Open Data

Tim Davies¹ and Ania Calderon²

¹Global Data Barometer, Stroud, UK

²Open Data Charter, London, UK

Keywords

Open data · Open government · Anti-corruption · Public sector information · Data sharing · Data spectrum

Definitions

Open data can be described in terms of three key properties: data that is accessible (usually online), data that is machine-readable (in structured data formats that can be explored without proprietary software), and data that is available for wide reuse (usually indicated by clear open license statements). The lobby for open data emerged early this century, challenging restrictions on access to, and free use of, government held data. A broad-based movement brought together a rare alliance of stakeholders from civil society, the private sector, and civic technology. Together they argued that greater social, economic, and political value could be won from data resources if they were made available as open data, providing a platform for new services, business models, innovation, and citizen involvement in government.

The Open Definition, a collaboratively drafted statement published online, and modeled on the Open Source Definition (<https://opensource.org/osd>), played a key role in this early movement. It enabled alignment of different stakeholders behind a common cause of calling for diverse datasets to be published so they could be “*freely used, modified, and shared by anyone for any purpose*,” (<https://opendefinition.org/>) and it provided detailed criteria that were used to judge whether dataset, data license, or initiative should be ruled in or out of the open vs. closed data binary.

The last decade saw a rise of open data initiatives launched by governments across the world, and in numerous sectors. Over this time, the open data agenda has broadened to also address the opening up of data held in the private sector and civil society, and has adapted to recognize growing concerns about data privacy, data governance, and the ethical implications in design of datasets. In this chapter we explore the spread of open data policies both as the product of successful lobbying and advocacy, and as a tool within the public policy toolbox for addressing complex global challenges.

In this chapter we focus on data linked to public policymaking. However, we note there has also been a parallel movement for open data in science, linked closely with Open Access and

Open Source movements, pushing for publication of research datasets alongside articles, and contributing to the creation of large-scale data infrastructures for scientific data sharing. While the technical definition of open data remains constant across public policy and science, this chapter does not directly address open data in science.

A Brief History of Open Data Advocacy

The emergence between 2007 and 2013 of open data as a tool of public policy, the subject of national legislation, and a topic in international dialogue can be traced to the convergence of the interests of a number of different groups in securing a more liberal regime for access to government held data and information. A primary focus on technical issues of data licensing and data formats allowed a temporary point of consensus between groups with divergent agendas, including an established business-centered Public Sector Information (PSI) community, a more emergent grassroots civic technology movement, and an open government movement that has sought to combine notions of citizen participation, with ideas of efficient public service delivery through data and technology.

When in November 2003, the European Union adopted the first PSI Directive, it was responding in part to comparisons drawn between US public data policy: where certain federal dataset were free of copyright, and provided for reuse, and approaches in Europe and the rest of the world, where state-held copyrights and database rights were commonly used to restrict access to government-generated data (Fitzgerald 2010). Citing the business ecosystems that had grown up around US data sources such as weather data, or global positioning system (GPS) data, early advocacy focused on the potential economic benefits from a switch from “trading fund” models in which government agencies sought to profit from their data holding, to provision of data at marginal cost and under consistent rules. The 2006 Measuring EU Public Sector Information Resources (MEPSIR) study put the value that could be unlocked from liberalizing reuse of EU PSI at

€27bn (Dekkers et al. 2006), and one subsequent widely referenced study from management consultancy firm McKinsey claimed that upwards of \$3tn/year could be “unlocked” worldwide from greater publication of open data across just seven sectors (Manyika et al. 2013). Following the global financial crash of 2007/8, such studies (and the consultancy firm visits to governments around the world to promote them) were particularly influential in securing support from Ministries of Finance for open data policies that might otherwise have been dismissed as a low-priority and potentially costly bureaucratic reform.

Around the time that the first PSI directive was being transposed into European national law, an emerging civic technology community was starting to articulate clear principles for open government data, calling for access to the “raw” data behind government figures, and seeking free access to infrastructural datasets such as national mapping and postcode lookups, to power new online civic engagement platforms. The “Eight Principles of Open Government Data” set out at a meeting in the United States in 2007, called for all public data to be made available, describing this as any government “data that is not subject to valid privacy, security or privilege limitations.” (<https://opengovdata.org/>) Such principles found their political moment in both the United States and the United Kingdom in 2009/10, as President Obama, on coming to power, issued a Memorandum on “Transparency and Open Government” (<https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>) that resulted in the launch of the data.gov open data portal in May 2009. The UK followed suit in early 2010, creating its own open data portal in January 2010, spearheaded by web inventor Sir Tim Berners-Lee, who had been invited by then UK Prime Minister Gordon Brown to put forward actions that might rebuild public trust in light of a damaging parliamentary expenses scandal (https://webarchive.nationalarchives.gov.uk/20100105080918tf_/http://www.number10.gov.uk/Page20595).

In this period, the connection in public discourse between open data and open government was cemented. That link was carried forward in the launch of the multi-stakeholder Open

Government Partnership in September 2011: now an association of over 90 national and subnational government members committed to increase the access to data, promote use of new technologies for openness and accountability, and to support civic participation (<https://www.opengovpartnership.org/process/joining-ogp/open-government-declaration/>). Open government data initiatives rapidly spread to countries in the global south, particularly as the World Bank put backing behind open data initiatives, funding initial work in Kenya and Moldova and subsequently rolling out support to a large number of countries (Hammer 2019). While in some cases, open data initiatives were successfully integrated into existing national agendas around data policy, in other cases, the initial momentum built through external funder support was not maintained. As an evaluation of lessons from open data leaders found:

The challenge of an initiative being championed or supported externally, is that it does not necessarily understand local context or dynamics. Emphasis was placed on supporting a champion, at the expense of a whole-of-government approach. Philip Thigo, Government of Kenya (<https://drive.google.com/file/d/1CtuQbOPNWmCsEvdV-Wlmj4Yp4N2ctCJ5/view>)

By 2013, the Open Data Barometer study identified at least 40 established national government open data policies, and by 2017 the agenda had reached more than 55 national governments (<http://www.opendatabarometer.org>) and many hundreds of cities and localities. Across this period, open data also found a place on the agenda of the G8, G20, OECD, and as a component of a growing number of sectoral governance initiatives, such as the International Aid Transparency Initiative (IATI), Open Contracting Partnership, and Extractives Industry Transparency Initiative (EITI). The 2013 G8 members signed the G8 Open Data Charter, introducing the concept of “high value datasets” that countries agreed should be “open by default.” The principles of the G8 Open Data Charter were revised and opened up to wider participation with the launch in May 2015 of the International Open Data Charter at the 3rd International Open Data Conference, a significant gathering of the growing, and cross-sector, open data community.

The important role of civil society advocacy in the development of the open data agenda cannot be overstated. Wilson writes that “Civil society has been a driving force in advancing open data agendas, especially through standard setting, awareness raising, and defining public expectations” (Wilson 2019), though also noting the tensions this has at times placed civil society under. Most notably, conflict between right to information and open data communities, and frequent ambiguity created by the blurred lines between nonprofit campaigners and for-profit enterprises active in open data advocacy. Shifts to traditional roles and practices also created tensions within the state, where National Statistical Offices were rarely engaged early on in open data initiatives, and have at times seen open data measures as a threat to the integrity of statistical processes. Government IT functions, more used to working through risk management and closed contracts with large systems integrator firms, have also struggled in many cases to engage with the more start-up, small-scale, and open-source ethos of open data communities.

There are different views on the extent to which the current state of open data reflects the co-option of civil society to support a neoliberal data agenda (Bates 2013), or the successful confluence of private and public interests in adjusting data management regimes. However, an embedding of the open data discourse is evident in the latest EU directives. While the 2003 PSI Directive made no mention of the term “open data,” by its revision in 2013 the language of open government data was fully embedded within its preamble, and the most recent revision in 2019 fully adopts an open data framing, titled the “Open Data Directive (Directive (EU) 2019/1024 on open data and the re-use of public sector information).” It is clear, however, that contemporary debates on open data are moving from general questions of license and format to discussions of the specific dataset and data infrastructures that should be made available.

The EU’s latest open data directive also seeks to formalize agreement on a list of high-value datasets that should be published to common technical standards. This has opened the door for significant lobbying efforts both for and against

certain datasets being included from those with commercial and civic interests in data availability, or in maintaining their privileged market position from reselling access to data. Similar debates are playing out across the world, and in a number of cases it is still government agencies who hold particular datasets putting up the greatest resistance to open data efforts, conscious that their relative autonomy may be at risk if they can no longer treat data as an independent source of revenue.

Recent debates around open data also play out in a considerably more complex landscape than the discussions c. 2010. Concerns about threats to privacy as a result of data collection and sharing, rising up the agenda after the Snowden revelations (chapter ► [“Hackers and Whistleblowers”](#)), have combined with a recognition that much data of public value is not held by the state, but is collected by private actors, particularly large information technology firms. Even where they can be made open, the resources and capabilities to use these big datasets (chapter ► [“Big Data”](#)) are often out of reach of the independent civil society networks that led early open data advocacy. Tools such as the Open Data Institute’s Data Spectrum (<https://theodi.org/about-the-odi/the-data-spectrum/>) have been developed to situate open data as one among a number of regimes for the management of data, each appropriate in different contexts. Data collaboratives (chapter ► [“Data Collaboratives”](#)) are one of a number of approaches being developed to support data sharing in contexts where a full application of open data principles cannot be achieved either on grounds of privacy or security concerns, or because of the need to balance public and private interests.

While open data movements have been subject to criticism for, on the one hand a naivety about the economics of data production and the incentives of data owners (Peled 2014), and on the other, a failure to look beyond the technocratic focus on the release of datasets to fully consider what each dataset represents (Taylor 2017), the broad open data community appears to be adapting to these new contexts. Open data platforms are morphing into data-sharing platforms, able to support multi-stakeholder collaboration on data, while retaining a preference to make data

open when it can be. And international initiatives such as the Open Data Charter (where one coauthor of this piece works) are adopting refreshed strategies that emphasize “publication with purpose” rather than the earlier narrative of “raw data now.” While advocacy and initiatives to create national open data policies may be becoming less common, open data is being adopted as a policy tool across a wide range of areas.

In the following section we briefly address some of these ongoing trends.

Issues in Open Data

Open Data as an Engine of Public Sector Innovation

The 2008 “Apps for Democracy” challenge run by the District of Columbia government set the template for numerous hack-day and app competition events that have run since: inviting entrepreneurs and engaged citizens to work with government data and to create new technical solutions to public problems, or to develop new businesses based on government data. These open data-driven activities have created space for increased dialogue between officials and those outside government, as well as contributing to upskilling of government officials, and supporting the development of more forward-looking technology and data teams within government (Landry 2019; Eaves et al. 2019). While O’Reilly’s 2010 vision of open data enabling “government as a platform” (O’Reilly 2011) is far from fully realized, there are numerous examples of data release allowing creation of citizen and business-focused tools that would have been unlikely to emerge from within the public sector itself (Young and Verhulst 2015).

In a number of countries, open data initiatives, often established by the executive branch of government and under prime ministerial or presidential authority, have been able to send out teams to work across governmental silos, both centralizing access to datasets from different departments and contributing to the rationalization of data infrastructure across government, while also supporting the development of data analytical

capacity in individual ministries. For example, in Mexico, “Data Squads” of external technical experts unlikely to be drawn into public service were embedded in government agencies under intensive 6-week programs to help identify and overcome obstacles in opening high-quality data (Puttick et al. 2014).

Today, however, open data is not the only form of data held up as a key tool of public sector innovation. In 2020, government responses to the COVID-19 pandemic have increasingly turned to big datasets from the private sector, often containing granular and highly sensitive personal information, as a potential tool to guide public policy responses. And where once governments were discussing open data strategies, open data can often now be found subsumed as a marginal component of artificial intelligence agendas seeking to harness data science techniques. This may give rise to a renewed phase of open data advocacy, linked to arguments for transparency and accountability of automated decision making.

Data, Democracy, and Power

In “The State of Open Data,” looking back over a decade of open data practice, Davies concludes that “Data is seen as a source of power, and advocacy for open data as a strategy to redistribute that power” (Davies et al. 2019) going on to note that “While open data work has, for a long time, been able to maintain a broad-based appeal to actors across government, civil society, and the private sector, future projects will involve negotiations over the shape of data infrastructure that may make keeping everyone on board in a broad coalition much more challenging.” This reflects a concern that, as open data advocacy moves from calling for access to data, to calling for the sustained availability of certain datasets as part of data governance frameworks, and for specific standardized fields within those datasets, the interests of different groups are likely to diverge.

These tensions are amply illustrated by Cobham and Jansky’s description of more than a decade of campaigning by Tax Justice Network and other partners for detailed disclosure of Country by Country reporting of the activities, revenue, and taxes by multinational firms: data seen as vital

to control tax avoidance activities (Cobham and Jansky 2020). Initial proposals put forward by the Tax Justice Network in 2003 for an accounting standard for country by country disclosures were, a decade later, able to influence OECD standards for reporting, adopted following G20 pressure. However, the OECD standard was only for private firms to report data privately to their home country tax authorities, and fell short in a few respects on specific disclosures advocated by civil society, failing to require data fields on employee pay, tangible assets, and intra-group purchases. The Tax Justice Network is hopeful that now that data is flowing, albeit privately, the foundations are laid for this data to be made open in future and to support greater scrutiny and democratic oversight of multinational taxation. This is not, however, a done deal. While some voluntary initiatives have seen corporations pledge to publish open data on their Country by Country reporting, others are likely to lobby against any new disclosures, and continue to oppose reporting as structured open data.

Struggles over access to private sector data are also important in the context of increasing government reliance on proprietary datasets to guide public policy. Where early open data advocacy drew upon the idea that the data upon which decisions are made should be subject to public scrutiny, this grows more challenging when that data is not actually managed by the government, but instead belongs to private actors and is only made accessible to the government for specific uses (without giving governments the rights to share that data more widely). There is a risk in these cases that policy may appear to come from a black box as the data on which decisions are made cannot be shared at a granular enough level. Ultimately, even “public data” can no longer be assumed to originate only in the public sector, particularly when public services and the data that goes with them are contracted out (<https://datajustice.github.io/report/>).

Open Data as an Anti-corruption and Accountability Tool

One area in which open data has become firmly established as a public policy tool is in the field of

anti-corruption. In 2015, G20 governments agreed on a set of G20 Anti-Corruption Open Data Principles (<http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf>), aiming to encourage countries to make more than 30 crucial datasets, outlined in research from the Open Data Charter (<https://open-data-charter.gitbook.io/open-up-guide-using-open-data-to-combat-corruption/>), available as open data to support anti-corruption efforts. These datasets included data on public procurement, beneficial ownership, politically exposed persons, and lobbying registers. Publishing this data, it was argued, would allow government and civil society actors to track the use of public resources and contracts, to follow the sources of political financing, and make it easier to prevent criminal activities like bribery and nepotism. However, countries have been slow to publish the datasets covered in the principles, or to create joined up data infrastructure (<https://www.transparency.org/en/news/open-data-promise-but-not-enough-progress-from-g20-countries#>).

Open data has also been applied as part of efforts to increase lobbying transparency. In the European Union, for example, Transparency International's Integrity Watch platform (<https://www.integritywatch.eu/>), based on open data on lobbying, now allows citizens to see that Google spends upwards of €3.5 m on lobbying the European Commission each year, and Airbus has met with senior officials more than 20 times in just 6 months in late 2019, early 2020. A number of national, regional, and city governments also publish lobbyist registers, such as the city of Austin, Texas, which provides detailed open data on registered lobbyists, their expenditures, and the questions they have been asking (<https://data.austintexas.gov/browse?q=lobbyists>). A set of standards for lobbying transparency, incorporating open data principles, were put forward in 2015 by a coalition of organizations under the banner LobbyingTransparency.net. However, reliable open data on lobbyist engagement with government decision making remains the exception rather than the rule across the world. More recently, attention has also turned to transparency of political content on major social media platforms, with platforms such as Facebook creating a

public (though not fully open data) "Ad Library" listing paid for political adverts for a number of countries, and providing details of spending totals and targeting of these (Gorwa and Ash 2019). These self-regulatory efforts using transparency and open data may be seen in part as an effort to avoid more formal regulation that may involve states adopting targeted transparency provisions to require private sector data disclosures.

Overall, there remains considerable work to be done to build a full open data infrastructure for accountability, anti-corruption, and political integrity. During the COVID-19 crisis, open data supported detection of a number of cases of fraud and corruption, but it is likely many more went undetected because of a lack of integration between data on contracts, on companies, their beneficial owners, and the interests of public officials. Pilots of the Open Data Charter's guide on anti-corruption in Mexico and Colombia found that breaking down these barriers required a focus on concrete problems that could be solved using open data, establishing clear incentives to share it, and embedding data skills for its use across government and civil society anti-corruption reformers.

Open Data and Other Global Challenges

There are few public policy challenges, from combating illicit financial flows or addressing the effects of climate change to managing a global pandemic or improving the business competitiveness of states, where data does not play a pivotal role. This does not necessarily mean that everything should be *open* data. Indeed, as a review of open data engagement across 16 sectors found, "sectoral engagement with open data is far from inevitable but instead relies on the right combination of advocacy, infrastructure, and backing at key opportunity points" (Davies et al. 2019: Section 1 Introduction).

In some sectors large-scale government-backed networks, such as the Global Open Data on Agriculture and Nutrition (GODAN) initiative, have brought together private, academic, non-profit, and government stakeholders to articulate a shared agenda for action on publication and use of open data to support global food security. In

other sectors, such as telecommunications, open data activities to improve global connectivity have been more ad-hoc, as established players in consolidated markets have less interest in increased transparency or efforts to create a more level playing field through sharing data. Certain initiatives, such as the Extractives Industry Transparency Initiative (EITI), began with relatively limited emphasis on data and open data, but have increasingly come to prioritize data as a tool to secure and monitor improved implementation of policy commitments.

Established and emerging open data initiatives vary substantially with respect to their sources of data, their local or global scope, and the extent to which they are technology led, or represent a data component embedded within a wider policy agenda. In some cases, contemporary open data initiatives involve creating “counter data” to challenge or substitute for government or private sector data sources. In other cases, they involve standardizing and aligning data sharing between different stakeholders. Significant questions also remain over how far open data initiatives assume certain levels of technical capacity, and how applicable they can be to states with less developed data infrastructures.

Conclusion

The rapid spread of open data policies in governments across the world, and as a tool for addressing complex global challenges, is the product of the coming together of a heterogeneous mix of agendas to form a broad-based movement of diverse stakeholders from civil society, the private sector and civic technology. While there are numerous examples of open data supporting improvements in service delivery or creation of citizen and business-focused tools, providing a platform for new services, innovation and citizen involvement in government is far from fully realized. As open data teams were established in government and a new wave of technology and competence in technology was driven into policymaking, these have often confronted the

power structures that data reflect, only sometimes successfully disrupting entrenched interests.

Cross-References

- [Big Data](#)
- [Data Collaboratives](#)
- [Hackers and Whistleblowers](#)

References

- Bates, J. (2013). The domestication of open government data advocacy in the United Kingdom: A neo-gramscian analysis. *Policy & Internet*. <https://doi.org/10.1002/poi3.25>.
- Cobham, A., & Jansky, P. (2020). *Estimating illicit financial flows: A critical guide to the data, methodologies, and findings*. Oxford, UK: Oxford University Press.
- Davies, S. W., Rubinstein, M., & Perini, F. (Eds.). (2019). *The state of open data: Histories and horizons*. Cape Town/Ottawa: African Minds and International Development Research Centre.
- Dekkers, M; Poleman, F; te Velde, R; de Vries, M. 2006. *MEPSIR – Measuring European public sector information resources*. <https://ec.europa.eu/digital-single-market/en/news/mepsir-measuring-europeanpublic-sector-information-resources-final-report-study-exploitation-0>
- Eaves, D., McGuire, B., & Carson, A. (2019). Open data around the world – North America, Australia, and New Zealand. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: Histories and horizons*. Cape Town/Ottawa: African Minds and International Development Research Centre.
- Fitzgerald, B. F. (2010). *Access to public sector information: Law, technology and policy*. Sydney, Australia: Sydney University Press
- Gorwa, R., & Ash, T. G. (2019). Democratic Transparency in the Platform Society. SocArXiv. October 4. <https://doi.org/10.31235/osf.io/ehcy2>.
- Hammer, C. (2019). Open data stakeholders - Multilateral organisations. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: Histories and horizons*. Cape Town/Ottawa: African Minds and International Development Research Centre.
- Landry, J. (2019). Open data and urban development. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: Histories and horizons*. Cape Town/Ottawa: African Minds and International Development Research Centre.
- Manyika, James, McKinsey Global Institute, and McKinsey and Company. 2013. *Open data: Unlocking innovation and performance with liquid information*.

- O'Reilly, T. (2011). Government as a platform. *Innovations: Technology, Governance, Globalization*, 6(1), 13–40.
- Peled, A. (2014). *Traversing digital babel: Information, e-government, and exchange*. Cambridge, MA, USA: MIT Press.
- Puttick, R., Baeck, P., & Colligan, P. (2014). *The teams and funds making innovation happen in governments around the world*. London: Nesta & Bloomberg Philanthropies.
- Taylor, L. (2017). What Is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*: <https://doi.org/10.1177/2053951717736335>.
- Wilson, C. (2019). Open data stakeholders – Civil society. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: Histories and horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre.
- Young, Andrew, and Stefaan G. Verhulst. 2015. *The global impact of open data: Key findings from detailed case studies around the world*, O'Reilly Media, Inc. Sebastopol, CA, USA.