's policies and practices for transport data collection, management, sharing and use.

Builds on the ODRA-style ecosystem approach

Focuses on a city’s transportation challenges and proposes innovative data-enabled solutions, drawing on global case studies.

Offers recommendations to transport agencies for deriving value from open, semi-open or closed datasets.

Covers both government-held and relevant private sector data.

Format similar to an ODRA – a team of 2-3 data and transport experts conducts a study and engages stakeholders in co-creating solutions over a 1 week-long mission. Recommendations are delivered in a solution-focused action plan.
TRANSPORT DATA INNOVATIONS – ASSESSMENT

01 Leadership and Strategy
02 Policy Framework
03 Transport Agencies and Related Companies
04 Data Collection and Management
05 Actual and Potential Use
06 Data Reuse Ecosystem
SOURCES OF TRANSPORT DATA INNOVATIONS

GOVERNMENT
- SURVEYS, SCHEDULES, ROUTES
- ENVIRONMENT SENSORS
- SMART CARDS
- CCTV CAMERAS
- GPS
- ELECTRONIC ROAD TOLLS

PRIVATE SECTOR
- TRANSACTION DATA
- CROWDSOURCED
- MOBILE PHONES
- IN-VEHICLE NAVIGATION SYSTEMS

Open government data
Out of the box