THE 7-STEP APPROACH FOR SETTING UP AN OPEN DATA INITIATIVE TO IMPROVE MFDR PRACTICE IN AFRICA

From the Secretariat of the African Community of Practice on Managing for Development Results at the African Capacity Building Foundation (ACBF)

Guide N° 6

SYNOPSIS

Open data is defined as data that are made available to the public free of charge, without registration or restrictive licenses, for any purpose whatsoever (including commercial purposes and excluding harmful purposes), in electronic, machine-readable formats that are easy to find, download and use. Open data is radically opening up government for greater accountability, public service improvement and economic growth. Open data is used across all development sectors such as administration, business, health, education or agriculture. This guide provides specific instructions on how to set up an open data initiative. First it outlines the requisite steps for setting up an open data initiative. The specific steps are as follows: (1) Engage stakeholders, (2) Decide on what data should be public, (3) Apply for an open license, (4) Choose appropriate data formats, (5) Make data discoverable, (6) Review, Learn, and improve, and (7) Linking MfDR pillars to open data.

Introduction

Data is distinct pieces of information, usually formatted in a special way. Data can exist as numbers, text, bits in electronic memory, facts stored in human memory. Data is collected and analysed to create information suitable for making decisions. Open data refers to data that are freely available, to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control[1]

Historically, open data was originally rooted in the scientific community. Researchers and scientists were the first to recognize and advocate for the openness and sharing of data. As early as 1942, Robert King Merton explained the importance of making results of research freely accessible to all. Information technologies have also given a new breath to this philosophy open science. In her research, the 2009 Nobel Prize of Economics Elinor Orstrom showed how information is very similar to public goods, because their use by one person does not impede their use by others. However, these are public goods of a new kind: not only their use doesn't deplete the common stock, but it enriches it.[2]A number of events (as the Sebastopol meeting

---

[1] Auer, S. R.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. (2007). "DBpedia: A Nucleus for a Web of Open Data". The Semantic Web. Lecture Notes in Computer Science 4825. p. 722.doi:10.1007/978-3-540-76298-0_52.ISBN 978-3-540-76297-3

[2] http://www.paristechreview.com/2013/03/29/brief-history-open-data/

in 2007 that defines the concept of open public data) and political commitment (as the Memorandum on Transparency and Open Government signed by US President Barack Obama in 2008) paved the way for full recognition and new dynamics of open data throughout the world. Today open data movement faces a number of challenges related to availability, interoperability and ownership to mention a few. Despite these challenges open Data is creating impact and has become a major element in the development agenda.

The World Bank identified four broad types of benefits of Open Data which are: fostering economic growth and job creation; improving efficiency and effectiveness of public services; increasing government transparency, accountability and citizen participation and facilitating better information sharing within government. Several initiatives of Open Data implemented around the world evidenced the added value of using Open Data to drive development[3]. Opening up data in Africa will contribute significantly towards transparent and accountable governance that energizes democracy. In line with this, it is clear that Open Data can help to advance MfDR practice, but African governments and institutions still lack the capacities and resources to take advantage from it.

This guide aims to address these capacity needs. It provides understanding about open data concept and delivers instructions on how to initiate and manage an open data initiative. This guide does not intend to provide all the details about developing an open data policy to the extent of an Open Government initiative, but it ensures everyone using it has the required knowledge to successfully make data open. The guide targets government officials, civil society, private sector, non-governmental organizations, ordinary citizen and any other stakeholders who are either on the demand or supply side as regards to data.

What is Open Data[4]?

The Open Definition[5] defines open data as data that anyone can access, use or share. To integrate all features, open data is defined as data that are made available to the public free of charge, without registration or restrictive licenses, for any purpose whatsoever (including commercial purposes and excluding harmful purposes), in electronic, machine-readable formats that are easy to find, download and use. The distinguishing feature of Open Data is not its size or complexity, but the fact that it is made available as a public good[6].Open data should not be misunderstood with "closed data" that "only data owners or people within an organisation can access, for reasons such as privacy, commercial sensitivity and security". Open data also differs from "shared data"[7] that is "shared with specific people and organisations for a specific purpose: to provide services, connect information and contribute to research".

Design of an Open Data strategy

Before you start

Ensure strong leadership from decision and policy makers: The decision to launch an open data initiative may not be welcomed by everyone; both people from inside who have a culture of secrecy and people from outside who use their access to that data as a comparative advantage to advance in their business interests. Moreover, opening data requires a minimum of resources for which a clear will and support from the organization is needed. Therefore, a strong leadership is required from those who are expected to effectively address these issues. In the case of a government, this person may be a high official from a ministry (or the President himself) or a head of a governmental agency.

---

[3] Examples at www.dataimpacts.org

[4] http://theodi.org/what-is-open-data
[5] Open Definition: http://opendefinition.org/od
[6] http://blogs.worldbank.org/ic4d/new-discussion-paper-how-open-data-can-drive-sustainable-development
[7] http://theodi.org/what-is-open-data

Review existing policy/legal framework: Issues related to data have legal implications. Personal data that a citizen stores in a public or private administration should be treated in accordance with clear predefined policies. Therefore, open data policies should be informed by existing provisions which ensure how and what information is accessed. To that end an analysis of the existing policy/legal framework is important before an open data initiative is launched.

Assess available resources: Despite running an open data initiative does not necessarily require a huge amount of financial resources; it demands a number of resources in terms of human skills (e.g. capability to collect and structure data), infrastructure (e.g. availability of a website and database). Consequently, a prior assessment of available resources will help to determine what the strengths of the organization are in opening data up and what gaps need to be addressed.

Step 1: Engage stakeholders

This is also a prerequisite in launching an open data initiative. It is essential to map the roles and added value of all stakeholders that may be interested in the initiative. Relevant stakeholders may involve data providers, data processors and data users. This is key to the acceptance of the initiative and essential for its success and sustainability. The following questions may be asked to delineate what relevant stakeholders to consider: Who makes and influences policies and decisions? - Who provides financial and technical resources? Who are the direct and indirect beneficiaries? Who needs special attention? Who are the potential re-users of data?

Step 2: Decide on what data should be public

Obviously one of the first steps in developing an open data initiative is to decide on what data your institution is going to open up. Not all data is relevant to the targeted end users or re-users. Therefore, it is important to ensure your institution is well aware of the data that are the most likely to

be useful to the potential beneficiaries. This also requires that a comprehensive inventory of available data is performed in early stage of the process. This inventory is not mandatory especially when you plan to go quickly open, but it is useful to have the list of available datasets which could be open. Some datasets that are clearly relevant for citizens are presented below (table 1).

Table 1 : Example of datasets

Dataset	Relevance
Parliamentary data and legal data	Of use to parliamentarians themselves, watchdog organizations, media covering parliament and voters
Public expenditure and budgeting data	Of interest for example to public sector and civil society oversight organizations, media, donor institutions
Environmental data	Of interest to affected citizens, research institutions, commercial service providers, environmental CSOs, public health administration
Demographic data	Of use to planning and development agencies, research institutions, public and commercial service providers
Socio-economic indicators	Of interest to research institutions, public and private service providers, planning and development agencies
Healthcare data	Of use to citizens, research institutions, watchdog organisations, public and commercial health care providers

Geographical data	Of use to planning and development agencies, commercial companies, end users relying on geographical applications such as navigation and map applications
Local transportation data	useful for public and private transport service providers as well as for consumers

Source: United Nations, 2013

This list is not exhaustive. Given the nature of innovation, re-users can develop new, creative applications, and services based on Public Sector Information that had not been foreseen earlier, underlining the need for opening up as much data as possible.

Box 1 : Start Small

Look for what your organization gathers and produces data about. It can be related to science, culture, environment, finance, sport, transportation. Numerous organizations have started their Open Data program with only a few datasets. Therefore, pick a topic that will allow you to ship a first dataset within days. Your first users will give feedback on subjects that are important to them. Limiting data volume helps you think like a start-up. You can share your data and get early feedback. You can evangelize your new platform to your audience. You can change your approach if needed and be focused on quality and on learning good practices.

Source: http://opendatasoft.com

After deciding what data your organization is going to make public, you will need to structure and give some context to your data. Context will give re-users a better understanding of your data. This context is usually contained in metadata. This will help users know what your data contains, how to interpret it and if it should be trusted. Structuring data is discussed under the section on *Choose appropriate data formats*. Finally, as mentioned above, open data policy should be comply with pre-existing legislation and directives about access to public information and use of data that is sensitive for privacy, security or other reasons.

Step 3: Apply an open license

If you want your data to be "open", you should put a license. We recommend licenses that conform to the Open Definition and marked as suitable for data. This list (along with instructions for usage) can be found at: http://opendefinition.org/licenses/. You can also find instructions on applying an open data license on the Open Data Commons site: http://opendatacommons.org/guide/

Box 2 : Picking a license

Explicit licensing is mandatory even if you wish to put your data in the Public Domain. You'll find below a list of the most common licenses for Open Data:

Open Database License: share-alike & attribution

Public Domain Dedication and License

Open Data Commons Attribution License

Don't hesitate to use the latest version of Creative Commons licenses - http://creativecommons.org/ . They are international. If you ask for attribution, keep it as minimal as possible: name of your organization and a link to the original dataset for instance.

Source: http://opendatasoft.com

Step 4: Choose appropriate data formats

Machine-readable data[8]

Technical openness should be ensured while developing an open data policy. This means data must be released in open formats that are easy and efficient to reuse via technology. Plainly, "open formats" refer to a rolling set of "open standards," often defined by standards organizations that store information in a way that can be accessed by proprietary or non-proprietary software means. A common example cited is CSV in lieu of XLS for spread sheets (the former being accessible via a wider variety of software mechanisms than the latter). We recommend that as much as possible you release the data both in proprietary and non-proprietary formats.

Machine processable data[9]

One step beyond machine-readable data is structured data (or machine-processable data), a format intended to ease machine searching and sorting processes. While formats such as HTML and PDF are easily opened for most computer users, these formats are difficult to convert the information to new uses. Providing data in structured formats, such as JSON and XML, add significant ease to access and allow more advanced analysis, especially with large amounts of information. It is important to recognize that not all data is amenable to these formats. Data only available in the form of text is a good example. If these data cannot in any way be converted into a structured format, it is advised to publish them in their original formats provided this is of interest for users.

Box 3 : Is your data 5-star?

The inventor of the World Wide Web, Tim Berners-Lee, proposed a 5-star model for linked Open Data. Each successive step in the 5-star model builds upon the previous one in ways that are not mutually exclusive.

☆ Make your data available on the Web (whatever format) under an open license

☆☆ Make it available as structured data (e.g., Excel instead of image scan of a table)

☆☆☆ Make it available in a non-proprietary open format (e.g., CSV as well as Excel)

☆☆☆☆ Use URIs to denote things, so that people can point at your data

☆☆☆☆☆ All of the above and link your data to other data to provide context

Another key feature of open data is its availability in bulk. Data is available in bulk if the entire dataset can be downloaded easily and efficiently to a user's own system. Conversely it is non-bulk if one is limited to getting small parts of the dataset, for example, is you restricted to a few elements of the data at a time and therefore requires thousands or millions of requests to get the entire dataset[10]. In some cases bandwidth and memory size limitations can prevent users to get the data in bulk. The solution is to configure an Application Programming Interface (API). For data, this is usually a way provided by the data publisher for programs or apps to read data directly over the web instead of downloading them first.

The app sends the API a query asking for the specific data it needs, e.g. the number of adults in a particular city. This allows the app to use the data without downloading the whole dataset, saving bandwidth and ensuring that the data used is the most up-to-date available.

---

[8] Guidelines for Open Data Policies, Sunlight Foundation
[9] Ibid

[10]
http://opendatahandbook.org/glossary/en/terms/bulk/

Step 5: Make data discoverable

If your data is open but it is quite difficult for users to locate and access it, then your data is not truly "open". You should make sure that your data is online and easy to find. You can start by putting your dataset on your website. At the beginning do not waste time and money in setting up your own dedicated portal. There are a number of tools which are live on the web that are specifically designed to make data more discoverable.

DataHub (https://datahub.io), a catalog and data store for datasets from around the world. You can also consider http://ckan.org/ which is fully-featured, mature, open source data management solution. CKAN provides a streamlined way to make your data discoverable and presentable. Each dataset is given its own page with a rich collection of metadata, making it a valuable and easily searchable resource. Finally, if users are required to sign up before using your data, then it is not truly open.

Step 6: Review – Learn – Improve

Like any initiative, implementing an open data policy should be considered as an iterative process that requires attention to its quality and efficiency. Therefore, it is important to establish a culture of reviewing and learning from what have worked or not. An essential part of this exercise is the collection of feedback from users to determine how the open data policy can evolve while maintaining success and sustainability.

Step 7: Linking Open Data and MfDR pillars

In this sub-section, we are giving examples of how open data contributes to the Managing for Development Results pillars.

Table 2 : Linking Open Data and MfDR pillars

MfDR Pillar	Open Data…
Strategic Planning	- Promote data sharing among government institutions to make planning more strategic and participatory - Make much more data available to set high-priority social change objectives and baseline for indicators
Results-Based Budgeting	- Ensure that government budget systems are more transparent and accountable to the public - Provide the public with comprehensive and timely information on the government's budget and financial activities and opportunities to participate in decision making
Institutional Capacity for Implementation	- Provide innovative way of managing programs and doing business
Monitoring and Evaluation, Statistical Capacity and Information Systems	- Provide data for effective M&E work to track results and learn lessons - Enable double-checking of achievements claimed by government institutions - Enhance the efficiency of national statistical offices
Leadership	- Offer the opportunity to leaders to show their level of commitment,

		openness and transparency
Accountability and Partnership	-	Provide sufficient information to civil society organizations, parliament to evaluate the work of government
	-	Provide opportunities to citizens to be able to access view and interrogate data relating to government performance and fiscal processes.

Conclusion

The data revolution is poised to transform the way governments, citizens, and companies do business. Open data is providing even more opportunities for all stakeholders to improve how they do business. This guide provides technical instructions on how to set up an open data initiative. If well developed, open data initiatives provide avenues for governments to open up their operations and improve transparency. Furthermore, it can serve as a strong tool for civil society to monitor public management so as to improve leaders' accountability. Therefore it is essential that African countries build capacities to develop open data initiatives. This guide intends to shade light on this process.

Reference

Open Data Policy Guidelines – Sunlight Foundation http://sunlightfoundation.com/opendataguidelines/

Open government data toolkit from the World Bank - http://data.worldbank.org/open-government-data-toolkit

Ten Open Data Principles - http://sunlightfoundation.com/policy/documents/ten-open-data-principles/

The Datahub - http://ckan.org/

The Genome Project – http://genome.gov

The Open Data Handbook - http://opendatahandbook.org/

The Open Data Institute – http://theodi.org

The Open Definition - http://opendefinition.org/

The open source data portal software http://ckan.org/

United Nation. 2013. *Guidelines on Open Government Data for Citizen Engagement.* P.29 New-York. United Nations.

World Bank. 2014. *Open data challenges and opportunities for national statistical offices.* Washington, DC. World Bank Group. http://documents.worldbank.org/curated/en/2014/07/20467305/open-data-challenges-opportunities-national-statistical-offices

..

Acknowledgement