

Introduction

A decade ago, open data was more or less just an idea, emerging as a rough point of consensus for action among pro-democracy practitioners, internet entrepreneurs, open source advocates, civic technology developers, and open knowledge campaigners. Calls for “open data now” offered a powerful critique of the way in which governments and other institutions were hoarding valuable data paid for by taxpayers – data that if made accessible, could be reused in a myriad of different ways to bring social and economic benefits and democratic change.

Ten years on, open data is much more than just an idea. First, it was a movement, and then a label applied to vast quantities of data from genomics and geospatial data to land registers, contracting, and parliamentary voting. Today, it’s a term found on government portals, in global policy documents, and in job descriptions. Thousands of businesses around the world owe their existence or their growth to the release of open government data, and hundreds of civil society organisations have embraced open data as a key element of their social change toolkit.

For a while, it may have been possible to identify a cohesive open data movement united by shared interests, working simply to gain access to more data and establishing the principle that government data should be open. However, as the movement has evolved, stakeholders have turned their focus to linking data use to specific needs and to questions of how to quantify the return on investment in advancing open data. Within this fast growing and organic open data movement, an ever-increasing number of networks and communities of practice have become more diverse, fluid, and cross-sectoral.

So what is the open data movement today? What has it achieved over the last decade? Answering these questions is at the core of this publication. It is a collective effort to explore what we can learn from the past, to identify how to build on the investments made to date, and to look at how open data policy and practice have started to address challenges such as mainstreaming and sectorisation.

Exploring these questions is not just important for historical purposes. It can yield important insights on how best to move forward. This publication is also an invitation to identify the issues that may sustain this broad coalition into the future. We believe that a deep reflection about the movement, even a reflection on whatever cracks have appeared or on the gaps between promise and reality, provides a vital opportunity to discuss where realignment and rethinking are needed.

This collection of essays is the product of an 18-month journey that has brought together almost 70 authors, supported by over 200 other contributors, to produce 37 short chapters on the current state of open data from a range of different perspectives, offering the most comprehensive attempt to explore the breadth and depth of the open data field to date.

Histories and horizons

Ten years may seem like a short period of time, but, when technology is involved, it constitutes a generational age. Institutional memories are curiously short, and in the cultural context of open data where amateurs are often welcome and professional barriers to entry are low, it is easy for work to proceed with little awareness of the past. This last decade has seen many succeeding phases of activity, so we have encouraged our authors to take a comparatively long view (when set against other contemporary writing on open data) to document the past in order to lay stronger foundations for future research and action.

We have also sought to understand open data as a global movement. Although some accounts have a tendency to focus on the North American or European roots of open data, tracing histories back to the launch of data.gov under Barack Obama's presidency, open data practice has been shaped by interventions from across the globe. To gain a vantage point on open data as a global movement, this collection draws upon the editors' engagement with the Open Data for Development (OD4D) network¹ which has been closely engaged in regional networks in the Global South and involved in a range of global initiatives, including the Open Data Barometer (ODB), the Open Data Charter, the Open Government Partnership (OGP) Open Data Working Group, the Impact Map, and the Open Data Leaders Network.

Since 2015, OD4D has also been the permanent co-host of the International Open Data Conference (IODC), and the editors of this volume have been involved in preparing conference reports, including shared roadmaps for action, for the third, fourth, and fifth IODC meetings. We have seen how, over its five editions, IODC has shifted from a focus on open data, in and of itself, toward discussions that are thematic, sectoral, regional, and issue oriented, fostering critical debates on open data. The conference tracks and sessions at IODC have ultimately provided many of the chapter titles in this book, reflecting the many subcommunities of the open data field that have emerged. The debates at IODC over the last nine years also provide a useful proxy for debates across the wider field of open data, so a survey of the IODC conferences offers us one route to explore, in broad strokes, a history of how the focus of the open data movement has evolved.

The evolution of a movement

The first IODC was hosted by the United States Department of Commerce and took place in November 2010 in Washington, DC.² At the same time, in London, a civil-society led conference, the Open Government Data Camp, was taking place.³ These parallel events captured the growing excitement about open data from both governments and civil society and marked the end of a

year in which open data had moved from idea to initiative and from inception to the earliest stages of institutionalisation. Over time, the boundaries between government and civil society networks have become more fluid with both positive and negative effects. The focus of these early events was on showcasing the platforms that had been built and discussing the potential for open data across sectors. However, even at this early stage, questions were being asked about how the impact of open data might be tracked, and whether bold claims being made on the transformative potential of open data could actually be realised.

By the time of the second IODC, hosted by the World Bank in July 2012,⁴ the question of how to measure emerging impact was firmly on the agenda. At this point, open data was being discussed in the context of international development and the movement had broadened to include a number of open data leaders from developing countries. Yet, while many of the projects profiled were still platform-focused, it was becoming clear that simply releasing data was not enough and that the quality of data available was far from perfect. Early discussions turned to whether the potential returns of open data had been overstated and how to deal with the growing gap between rhetoric and reality. That early sense of an impact gap still pervades many of the chapters in this collection with several authors exploring the various reasons that could explain less than promised progress on transformative use. However, we note that the perception of an impact gap is rarely reflected by a similar level of difficulty in sourcing case studies of open data use, raising questions about the perception and the reality of progress on open data, as well as the influence of early conceptual models for open data impact on current critical practice.

By the time of the third IODC in Ottawa in May 2015, the focus had moved to an examination of how open data ideas and practices were developing in different sectors and regions.⁵ The conference captured a period of dramatic regional and sectoral growth of open data activity with increasingly diverse representation from across the globe. There was growing recognition that opening data alone was not enough to create impact. Instead, as many of the chapters in this collection explore, to secure outcomes from open data, clear goals need to be established and a series of strategic interventions identified. Policy design, intermediaries, and capacity building were all on the agenda. As more stories of open data in use to solve specific problems were shared, there was a growing recognition that impacts secured in one context or sector may not automatically translate to another. And with this recognition came an understanding that, rather than a single open data movement, there may be many overlapping, interwoven movements, drawing on particular elements of open data to address many different agendas.

The third IODC also made explicit the potential links between open data and sustainable development, highlighting that open data was no longer the only data game in town. Instead, in the context of international development, open data now had to find its place alongside renewed efforts to build the capacity of long-established statistical agencies, as well as newer initiatives seeking to tap into the potential of big data from proprietary private sector providers.

The fourth IODC, held in Madrid in October 2016, was framed in terms of “Global Goals, Local Impact”, reflecting increased consolidation of global advocacy and a continued focus on shared global principles, which was evolving in parallel with the growth of subnational and thematic initiatives.⁶ Although the open data agenda had matured and become well-established as part of global policy-making, discussions explored concerns that it risked becoming a niche issue, destined to be the focus of only a small group of the “usual suspects”. Issues of privacy,

gender equity, diversity, inclusion, and Indigenous data rights, all competed for space on the agenda, along with a new space for more critical discussion of how open data impact might be realised and the potential for more nuanced approaches to open data practice.

These critical threads continued into the fifth IODC that was held in Buenos Aires in September 2018.⁷ New on the agenda were discussions related to artificial intelligence (AI), and the conference saw a stronger focus on data standards and open data infrastructure. Although these later issues have long been discussed by a small but dedicated element of the open data community, there was increased recognition that they are not just technical issues. They also involve questions of data governance with political choices embedded in the use of data standards and structures, having substantial consequences for who can use and benefit from data.

In 2018, for the first time, the IODC agenda also featured a session on “Open Data Under Threat”, capturing a sense that continued progress was by no means assured. Against the backdrop of a deepening crisis of diminishing government support for openness around the world and much more public debate around the positive and negative potential of technology, concerns voiced over open data were no longer solely about a perceived impact gap. They also involved a deeper questioning of when and where openness can be safely practised and whether open data should be a priority for donors, advocates, and activists in the future.

A look at the 2018 IODC agenda also illustrates sectoral and regional sessions going deeper into the specific concerns of their fields and localities. In this, we find a reflection of the increasing diffusion of open data ideas, representing both a marker of success but also a potential risk to any future coherence of open data activity. In putting together this collection, while drawing on the OD4D network and IODC as a starting point, we have been conscious of the need to move beyond to capture wider activity on open data and to explore how an early open data movement has now become many overlapping movements. By working with a diverse community of authors, encouraging them to draw on both published literature and their own domain-based networks, as well as on wider online outreach to the community, we have looked to capture insights into the open data world from far beyond the core IODC community.

Taking stock

Culture and temperament inevitably shape any qualitative review of progress. As with any invested community, a substantial number of people and organisations engaged with open data have a tendency toward critique. For many, the idea that data should be open was ultimately born out of a critical opposition to the way governments were handling data and an ambitious imagining of an alternative future in which access and capacity to gain benefit from data is more evenly distributed. Coupled with the differences in pace between rapid technological change and comparatively glacial governmental reform, this critical approach combined with well-meaning ambition can lead to the progress of the last decade being underplayed. Challenges on the horizon ahead can too often serve to mask the steps that have been taken in order for those challenges to become visible.

In looking across the chapters that follow, we are struck by the extent to which open data ideas have become established across the globe. For instance, in Chapter 28 (Multilateral organisations),

Hammer describes how, from 2010 onward, global development banks have integrated open data into their own methodologies, helping to popularise open data initiatives in developing and developed countries. In Chapter 29 (Private sector), Gurin, Bonina, and Verhulst illustrate the private sector's widespread use of open data with examples from Asia, Africa, Latin America, Europe, and America. And since the Sustainable Development Goals were adopted in 2015, robust, comparable, and open data has been emphasised as a critical tool to both inform and monitor development efforts. Across the entire section on Open Data Sectors and Communities, examples of open data being used to drive socioeconomic benefits or to shape policy debates are too numerous to mention here.

The adoption of open data as a central tool used in a number of major global policy initiatives of the last decade is particularly notable. The OGP, the International Transparency Initiative, the Extractives Industry Transparency Initiative (see Chapter 8: Extractives), and the Global Legal Entity Identifier Foundation which was created to respond to the last financial crisis (see Chapter 3: Corporate ownership), have all embraced open data within their work. Within the OGP in particular, commitments related to open data have been some of the most popular and successful.⁸ As Chapter 17: Algorithms and artificial intelligence explores, even as public attention shifts from open data toward a new wave of excitement about AI, open data ideas appear firmly established as a foundation for governmental AI policy.

So why is the current period for open data one of re-evaluation, rather than of celebrating progress? Put simply, the adoption of open data as part of the global development toolbox has opened it (rightly) to substantial scrutiny. How quickly are efforts to open up data leading to change? What is the return on investment from open data-related reforms? What are the factors that shape whether or not open data leads to impact? And finally, how does work on open data interact or integrate with other core issues of sustainable development, such as gender equity, Indigenous rights, and good governance? Questions such as these have received increasingly detailed attention over the last few years. Although hardly any of these questions have simple answers, by looking at both progress and challenges, this volume seeks to bring together evidence, examples, and analysis that can support efforts to address them more clearly than before.

Looking to the future: An impending identity crisis

For all the steps forward described above, as we look to the horizons of open data, we are confident in stating that policy excitement about open data has peaked. Ten years in, we are past the peak of a hype cycle and past the point where promise has to give way to evidence of practical impact. As a result, many open data communities are fast approaching their difficult teenage years with a deepening identity crisis.

Over the last decade, debates around the role of data in society have moved to centre stage, but arguments for openness now have to share the spotlight with newer excitement over the economic potential of big data, machine learning, and growing fears about the negative impacts of data stemming from data-driven manipulation of politics or the corporate invasion of personal privacy. Although early narratives around open data may have been able to present increased access to data as an unalloyed public good, contemporary advocacy must confront a much more

complex landscape in which power, politics, and the question of who gains or loses from unfolding regimes of data access cannot be ignored.

This presents a number of key challenges with which the following chapters attempt to grapple. As open data has spread globally, the way in which open data ideas have manifested across different sectors, communities, countries, and stakeholder groups has increasingly varied. Regional distinctions of emphasis have developed, with, for example, some downplaying the importance of open licences (see Chapter 37: Sub-Saharan Africa) and others talking of innovation rather than of openness in order to avoid political resistance (see Chapter 34: Middle East and North Africa). As sectoral efforts deepen, it is domain or subject matter experts, rather than data specialists, who drive activity forward, so that the challenges of creating cross-sectoral linkages and building shared data infrastructure become even greater. Increased emphasis on inclusion places a substantial demand on problem-centred initiatives, which, in light of low levels of data literacy, must choose whether to focus on data for expert communities or to actively pursue the promise of open data as a tool of wider popular empowerment. When the focus shifts from calling for access to data to creating data infrastructure and putting data to work, the divergent goals of those who formed an initial open data movement come clearly into view and managing the tensions that emerge can be complex.

It was in mid-2017, as these tensions were becoming more apparent, amid a sense that overall momentum for open data may be faltering, that the State of Open Data project was conceived. Our objective:

To critically review the current state of the open data movement, assessing its progress and effectiveness in addressing challenges related to social and economic development and democratisation around the world.

Based on such a broad stock-taking of open data activity, we may not be able to fully resolve questions about the future of open data, but we can provide an account that helps practitioners, policy-makers, and community advocates to step back from their own position to gain a view of the wider landscape. By doing this, we hope to offer a rich and timely perspective and the groundwork for constructive debates that will shape the next decade of open data.

A collaborative review: The approach

The open data field already benefits from a number of semi-regular quantitative studies of the progress of open data, such as the ODB⁹ and Open Data Index,¹⁰ both also supported by OD4D. To complement these, the approach to the State of Open Data project was designed, from the outset, to be more qualitative and narrative in style, involving a five-stage process.

1. **Selection.** Working with the OD4D network, potential chapters were identified based on open data communities, regions, stakeholder groups, and cross-cutting issues. Authors were then invited to lead on creating these chapters. The introduction to each section of this book provides details on the selection of chapter topics.

2. **Engagement.** Authors were asked to create an initial “environment scan”: a community brainstorming of issues, evidence, key actors, and events related to their topic. Scans were posted online for public comment and additions to gather more examples, case studies, articles, and input from beyond the authors’ own networks.
3. **Writing and review.** Responding to a common set of questions and prompts, authors then completed full chapter drafts, drawing on the input received from the environmental scans. These draft chapters were sent for peer review by independent reviewers and by members of our editorial board. Reviews were sent to authors who completed chapter revisions based on the input received.
4. **Public drafts and discussion.** Public drafts for the majority of the chapters were posted online ahead of the IODC in Buenos Aires in September 2018, where emerging themes were discussed. Panel discussions on themes from the work were also held at the OGP Summit in Tbilisi, Georgia, followed by additional opportunities for revision.
5. **Synthesis and recommendations.** Based on a collective review of all chapters, the editors have worked to draw out key findings and recommendations, which are summarised in section introductions and the book’s conclusion, including recommendations for research, funding, policy-making, and practitioner communities.

The authors and contributors to this project have been drawn from a wide range of backgrounds. Some have been active in the open data field for many years, while others are relative newcomers. Some are advocates and activists, while others are observers or academics. Some are open data generalists, while others specialise in a particular field. Many draw upon a range of different roles and positions.

When considering all of the authors, contributors to the environment scans, independent reviewers, and the editorial board, input has been received from over 220 individuals from around the world. Representing the diversity of the open data community with regard to gender, diversity, and global inclusivity has been the key principle underlying our approach to this volume. The goal was to achieve a 50–50 gender split in terms of authorship, although we fell short of this with a 58–42 split in favour of men.

Definitions and scope: Open government data

Our focus in this volume is primarily, but not exclusively, on open government data. That is, data which traditionally originates from governments, is created or used during the business of governing, or is created or published at the request of governments. We have intentionally adopted a broad definition here, cognisant that, over recent years, the traditional monopoly of national-level governments both in data collection and in being a primary site of governance has been eroded. For example, satellite imagery data from

private companies or crowdsourced data from citizen scientists can all fall within the broad landscape of open data either traditionally collected by governments or used for governing. Similarly, data that results from academic research networks, but which informs public decision-making and action, forms a component of some chapters within this volume. However, reflecting the way that communities of practice around open data are generally organised, we have mostly stayed away from looking at open data in terms of open science or evaluating the extent to which different scientific disciplines and communities are approaching data sharing, access, and openness. This is well addressed in other work.¹¹

When it comes to defining open data, we draw upon the widely used definition of open data as data that is accessible, machine-readable, and free of licensing restrictions on reuse. However, we apply the definition heuristically rather than legalistically. This recognises, for example, that in some countries and contexts, the lack of a fully “open licence” is less of a barrier to reuse in practice than in others, or that, at times, data may not be provided in machine-readable formats at source but has been easily converted for reuse by intermediaries. Rather than rule out such cases from exploration on a technicality, they are included in the scope of this study with their limitations noted where relevant.

Targeting the core stakeholders

One of the notable features of open data is the way in which it has been adopted and shaped by so many different stakeholders. Unlike “big data”, for example, which appears to be primarily a corporate concept marketed to governments and civil society, networks around “open data” have always been much more diverse, fluid, and cross-sectoral. More than anything, this breadth and fluidity lies at the root of the impending identity crisis of the open data movement. For a long time, it may have been possible to manage the tension between different interests via a short-term focus on simply gaining access to more data. However, when stakeholders turn their focus to data use and the need to quantify the return on their investment of time and resources, a broader open data coalition is much harder to sustain. Determining what the open data movement can (and should) yield moving forward, how to maximise every investment made, and how to take on the challenges of mainstreaming and sectoralisation simultaneously, is at the core of the movement’s identity crisis. The cracks that may appear need not lead to crisis. Rather, they should serve to highlight in relief where realignment and rethinking are needed for the future.

In editing this collection, we have sought to work with all of the authors to address the needs of four main groups: researchers, funders, policy-makers, and practitioners.

For researchers, each chapter draws upon available academic and grey literature, providing detailed citations and suggesting further reading. The hope is that researchers will use these chapters as a primer on open data within particular contexts to identify critical research gaps in need of further attention. In particular, the inclusion of further reading is designed to assist the use of these chapters in a teaching context.

For funders, we have sought to highlight key organisations and stakeholders in each sector and region and to point out instructive examples of what is being done with open data, noting, where appropriate, gaps in the available resources needed to develop new ideas or to scale what works in more locations for larger impact. A dedicated chapter on donors and investors (see Chapter 25) also considers the need for greater coordination of funding, and, as with most chapters, points to current areas of underinvestment, particularly around the infrastructure needed for sustainability and high-quality data delivery, as well as capacity building, to create a widespread culture of data use.

For policy-makers, we have encouraged authors to address both progress and challenges in the implementation of open data. In many cases, you will find more on the persistent challenges, reflecting not so much a lack of progress but rather the shared critical and progressive mindset of our authors who seek ambitious social change through the application of open data. We have sought, however, to keep chapters focused on a relatively small number of issues, prioritising those that most deserve policy attention at present.

For practitioners interested in detail on open data projects, whether focused on data publication or use, we have sought to provide them with both critical reflection and inspiration. The hope is that by reviewing chapters related to a specific sector from multiple perspectives, practitioners will discover new ways of framing old problems and practical ideas about how to move forward in using open data as a tool of entrepreneurial development or social progress.

Crucially though, we do not know how many of the readers of these essays will, in the future, associate themselves with the label of “open data practitioner” or “researcher”, or whether they will simply perceive their role as someone who engages with open data as one tool among many. This is perhaps core to the identity crisis the movement may be currently experiencing and to the corresponding adjustments that open data communities will need to make in the second decade of open data. Is there still a need for a sustained movement that identifies the technical and licensing regime around open data as its core objective? What ethical and normative approaches need to be integrated into any future engagement with open data? Is it a good thing for the debate to move on from openness to adopt other narratives related to “good data”,¹² “data justice”,¹³ or “data rights”¹⁴? We will return to these questions after our review of the state of open data offered in the following chapters, when we will be better placed to discuss what stands to be gained or lost in the years ahead.

Endnotes

- 1 <https://www.od4d.net/>
- 2 <https://web.archive.org/web/20101128112407/https://www.data.gov/conference/>
- 3 <https://web.archive.org/web/20101218004117/http://opengovernmentdata.org/camp2010/>
- 4 <https://web.archive.org/web/20120821060331/http://www.data.gov/communities/conference>
- 5 <https://web.archive.org/web/20150716100733/http://opendatacon.org/>
- 6 <https://web.archive.org/web/20161104210854/http://opendatacon.org/>
- 7 <https://web.archive.org/web/20190127062831/https://www.opendatacon.org/>

- 8 Khan, S. & Foti, J. (2015). Aligning supply and demand for better governance: Open data in the Open Government Partnership. *Open Government Partnership*, 1 January. <https://www.opengovpartnership.org/resources/aligning-supply-and-demand-better-governance-open-data-open-government-partnership>
- 9 <https://opendatabarometer.org/>
- 10 <https://index.okfn.org/>
- 11 Digital Science. (2018). *The state of open data report 2018*. London: Digital Science and FigShare. https://digitalscience.figshare.com/articles/The_State_of_Open_Data_Report_2018/7195058
- 12 Daly, A., Mann, M., & Devitt, S.K. (2019). *Good data*. Amsterdam: Institute of Network Cultures.
- 13 Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2). <https://doi.org/10.1177/2053951717736335>
- 14 Tisne, M. (2018). It's time for a bill of data rights. *MIT Technology Review*, 14 December. <https://www.technologyreview.com/s/612588/its-time-for-a-bill-of-data-rights/>