

Open Data Readiness Assessment: Methodology Lessons Learned

Andrew Stott

WebEx
31 March 2016

@dirdigeng
andrew.stott@dirdigeng.com



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 “rubrics”
 - 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 available⁴⁵. The overview of key datasets is illustrated below.

Dataset	Score	Breakdown	Location (URL)	Information
Transport Timetables	5%		'15	Propose Revisions
Government Budget	45%		'15 http://minfin.tj/in...	Propose Revisions
Government Spending	45%		'15 http://minfin.tj/	Propose Revisions
Election Results	0%		'15 http://www.osce...	Propose Revisions
Company Register	75%		'15 http://www.ando...	Propose Revisions
National Map	30%		'15 http://mapfactor...	Propose Revisions
National Statistics	100%		'15 http://stat.tj/	Propose Revisions
Legislation	45%		'15 http://mmk.tj/	Propose Revisions
Postcodes / Zipcodes	0%		'15	Propose Revisions
Pollutant Emissions	75%		'15 http://meteo.tj/	Propose Revisions

Key: Yes No Unsure No data

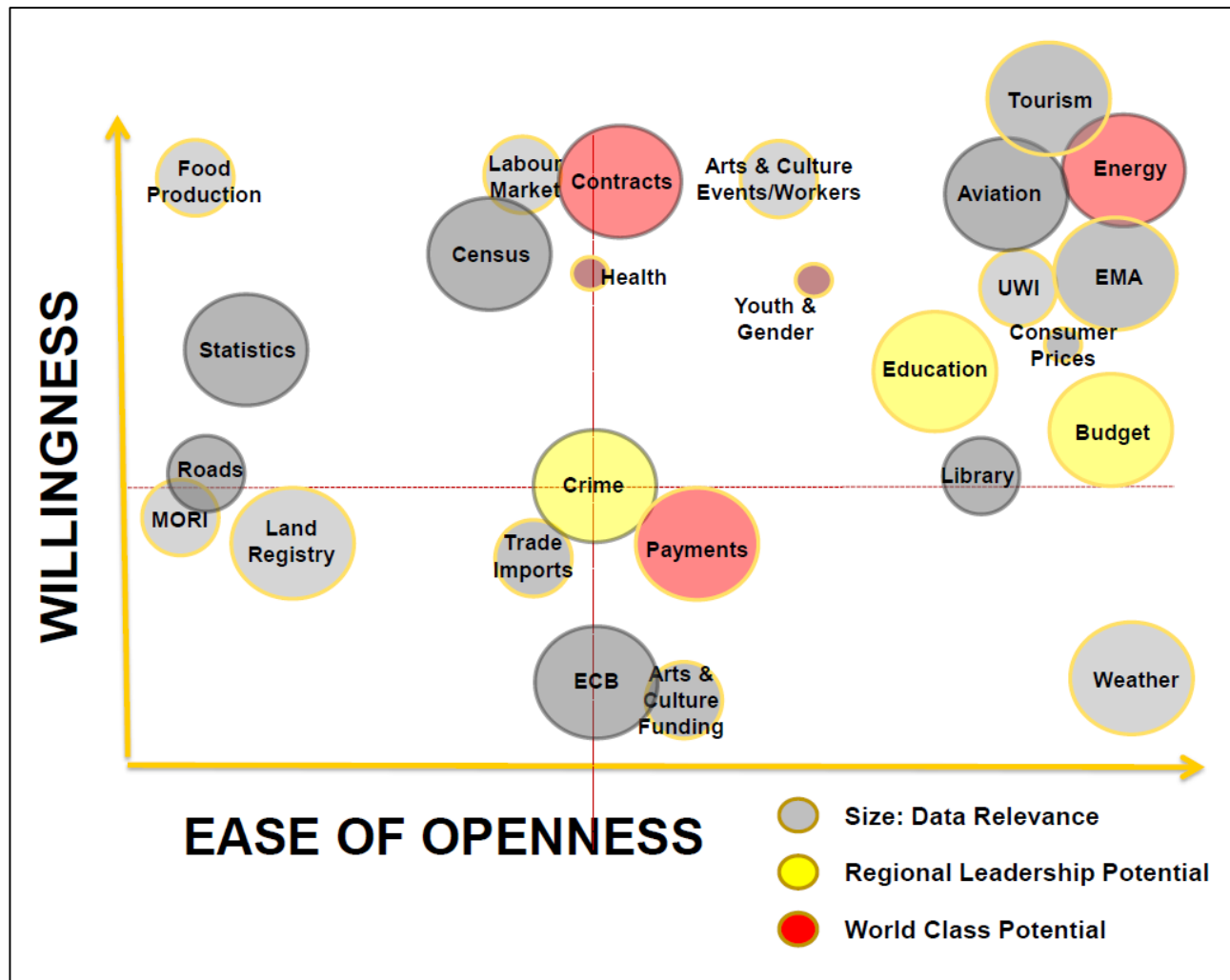
Datasets: Fully structured analysis

Dataset	Environmental monitoring: levels of pollution, energy consumption	
Availability	<p>The Ministry of Environment and Sustainable Development manages the Environmental Information System of Colombia (SIAC [90]). It is an informative portal for consultation of environmental indicators redirected at other services such as the IDEAM, the Institute for Coastal and Marine Research (Instituto de Investigaciones Costeras y Marinas (INVERMAR [91])) and the Information System on Biodiversity of Colombia (Sistema de Información sobre Biodiversidad de Colombia (SIB [92])).</p> <p>Nonetheless, an advanced search of the SIAC portal, to see what documents are being published in Excel, shows 249 results corresponding to data from the system of indicators. The search carried out on Google is: site:www.siac.gov.co filetype:xls. These documents cannot be directly located on the web portal.</p>	
Feasibility of openness: Political	The sphere of competence of the SIAC 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	Institutionally, 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

Ref	Dataset	Feasibility	Key Benefits & Risks	Recommendation
D32	Road Accidents	<p>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 co-ordinates of the location of the accident which is then recorded with the other data. Data is not available publically but is aggregated in some statistics.</p> <p>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.</p>	<p>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.</p> <p>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.</p>	QUICK WIN

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

