

Open Data: Differences and Implications across Countries

Lay Lian Chuah and Norman V. Loayza

The pros and cons of making data more accessible to the public have been widely debated. Proponents of open data argue that it is good practice for governments because it fosters transparency, promotes greater participation, and encourages sharing of ideas, which is important in building a research-oriented culture. Others, however, are less convinced of the merits of open data. This research policy brief finds that there is a relationship between accessibility of data and income levels of a country, and between data availability and the productivity and quality of economic research.

The Merits of Open Data

Advocates of open access to public data argue that open data nurtures research; spurs the sharing of ideas; and helps individuals, firms, and policy makers make informed decisions that lead to improved outcomes. While the benefits of open data seem compelling, trust and privacy issues, as well as lack of expertise, resources, and technological capabilities, continue to act as barriers to open data practices.

While commitment to open data initiatives and efforts made to achieve them vary widely around the world, many countries are making great strides in making data as accessible as possible (Neubauer 2013). However, other countries are less convinced of the merits of providing complete access to data. This brief discusses and provides empirical evidence on two key questions related to the merits of open data: Is there a relationship between accessibility of data and income levels of a country, and is there a relationship between data availability and research productivity and quality.

Two international assessments of data openness around the world—Open Data Barometer (ODB) and Open Knowledge International (OKI)—define open data as public information that can be “freely used and shared by anyone for any purpose.” Data quality is equally important. Together, the availability, accessibility, and quality of data determine the usefulness and usability of data. Users will find significant value add in having accurate,

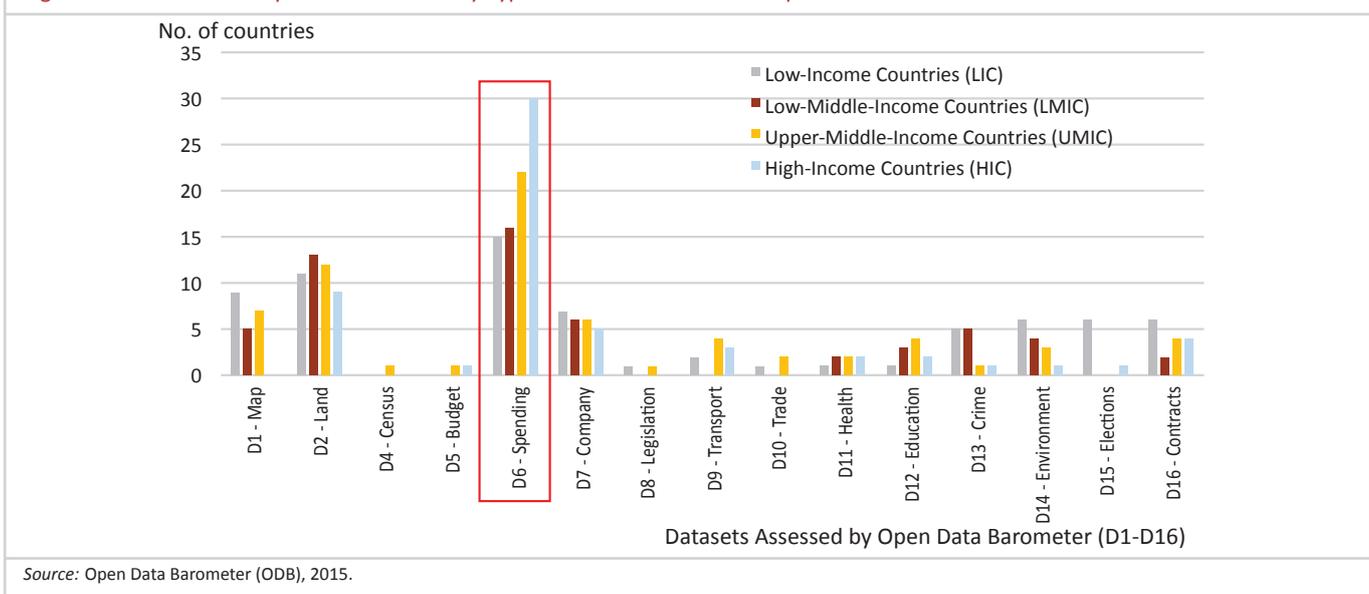
complete, easy to interpret, and timely data to solve problems or make decisions.

While in many countries the national statistical agencies may not be the only custodians of key public data, the national government is nonetheless responsible for ensuring the open publication of such data. Some data-collecting agencies consider data protection to be important in maintaining the trust of the establishments they survey and in eliciting truthful responses from them. However, data that are inaccessible represent a locked resource from which value cannot be fully extracted. By unlocking data, the government can leverage on the creative and rigorous policy recommendations from the research community for its policy analysis and planning.

How accessible are data in different countries?

ODB and OKI assess the state of open data initiatives globally. The two sets of rankings differ in their methodology and coverage. ODB covers 92 countries and 15 types of datasets. It computes a country’s ranking based on three dimensions: a country’s readiness to support and respond to the positive outcomes from open data initiatives; the implementation of open data practices; and the impact of open data on governments, societies, and the economies. ODB also combines experts’ opinions, technical assessments of data supply, and secondary data for the construction of its open data rankings. OKI covers 122 economies, 13 types of datasets, and measures openness based solely on data accessibility.

Figure 1. Data Accessibility across Countries by Type of Data and Income Group

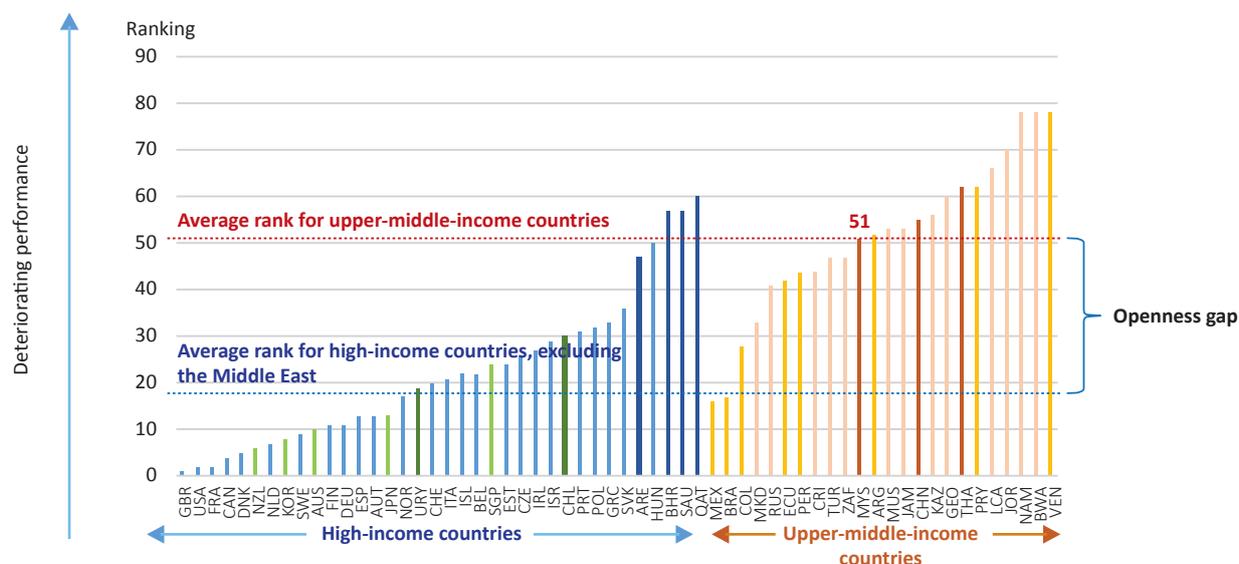


Affiliation: Development Research Group, the World Bank.

Acknowledgement: Kenneth Simler and Nancy Morrison contributed to this brief with insightful comments and suggestions.

Objective and disclaimer: Research & Policy Briefs synthesize existing research and data to shed light on a useful and interesting question for policy debate. Research & Policy Briefs carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions are entirely those of the authors. They do not necessarily represent the views of the World Bank Group, its Executive Directors, or the governments they represent.

Figure 2. Openness Gap in Data between Upper-Middle-Income Countries and High-Income Countries



Source: Open Data Barometer, 2015

Note: The openness gap is the difference between the average rank for upper-middle-income countries and high-income countries.

Dark blue bars are high-income countries in the Middle East; sky blue bars are high-income countries in Europe and North America; light green bars are high-income countries in the Asia Pacific; and dark green bars are high-income countries in Latin America. Orange bars are upper-middle-income countries in Latin America. Pale orange bars are upper-middle-income countries in Central Asia, Africa, and the Caribbean. Red bars are upper-middle-income countries in Asia.

According to ODB’s 2015 assessment, 55 percent of the countries surveyed have an open data initiative in place; however, only 10 percent of government data is freely accessible. Twenty-six of the top 30 countries in the ranking are high-income countries. Half the open datasets are found in just the top 10 member-countries of the Organisation of Economic Cooperation and Development (OECD), while almost none are in Africa. The rankings reveal a large openness gap between high- and low-income countries. Similar patterns also emerge in the OKI rankings for 2015. High-income countries occupy 23 out of the top 30 places. Taiwan, China tops the 2015 ranking and became the first non-European economy to be placed in the top three.

Data accessibility can be assessed at three levels: (1) public data do not exist; (2) public data do exist but are somehow inaccessible; and (3) public data exist and are accessible. According to ODB, only two of the 92 countries in their rankings—the United Kingdom and Japan—make all of their existing 15 types of datasets available to the public. Slightly more than half the countries have at least one or two types of datasets that exist but are not available to the public (figure 1). Of the 1,380 datasets for 92 countries surveyed, only 18.7 percent (256) of datasets that exist are not accessible to the public; they mainly cover government spending and land data.

The first 32 spots in the ranking are mainly filled by high-income countries, excluding countries of the Middle East. The average ranking for high-income countries is 18 (figure 2). The upper-middle-income countries tend to occupy the 16th to 78th spots. However, Latin American countries such as Mexico, Brazil, and Colombia (in the top 30) perform more favorably than other countries at similar income levels. The average ranking for upper-middle-income countries is 51. Thus there is an “openness gap” of 33 notches between the upper-middle-income countries and the high-income countries. ODB attributes this gap mainly to the lower scores for the sustainable publication of data, discoverability of data, and links to key datasets.

In Asia the majority of the upper-middle-income countries are concentrated in the 40th to 80th spots, according to the OKI 2015

rankings which shows Malaysia and China falling outside the upper-middle-income band. The openness gap between the upper-middle-income and high-income countries in the region is 44 notches—wider than the corresponding gap in non-Asian countries. The disparity in open data initiatives for countries in Asia is higher than for countries outside this region.

Is there a relationship between accessibility of data and income levels?

The positive relationship between data access and income levels suggests that greater access to data allows people to be more informed and efficient in solving problems, which leads to better outcomes. However, it could also be the case that the causality runs the other direction: that is, higher-income countries have the resources to invest in open data initiatives.

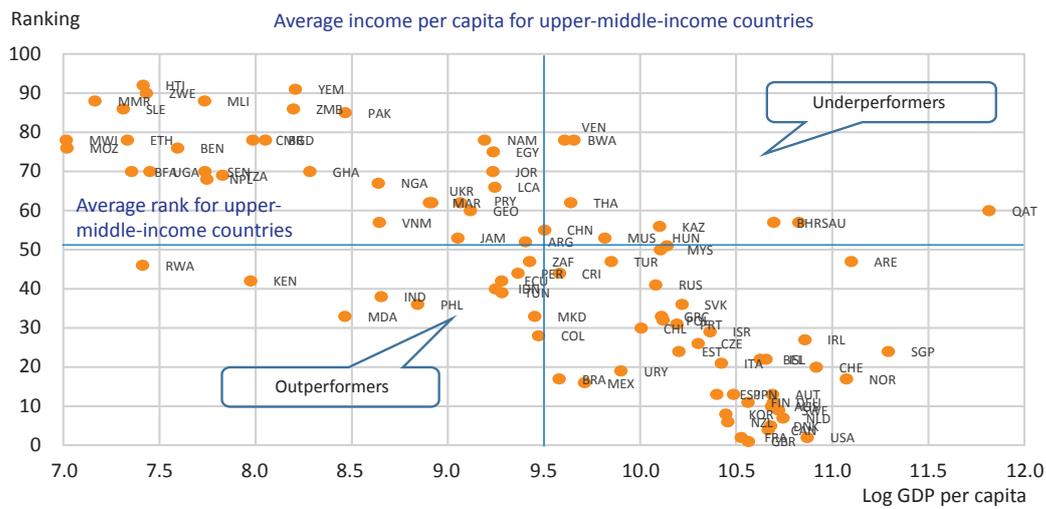
The correlation between a country’s ranking in data openness and GDP per capita is -0.75 (figure 3). Using the averages of the upper-middle-income rankings and GDP per capita as benchmarks suggests that countries such as Malaysia, Thailand, and Venezuela (right upper quadrant) have underperformed with respect to countries in the similar GDP band. Among the upper-middle-income countries, Ecuador and South Africa are slight outperformers.

Is there a relationship between data availability and research productivity and quality?

The field of economics has evolved over the past several decades toward greater emphasis on empirical work. While economic theory provides a conceptual framework, better data facilitate more rigorous testing of theories and assessment of their relevance (Einav and Levin 2014, Jin 2009). In a growing number of cases, more granular data are needed (McGuckin 1993); that is disaggregated data coming from household, labor, and firm surveys and censuses, for instance.

Until the mid-1980s, the majority of papers published in the top three economic journals—American Economic Review,

Figure 3: Data Accessibility and Income per Capita



Source: Open Data Barometer, 2015, and the World Development Indicators (WDI) database.
 Note: Log GDP per capita is in terms of purchasing power parity (PPP) in constant 2011 international dollars.

Journal of Political Economy, and Quarterly Journal of Economics—were theoretical, according to a review of publication patterns from 1963 to 2011 (Hamermesh 2013). However, the share of empirical papers in top journals has climbed to more than 70 percent in 2011. These empirical papers use data that have been assembled by public agencies, obtained directly by the authors, or generated through controlled experiments. Importantly, granular data is being used to pose new questions. As a result, it enables new research designs that can offer insights in the consequences of different economic policies and events.

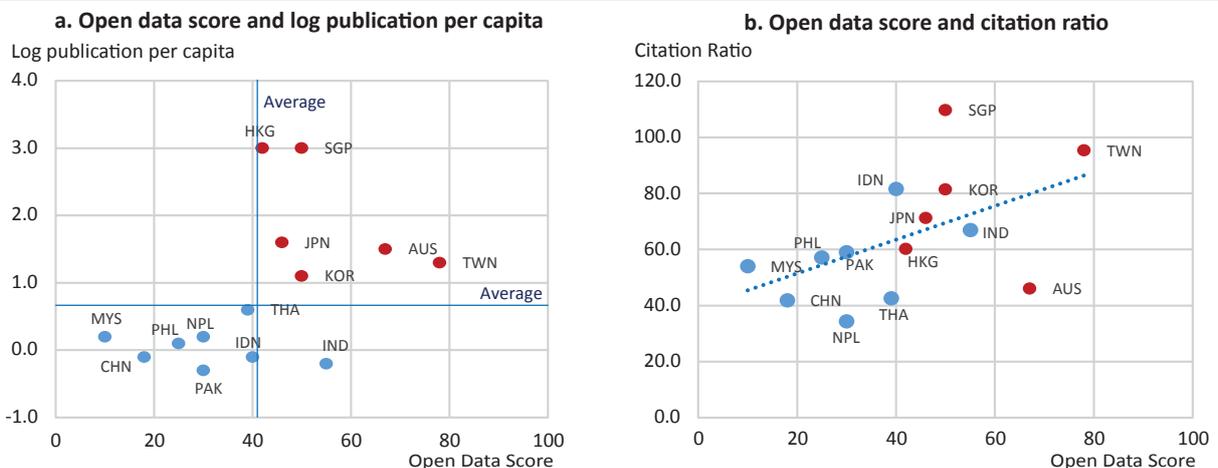
We consider the link between countries’ quality of research and the open data ranking compiled by OKI for countries in the Asia Pacific region. For a proxy for quality, the analysis uses the number of publications and citations of journal articles published in the top 10 high-impact economic journals in the past 10 years by countries in the sample. This approach is adapted from Aizenman et al (2011). The high-impact journal articles are selected based on their ranking by a database dedicated to economic research, IDEAS, which computes citation counts and various measures of a publication’s impact. The articles published by the top 10 journals

between 1995 and 2015 are summed and divided by the average population during the same period to yield a citation ratio. Because the ratios are very small, they are multiplied by 10 million before converting into logs (figure 4).

As shown in panel a, there is a positive correlation of 0.46 between the OKI’s open data scores and log publication per capita of the Asia Pacific economies. High-income countries (shown by red dots) with higher open data scores have a higher publication per capita (1.9 average, compared to the overall average of 0.8).

The association between the citation ratio and open data is positive (+0.51). This is consistent with the idea that data openness is needed to produce quality research (panel b). High-income countries tend to produce higher quality research. For every article published by high-income countries, an average of 77.3 citations are registered, compared to 46.1 citations for upper-middle-income countries; 66.1 citations for lower-middle-income countries; and 34.3 citations for the only low-income country in the sample, Nepal. The results suggest that quality publications may be associated with open data and in turn appear to be associated with higher income levels.

Figure 4. Data Openness is Positively Correlated with Publication per Capita and Citation Ratio



Source: Open Knowledge International (OKI) and IDEAS.
 Note: Higher open data scores result in better rankings. TWN = Taiwan, China.
 Red dots = high-income countries. Blue dots = upper middle, low-middle and low-income countries.

Box 1. How Data Users Perceive the Accessibility and Quality of Public Data in Malaysia

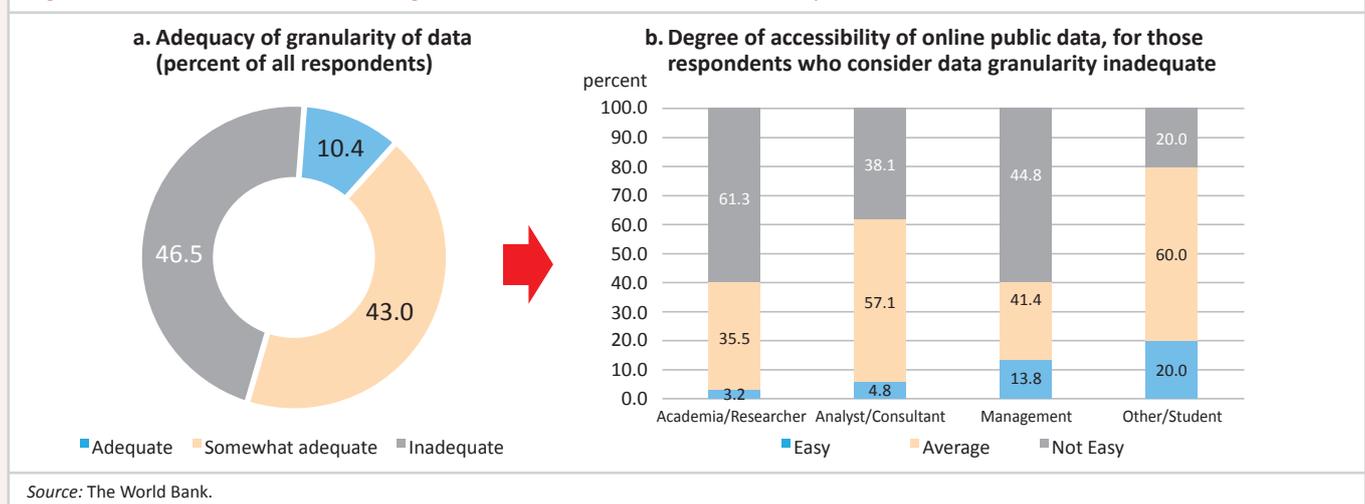
A survey conducted in September 2016 for this Brief set out to answer questions on the accessibility and quality of public data. The three-week survey sampled 831 respondents of whom about one-quarter (28 percent) responded. Nearly half the respondents (46.1 percent) are in the government or public administrative sector, while 18.3 percent work in education, 16.5 percent work in finance, and 11.7 percent work in nonprofit organizations, including think tanks. By profession, 39.8 percent are either analysts or consultants, 28.1 percent are in management positions, and 24.2 percent are academicians or researchers.

Almost 60 percent of the respondents have at least 11 years of work experience. An overwhelming majority (95.7 percent) are comfortable using computers. The majority of the respondents (70 percent) have found publicly available data relatively easy to

access online. Only 30% of the respondents found it to be difficult. Approximately half of the respondents also consider the quality and format to be average.

On the other hand, most respondents (89.5 percent) reported that the data were not adequate in terms of granularity needed for rigorous economic research (figure B1.1a). Granularity corresponds to disaggregated data, at the individual, household, worker, or firm levels. Of those who consider that access to sufficiently disaggregated data to be inadequate, 68 percent work in professions that use data intensively, including academicians, researchers, analysts, and consultants. Among the academicians and researchers who consider data to be inadequate, 61.3 percent also consider public data not easily accessible (figure B1.1b). About 61.3 percent of the respondents who indicated that the data were not granular enough also found all data to be of average quality. More than 75 percent of the respondents agree or strongly agree that availability of data contributes to research capacity in Malaysia.

Figure B1.1 Access to Granular Data (e.g. household and firm-level data) in Malaysia is a Constraint to Research



Conclusion

Public data are an asset. Making them available, usable, and discoverable—that is, “open”—promotes efficiencies, increases transparency, creates economic opportunities, and increases people’s participation in the contribution of ideas (www.data.gov). However, some countries, concerned about confidentiality and data abuse, remain unconvinced about the benefits of making data more publicly accessible. These diverging opinions are reflected in the uneven progress in making data accessible to the public. Based on the rankings provided by ODB and OKI, high-income countries appear to have made better progress in terms of readiness to adopt and implement open data practices.

Data are vital for research. The analysis shows that there is a positive association between research productivity and data accessibility. The association is also positive between quality of research and data accessibility.

The movement towards open data is gaining traction. Countries need to be prepared to devote resources to strengthen data management and programming capacities, plan for unintended consequences, and engage communities to harness the potential of open data. A challenge is to develop methods for researchers to access data in ways that respect privacy and confidentiality.

References

Aizenman, J., H. Edison, L. Leony, and Y. Sun. 2011. “Evaluating the Quality of IMF Research: A Citation Study.” Background Paper BP/11/01, Independent Evaluation Office, International Monetary Fund, Washington, DC.

Einav, L., and J. Levin. 2014. “Economics in the Age of Big Data.” *Science* 346 (6210): 715–821.

Hamermesh, D. S. 2013. “Six Decades of Top Economics Publishing: Who and How?” *Journal of Economic Literature* 51 (1): 162–72.

Jin, J. C. 2009. “Economic Research and Economic Growth: Evidence of East Asian Countries.” *Journal of Asian Economics* 20 (2): 150–55.

McGuckin, R. H. 1993. “The Importance of Establishment Data in Economic Research.” Working Paper 93-10, Center of Economic Studies, U.S. Census Bureau, Washington, DC.

Neubauer, M. 2013. “New Report Highlights Successes and Challenges of Worldwide Open Data Policies.” Techpresident.com, November 1. <http://techpresident.com/news/wegov/24480/new-report-highlights-successes-and-challenges-worldwide-open-data-policies>.

Open Knowledge International. Global Open Data Index. <http://2015.index.okfn.org/>.

Open Data Barometer. <http://opendatabarometer.org/barometer/>.