

The Lack of Decentralization of Data: *Barriers, Exclusivity, and Monopoly in Open Data*

Carla Hamida¹, Amanda Landi²

¹Bard College at Simon's Rock, Great Barrington, USA, chamida16@simons-rock.edu

²Bard College at Simon's Rock, Great Barrington, USA, alandi@simons-rock.edu

ABSTRACT: Recently, Facebook creator Mark Zuckerberg was on trial for the misuse of personal data. In 2013, the National Security Agency was exposed by Edward Snowden for invading the privacy of inhabitants of the United States by examining personal data. We see in the news examples, like the two just described, of government agencies and private companies being less than truthful about their use of our data. A related issue is that these same government agencies and private companies do not share their own data, and this creates the openness of data problem. Government, academics, and citizens can play a role in making data more open. In the present, there are non-profit organizations that research data openness, such as Open Data Charter, Global Open Data Index, and Open Data Barometer. These organizations have different methods on measuring openness of data, so this leads us to question what does open data mean, how does one measure how open data is and who decides how open should data be, and to what extent society is affected by the availability, or lack of availability, of data. In this paper, we explore these questions with an examination of two of the non-profit organizations that study the open data problem extensively.

KEYWORDS: centralization, data openness, government, institutions, open data principles, privilege, marginalized groups, methodology

Introduction

In economist Friedrich A. Hayek's infamous *The Use of Knowledge in Society* published in 1945, he claims that relying on a central board and third-party information to make impactful decisions is nonoptimal, and as a consequence,

"We need decentralization because only thus can we insure that the knowledge of the particular circumstances of time and place will be promptly used. But the "man on the spot" cannot decide solely on the basis of his limited but intimate knowledge of the facts of his immediate surroundings" (Hayek 1945, 524-525).

Hayek's words echo to fields beyond economics, and to the present. He raises an important matter asking to what extent do central boards, which are today's governments, should intervene. While relying on individual institutions at the time and place to decide how to solve issues may be a great idea, as they are the ones who understand the problem at hand, this individual institution does not necessarily possess the information that may affect other people in society. Hayek wrote his paper at a time when information did not move as quickly as today, before the Internet. We are now living in a time where data can be more open and transparent, in order to solve problems collectively as a society, instead of as disparate institutions.

Control of data is an elitist phenomenon. Just as access to quality education is a privilege in many societies in this world, so is access to data. Knowledge and information give autonomy to people; people will be able to make informed decisions by considering external factors that may concern them. The access to much of this "knowledge" is restricted to a specific set of people, and so it is not shared outside these "gatekeepers." We can improve how we make decisions and the autonomy of individuals by sharing this data. We can make this information more open and transparent for everyone, regardless of the identities they belong to.

In this paper, we will first list the open data principles and review the backgrounds of the organizations working to make data more open. Then in the following section, we will discuss the findings of the research of those organizations. In section 3, we analyze their measurement methodologies, and talk about their lack of inclusivity. For the fourth section, we highlight the

monopolization of data by private companies and expand on the need to decentralize the regulation for data openness.

Section 1. The Open Data Charter & Open Data Principles

At the Group of Eight 2013 summit, the eight governments recognized the significance of this need for open data. Thus, the Open Data Charter (ODC) was founded in 2015, and it is a collaboration of more than 70 governments, experts, and organizations whose sole goal is to make governmental data more available and accessible to citizens of the world. In 2015, the Open Data Charter proposed six principles for how governments should publish data in order to be counted as open. Those are:

- “Open by default”
- “Timely and comprehensive”
- “Accessible and usable”
- “Comparable and interoperable”
- “For improved governance and citizen engagement”
- “For inclusive development and innovation” (ODC 2015)

These six principles guide several open data measures created by organizations that belong to the Open Data Charter. Some organizations include Open Data Barometer, Global Open Data Index, Open Data Inventory, Open Useful Reusable Government Data (OURdata), and European Open Data Maturity Assessment. Despite the fact these organizations follow the Open Data Charter, they have created measures that each assess a government’s data openness score using different methods.

The Open Data Charter website itself states that Open Data Barometer (ODB) follows the Charter’s principles most. ODB is produced by the World Wide Web Foundation. Each year ODB produces a global report analysis on the progress and development of data openness by measuring each government’s data openness. Additionally, ODB shows which countries have improved or worsened in a year’s time. On the other hand, Global Open Data Index (GODI 2017c) is funded by an organization called Open Data For Development (OD4D), which is sponsored by the government of Canada, the United Kingdom, and the World Bank (OD4D, n.d.). GODI measures a government’s openness by mostly assessing based on the Charter’s first three principles. Each year they try to improve their methodology on how to measure openness. Other than measuring how open governments are with their data, ODB and GODI research the barriers of open data and create recommendations.

Section 2. Findings from ODB & GODI

In this section we discuss the findings of ODB and GODI. We decided on these two organizations, because the Charter’s website shows how ODB follows each open data principle and GODI incorporates a lot of user engagement in their research. One of the main concerns in open data is that there are often data gaps (GODI 2017c). For instance, Indonesia’s annual total government spending is available online. However, since the data does not include dates of transactions, individual records of transaction, and other standard characteristics required by GODI, GODI assumes that Indonesia’s government spending is not available, therefore creating a data gap. Different datasets for the same category are more granular than others. These data gaps are often caused by the fact that there is a lack of standardization for each data category or that some governments lack the resources to “modernize” their information systems (GODI 2017c).

In addition to these data gaps, governments have the tendency to publish their data in different locations and different formats. Often at times the data is not raw, and has been converted to some graph, which means the data is not modifiable. Since data is often not standardized and there are confusing licenses, this also prevents data from following the open definition (GODI 2017c).

Despite these financial and organizational setbacks, there are existing implementations of alternative solutions for developing countries. *Check my Barangay* is an example of how community members can monitor government budgets in the Philippines. A “barangay” is the smallest administrative division in the Philippines, and the country consists of 42,029 Barangays. Part of the project includes having regular meetings where community members discuss their barangay’s budget

and what they want to improve in their communities. Although a community member may not have access to internet and can access the data portal for the budgeting plans, members are still given SMS notifications. They use mobile technologies as an alternative to the internet (ODB 2017).

ODB creates a large emphasis in their most recent global analysis report that countries need to be more inclusive in their data collection, data design, data access, and data use, by involving marginalized groups in each process, in order to create more transparent and coherent data (ODB 2017). They suggest that governments can include marginalized groups in data collection by investing in disaggregation of data by income level, sex, age, by creating better indices that measure diversity and stratification in society better (ODB 2017). When designing the format in which the data is published, governments need to keep in mind that there are marginalized groups that may not understand the meaning behind the numbers. In addition, when designing how to collect the data, governments must also keep in mind that marginalized groups must also be represented in the future numbers that they publish. Governments need to establish inclusive design processes for policies. They should also invest in low cost and accessible internet access for these marginalized groups. Access and representation in data should not be a privilege, rather it is a right. Finally, we need to ensure that the data publish can actually create a difference and help people. Governments need to invest in processes that enable these marginalized groups to use the data and have a voice in policy-making.

Section 3. The Lack of Inclusivity in Open Data Measurement

Open data measurement indices, such as GODI and ODB, gather their data and measure data openness using their own different methodologies. For example, any person can be part of the GODI community, and does not need to be an expert in the field. GODI uses a snowball sampling approach, which means that they gather participants by having their previous participants suggesting them (GODI 2017a). The reviewers limit themselves to only data publication of national government data. They have a review phase where their members fill in a survey, as well as a dialogue phase where any member can discuss and debate about GODI's results, and during that time GODI may change their results (GODI 2017a). A different approach would be collecting data through expert surveys and existing data from secondary sources, which is what ODB does (ODB 2017). They collect specific data in order to measure the readiness, implementation, and impacts of data openness (ODB 2017).

Marginalized groups are not well represented in these databases, for example, despite ODB's large emphasis on inclusivity for marginalized groups in their latest report, it is not translated into their methodology. They gather data mostly through asking "To what extent ..." questions to experts in the field and these experts answer on a scale from one to ten (ODB 2017). It is reasonable to ask these types of questions to 'experts'. However, there is more room to create a more inclusive space for marginalized groups. We can include marginalized groups in data collection, and include them in assessing data openness, or at least try to. One of the questions included in the expert survey is "was it easy to find information about this dataset?" (ODB 2017). Alternatively, we can try to ask these types of questions to the average person instead of only experts. The question is only worth five points. How easy it is to find data is also as important an open data problem as to whether or not it is open license. If the data is open licensed but difficult to find, then it would be a waste. The easiness of finding the data should be worth more than five points. ODB should have a more diverse group of participants other than only experts.

GODI on the other hand includes a different group of participants. The fact that GODI allows for any person who has access to the internet to become a member of their community, which allows for them to be a part of the Global Open Data Forum and discuss GODI's results, is a step ahead from ODB. However, most people may not know where to look, unless they have an interest in data openness or data science. Nonetheless, if ODB and GODI want to create be more inclusive of all groups, they need to consider people who do not have immediate internet access. The information they gather contains response bias, and people with less privileged circumstances are unable to access the data or realize that they have a right to it.

Another concern with the surveys is their reliability. For example, some of GODI's rankings are inconsistent with western media depiction of those countries. Taiwan ranked first, Colombia ranked fourteenth, while Sweden ranked twenty-first (GODI 2017b). Colombia scored 100% for government

budget, government spending, and election results (GODI 2017b). At the same time, there have been many new articles that have said how the Colombian government has some corruption scandals. Prosecutors say that Colombia's current president accepted bribery for his re-election campaign from Odebrecht, a Brazilian construction company. Media outlets such as the New York Times (Zarate 2017) and Reuters (Murphy 2017) in 2017 published articles on the Odebrecht scandal. Dejusticia, a Colombian human rights organization, reported in 2014 on average 50 trillion pesos are lost to corruption annually (Garavito 2014). On April 1, 2017, 16,000 Colombians marched on the streets of Bogota as a protest against the Odebrecht bribery scandal (Diaz 2017). The news articles and the citizens of Colombia are inconsistent with Colombia's GODI score. However, the reviewers for each survey must provide the link for the published data on the Colombian government website. Since the numbers exist online, it does not make sense for the government to be transparent with their data and at the same time be corrupt. Either the published data is false or the media and public are mistaken about the government's activities.

GODI's method of measuring data openness only measures how available the data is, and does not consider its impacts. The GODI measure of transparency does not reflect whether or not the data published is accurate. ODB creates an emphasis on the impact the availability of the data can give and that governments should publish data that will make citizens trust in the government. If the data is published and creates no positive impact or the citizens do not trust it, then the publication of the data is pointless. That is the case with Colombia.

ODB and GODI included many data categories for their indices. None of the organizations that follow the Charter included data categories pertaining to minority groups. The data needs to be granular. For instance, it is not enough to publish crime statistics data that include annual returns on levels of crime, location of crime, and time of crime. It would be more helpful to include details like the income level of the criminal. Other data that pertains to gender, ethnicity, or religion should also become more transparent. We need to avoid falling into Simpson's Paradox; this happens when there is a trend in different groups, but when we combine the groups together, the trend disappears or inverses (Pearl 2013). The data collected and published needs to be able to help us understand the problems that lie with different minority groups across places, in order for us to tackle those problems in a better and efficient way.

Section 4. The Monopoly of Data

Pressures to make data more open are mostly from outside the government. Research done by the Netherlands Organization for Applied Scientific Research (TNO) in 2011 shows that citizen initiatives and market initiatives are drivers for governments to become more transparent with their data (TNO 2011, 8). We need more user driven innovations as well as NGOs, entrepreneurs, and journalists to pressure governments to develop open data policy.

Since there are governments that sell their data, one could argue that even governments act like private institutions. In Estonia, Denmark, and the United Kingdom, the income of several government organizations are based on the data they sell, which – as a consequence – makes them reluctant to publish the data for free (TNO 2011, 8). In many situations data acts like a commodity; a commodity in which it can be purchased with the privilege of money. The centralization of information occurs due to this privilege. Even in the present, as Hayek said, a lot of the detailed information of a problem is centralized with one group. Since no other people have access to that information, we have not solved these problems or they were too late to solve.

Most of the conversation on data openness of organizations such as GODI and ODB focus more on governments. We need to remember that in existing countries there are many private institutions that collect data on many individuals. Private institutions within each country should have the obligation to share this data, or be transparent with what they are doing with that data. Friedrich A. Hayek claims that a centralized governmental system would not work because information is too dispersed, which makes Hayek create the assumption that a decentralized system, capitalism, would be better. Hayek contradicts himself because in an existing capitalist system, knowledge, information, and data are not decentralized. This is certainly true in the field of science. In the digital morphology field, a lot of research and the

potential for this field to grow has been held back due to the fact that institutions that have the digital data are not sharing it, or the data published does not have sufficient quality for it to be reused or modified as researchers need it to. Some researchers in the field of anthropology, paleontology, and biology have been reusing old data instead. This has prevented them from visualizing and analyzing the new discoveries in three or four dimensions (Davies et al. 2017).

This centralization of knowledge does not only happen in academia and government. Wall Street men, such as Michael Burry and Steve Eisman, who had prior knowledge of the subprime mortgage loans possible default, took advantage of America's economy collapsing and reaped billions from the 2008 financial crash. They were in a position of privilege to make decisions to benefit themselves, meanwhile, 1.2 million Americans were evicted from their homes. The crash affected lives across the world, including factory workers in China, since America is one of China's main importers for goods (Ferguson 2010).

Similar to the Charter's principles, there has been an effort to create open data principles within the field of science by the University of Cambridge and Open Knowledge Foundation, which is called the Panton Principles. When data is published, a statement containing the expectations of how the data will be reused and repurposed by the publishers should also be included. Secondly, the data should have the appropriate copyright waiver or license in order to let other people access and reuse the data. The data should also certainly follow the open data definition. In addition, it is the utmost importance that the data be published in a public domain through the Public Domain Dedication and Licence or Creative Commons Zero Waiver, to follow the Science Commons Protocol for Implementing Open Access Data and the Open Knowledge/Data Definition (Pollock n.d.).

It would be great if a set of open data principles were created in other fields as well, ones that are not only controlled by government, so that whether or not a government has more or less strict government policies, the private institutions must still carry the obligation to be transparent. Private institutions such as insurance companies, tech companies, financial institutions need to create principles in their own fields to foster a consensus on to what extent do they need to share their data in order to be transparent with what their company is doing, and at the same time protect the privacy of their customers and employees.

Conclusions

In this paper, we have highlighted efforts by the Open Data Charter's non-profit organizations to define open data and to rank the data transparency of various governments. Unfortunately, there are flaws with some of the measures in place. We implore all organizations and governments to continue their efforts and to improve upon them because data openness, as discussed in section 3, is an immediate human right for marginalized groups. The open data discussion is still exclusive, as there is still much elitism surrounding access to data. Many of the perspectives that are publicized are one-sided, and come from a privileged place. This response bias prevents different information, that may help innovation, from being shared. At the same time, many information is inhibited, and are exclusive to a group of people, whether it is data from academia, private companies, or government. While the action needed should be expedited, an effective decentralization of data sources is a long-term process.

When measuring the openness of data, we cannot only rely on expert opinions or of people who have the privilege to access the Internet. We need to create a system of collecting information and data that includes marginalized groups, which can include making surveys or access to data free. We can explore existing platforms or create new ones for open data in developing countries, using the existing technologies that they have. There are still ways to make data more open without the Internet. For instance, *Check my Barangay* is using SMS notifications to keep community members updated. Otherwise, developed nations should create updated and granular data platforms that are not dispersed. Further research in the open data problem would be creating technological structures that can ensure that information published follows "the standard", and that everyone has the access to view the data. The answer does not lie in how much government should intervene, as every institution should have the responsibility to be transparent with their data and what they are doing with their data, without

jeopardizing anyone's privacy or security. We need to think of solving the open data problem more as a collaboration. After all, pressures to make data more open come from citizen initiatives.

References

- Davies, Thomas G., et al. 2017. "Open data and digital morphology." *The Royal Society of Publishing*. Accessed July 21, 2018. <http://rspb.royalsocietypublishing.org/content/284/1852/20170194>.
- Diaz, Fabrio Andres. 2017. "Colombians are fed up with corruption, and everyone seems to be under investigation." *The Conversation*. Accessed July 21, 2018. <https://theconversation.com/colombians-are-fed-up-with-corruption-and-everyone-seems-to-be-under-investigation-75173>.
- Ferguson, Charles, director. 2010. *Inside Job*. Sony Pictures Classics.
- Garavito, César Rodríguez. 2014. "Gracias, Juan Ricardo Ortega." *Dejusticia*. Accessed July 21, 2018. <https://www.dejusticia.org/column/gracias-juan-ricardo-ortega/>.
- Global Open Data Index (GODI). 2017a. "Methodology." Accessed July 21, 2018. <https://index.okfn.org/methodology/>.
- Global Open Data Index (GODI). 2017b. "Place overview." Accessed July 21, 2018. <https://index.okfn.org/place/>.
- Global Open Data Index (GODI). 2017c. "Insights." Accessed July 21, 2018. <https://index.okfn.org/insights/>.
- Hayek, F.A. 1945. "The use of knowledge in society." *The American Economic Review* 35 (4): 519-530. Accessed July 21, 2018. <https://www.jstor.org/stable/1809376>.
- Molly, Jennifer C. 2011. "The Open Knowledge Foundation: Open data means better science." *Public Library of Science*. Accessed July 21, 2018. <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001195>.
- Murphy, Helen. 2017. "Colombia president's campaign allegedly took Odebrecht cash -official." *Reuters*. Accessed July 23, 2018. <https://lta.reuters.com/article/marketsNews/idLTAL1N1FS26M>.
- Open Data Barometer. 2017. "ODB Global Report Fourth Edition." *World Wide Web Foundation*. Accessed July 21, 2018. <https://opendatabarometer.org/4thedition/report/>.
- Open Data Charter (ODC). 2015. "Principles." Accessed July 21, 2018. <https://opendatacharter.net/principles/>.
- Open Data for development (OD4D). n.d. "About." Accessed July 21, 2018. <http://od4d.net/about/>.
- Open Knowledge International. n.d. "The Open Definition." *Open Knowledge International*. Accessed July 21, 2018. <https://opendefinition.org/>.
- Pearl, Judea. 2013. "Understanding Simpson's Paradox." University of California, Los Angeles. Accessed July 21, 2018. http://ftp.cs.ucla.edu/pub/stat_ser/r414.pdf.
- Pollock, Rufus, et al. n.d. "Panton Principles." Panton Principles. Accessed July 23, 2018. <https://pantonprinciples.org/>.
- The Netherlands Organisation for Applied Scientific Research (TNO). 2011. "Open data: an international comparison of strategies." *European Journal of ePractice*, no. 12 (March / April): 1-13. <http://unpan1.un.org/intradoc/groups/public/documents/UN-DPADM/UNPAN046727.pdf>.
- Zarate, Andrea. 2017. "Corruption scandal ensnares leaders of Peru and Colombia." *The New York Times*, February 7, 2017. Accessed July 23, 2018. <https://www.nytimes.com/2017/02/07/world/americas/perus-attorney-general-seeks-to-arrest-ex-president-toledo.htm>.