DO OPEN DATA AND TRANSPARENCY PROVIDE WHAT WE EXPECTED?

By Jeff Myers[1] Anand Trivedi, and Oludotun Babayemi

Open Government websites have become more common in countries throughout the world, and have generally been established to promote transparency, public engagement, and government accountability. But, since most transparency websites allow users to access data without requiring the user to identify themselves, very little is known about the users and uses made of the data, and whether those public engagement and accountability purposes are being achieved. In short, it is hard to tell if we are getting what we expected.

This paper explores what we are getting – across a selected group of different countries. Its prominent conclusions are:

- Public engagement and government accountability seem to be secondary and perhaps incidental results from Open Government initiatives
- The most frequently accessed datasets suggest that open data is best suited to use by:
    - Business, to use in setting strategy and finding opportunities for business growth
    - Governments, to evaluate performance of their programs, and
    - Media, researchers and activists, to understand where societal problems exist

BACKGROUND

Open government has gained increasing focus over the past two decades. The Open Government Partnership, for example, was formed in 2011. It has grown from 8 original member countries to include 75 countries now, as well as multiple municipal government entities. The Open Government Partnership was formed to "...promote transparent, participatory, inclusive and accountable governance."[2] The primary open data website in the US, Data.gov, "...aims to make government more open and accountable. Opening government data increases citizen participation in government, creates opportunities for economic development, and informs decision making in both the private and public sectors."[3] From the experience of the authors, we recognize that many governments tend most quickly and willingly to share data that is readily available, reliable and of good quality. We recognize that governments often feel compelled to make some investment spend some time to gather it, clean it, and ensure its accuracy.

Perhaps consistently with that need for effort, there is a bit of a supply-side perspective – governments share what they have – but there may not be much effort exerted to identify what potential users want – i.e., the demand side.

---

1. Jeff Myers, the corresponding author, and may be reached at JMyers@reisystems.com
2. See: https://www.opengovpartnership.org/about/
3. See: https://data.gov/open-gov/

METHODOLOGY

We wanted to explore that "demand side" for open data, in order to test the strength of the hypothesis that open data and government transparency are focused on public engagement and government accountability. We did so by exploring which datasets are actually used – or at least accessed – and by developing our own hypotheses as to the users and use cases for data sets that are most frequently accessed in each of our respective countries, exploring how the most frequently accessed datasets could provide value, what kind of value, and who might receive that value.

Each of the authors has experience with open data and government transparency in a separate country, including Nigeria, India, Finland and the United States

To prepare this analysis and the accompanying OpenGovWeek 2024 discussion, we each identified the most frequently accessed datasets from our own country – typically the information about most frequently accessed datasets is available from a country's open data portal. However, since most/all datasets are downloaded anonymously, it was not possible to be certain who are the users, and to comprehensively assess what uses are made of open data.

For the ten most frequently accessed datasets, we hypothesized the likely users, and the purposes or use cases for the data set that might provide significant value. The full list of most frequently accessed datasets, hypothesized use cases, and hypothesized users for each country is provided in Appendix A below.

RESULTS

From this analysis of the most frequently accessed datasets, we discovered several things common across all four countries and across the datasets that are frequently accessed.
- Commercial users and use cases are quite common – in fact, they seem to be the most prevalent;
- Personal decision-making seems to be the second most prominent use of open government data – although our experts perceive that media is likely to be the direct user of the datasets; which media then provides an interpretation of the data's meaning to the public. For example, an individual citizen may not regularly download crime statistics, but a newspaper or television report may review them, publicize trends, and berate government officials in regions where crime is increasing;
- Government itself may use open data - to evaluate its programs, their design, and their effectiveness; government may also use open data to identify a need for new programs or changes to existing programs;
- Activists may use some data sets to determine where they want to allocate their efforts/attention
- Research and analysis (by academia, and by government agencies) are significant users of open data – perhaps informing others amongst the users noted above

- Government accountability and anti-corruption efforts seem to be a relatively infrequent use case, based upon the data sets that are most frequently accessed. Few of the data sets that are frequently accessed appear to help expose government corruption. (Only one data set from one of the four countries related to procurement.)

One caveat is that some datasets that provide a big benefit (e.g., weather data from the US) may not be frequently requested and accessed via an open data portal because users (such as a smart-phone's weather app) have established automated recurring data feeds, often directly to the originating data source. Thus, the weather app may not access the open data portal, even though it makes more frequent and perhaps valuable use of government data than do many other users.

CONCLUSIONS

1. The most frequently accessed data sets seem substantially unrelated to those purposes of public engagement and government accountability. Perhaps other data sets would better advance those purposes.

2. There has been no clear connection between open data / transparency and improved public engagement or government accountability. It may well be that better distribution and stronger marketing of data could help achieve the purposes of public engagement and accountability – but there does not seem to be good evidence as to which data might best contribute to those purposes in any case. This could prove a fruitful area for future research.

3. Open data sets present potential significant value to commercial organizations, and could be a substantial contributor to economic growth – but have rarely been recognized as offering such value. Further, government agencies don't attempt to measure the value of open data to commercial organizations, nor to invest in it to maximize that value.

4. Little energy has been dedicated toward identifying and understanding the "demand side" for open data. None of our respective countries can readily say with confidence who are the most frequent users of open data, what uses do they make, and what value do they receive from open data. Consequently, significant valuable uses of open data may exist, but not be captured – because even if one country offers that data, the country itself may not know who is receiving how much value, while other countries may not be aware of that potential value at all.

RECOMMENDATIONS

Based on this analysis and discussion of most frequently accessed data sets, we recommend that The Open Government Partnership:
- Commission research to identify users of open data, and to characterize and estimate the greatest sources of value from data which has been exposed

- Explore best practices for exposing data that has been demonstrated to increase public engagement, and that has been demonstrated to increase trust in the public sector. Those best practices may include choices such as:
    - *Which* data to expose;
    - How to communicate availability of the data;
    - How to identify consumers of open data; and
    - How to gather input about what data is desired and what data can support the most valuable uses.

ABOUT THE AUTHORS:

- Jeff Myers. Jeff is a Senior Director at REI Systems, Inc. (REI Systems developed Data.gov for the U.S. General Services Administration and the US Federal Government.)

- Anand Trivedi. Anand is a Principal at REI Systems, Inc.; he is also the former Director of the Government of India's Evidence, Data, and Digitization portfolio

- Oludotun Babayemi. Olu co-founder of Connected Development (CODE), an Abuja-based organization that uses online and offline tools to promote accountability and transparency in Nigeria. Through CODE's Follow the Money project, he and his team advocate for and track the disbursement of funds for local communities. Olu is also the founder of Cloneshouse Nigeria, a Monitoring, Evaluation and Learning organization. He is proud to have been a Reagan-Fascell Democracy Fellow at the U.S. National Endowment for Democracy

ACKNOWLEDGEMENTS:

Most Frequently Accessed Datasets, Potential Users and Potential Purposes / Use Cases

Top USA OpenGov datasets

From Data.gov, ranked as of May 20, 2024	Potential Users and Potential Purposes
1. Electric Vehicle Population Data (3175 recent views) Number of Battery Electric Vehicles (BEVs) and Plug -in Hybrid Electric Vehicles (PHEVs) that are currently registered by the Washington State Department of Licensing .	*Users:* Business, Personal, Gov't *Purposes:* Estimate market demand, regulation
2. Crime Data from 2020 to Present (2325 recent views) City of Los Angeles — Incidents of crime in Los Angeles .	*Users:* Real Estate, Personal, Gov't. *Purposes:* Choose residence location; Target police efforts
3. FDIC Failed Bank List (2201 recent views) Federal Deposit Insurance Corporation — Banks which have failed since October 1, 2000.	*Users:* Gov't, Academia, Business *Purposes:* Bank risk analysis, Research; Bank investor decisions
4. Dynamic Small Business Search (2107 recent views) Small Business Administration — Small business registration profiles populate DSBS, a tool government purchasers (contracting officers) use to identify potential small business contractors for upcoming contracting opportunities. Small businesses can also use DSBS to identify other SBs for teaming and joint venturing.	*Users:* Gov't, Business *Purposes:* Gov't purchasing decisions (to favor small business), small business partnering decisions
5. Motor Vehicle Collisions - Crashes (1459 recent views) City of New York — Details on crash events from all police reported motor vehicle collisions in NYC where someone is injured or killed, or where there is at least $1000 worth of damage.	*Users:* Gov't, Automobile mfgr, Insurance, Personal *Purposes:* Safety rules & aid, auto design, insurance rate setting

Top USA OpenGov datasets (cont'd)

From Data.gov, ranked as of May 20, 2024	Potential Users and Potential Purposes
6. Supply Chain Greenhouse Gas Emission Factors (1387 recent views) U.S. Environmental Protection Agency — The datasets are comprised of greenhouse gas (GHG) emission factors for 1,016 U.S. commodities.	*Users:* Gov't, Business, Activist *Purposes:* Project GHG levels, identify targets for GHG reduction, revise product designs
7. Walkability Index (1277 recent views) U.S. Environmental Protection Agency — Characterizes every Census 2019 block group in the U.S. based on its relative walkability. Walkability depends upon characteristics of the built environment that influence the likelihood of walking being used as a mode of travel.	*Users:* Gov't, Activist, Personal *Purposes:* Develop transit policy, encourage mass transit use, choose location of residence
8. Lottery Powerball Winning Numbers: Beginning 2010 (1081 recent views) New York State Lottery past Powerball results.	*Users:* Personal *Purposes:* Choose gambling tactics
9. Fruit and Vegetable Prices (1018 recent views) USDA — How much do fruits and vegetables cost? USDA Economic Research Service estimated average prices for 153 commonly consumed fresh and processed fruits and vegetables.	*Users:* Gov't, Business, Personal *Purposes:* Measure inflation factors, set pricing, choose products to purchase.
10. Death rates from suicide (952 recent views) U.S. Dep't of Health & Human Services — Data on death rates from suicide, by sex, race, Hispanic origin, and age.	*Users:* Gov't, Health Care, Activists *Purposes:* Research suicide, design prevention techniques

Top Nigerian OpenGov datasets

Nigeria - as of May 23, 2024	Potential Users and Potential Purposes
1. Economic Indicators: These datasets cover various economic parameters such as GDP growth, inflation rates, employment data, and more. These indicators are published by the National Bureau of Statistics at https://nigerianstat.gov.ng/ . Researchers and policymakers use these datasets for economic analysis & policy formulation	*Researchers, Lawmakers, Media, Investment decisions,*
2. Company registers: List of all registered companies in Nigeria. Does not include detailed financial data such as balance sheet, etc.. Published by the Corporate Affairs Commission at https://search.cac.gov.ng/home Open data from company registers may be used for many ends: e.g. enabling customers and businesses to see with whom they deal, or to see where a company has registered offices	*Contract Buyers, tax agencies, economic crime agencies; justice sector agencies; job seekers*
3. Financial and Budgetary Data: These datasets cover government budgets, revenue collection, expenditure patterns, and fiscal policies. The Nigeria budget data is published by the Nigeria Budget Office of the Federation at https://budgetoffice.gov.ng/	*Government Contractors, Lawmakers, Researchers, Civil Society Organizations, International Development Agencies* Predict government revenue, and expenditure, cost benefit analysis, showing where money is spent on, how public funds develop over time, and why certain activities are funded.

REI SYSTEMS

Top Nigerian OpenGov datasets (cont'd)

Nigeria - as of May 23, 2024	Potential Users and Potential Purposes
4. Election Results: Election data informs about voting outcomes and voting process. What are electoral majorities and minorities? How many votes are registered, invalid, or spoilt? This data set is published by the Independent National Electoral Commission at https://inecelectionresults.ng/ The datasets here are mostly assessed during an election period	*Lawmakers, Researchers, Political Parties, Civil Society Organizations, Independent election monitors, media, citizens* *Analyse election results, validate polling results, provide near to real-time results during elections.*
5. Procurement data: All tenders and awards of the national/federal government aggregated by office. It includes procurement planning or other procurement phases. Monitoring tenders helps new groups to participate in tenders and to increase government compliance. The data set is published by the Bureau of Public Procurement on the Nigeria Open Contracting Portal (NOCOPO) at https://nocopo.bpp.gov.ng/	*Private sector, government agencies, suppliers, contractors, consultants, citizens* Open procurement data may enable fairer competition among companies, track procurement processes, allow to detect fraud, analyze government spending
6. Health and Demographics Data: Datasets related to health indicators, disease prevalence, vaccination coverage, and population demographics are crucial for public health research and planning. Published by the National Bureau of Statistics at https://nigerianstat.gov.ng/ and the Nigeria open data portal Health Atlas at https://nigeriahealthatlas.opendataforafrica.org/	*Researchers, Lawmakers, International Development Agencies, Private sector health providers, Small Businesses, Media* Targeted health Project interventions; impact assessment of health interventions

Top Nigerian OpenGov datasets (cont'd)

Nigeria - as of May 23, 2024	Potential Users and Potential Purposes
7. Legislation/Laws: Datasets of all national laws and statutes, which might also include information on legislative behaviour e.g. voting records. This helps citizens to track the law their representatives is proposing, and how they are voting at the national assembly. The datasets is published under resources at https://nass.gov.ng/	*Lawyers, Lawmakers, International Development Agencies, Justice Sector Agencies, Civil Society Organizations, Media* Legitimize the work of lawmakers, encourage feedback from citizens
8. Population Census: offers standardized datasets from Nigeria's extensive administrative data ecosystem. The Primary Population Census 2006 dataset is published by the Nat'l Bureau of Statistics Nigeria open data portal: https://nigeria.opendataforafrica.org/ Though conducted over 15 years ago, population estimates are drawn from the 2006 census dataset.	*Researchers, Lawmakers, Media,* Essential for planning across all sectors.
9. Crime and Safety: Crime statistics, law enforcement data, and safety-related information are essential for policymakers, law enforcement agencies, and researchers. The National Bureau of Statistics publishes occasional crime statistics at https://nigerianstat.gov.ng/	*Private investigators, Realtors, citizens, businesses, investors* Make decision on residence, targeted planning by the police

Top Nigerian OpenGov datasets (cont'd)

Nigeria - as of May 23, 2024	Potential Users and Potential Purposes
10. Education Statistics: Datasets related to school enrollment, literacy rates, educational infrastructure, and student performance are valuable for educational research and planning. Recently, the National Bureau of Statistics at https://nigerianstat.gov.ng/ started publishing national exams results while the Ministry of Education published school enrollement rates on the Education Management Information System page at https://nemis.gov.ng/	*Parents, Researchers, Lawmakers, International Development Agencies, Private sector, Small Businesses,* Targeted education Project interventions; impact assessment of education interventions; parents decision on children schools

Top Indian OpenGov datasets

From past 12 mo, ranked as of May 29, 2024	Potential Users and Potential Purposes
1. All India Pincode Directory Contains all the pin -code list across India with geocodes and other relevant information like Office Name, Office Type, Delivery Status, etc.	*Users*: *Citizens, Businesses, Governments* *Purposes* : E-commerce/Logistics companies, Government Departments, Market research agencies, Emergency services, Address verification by banks/utility companies, etc.
2. Company Master Data Master details of any company registered with Registrar of Companies (RoC). Data contains various information like Corporate Identification Number(CIN), Company Name, etc.	*Users*: *Governments, Investors, Businesses, C itizens* *Purposes* : Regulatory compliance, Due diligence & background checks, business research/consulting, credit agencies, fraud detection, etc.
3. Total and Newly Registered Motor vehicles New and Total Number of Registered Motor Vehicles in India.	*Users*: *Governments, Businesses* *Purposes* : Insurance, Traffic Mgmt., Transport planning, Socio - economic and demographic profiling, research and academia
4. State and District -wise Scheduled Caste and Scheduled Tribe Population : Granular information on SC/ST population by sex and residence.	*Users*: *Governments, Researchers, Academia* *Purposes* : Policy/program planning, social and economic research, political analysis/strategy, academic research, etc.
5. Reserve Bank of India – Database of Indian Economy Macroeconomic data with indicators about the state of the Indian economy including data on financial sector	*Users*: *Governments, Investors, Businesses, C itizens* *Purposes* : Monetary policy, Fiscal policy, Program planning and implementation, financial sector analysis, media, etc.

Source: https://data.gov.in/analytics/ogdindia/All/download_views_block

Top Indian OpenGov datasets (cont'd)

From past 12 mo, ranked as of May 29, 2024	Potential Users and Potential Purposes
6. Crop Production Statistics The data refers to district wise, crop wise, season wise and year wise data on crop covered area (Hectare) and production (Tonnes)	*Users*: *Agri-businesses, Governments, Academia* *Purposes*: *Agricultural policy planning, food security and distribution, agricultural marketing, agri-finance/insurance, etc.*
7. All India Consumer Price Index The data refers to group wise all India Consumer Price Index for Rural & Urban with base year 2010.	*Users*: *Governments, Businesses, Investors, Academia* *Purposes*: *Policy/program planning, Macro-economic research, labour policy/wage determination, etc.*
8. India's Bhuvan Geo-Platform 2D and 3D maps with details like the environment and other natural features.	*Users*: *Governments, Businesses, Citizens* *Purposes*: *Businesses, Urban Local bodies, Disaster Management, Academic research, etc.*
9. Indian Railways Timetable : **C**ontains train wise departure and arrival times at various stations. It also provides information of the route, distance covered , source station and destination station etc..	*Users*: *Citizens, Businesses* *Purposes*: *Train status enquiry, Logistics, etc.*
10. Import Export Classification (HS Code and Import Policy) Refers to details on Import Export Classification, ITC (HS) Code and Import Policy.	*Users*: *Businesses, Governments, Financial Institutions, Academia* *Purposes*: *Import and export businesses, customs and border officials, trade policy, supply chain management, academic research, etc.*

Source: https://data.gov.in/analytics/ogdindia/All/download_views_block

Top Finnish OpenGov datasets

From past 12 mo, ranked as of May 29, 2024	Potential Users and Potential Purposes
1. Finnish Names Digital and Population Data Services Agency; first names and surnames in the Population Information System	*Users:* citizens, journalists, researchers *Purposes:* finding out names that are allowed, seeking inspiration for a baby's name, reporting statistics on most popular names, sociological research
2. Building addresses and electoral districts Digital and Population Data Services Agency; building ID code, building category, co-ordinates, address, electoral district of building	*Users:* citizens, businesses, public sector *Purposes:* sending marketing mail, finding a certain or types of certain building for official purposes, snooping/curiosity
3. Addresses, postal codes and WGS84 coordinates of buildings Digital and Population Data Services Agency; roughly same as above but with WGS84 coordinates and postcodes separated out	*See above*
4. War Victims of the Second World War National Archives of Finland; information on all military casualties in Finnish Army, Navy and Air Force during World War II.	*Users:* citizens, historians *Purposes:* curiosity, genealogical research, historical research
5. Admission criteria for upper secondary schools in Helsinki Finnish National Agency of Education; entrance limits for getting to upper secondary schools (comprehensive school grade average)	*Users:* citizens *Purposes:* figuring which upper secondary schools to apply to, and which upper secondary schools are "elite"

Top Finnish OpenGov datasets (cont'd)

From past 12 mo, ranked as of May 29, 2024	Potential Users and Potential Purposes
6. Application results for upper secondary schools in Helsinki Finnish National Agency of Education; number of applicants, number of selected candidates, as well as the entrance limits	*Users:* citizens *Purposes:* figuring which upper secondary schools to apply to, and which upper secondary schools are "elite"
7. Registry of businesses allowed to sell or serve alcohol National Supervisory Authority for Welfare and Health	*Users:* regulatory bodies, citizens, businesses *Purposes:* checking that a business is allowed to serve alcohol (while carrying out other regulatory checks, e.g. food or worker safety); figuring size & scope of competition in a geographical area
8. Road traffic accidents Finnish Transport Infrastructure Agency; details of every road traffic accident in a given year (e.g. category, location, type of road/intersection, weather, etc.)	*Users:* journalists, regulatory bodies, researchers *Purposes:* reporting stories on traffic statistics, finding dangerous roads, curiosity
9. Data from OpenProcurement.fi State Treasury of Finland; data includes the Finnish central government's procurement related invoices from the beginning of 2016, as well as some cities' invoices	*Users:* journalists, activists, researchers *Purposes:* finding out where taxpayer euros go
10. War Victims in Finland Between 1914 and 1922 National Archives of Finland; war-related deaths in Finland from the beginning of WWI to the end of Kinship Wars in 1922.	*Users:* citizens, historians *Purposes:* curiosity, genealogical research, historical research