Sri Lanka's Open Data Portal

Current Status and Opportunities for Improvement

Viren Dias, Ramathi Bandaranayake

LIRNEasia, 12 Balcombe Place, Colombo, Sri Lanka

(viren, ramathi)@lirneasia.net

LIRNEasia is a pro-poor, pro-market think tank whose mission is *catalyzing policy change through research to improve people's lives in the emerging Asia Pacific by facilitating their use of hard and soft infrastructures through the use of knowledge, information and technology.*

Contact: 12 Balcombe Place, Colombo 00800, Sri Lanka. +94 11 267 1160. info@lirneasia.net www.lirneasia.net

Introduction

Open data, as defined by the Open Knowledge Foundation (2015), is data that "can be freely used, modified, and shared by anyone for any purpose." The term "open data" first appeared in 1995, when it was published in a document produced by the National Research Council (1995) in America.

Importance of Open Data

For data to be useful, it needs to be of high quality: up to date, accurate, complete, and well documented. However, collecting high-quality data is an arduous task. Whilst the composition of resources required may change from one methodology to another, high-quality data collection will always be very resource intensive:

- Traditional forms of data collection, such as in-person surveys, typically require: meticulous planning to ensure an unbiased sample, manpower to recruit sufficient respondents for a representative sample, and training of personnel to ensure quality responses.

- In recent years, owing to the digitization of day-to-day activities, data collection in the form of automated tracking has become predominant. This form of data collection requires: pre-planning for what data to collect anticipating its potential use, manpower to write the necessary code, infrastructure to store the large volumes of data, and post-processing to clean the collected data to generate a useful dataset.

This is where open data plays an integral role. Open data enables research by organizations or individuals that may not have had the necessary resources to gather the required data. This research could be novel, or a reproduction of existing research—which has value in ensuring robust research practices. Additionally, open data provides several ancillary benefits for researchers, such as: long-term and safe storage of data at no cost to the researcher; and increased visibility, reputation, and potential for collaboration. For the respondents, open datasets minimize the negative impact of the data collection process, as it only needs to be conducted once for multiple groups of researchers. At a minimum it results in a time-saving, and in extreme cases it can prevent a respondent from having to relive a traumatic experience more times than necessary.

Challenges of Open Data

However, there are several issues associated with open data as well, which revolve around privacy concerns. Once a dataset is made publicly available, its use is difficult to control. Although privacy guarantees such as differential privacy have long been established in literature, satisfying these guarantees result in a loss of information, which diminishes the utility of the dataset (Dwork et al., 2006). Furthermore, for a published dataset to be useful,

it typically requires additional time and effort by researchers to mold the dataset into a generally useful format, and document the collection methodology and the description of the variables in the dataset.

Status of Open Data in Sri Lanka

The need for open data in Sri Lanka had been established following the United Nations E-Government Survey in 2012 (Department of Economic and Social Affairs, 2012). This set in motion two government-led initiatives: the Open Data Portal of Sri Lanka (ODPSL, http://www.data.gov.lk/), and the National Data Sharing Policy (Information and Communication Technology Agency [ICTA]) To affirm their commitment to the open data movement, the Government of Sri Lanka joined the Open Government Partnership (OGP), and made pledges pertaining to open data in their 2016–2018 and 2019–2021 action plans (ICTA, 2015; ICTA, 2019).

Open Data Portal

The ODPSL (http://www.data.gov.lk/) is an online platform for sharing datasets. The datasets span 11 categories, namely: agriculture and livelihood, tourism and leisure, travel, information technology and cybersecurity, demography, economic, employment and skills, industry and investments, infrastructure, national security and safety, and transport. The data is typically collected by the public sector, but anyone can contribute to the portal after their dataset undergoes a review process.

The website itself is fairly well-made, with functionality similar to that of other national open data portals. This includes, but is not limited to: forms for uploading and suggesting datasets; search tools for finding datasets by category or tags; basic tools for previewing datasets online in tabular, graphical and geographical formats; and social features such as forums and comments. The primary flaw of the portal is the datasets: they are limited in number with just 136 available as of the writing of this brief, a significant portion are not in machine-readable format, and documentation of collection methodologies and variable descriptions were severely lacking. Perhaps as a consequence of the poor quality and quantity of the datasets, the social features of the website are simply not used, with the forums absent of any posts. A more alarming observation is the presence of three datasets of beekeepers in the Kandy, Nuwara Eliya, and Matale districts, containing personal information such as: full name, complete address, and telephone number (ODPSL, 2018a; ODPSL, 2018b; ODPSL, 2018c). Whether consent was obtained or not, this detail of personal information would not be required for any form of research, and illustrates a clear flaw in the dataset review process.

A study by Máchová and Lnénicka (2017) evaluated the quality of 67 national open data portals, including that of Sri Lanka. The benchmarking methodology involved 28 metrics evaluated on a five-point Likert scale, resulting in a score ranging from 28–140. The top three portals belonged to the United Kingdom, India, and the United States, with scores of 129.6, 126.0, and 124.6. Sri Lanka's portal was ranked at 47, with a score of 90.0.

National Data Sharing Policy

Much of the data generated in Sri Lanka exists in silos within various government departments and offices. As such, each individual section of the government does not have the complete picture, which may result in poor decision making. To combat this issue, ICTA published the National Data Sharing Policy. The document defines its primary objectives as: encourage proactive sharing and updating of high-quality datasets, facilitate access to datasets for all stakeholders, and protect the rights of the data producers. Additionally, the document outlines its data-sharing principles, which includes among others: transparency, protection of intellectual property, protection of privacy, non-proprietary and machine-readable formats, universal accessibility, high data quality, and fostering a culture of open data.

Open Government Partnership

Sri Lanka joined the OGP in 2015. Under this, Sri Lanka has made commitments to open data among others. For the 2016–2018 action plan cycle, Sri Lanka pledged to "promote Open Data using ICT platforms and ensure citizens get its benefits using similar technologies," and identified four milestones: revamp the ODPSL, conduct a survey to determine citizen's demand for government data, hold an open consultation on data and services classification, and increase the number of datasets available on the ODPSL from 89 to 200 by the end of 2016 and 500 by July 2018 (ICTA, 2015). However, by the end of this action plan cycle, the government failed to meet three milestones, having only completed the first milestone in 2017 (OGP, 2018).

The 2019–2021 action plan cycle is currently active. In it, Sri Lanka has only identified a single milestone related to open data: to increase the number of datasets available on the ODPSL from 300 to 2000 (ICTA, 2019). However, as of the writing of this brief, only 136 datasets are available.

Recommendations

As identified previously, the Open Data Portal of Sri Lanka consists of poor quality datasets that are few in number. We believe that this state is the result of the lack of incentives and the amount of effort required to publish high-quality datasets. As such, our proposed recommendations revolve around tackling these root causes.

Our first recommendation is a stricter review process, coupled with a more structured submission process. This would involve documenting and publishing a set of hard, unambiguous criteria, which will aid reviewers in assessing the quality of the submitted datasets in a consistent manner, and guide contributors to make quality submissions. This will undoubtedly raise the barrier for contributing datasets, which is where a more structured submission process plays a role. In its current state, the submission form only has two fields pertaining to the dataset: dataset name, and dataset description. The recommendation is to revamp it to include additional fields referring to specific dataset attributes—such as collection methodology, names and descriptions of each variable, and additional notes—which will guide contributors to make submissions that meet the required criteria. Technical support should also be provided to departments that lack the relevant capacities to publish high quality datasets.

Our next recommendation would be to increase incentives to contribute datasets. For instance, India (whose open data portal (https://data.gov.in/) was ranked second out of out of 67 portals assessed worldwide by Máchová and Lnénicka (2017)), holds the Digital India Awards (https://digitalindiaawards.gov.in/)—a biennial award ceremony celebrating the contributions of government organizations toward open data, among others. India's portal boasts nearly 500,000 datasets as of the writing of this brief. Such incentives could help increase the number of open datasets.

Our final recommendation is to add functionality to the ODPSL to host data-related competitions. This is similar in vein to Kaggle (https://www.kaggle.com/) and would serve as an incentive for the private sector to contribute to open data. Such competitions have been held in the past, such as the famous Netflix Prize (https://www.netflixprize.com/), and the hosting firm benefits from an outside approach to solving their problems, as well as scouting for future hires.

References

Open Knowledge Foundation. (2015, November). *Defining Open in Open Data, Open Content and Open Knowledge*. https://opendefinition.org/

National Research Council. (1995). *On the full and open exchange of scientific data*. National Academies Press. https://doi.org/10.17226/18769

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In S. Halevi & T. Rabin (Eds.), *Theory of Cryptography* (Vol. 3876, pp. 265–284). Springer Berlin Heidelberg. https://doi.org/10.1007/11681878_14

Department of Economic and Social Affairs. (2012). *United Nations E-Government Survey 2012: E-Government for the People*. United Nations. https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2012

Information and Communication Technology Agency. (n.d.). *National Data Sharing Policy*. Ministry of Telecommunication and Digital Infrastructure, Government of Sri Lanka. https://www.nsdi.gov.lk/sites/default/files/National%20Data-Sharing-Policy%202013-Draft.pdf

Information and Communication Technology Agency. (2015). *Sri Lanka National Action Plan 2016–2018*. Ministry of Telecommunication and Digital Infrastructure, Government of Sri Lanka. https://www.opengovpartnership.org/documents/sri-lanka-national-action-plan-2016-2018

Information and Communication Technology Agency. (2019). *Sri Lanka National Action Plan 2019–2021*. Ministry of Telecommunication and Digital Infrastructure, Government of Sri Lanka. https://www.opengovpartnership.org/documents/sri-lanka-action-plan-2019-2021

Open Data Portal of Sri Lanka. (2018a). *මහනුවර දිස්ත්‍රික්කය මිමැසි,මල්,ආරක්ෂිත ගෘහ* [Kandy District Beekeepers]. http://www.data.gov.lk/dataset/මහනුවර-දිස්ත්%E2%80%8Dරික්කයමිමැසිමල්ආරක්ෂිත-ගෘහ

Open Data Portal of Sri Lanka. (2018b). *නුවරඑළිය දිස්ත්‍රික්කය මිමැසි,මල්,ආරක්ෂිත ගෘහ* [Nuwara Eliya District Beekeepers]. http://www.data.gov.lk/dataset/නුවරඑළිය-දිස්ත්%E2%80%8Dරික්කය-මිමැසිමල්ආරක්ෂිත-ගෘහ

Open Data Portal of Sri Lanka. (2018c). *මාතලේ දිස්ත්‍රික්කය මිමැසි,මල්,ආරක්ෂිත ගෘහ* [Matale District Beekeepers]. http://www.data.gov.lk/dataset/මාතලේ-දිස්ත්%E2%80%8Dරික්කය-මිමැසිමල්ආරක්ෂිත-ගෘහ

Máchová, R., & Lnénicka, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research*, 12(1), 21–41. https://doi.org/10.4067/S0718-18762017000100003

Open Government Partnership. (2018). *Promote the Open Data Concept and Delivering the Benefits to Citizens Through ICT (LK0006)*.
https://www.opengovpartnership.org/members/sri-lanka/commitments/LK0006/

Open Government Partnership. (2021). *Open Data (LK0029)*.
https://www.opengovpartnership.org/members/sri-lanka/commitments/LK0029/