

AidData: The Open Data Movement
By Alex Atkins and Jess Percic

Introduction

In a memorandum for the heads of executive departments and agencies, President Obama states that “Openness will strengthen our democracy and promote efficiency and effectiveness in Government.” The openness to which he refers is the Open Data Movement. This movement, like other “Open” movements, such as open source, is grounded in the belief that certain forms of information should be made available at no cost to members of society and with no restrictions on its use. While many open data initiatives have taken shape, this paper looks specifically at Open Government Data (OGD). We present an overview of the existing literature, discuss results from our OGD analysis, and, finally, offer recommendations to help influence future policymaking. Overall, there seems to be a general consensus for the widespread enactment of OGD initiatives; however, further research is still needed to identify the specific consequences, both positive and negative, of implementing such policies.

Background

The Open Definition lays out principles that define “openness” in relation to data and content. Open simply means “anyone can freely access, use, modify, and share for any purpose” (Open Definition, n.d.). Government data is considered open if it is made public in a way that complies with the 8 Principles of Open Government Data. These principles were compiled by a group of thirty open government advocates gathered in Sebastopol, California on December 7-8, 2007. The 8 Principles of Open Government Data stipulate that data is:

- 1) **Complete.** All public data is made available. Public data is data that is not subject to valid privacy security or privilege limitations.

- 2) **Primary.** Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.
- 3) **Timely.** Data is made available as quickly as necessary to preserve the value of the data.
- 4) **Accessible.** Data is available to the widest range of users for the widest range of purposes.
- 5) **Machine Processable.** Data is reasonably structured to allow automated processing.
- 6) **Non-Discriminatory.** Data is available to anyone, with no requirement of registration.
- 7) **Non-Proprietary.** Data is available in a format over which no entity has exclusive control.
- 8) **License-Free.** Data is not subject to any copyright, patent, trademark or trade secret regulation; however, reasonable privacy, security and privilege restrictions may be allowed (opengovdata.org, n.d.).

These 8 principles are important for establishing a common framework for the execution of OGD initiatives. In the following section, we explore how this looks in practice. Making government data available to the public can improve civic participation and foster business development, but researchers document a number of barriers that have prevented these types of initiatives from taking off.

Drivers & Barriers

In order to better understand the drivers and barriers to open data policy, Huijboom & Van den Broek (Huijboom & Broek, 2011) conducted a survey among policy-makers and experts in five countries: Australia, Denmark, Spain, the United Kingdom, and the United States. Results show the Top 10 drivers include: 1) strategies and experiences in front runner countries, 2)

political leadership, 3) regional initiatives, 4) citizen initiatives, 5) market initiatives, 6) emerging technologies, 7) European legislation, 8) thought leaders, 9) possibility of monitoring government, and 10) budget cuts. The authors observe that while drivers lie predominately outside government, the barriers are within government organizations. The Top 10 barriers consist of: 1) closed government culture, 2) privacy legislation, 3) limited quality of data, 4) limited user-friendliness/info overload, 5) lack of standardization of open data policy, 6) security threats, 7) existing charging models, 8) uncertain economic impact, 9) digital divide, and 10) network overload. Differences between drivers and barriers for open data are seen among Anglo-American countries and other countries.

A common issue with OGD initiatives is the type of data that is made available. The *Open Data Barometer* global report (Davies, Open Data Barometer, 2013) report found that 55% of nations have OGD policies, but only 7% of datasets are machine-readable and under open licenses. This is a significant barrier to OGD initiatives. Even in nations that have lofty OGD goals, it is uncommon for governments to make politically important datasets available, particularly ones that are crucial to holding governments and companies accountable. The research presented in this paper also found that, while much of data from statistical authorities is online, it is often released only in aggregated forms and with unclear or restrictive licenses. Additionally, government support for innovation is limited; there seems to be an emphasis on releasing data, but little support for the reuse of data.

Zuiderwijk et al. (Zuiderwijk, Janssen, Choenni, Meijer, & Alibaks, 2012) examine the socio-technical impediments of open data. While there are potentially many benefits to gain from making data publically available, opening and using the data presents numerous impediments which can have both a socio and technical nature. Drawing on the existing literature, the authors

identify 118 impediments, which they analyzed and grouped into ten categories: 1) availability and access, 2) find ability, 3) usability, 4) understand ability, 5) quality, 6) linking and combining data, 7) comparability and compatibility, 8) metadata, 9) interaction with the data provider, and 10) opening and uploading. Their analysis shows that little attention is paid to the user perspective, even though users are responsible for generating value from the data. The authors speculate that impediments will continue to evolve over time. As new policies and infrastructures resolve the current impediments, new ones will likely arise.

Janssen et al. (Janssen, Charalabidis, & Zuiderwijk, 2012) take a closer look at the benefits and adoption barriers to open data. They note that a gap exists between the consumer's expected benefits and barriers to usage. Moreover, findings reveal that users adhere to a conceptually simplistic view when it comes to open data; there is a tendency to associate publication of data with uses and benefits. For this reason, Janssen et al. formulate five myths reflecting the gap between the benefits and barriers of open data: 1) the publicizing of data will automatically yield benefits; 2) all information should be unrestrictedly publicized; 3) it is a matter of simply publishing public data; 4) every constituent can make use of open data; and 5) open data will result in open government. Myths (1) and (2) are intended to convince data providers to make their data publically available, while ignoring barriers and the heterogeneous nature of open data. Myths (3) and (4) illustrate how the user view is often neglected and that using open data may be easier said than done. The final myth suggests the transformative nature of open data is more difficult to achieve than otherwise thought. Taken together, these five myths are said to form the basis of policy-making.

Motivation

In their cross-country comparison of strategies, Huijboom & Van den Broek (Huijboom & Broek, 2011) discover three primary reasons for publishing government data: to increase democratic control and political participation, to foster service and product innovation, and to strengthen law enforcement. Many of the countries studied maintain that publishing government data can empower citizens to exercise their democratic rights. A number of governments highlight the opportunities for innovation spurred by open data. Still, some view open government data as a means to involve citizens in and strengthen policing and law enforcement.

Why should the government and/or private organizations be interested in making data available and usable? From the government's perspective, making data available could benefit efficiency in decision making, but governments often have limited resources for data analysis. The private sector, on the other hand, has an interest in monetizing on data, which the paper refers to as commercializing government data. Making data available to the public is also part of a transparency initiative on behalf of governments that is often called for by individuals. Due to this interest in transparency, we could potentially see an increased role of private sector and civil society advocating for the widespread publication of government data for focused analysis and commercialization. Not all governments want or have an interest in transparency; however, the private sector and NGOs can place pressure on governments by collecting data through news articles, crowdsourcing, and physical copies of publicly available data. Subsequently, they can aggregate this data and make it available to the public. The Aiddata initiative for Tracking Chinese Development Finance is a perfect example (www.china.aiddata.org).

Maeda & Torres (Maeda & Torres, 2012) pose an interesting question regarding the importance of open data, with a particular focus on environmental data. However, the questions

and answers presented in this paper are highly relevant to the open data initiative as a whole. A main driving force behind open data, as identified in this paper, is that data being generated with public money should be used in a way that results in the most benefits to the public. This belief emphasizes that access to data should not be limited, so that there is a greater possibility of applications and uses for data (Maeda & Torres, 2012). What do research institutions from developing countries have to gain from open data policies? In fact, open data could negatively impact those countries because they have less access to state of the art technology used to process the data. Therefore, institutions from developing countries may simply provide data, while wealthy institutions use their funds and the data to generate the new research and scientific findings. The strategy for overcoming this problem, according to the paper, is cooperative networks between institutions from different nations (Maeda & Torres, 2012).

Stakeholders

Researchers agree there are three primary stakeholders of OGD initiatives: private institutions, citizens, and public administrations. The main value they receive is improving government accountability, transparency, responsiveness and democratic control. It can promote citizen self-empowerment, social participation, and engagement. Overall it fosters innovation, efficiency and effectiveness in government services (Ubaldi, 2013).

In a typical OGD initiative, the public administration is the primary data producer, collecting the initial data and structuring its release. After its release, the government is able to acquire feedback and input from its users about ways to improve the data. One of the major fears about OGD initiatives is that they will open the government to criticism. However, this doesn't tend to be an issue in practice. The data may be analyzed by stakeholders to garner new insights

that can be valuable to a public administration. OGD initiatives also allow other public agencies to view data that was not easily accessible to them before. This can facilitate dialogue and collaboration between agencies.

In this case, private institutions include research institutions, non-governmental organizations, and private companies. These institutions are adept at collecting OGD and modifying it to produce increased socio-economic value. This can be done through research, policy recommendations, or creating specialized data sets.

The average citizen or end-user of OGD portals benefit from the knowledge produced by these initiatives. This enhanced transparency allows users to garner new insights as well as provide feedback to the government. It creates a dialogue between government and citizens that has a great deal of potential for improved services and accountability.

However, the roles of various stakeholders in this process are not static. While the government will likely continue to be the primary funder and maintainer of OGD since they view it as a “commodity to generate revenue or to cover the cost of data production”, they may not design or deliver the services themselves (Ubaldi, 2013). The most interesting trend is that roles which previously fell solely under the purview of the government are now being played by a variety of different stakeholders: professional, politician, practitioner, civil servant, expert, consumer and citizen (Ubaldi, 2013). It is important for governments to realize this shift is occurring and rethink how it approaches OGD. Primarily, a data centric approach to OGD where the government is only concerned with releasing more and more data is no longer sufficient and can create a bottleneck in innovation, sustainability, and value. Ubaldi recommends that governments create publicly accessible infrastructure that allows users access to raw data. It must focus on improving the most relevant datasets to stakeholders that are identified through public

consultation. Finally, governments should place an emphasis on data releases in accessible formats that encourage re-use (Ubaldi, 2013).

Private stakeholders tend to specialize in collecting government data and delivering it to citizens in a format that will create additional value for them and increases re-use. In addition to repackaging data, they “constantly create and reshape the tools that individuals use to find and leverage public data” (Ubaldi, 2013). According to the literature, it is important for governments to put forth a strong effort to encourage the private party utilization and innovation of public data. Governments could take an additional step by encouraging a minimum standard for the publication of data in order to guarantee quality or effective use (Ubaldi, 2013).

Key Players

The *Open Data Barometer* global report (Davies, Open Data Barometer, 2013) provides information about OGD practices and a ranking of countries. The countries that are implementing OGD initiatives tend to be developed countries working towards building national data infrastructures, but also far from embedding OGD practices across government. The highest-ranking countries are the United Kingdom (1), the United States (2), and Sweden (3). Mid-ranking countries have made datasets available, but do not make key datasets available; examples of key data sets include Land Registries and Company Registries because they can be used to hold governments and companies accountable. Low ranking countries typically do not engage in open data. The nations that rank lowest on OGD initiatives are developing nations because they tend to lack the infrastructure and resources to make digitized government data available. If any data is available, it’s likely scanned documents or PDFs with little or no

organization. The burden falls on private companies to collect this data and format it to be usable.

Realizing the Vision of Open Government Data (Gigler, Custer, & Rahmetulla, 2011) focuses on how governments in developing countries do not always have the necessary resources to collect and make available their data. It identifies democratic and developed nations as the most likely candidates to have Open Government Data initiatives; however, governments face struggles implementing OGD, mainly with poor ICT infrastructure, lack of resources, and limited demand from civil society or within the government itself. These issues are overcome by the government focusing on publishing data and partnering or collaborating with the private sector, often development banks, to release the data set in a useable format. As part of a controversial practice, donors and development banks can influence developing countries to publish their data by placing restrictions on development finance.

Case Studies

In the United Kingdom, transparency centers on Data.gov.uk, the public data catalogue that points to thousands of datasets downloadable under permissive open government license. The datasets are often in comma-separated value (CSV) format or spreadsheets, but there is potential for improving their utility by linking them using structured machine-processable formats. Resource Description Framework (RDF) is the format most integrated into current thinking about future generations of the Web, as its use of Uniform Resource Identifiers (URIs) allows data to be identified by reference and linked with other relevant data by subject, predicate, or object. The use of Semantic Web standards in open government data was pioneered by Advance Knowledge Technology; this vision is referred to here as the linked-data Web (LDW).

Shadbolt et al. (2012) are interested in learning how to build or reuse ontologies easily and appropriately for particular application. Second, they want to see query methods that scale across the unbounded Web, not just within small islands of well-structured data. Third, the authors address the need for visualization and browsing tools, and finally, the need to populate the LDW to increase the network effects of large-scale linking.

OGD will play an important role in LDW: its quantity will help deliver the network effects expected from the LDW, its origin is clearer than that of many other types of data, and it is often seen as high quality, trustworthy, and neutral. Shadbolt et al. use the EnAKTing¹ approach to develop an integrated account of how to bring OGD into the LDW, focusing on the populations of the LDW with OGD from Data.gov.uk to analyze the stated issues. Their research elicits a number of lessons for government, the technical community, and citizens. We'll start by going through the lessons for government. First, although hard-to-link formats are a problem, the regulatory setting of reusability is crucial. Data requires an open license to be counted as open data. Second, few systems other than Data.gov, Data.gov.uk, and Eurostat provide direct access to catalogues in raw formats; usually, they need to be discovered by HTML search. National PSI portals, where available, provide powerful and rich metadata, though regional data is less well described. Third, metadata standards tend to be tailored to the information needs of the primary intended users of the information assets. These metadata standards should be more widely adopted, and in the context of LDW they need to be directed toward all data users, rather than just the immediate communities who generated the data. Fourth, there are often temporal or geographical gaps between datasets that limit the usefulness of linking them, or even methodological gaps that render the data incommensurable. Versioning on the LDW is a vital

¹ EnAKTing project (www.enakt.org), funded by the UK's engineering and Physical Sciences Research Council (EPSRC)

topic. Pan-national interpretation and comparison is particularly challenging. Where URIs are produced by governments, they and others should seek to reuse them as a way of becoming the essential join points of the OGD digital infrastructure.

Speaking to the technical community, Shadbolt et al. heed the following advice. There are many steps in the creation of linked data from OGD that would benefit from better tools and, where possible, automation. The perennial user interface is critical to OGD's value. If few people are able to access and analyze the data, then OGD's contribution to democratic debate will be reduced accordingly. Second, further research is needed to identify join points that let the data consumer assemble a coherent picture across datasets, and that facilitate data normalization to detect and correct variations in nomenclature. Examples include geography, time and origin. Knowing the join points can help governments to better prioritize resources. Third, the EnAKTing work is deliberately lightweight and pragmatic in the way it relies on well-known ontologies. The authors use Scovo for statistical data because of its relative simplicity and widespread adoption. This design decision is practical and should be made on a case-by-case basis. Fourth, co-reference resolution remains a serious problem. Shadbolt et al. adopt the *SameAs* context-relative approach, which suits a data- and citizen-centric approach. Lastly, they argue that migrating OGD to the LDW should be accompanied by quick development of websites supporting data-centric interactions. For citizens, transparency will create opportunities to hold their governments accountable, and to develop and use innovative services. The more citizens take advantage of these new opportunities, the more likely it is that OGD will continue to be released (Shadbolt, et al., 2012).

Whitmore (2014) explores the use of open U.S. Department of Defense contracting data to identify patterns of spending activity that can be used to predict future military engagement. The

idea behind this theory is that governments must spend large sums of money mobilizing and equipping soldiers in preparation for deployment. So, an understanding of government spending patterns might yield insight into future military conflict. Over the past decade, the United States has been involved in a series of military conflicts including the wars in Afghanistan and Iraq. The process of preparing for and engaging in war requires the U.S. Department of Defense to pay billions of dollars to private defense contracting companies for the procurement of goods and services. Given the extent to which the U.S. D.O.D. utilizes military contractors, it seems reasonable to consider the contract granting activity of U.S. D.O.D agencies as a proxy measure for classified operational activities. If there is a correlation between contracting activity and classified operational activities, then large increases in military contracts might be a leading signal for future military conflict. The author uses defense contracting data found on the [usaspending.gov](https://www.usaspending.gov) data portal to explore this relationship.

The research taken in this article follows a two-stage approach. The first stage involved the exploration of the research question in the context of a specific case—namely, the U.S. invasion of Iraq in 2003. This portion of the analysis showed that the use of U.S. D.O.D. contracting data to predict future war has promise; however, a number of problems with the data and online portal prevented conclusive, generalizable results. The second stage classified the data and systems problems encountered when performing the analysis according to an established analytical framework for open data barriers along with specific examples of how these barriers manifest in the research. In particular, two categories of barriers presented significant impediments to the use of the data for research purposes: task complexity and data quality.

Specific instances of task complexity barriers in the [usaspending.gov](https://www.usaspending.gov) portal included massive datasets available in difficult to-use formats, interactive portal features that hid much of

the data available in the download, a lack of information about the connection of the downloaded data to its sources, and data that changed over time. The information quality issues that were encountered include missing values (especially foreign keys), inconsistent data which suggest problems with the source database structure, and a very limited time range for which the data were available. It is important to note limitations of the research. This article addresses several open data barriers derived from the research framework article but did not address others. Whitmore takes this approach since his interaction with usaspending.gov portal was as a user, rather than a provider of the data. Therefore, this article does not address the supply-side barriers that the government agency managing the usaspending.gov portal may face and, as a result, the comprehensiveness of the barrier analysis for the usaspending.gov data and portal is limited.

Another limitation of the paper is its reliance on a single case study, the 2003 invasion of Iraq. Given the limited time frame for which data were available on the usaspending.gov portal, a more comprehensive examination of many cases was not feasible. The use of only a single case, however, made it impossible to validate or generalize the findings in a way that would meet common social science research standards. Nevertheless, Whitmore's findings provide further empirical, case-based evidence in support of open data barrier frameworks proposed in the literature. Furthermore, this study extends the literature by identifying specific examples of the manifestation of these barriers encountered when working with the usaspending.gov data portal and demonstrating some of the specific challenges to research validity and generalizability presented by these open data barriers (Whitmore, 2014).

Frameworks

A great deal of the literature on OGD initiatives seeks to develop frameworks to conceptualize trends and to address common problems. A framework defines dimensions that describe the relationships between the government and stakeholders. It suggests potential implementations that strengthen these relationships and increase the value generated from the initiative. While some frameworks attempt to address ongoing OGD initiatives, others lay out a framework for establishing new initiatives in developing or developed countries.

The data driven approach to OGD places an emphasis on publishing as much raw data as possible. This approach is not necessarily concerned with the quality or usability of the data, but merely making it accessible to the user. It often requires the least capacity to implement and, therefore, is common amongst developing countries. However, this does not guarantee the data will be used or even that it can be used. In fact, releasing large volumes of data “without background on why and how it is collected, how it is organized, and its intended use, leaves citizens with herculean tasks of determining its relevance and reliability” (Evans & Campos, 2013).

While this framework provides an important first step towards government transparency, it holds little benefit to the average user and requires a large degree of investment from private institutions to create a useable product. These initiatives are less likely to create lasting relationships between the government and its stakeholders. As a result, the initiative fails to produce valuable results within the government. One of the greatest benefits of OGD is that it motivates better data collection practices and record keeping within the government agency. Due to lack of resources, this is the approach most commonly employed by developing countries (Davies, Open Data in Developing Countries - Emerging Insights from Phase 1, 2014).

A second approach discussed in the OGD literature is focusing on the overall quality of data. Governments should determine potentially high impact datasets to release instead of trying to put out as many datasets as possible. Then, the government can ensure that those datasets meet a strict quality standard and can be iterated upon in the future.

A third approach is to focus on user involvement. This approach acknowledges the fact that citizens are a primary stakeholder. If the data does not attract users, then it is missing one of the primary benefits from the OGD movement. It also seeks to maximize sustainability by ensuring data has high reusability through increased collaboration with all stakeholders. By working with stakeholders, governments can continuously improve datasets based off of their feedback. This creates a sense of community and investment by stakeholders as well. This includes collaboration with other agencies, which is more likely to yield modular datasets that can be easily merged with external datasets. It is this approach that most literature promotes moving forward and is the most crucial for successful OGD initiatives in the future.

Ubaldi discusses an incremental model for making data completely open that can be generalized to developed and developing countries: 1) Determining what government data already exists; 2) Increasing data transparency; 3) Improving open participation; 4) Enhancing open collaboration; and 5) Releasing ubiquitous engagement. However, this model does not inherently address whether the data will be useful or even used (Ubaldi, 2013). In order to ensure value creation in data, Ubaldi discusses a user centric model. Key tasks include “identifying high-value, high-impact data for the public; improving and assuring data quality in terms of accuracy, consistency, and timeliness; fostering data demand and use by the various actors” (Ubaldi, 2013). Data that meets these standards include: “public data (e.g. trade data, fiscal data, health data, education data, transport data, weather data, census data, map data/geographic data,

crime data); micro-statistic (data the state needs to function itself); performance data on government programs and spending (data on mortality for a doctor or in public hospitals); personal data of public services' users; and public cultural information" (Ubaldi, 2013).

Most frameworks fall under the later approach and are most relevant to discuss today. One framework focuses on government accountability and provides valuable context for the relationships between stakeholders. The ideas presented in this framework can be found across the literature. This framework focuses on the potential of OGD to improve accountability within the government. It applies Koppell's accountability framework for the public sector to the specialized case of e-government. The authors identify several key accountability dimensions: transparency, liability, controllability, responsibility, and responsiveness. Transparency is determined by the availability of information regarding the governmental organization to outside stakeholders. Liability of the agency is determined by the ability to hold the organization and/or individuals accountable for their actions and providing positive or negative reinforcement depending on the situation. Controllability is determined by the influence that an external stakeholder has on the organization. Responsibility is determined by the formal rules, regulations, and moral standards that guide the organization's actions. Responsiveness is determined by how efficiently and effectively an organization meets the demands or requirements of its stakeholders (Aman, Al-Shbail, & Mohammed, 2013).

The framework suggests that six different elements must be considered for a successful OGD implementation: leadership support, organization size, internal audit, collaboration, IT strategic planning, systems integration (Aman, Al-Shbail, & Mohammed, 2013). Leadership support is crucial in tactical planning, minimizing risk, and resistance. It plays a key role in encouraging transparency and responsiveness in implementation. An organization's size

determines the capacity of the organization to meet the needs of stakeholders in volume and quality. An effective OGD initiative tends to invite increased interaction with stakeholders, requiring a more customer-centric approach moving forward. While this can improve the overall efficiency and effectiveness of the data, it is important that the organization has the resources available to meet these demands (Aman, Al-Shbail, & Mohammed, 2013).

Internal audits are important factors for increasing liability, responsibility, responsiveness, and transparency because they place oversight on the organization. These audits are useful for determining the strengths and weaknesses of existing implementations that may not be apparent to stakeholders and addressing them moving forward. OGD initiatives that aim to increase collaboration and data sharing have several advantages. They reduce redundancies across organizations, decrease barriers to data integration, and identify potential problem areas earlier. More generally, it tends to benefit performance, responsiveness, transparency, controllability and reduces costs. Information technology strategic planning is important for identifying responsibility and liability for specific services and information. It preemptively addresses many potential problem areas, allowing increased responsiveness and feedback moving forward. Systems integration allows data that was limited to a single scope to be analyzed in new context and by a wider audience. This inherently improves data sharing and transparency as well as the overall quality of service. It also increases the involvement of stakeholders and fosters a higher degree of responsiveness (Aman, Al-Shbail, & Mohammed, 2013).

Hanna identifies three key faults in OGD initiatives. Primarily, they fail to meet the needs of multiple key groups—particularly, businesses and civil society. According to Hanna, they are not responsive to development officials and public administration reformers who contextualize,

use, and supply content to the online portal. Additionally, they are not responsive to citizens and grassroots organizations, of which constitute the final consumer of information and the key to providing feedback for improvements of the service (Hanna, 2010).

In order to overcome these issues, any OGD initiative must collaborate with business and civil society organizations. They should try to utilize, complement, and improve technology and data produced by related efforts of stakeholders or other governments. This will increase the overall capacity of the public to use these services. These initiatives must be integrated with formal governmental efforts to improve operations. They must be responsive to citizens and grassroots organizations, utilizing their criticisms and preferences in order to improve the service (Hanna, 2010). More broadly, Hanna claims that OGD initiatives are likely to succeed if their primary mission is to achieve functional governmental objectives, such as public administration reform through streamlined data collection processes or economic reform requiring high quality data. As a side effect, we see increased technology transfer and transparency. OGD initiatives also have the benefit that these objectives can be furthered by citizen participation through social networks (crowdsourced data) or open source software that lowers costs and provides additional innovation (Hanna, 2010).

According to Ubaldi, it is important for governments to recognize the value of data that is created or shared by citizens. Crowdsourcing data can “enable real time data and information sharing, and to engage relevant actors outside public organizations to use them to create value” (Ubaldi, 2013). Therefore, it will be important for governments promoting OGD to “act as an open participation and collaboration platform, both online and offline, allowing stakeholders to contextualize data provision and, thus, increase the potential for value creation” (Ubaldi, 2013). This is made possible by providing resources or service, encouraging collaboration with different

stakeholders, standardizing and modularizing data across agencies so that data can be easily combined and re-used, and crowdsourcing feedback on content, services, and policies (Ubaldi, 2013).

Evans & Campos discuss an analytical framework for informing citizens. The authors point out that, as OGD initiatives have progressed, different agencies have developed different approaches to publishing data even though they should be integrating OGD principles “into their operations, procedures, and government structures” (Evans & Campos, 2013). It seems few agencies have considered the purpose of OGD and make little effort to answer questions about “why did agencies collect these data, how are the data used within the agency, of what relevance are these data to policy deliberations, of what relevance are the data to informing citizens about public policies, how are policymakers going to apply this information, and are the limitations of these data easily understood?” (Evans & Campos, 2013). It is, therefore, important to develop an analytical framework that will “result in rigorous selection of relevant content, a clear revelation of the content’s purpose and a logical framework in which the public user can navigate” (Evans & Campos, 2013). The author proposes leveraging “the skill sets of policy analysts, which include assessment of data and evaluation of policy options, may be particularly well-suited to formulating contextual frameworks for open government data initiatives” (Evans & Campos, 2013). In other words, agencies could contextualize data and information using basic elements of policy analysis (Evans & Campos, 2013). While this may not be applicable to all data or issues, it would make it easier for the general public to understand complex problems in the same way that the administration looks at them. In the absence of this additional context, it is difficult for people to draw meaning from basic statistics or line items (Evans & Campos, 2013)

Einav and Levin discuss OGD and data in general in the context of its application for economic analysis. As a result, the paper has a very innovative, user centric view of how government data should be presented and used, as well as various applications moving forward. Primarily, they note that even the United States underutilizes the data it collects. While some data is made available to the public, not all data is open and some is even restricted by law. The paper notes the need for sophisticated security measures, data quality assessments, and review before data is made public. Nevertheless, simply prohibiting its use and the potential for the development of the processes is not productive (Einav & Levin, 2014).

The paper also notes that major data sets are typically maintained separately in the U.S. Other developed countries have more advanced integration programs that allow researchers to easily “merge individual demographic, employment, and in some cases health data, for the entire population” (Einav & Levin, 2014). According to the authors, administrative data, in particular, is very valuable, as it “covers individuals or entities over time, creating a panel structure, and data quality is high” (Einav & Levin, 2014). Governments also have the capability to monitor economic activity. However, the Bureau of Labor Statistics still measure price inflation by sending out surveyors to stores to collect data manually (Einav & Levin, 2014). This suggests that, even in developed countries, methods of data collection are still antiquated and behind advances found in the private sector. Governments could instate policies and standards for manual collection of massive volumes of economic data, through integration with private sector sales systems or use of private sector tools that meet governmental standards (Einav & Levin, 2014). Furthermore, governments tend to be behind the private sector in utilizing their own data. If governments placed higher standards on the way agencies utilized data, agencies would have

to place higher standards on the way data is collected and level of integration required to do analysis (Einav & Levin, 2014).

The article also poses interesting questions about using government data and predictive models to inform or target government services at individuals the same way a private sector company might. For example, using the Medicare score to inform response to treatment and coverage or, alternately, targeting tax rebates to different demographics based on their marginal propensity to consume (Einav & Levin, 2014).

Haak advocates for a distributed data infrastructure because data sets are too large and diverse with security and licensing concerns that make a single solution impossible. However, creating “unified standards for exchanging data could enable a Web-based distributed network, combining local and cloud storage and providing public-access data and tools, private workspace ‘sandboxes’, and versions of data to support parallel analysis” (Haak, et al., 2012). This will allow various organizations to have control over different data and allow resources to be accessible and shared in a centralized location. This solution would require increased collaboration among governments and private sector organizations. While there are examples of this infrastructure (Common European Research Information Format), “these models need to be expanded to be international and interdisciplinary” (Haak, et al., 2012). This standard should be developed by a committee of major data providers with cost-benefit analysis in mind (governments, standards organizations, private vendors, etc.) (Haak, et al., 2012). These standards should look to incorporate and understand user incentives and the value of data sharing. They should cover grant and reporting systems, reporting activities, citations standards, non-publication research outputs, defining and maintaining data interoperability standards, and technology standards to name a few (Haak, et al., 2012). It should also incorporate a model for

secure data access, distinguishing levels of sensitivity. Since privacy laws vary across countries, linking data could cause levels to change or be different in different locations necessitating a system to manage this process. The authors recommend a model that hosts both types of data, but restricts access to data based on permissions (Haak, et al., 2012).

Ubaldi also discusses data reuse and adhering to open data standards in general. He points to an ecosystem model where the government identifies categories of actors and adopting policies that allow the establishment of OGD initiatives around relevant issues. He identifies three ecosystem categories: ecosystem of data producers, intermediaries, and users (Ubaldi, 2013). Ubaldi further discusses the role of Information Communications Technology (ICT) in producing OGD portals. This infrastructure allows OGD to be hosted and published. However, the author notes that a common problem is that governments place more emphasis on creating a national portal than developing the technical infrastructure to create and release reusable and high quality data. The advantage of this focus is increased publicity and access for the average citizen. Still, it is difficult for the government to produce a portal that meets the rapidly evolving standards of the private sector. For this reason, the private sector is more adept at handling the development of the portal (Ubaldi, 2013).

Not all of this information is applicable to initiatives in developing countries, because they lack the capacity to implement these changes. Without strategic planning, developing countries tend to follow a data driven approach. It also becomes difficult to publish usable data because records may not be digitized. However, OGD initiatives are important because they necessitate modern record keeping and transparency. This creates a venue for forming new relationships and insights that are crucial to improving administrative processes (Davies, Open Data in Developing Countries - Emerging Insights from Phase 1, 2014).

Another common issue in developing countries is that officials fear the open data process. This is due to concerns about criticism both externally or internally, loss of power, or simply a misunderstanding of how open data works. Efforts to assuage these concerns are crucial to lowering the barriers to OGD in developing countries. Conversely, concerns in many developed countries are not present in developing countries, such as licensing and privacy issues (Davies, *Open Data in Developing Countries - Emerging Insights from Phase 1*, 2014).

Overall, the literature on developing countries is lacking compared to developed countries. However, the Open Data Barometer dataset provides presents ample opportunity for quantitative analysis on various factors. These include quality assessment, impact areas, and readiness of stakeholders. For example, the data shows that the readiness of entrepreneurs and businesses to utilize OGD is typically lacking in developing countries compared to the readiness of government and/or citizens (Davies, *Open Data Barometer*, 2013). Future research could determine why private institutions are less responsive to demand for OGD initiatives in developing countries and how this gap could be filled.

Policy Recommendations

The literature on OGD initiatives tends to be fairly consistent. Newer papers advocate for user-centric data, collaboration with private institutions, and government-wide initiatives that seek to integrate data across agencies. Most of the literature refers to mature OGD initiatives in developed countries that still require significant effort moving forward. Future research should apply existing frameworks to developing countries. The literature would also benefit from in-depth analysis of case study literature for developing countries across multiple sectors.

Although the barriers to OGD initiatives are known, the literature does not contain ample on-the-ground tactics to overcoming these barriers. The most promising avenue is increased involvement of the private sector in producing technical assistance and services to lower these barriers. Therefore, it is crucial that these organizations develop processes that are designed to overcome these barriers and produce successful OGD initiatives.

Private institutions can synthesize experience and data to determine high impact areas and develop best practices, which could be implemented as a tailored product or service. For example, an institution that provides ICT and software services to OGD initiatives could institute its own best practices that the country government must adhere to. This might include requiring that non-sensitive data entered into a resource management software will automatically be published to an OGD portal after a designated amount of time. Once the agency agrees to this, they are obligated to put forth enough resources toward providing quality data in order to avoid potential repercussions. This instills a sense of urgency and liability that would otherwise not exist under an OGD initiative where the publication date was determined by the agency. This does not mean the data would forgo quality assessment. In fact, it is more likely to go through quality assessment under these circumstances because the publication of the data can no longer be shirked. This avoids a common problem with responsiveness to user feedback and publication that frequents small or underfunded agencies with low capacity.

The best way to gain on-the-ground insights about OGD initiatives, especially in developing countries, is through surveys or interviews, media analysis, and active engagement via stakeholder workshops and data use. AidData is in a unique position to further research in all three of these categories. AidData's institutional ties with Development Gateway, Reform Incentives, and TRIP give it unparalleled access to specialized survey software, registries for

political elites and academics, and contacts in developing countries. AidData should conduct surveys and interviews in these countries to gain new insights about the impact of the open data movement in developing nations, and to assess new areas for targeting its own services and data.

AidData can utilize its team of interns to conduct research on media analysis and capitalize on its institutionalized workshops, hackathons, and research to gain more insights into data produced by OGD initiatives and lessons learned. AidData can also promote standardized data collection methodology through the development of a ranking system for development finance based on quality, adherence to standards, and ease of integration from data producers.

Conclusion

The effects of open data policy remain uncertain. Studies are lacking at both the micro and macro levels. Even less evidence is available on the social and democratic effects of open data policy. While open data is an excellent opportunity to generate new research, there are significant barriers to developing nations that must be closely examined. In order to eliminate what could become an unfair system of data generators and data users, the scientific community should work towards rewarding data sharing initiatives. The government has utilized public competitions and hackathons, promoting and fostering a community of users and contributors, to encourage greater involvement and use of published data. But the government and the private sector still face challenges encouraging civil society to utilize the data and push research forward. There are similar issues with crowdsourcing efforts to collect data that governments do not make available or do not have.

Realizing the Vision of Open Government Data (Gigler, Custer, & Rahmetulla, 2011) suggests ways that governments can lead more successful OGD initiatives. Intermediaries, it

postulates, are necessary to take government data and turn it into platforms and products with social and economic value. Moreover, e-users must be equipped to access and work with the data in different ways. It is widely held that a more open government leads to a more transparent government. But openness and transparency do not always correlate. If government fails to publish data in a usable format (i.e. it is not machine-readable), then we are no better informed—and, hence, government is no more transparent—despite being “open.” The success of the Open Data Movement depends on the *quality* of data being published, not simply the publishing of data.

Bibliography

- (n.d.). Retrieved from opengovdata.org: <http://opengovdata.org/>
- Aman, A., Al-Shbail, T. A., & Mohammed, Z. (2013). Enhancing public organizations accountability. *International Journal of Conceptions on Management and Social Sciences*, 15-21.
- Davies, T. (2013). Open Data Barometer.
- Davies, T. (2014). *Open Data in Developing Countries - Emerging Insights from Phase 1*. Berlin, Germany: The World Wide Web Foundation.
- Einav, L., & Levin, J. (2014). The Data Revolution and Economic Analysis. *Innovation Policy and the Economy*, 1-24.
- Evans, A. M., & Campos, A. (2013). Open Government Initiatives: Challenges of Citizen Participation. *Journal of Policy Analysis and Management*, 172-203.
- Gigler, S., Custer, S., & Rahmetulla, H. (2011). Realizing the Vision of Open Government Data - Opportunities, Challenges, and Pitfalls. *Open Development Technology Alliance, The World Bank*.
- Haak, L. L., Baker, D., Ginther, D. K., Gordon, G. J., Probus, M. A., Kannankutty, N., & Weinberg, B. A. (2012). Standards and Infrastructure for Innovation Data Exchange. *Science*, 196-197.
- Hanna, N. K. (2010). *Transforming Government and Building the information Society: Challenges and Opportunities for the Developing World*. New York: Springer.
- Huijboom, N., & Broek, T. V. (2011). Open Data: An International Comparison of Strategies. *European Journal of ePractice*.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 258-268.
- Maeda, E. E., & Torres, J. A. (2012). Open Environmental Data in Developing Countries: Who Benefits? *National Center for Biotechnology Information*, 410-412.
- Open Definition*. (n.d.). Retrieved from <http://opendefinition.org/>
- Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. (2012). Linked Open Government Data: Lessons from Data.gov.uk. *IEEE Intelligent Systems*.
- Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance*, 1-61.
- Whitmore, A. (2014). Using Open Government Data to Predict War: A Case Study of Data and Systems Challenges. *Government Information Quarterly*, 622-630.
- Yu, H. M.-T. (2012). Designing Software to Shape Open Government Policy.
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-Technical Impediments of Open Data. *Electronic Journal of e-Government*, 156-172.

AidData: The Open Data Movement

By: Alex Atkins & Jessica Percic

OGD Stats

- ▶ According to the Open Data Barometer global report
 - ▶ Surveyed 77 countries
 - ▶ 55% of nations have OGD policies
 - ▶ 45% of countries with OGD policies saw no impact
 - ▶ Only 7% of datasets are machine-readable and under open licenses
 - ▶ Highest-ranking countries are the United Kingdom (1), the United States (2), and Sweden (3)

Drivers

- ▶ Success stories
- ▶ Influential political leadership
- ▶ Local / agency-wide initiatives spur national open data policy
- ▶ Citizen initiatives - best practices of user-driven innovations based on government data
- ▶ Market initiatives - NGOs, entrepreneurs, and journalists pressure governments to open up (i.e. China)
- ▶ Emerging technologies - technological trends allow for engagement and innovation based on government data
- ▶ Legislation - laws, such as the EU PSI Directive
- ▶ Thought leaders - experts and community leaders play an important role in putting open data on the political agenda
- ▶ Accountability - the urge to keep a check on government
- ▶ Budget Cuts - government savings were an incentive to publish data on public expenditures and involve citizens in determining where to make cuts

Barriers

- ▶ **Government Capacity**
- ▶ **Political Commitment**
- ▶ **Civil Society Freedoms**
- ▶ **Civil Engagement**

Impediments to Use

- ▶ Availability and access
- ▶ Findability
- ▶ Usability
- ▶ Understandability
- ▶ Comparability and compatibility
- ▶ Quality
- ▶ Linking and combining data
- ▶ Metadata
- ▶ Interaction with the data provider
- ▶ Opening and uploading

Phases of OGD Initiative

▶ Developed Nations

- ▶ Data collection initiatives for internal use already exist
- ▶ Launch OGD Initiative
- ▶ Publish select datasets that hold some value for user base
- ▶ Private sector collects and refines datasets
- ▶ Governments improve data collection and quality

▶ Developing Nations

- ▶ Data only exists in raw human readable formats
- ▶ Launch OGD Initiative
- ▶ Publish raw datasets that are available
- ▶ Private sector collects and refines datasets

Moving forward

- ▶ Private sector specialization
- ▶ Multinational cooperative networks
- ▶ Improve government data collection resources
- ▶ Increase citizen participation

Private Sector Partnerships

- ▶ Pre-packaged online portals for easily publishing raw government data
- ▶ Collecting, cleaning, and aggregating disparate datasets and producing human and machine readable data that holds economic and social value.
- ▶ Produce intuitive websites, visualizations, and simple analytical tools to make their data more accessible.

Citizen Involvement (Challenge.gov)

▶ Goals

- ▶ Improved government awareness of social problems
- ▶ More effective practices based on broad citizen experience
- ▶ Increased transparency and trust between government and citizens

▶ Issues

- ▶ Clearly defining a challenge and evaluating the best incentive strategy
- ▶ Designing clear goals for challenges in order to elicit the desired response
- ▶ Authoring a clear definition of the desired solution

Designing Infrastructure for Data Exchange

- ▶ International unified standards for data exchange
- ▶ Web-based distributed network
- ▶ Publicly accessible cloud workspaces
- ▶ Specialized versions of datasets
- ▶ Clear objectives for involved parties (governments, IGO's, NGO's, private sector, academics)
- ▶ Centralized cross-disciplinary data-registry

Research Methodology

- ▶ Online Journal Databases
- ▶ Search Engine Results
- ▶ Data aggregators
- ▶ Research Institutions

Evaluating Countries

- ▶ Launch Date
- ▶ IGO Involvement
- ▶ Data types
- ▶ Accessibility
- ▶ Competition
- ▶ Hackathon
- ▶ Development Aid

Analysis of Literature

- ▶ Literature assessing the aggregate OGDIs are in agreement, but lack quantitative analysis
- ▶ Most recommendations in literature are already being pursued by governments or private institutions
- ▶ Future aggregate research should utilize Open Data Barometer datasets
- ▶ Case studies tend to be sector specific and cover a broad scope

Literature on Developing Countries

- ▶ Government websites are difficult to navigate and government provided literature tends to be unhelpful
- ▶ Must rely on case studies from institutions or select aggregate literature
- ▶ Most literature is qualitative with the exception of Open Data Barometer
- ▶ Majority of literature is produced by opendataresearch.org

Policy Recommendations

- ▶ Conduct surveys of different country governments and private stakeholders across sectors to find common successes and failures.
- ▶ Synthesize this information into best practices, which could be implemented as a product or service.
- ▶ Determine high impact areas of OGD in development finance and tailor services to strengthen these initiatives.
- ▶ Create ranking system for development finance based on quality, adherence to standards, ease of integration.
- ▶ Promote integration of development finance data with related data, such as budget data.

AidData: The Open Data Movement

By: Alex Atkins & Jessica Percic