Why do some governments release data and facilitate its reuse more than others?

Results indicate that internet access, public sector corruption, the strength of civil society and GDP explain 70 percent of the differences in data openness.

Using the implementation datasets average indicator of data openness from the Open Data Barometer (ODB), this brief discusses the results of a regression analysis identifying country characteristics that explain differences in open government data (OGD) around the world. It begins with a description of the research question, data and model used to arrive at its results, follows with a summary and interpretation of those results and concludes with suggestions for further research.

Publicly available government data can empower marginalized groups, reduce corruption and improve public services (Badouard,2014; Duguay et al., 2019; Kucera and Chlapek, 2014; Pereira et al., 2017). However, it only offers these benefits if it can be reused and analyzed. Good data management practices are therefore essential for greater data openness (Baack, 2015; Badouard,2014; Geiger and Von Lucke, 2012). As shown in figure 2, most countries have released the majority of the data thought to have the greatest potential to benefit the public but only a small share of that data is machine-readable. For this reason, the analysis defines data openness based on the ODB implementation variable that considers, among other criteria, whether the data is available, is free to use, and is machine-readable.

Study Design

The study design uses cross-sectional, ordinary least squares (OLS) regression analysis and data from 2013-2016 to understand what accounts for differences in the levels of data openness (this dynamic map shows those differences).[i] This does not imply that the factors mentioned lead to government data openness; simply the analysis identifies characteristics that are closely associated with better OGD practices. To measure openness the analysis uses the implementation datasets average indicator included in the ODB dataset. This measure is selected because it considers the availability of government data as well as meticulously scores[ii] the data management practices used to create and publish the data. It also only considers data that is thought to be important for government accountability, private sector innovation and social policy.[iii] Using this measure ensures data openness is defined in such a way that increases the likelihood it catalyzes positive change.

The analysis evaluates the relevance of 7 factors that could impact data openness. These include: the strength of civil society[iv], public sector corruption[v], corruption within the executive branch of government[vi]; civil liberties[vii], government effectiveness[viii], fixed broadband internet subscription[ix]; and GDP[x] (see the endnotes for more details about the variables). The variables are standardized so that their means are 0 and their standard deviation is 1. A regression coefficient of 1 can be interpreted to mean that a 1 standard deviation increase in the x-variable is associated with a 1 standard deviation increase in the y-variable. Transforming the regressors in this way allows for the magnitude of the regression coefficients to be compared across variables. Lastly, robustness checks are conducted by modifying the specification of the model and changing the variables included to ensure that the results are reliable.

Findings

The findings indicate, in order of effect size from largest to smallest (with each regression coefficient appearing in parenthesis), that broadband access (0.34), level of public sector corruption (-0.29), GDP (0.29) and the strength of civil society (0.15) are related to government data openness with at least 95% statistical confidence. These variables explain 70% of the differences in data openness and confirms the relevance of

the x-variables included in the model.[xi] Surprisingly, corruption in the executive branch, government effectiveness and the establishment of civil liberties are not found to be relevant. The robustness checks (see figure 3) validates the results of the main model.

Lessons learned:

1. Broadband access is the most powerful explainer of government data openness.
    a. This may be because when more people can access data online, there is more demand and greater public pressure to release data.
    b. Alternatively, societies with greater internet access may have government officials that better understand the power of data and are more inclined to prioritize its publication.
    c. Regardless the reason, lowering the cost and increasing internet access should be a priority for organizations working to advance government data openness.
2. With less public sector corruption comes greater openness.
    a. When addressing corruption for greater openness, reform efforts should focus on government bureaucracies as opposed to the executive branch of government.
    b. Public sector corruption's relationship with openness could be driven by more corrupt bureaucracies being less interested in exposing their work to public scrutiny.
3. More effective governments are not necessarily more open.
    a. This may indicate that a commitment or desire for openness on the part of governments is more important than the overall levels of effectiveness.
4. Strong civil societies and data openness are closely linked.
    a. Civil society should know that their cumulative strength is associated with government data openness and may want to further prioritize OGD in their work.

Conclusion

Government data openness ranges greatly (see figure 1) and governments lag behind in ensuring that the data can be easily reused and analyzed (see figure 2), highlighting the need for continued progress towards better data management practices. By shedding light on which country characteristics relate to OGD, this study attempts to identify factors advocates of open data can target to advance the movement.

The results of this brief should be interpreted as preliminary. A number of questions remain. For starters, it is possible that the relationships identified are being driven by factors outside the model. For example it may be the case that cultural norms drive openness as well as levels of corruption. Similarly, it may be that reverse causality is driving the results. This would be the case if openness discourages corruption. Future research should address these questions.

Currently, the data is not well suited for panel data analysis. In particular, fixed effects models or a generalized method of moments (GMM) technique relies on variation within observations across time; but with only 5 years of data (if the leaders edition is included) and explanatory variables that change minimally from year to year (e.g. GDP, levels of corruption and government effectiveness), there will not likely be enough variation to identify statistically significant relationships. However, as more years of data become available, the use of a fixed effects model or GMM combined with lagged x variables could address some of the issues mentioned above.

Works Cited

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. Big Data & Society, 2(2), 2053951715594634.

Badouard, R. (2014). Open government, open data: l'empowerment citoyen en question. Open Data Big Data Sci. Donnees Editions Maison Sci. L 'Homme, 31-54.

Brandusescu, A., Iglesias, C., Robinson, K., Alonso, J. M., Fagan, C., Jellema, A., & Mann, D. (2018). ODB 4th Edition Global report.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., ... & Lührmann, A. (2019). V-dem codebook v9.

Duguay, R., Rauter, T., & Samuels, D. (2019). The Impact of Open Data on Public Procurement. Available at SSRN 3483868.

Geiger, C. P., & Von Lucke, J. (2012). Open government and (linked)(open)(government)(data). JeDEM-eJournal of eDemocracy and open Government, 4(2), 265-278.

Graham, Benjamin .A.T., Raymond Hicks, Helen Milner, and Lori D. Bougher. 2018. World Economics and Politics Dataverse.

Iglesias, C., Robinson, K., Brandusescu, A., Alonso, J. M., Fagan, C., Jellema, A., & Mann, D. (2016). Open data barometer 4th edition: research handbook-v1. 0 10th June 2016.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The worldwide governance indicators: methodology and analytical issues. Hague Journal on the Rule of Law, 3(2), 220-246.

Kucera, J., & Chlapek, D. (2014). Benefits and risks of open government data. Journal of Systems Integration, 5(1), 30-41.

Web Foundation (2013). Open Data Barometer – 1st Edition. Washington DC: *World Wide Web Foundation*. Available from: https://opendatabarometer.org/leadersedition/data/

Web Foundation (2014). Open Data Barometer – 2nd Edition. Washington DC: *World Wide Web Foundation*. Available from: https://opendatabarometer.org/leadersedition/data/

Web Foundation (2015). Open Data Barometer – 3rd Edition. Washington DC: *World Wide Web Foundation*. Available from: https://opendatabarometer.org/leadersedition/data/

Web Foundation (2016). Open Data Barometer – 4th Edition. Washington DC: *World Wide Web Foundation*. Available from: https://opendatabarometer.org/leadersedition/data/

Web Foundation (2016). Open Data Barometer - Open Data Barometer ODB Methodology - v1.0 28th April 2016 online]. Available from: https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-Methodology.pdf [Accessed 24th of March 2020]

Pereira, G. V., Macadar, M. A., Luciano, E. M., & Testa, M. G. (2017). Delivering public value through open government data initiatives in a Smart City context. Information Systems Frontiers, 19(2), 213-229.

[i] Although data from various years are used, the analysis can be considered cross-sectional because the time dimension is not considered and country observations from different years are treated the same as observations of different countries in the same year.

[ii] The scoring system is based on ten criteria: existence, online availability, machine-readable format, presented in bulk, free, is licensed, updated, sustainable, easy to find and has links to key elements. A more detailed description can be found in the ODB methodology guide.

[iii] The 15 datasets considered by the ODB indicator can be found in the ODB methodology guide.
[iv] The Civil Society Participation Index is constructed from a number of indicators that together capture the autonomy, strength and influence civil society has over government. The variable was originally collected from the Varieties of Democracy Project.
[v] The Public Sector Corruption measures the prevalence of bribery, theft and the misappropriation of public funds by government bureaucrats and members of the executive. The variable was originally collected from the Varieties of Democracy Project.
[vi] Executive Corruption Index measures the same as previous one but on members of the executive. The variable was originally collected from the Varieties of Democracy Project.
[vii] The Civil Liberties Index measures the extent governments infringe upon individuals political and personal liberties and the extent governments act violently against their own citizens. A higher number corresponds to more liberties and less violence. The variable was originally collected from the Varieties of Democracy Project.
[viii] Government effectiveness captures "the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies." The variable is taken from the World Bank's Governance Indicators.
[ix] The variable for internet access indicates the percentage of the population with broadband access. It was originally collected by the World Bank, World Development Indicators. This variable is also included in the ODB Scores dataset.
[x] GDP (constant 2205 US, logged) was collected from the World Bank, World Development Indicators.
[xi] The R-squared associated with the regression used is over 70. The R-squared measures the percentage of the differences in the outcome variable explained by the explanatory variables used in the model.