

**DATA
FOR
POLICY
2022**



Open Data Resource Guide

PREPARED BY OPEN DATA WATCH

Open data can be a powerful resource for informing policies, increasing transparency, and measuring progress. But making data open requires commitment, organization, and technical capacity. We look forward to discussing how to harness the full potential of open data at the **Data for Policy** conference. Ahead of the meeting, we offer this guide, grouped by common questions, to shape our discussion and serve as a reminder of the progress made to date, the persistent challenges in the open data space, and the opportunities for overcoming those challenges. Based on the conference’s discussion, we will revise and translate this list into an FAQ for general use.

If there are additional questions not covered in this initial edition, please let us know by email at info@opendatawatch.com.

CONTENTS

1. What is open data?	2
2. Why is open data important?	2
3. How can privacy be reconciled with open data?.	4
4. What does open data mean for AI and machine learning?	4
5. How has the open data movement evolved?	5
6. What are the preconditions for maximizing the benefits of open data?	6
7. What are the roles of different sectors of society with respect to open data?	7
8. What are some of the best use examples of open data?	8
9. Where do we see more and where do we see less progress on open data and why?	9
10. How can I or my organization contribute to the open data movement?	10

1. WHAT IS OPEN DATA?

The most commonly cited definition of open data says, “Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).” Or, in short, “open data and content can be freely used, modified, and shared by anyone for any purpose.” The full statement of **Open Definition 2.1** describes four elements of an open work: open license or status, access, machine readability, and open format.

Open Knowledge Foundation. n.d. “Open Definition.”

The *Open Data Handbook* elaborates on the elements of open data: availability and access, re-use and redistribution, and universal participation. Availability and access mean data are available as a whole, at reasonable reproduction cost, in a convenient and modifiable form, and preferably online. Re-use and redistribution require data to be available under terms that permit re-use and redistribution as well as intermixing with other datasets. Open data should therefore be interoperable. And universal participation means that everyone is able to use, re-use, and redistribute the data without restrictions.

The Open Knowledge Foundation further summarizes the definition of open data into the idea of interoperability, which denotes the ability to interoperate - or intermix - different datasets. Asking what kinds of data should be open, it states that the focus is on non-personal data. And for general government data, national security restrictions may apply.

Open Knowledge Foundation. n.d. “What is Open Data.” *Open Data Handbook*.

All kinds of non-personal data should be open, although for government data national security restrictions may apply. The same standards for open data can be applied to other types of information, including cultural works and artefacts, scientific research, government financial accounts and financial market information, census, surveys, and socioeconomic indicators, weather, climate, and other environmental data.

Open Knowledge Foundation. n.d. “What kinds of open data?” *What is Open?*

2. WHY IS OPEN DATA IMPORTANT?

Open Data for Official Statistics makes the political and ethical case for open data: “[C]itizens are entitled to the products of their government and to use that information to hold governments accountable. The benefits are realized through increased government efficiency and responsiveness to citizens’ needs. But there may also be substantial economic benefits to making data open.” It also makes an economic and financial case for open data, rooted in the theory of public goods. Furthermore, there are benefits that national statistical offices can realize by taking the lead on open data.

Open Data Watch. 2021. *Open Data for Official Statistics: History, Principles, and Implementation*.

The World Bank's Open Data Toolkit acknowledges that some expenditure of public resources and effort will be required, but the benefits far outweigh the costs. These benefits include transparency, public service improvement, innovation and economic value, and efficiency. Open data support public oversight of governments and helps reduce corruption by enabling greater transparency. Open data give citizens the raw materials they need to engage their governments and contribute to the improvement of public services. Public data, and their re-use, are key resources for social innovation and economic growth. Open data make it easier and less costly for government ministries to discover and access their own data or data from other ministries, which reduces acquisition costs, redundancy and overhead.

Data producers and users receive benefits from open data policies. (Governments can be considered both producers and users.) "For governments, ministries and supply-side organizations, policies provide guidance, instructions, requirements, and tools for implementing Open Data. ... For user groups comprised of citizens, civil society organizations, businesses, researchers, and data consumers, Open Data policies clearly define which data are or will be made public, how and where to acquire data, standards for providing data and metadata (which also foster accountability) and how to engage with the government or producing agency. ... An additional benefit of Open Data policies is the insight they provide into a government's internal procedures for managing the Open Data initiative, which helps consumers better understand the data ecosystem."

World Bank. n.d. "Benefits of Open Data." *Open Data Toolkit*.

Open data can help monitor and achieve the Sustainable Development Goals (SDGs). Open data are a key resource for fostering economic growth and job creation, improving efficiency and effectiveness of public services, increasing transparency, accountability, and citizen participation, and facilitating better information-sharing within government. Open data can help inform plans to achieve the SDGs and measure their outcomes. Open data are "a facilitator of standards, a tool for accountability and an evidence base for impact assessment."

Gurin, Joel et. al. (2015). "Sustainable Development Goals and Open Data." *World Bank Blogs*.

Open data increase transparency, participation, engagement, and opportunities for value creation by all sectors of society. This includes enhanced decision making and allocation of resources by government and other sectors, which lead to improved quality of services and more efficiency and innovation. Open data empower citizens to acquire knowledge from a wider range of sources to engage more effectively in civic life and to hold their governments to account.

Wessels, B. et al. 2017. "Open Data and the Knowledge Society." *Visions of open data*, Ch. 3. Amsterdam University press.

3. HOW CAN PRIVACY BE RECONCILED WITH OPEN DATA?

Teresa Scassa provides an overview of how privacy concerns impact open data. She frames it as a matter of finding a balance between transparency and privacy and highlights some of the challenges, including the commercial reuse of open data, and variance in anonymization techniques. This article also shares the ODI's Open Data Spectrum that shows the range of data from closed to shared to open. How each government provides training and resources to strike this balance will have a significant impact on a government's place along the spectrum. This article concludes that what is required for balancing privacy and transparency is "a process for determining whether the data can be adequately anonymized to protect privacy while furthering the release of open data."

Scassa, T. (2019) "Issues in Open Data - Privacy." 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.

Pittman and Appel's discussion of the collection of gender-related data includes a section on privacy protection and informed consent in the context of mitigating the challenges of big data. "The many possible interconnections of big data sources make it difficult to guarantee irreversible de-identification." They share the United Nations Sustainable Development Group (UNSDG) guidelines on data privacy, data protection, and data ethics concerning the use of big data, which emphasize the importance of the knowledge and proper consent of the individuals on which the data are collected.

Pittman, A., & Appel, D. (2021). "Feminist Data Collection: Building a Vision of an Inclusive System." *From Bias to Feminist AI*.

4. WHAT DOES OPEN DATA MEAN FOR AI AND MACHINE LEARNING?

An article by Deloitte discusses the opportunities and risks that open data could have for AI and machine learning. Large amounts of open data released by governments, nonprofits, think tanks, and private companies can be used to train AI models. "Opening up data for AI use can unlock huge value for society—from finding cures for lethal diseases, to combatting climate change, to effectively responding to crisis, the potential is immense." However, it also acknowledges that "using massive public datasets to train models can unintentionally undermine privacy or perpetuate encoded biases." The article argues that to harness benefits and minimize risks, "agencies should play an active role to protect the data from both intentional tampering and unintentional inaccuracies."

Austin, Tasha et. al. (2021). "Trustworthy open data for trustworthy AI: Opportunities and risks of using open data for AI." Deloitte.

Jennifer Cohen and Homi Kharas argue that “... spatial disaggregation and timeliness could permit a process of evidence-based policy making that monitors outcomes and adjusts actions in a feedback loop that can accelerate development through learning. Big data and artificial intelligence are key elements in such a process. Emerging technologies could lead to the next quantum leap in (i) how data is collected; (ii) how data is analyzed; and (iii) how analysis is used for policymaking and the achievement of better results. Big data platforms expand the toolkit for acquiring real-time information at a granular level, while machine learning permits pattern recognition across multiple layers of input. Together, these advances could make data more accessible, scalable, and finely tuned. In turn, the availability of real-time information can shorten the feedback loop between results monitoring, learning, and policy formulation or investment, accelerating the speed and scale at which development actors can implement change.”

Cohen, J. L., & Kharas, H. (2018). *Using big data and artificial intelligence to accelerate global development*. Brookings.

Lindsey Anderson anticipates that “Artificial intelligence (AI) will soon be at the center of the international development field. Amidst this transformation, there is insufficient consideration from the international development sector and the growing AI and ethics field of the unique ethical issues AI initiatives face in the development context. This paper argues that the multiple stakeholder layers in international development projects, as well as the role of third-party AI vendors, results in particular ethical concerns related to fairness and inclusion, transparency, explainability, and accountability, data limitations, and privacy and security. It concludes with a series of principles that build on the information communication technology for development (ICT4D) community’s Principles for Digital Development to guide international development funders and implementers in the responsible, ethical implementation of AI initiatives.”

Andersen, Lindsey. (2019). *Artificial Intelligence in International Development: Avoiding Ethical Pitfalls*.

5. HOW HAS THE OPEN DATA MOVEMENT EVOLVED?

A review of open data and official statistics by Open Data Watch finds the roots of the open data movement in the open-source, open science, and government transparency and accountability movements. Open data was also discussed in the open-source and computer science communities, which saw benefits to sharing code and data for reuse and made the case for open government data. National governments began implementing open data as official government policy. With national governments working to implement open data, there was renewed pressure on international organizations to do the same. The movement is ongoing, but open data has become a central part of international governance.

Open Data Watch. 2021. *Open Data for Official Statistics: History, Principles, and Implementation*.

The article by Open Data Watch (2021) provides a rough timeline of the open data movement that is reproduced here with minor changes:

Foundations of the open data movement:

1994: The United Nations Statistical Commission adopts the **Fundamental Principles of Official Statistics** (FPOS).

2003: **Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities** is signed.

Making the case for open government data:

2007: A **meeting of open government advocates** organized by Lawrence Lessig and Tim O'Reilly, makes the case for open government data as essential to democracy and sets forth initial open government data principles.

National governments implementing open data:

2009: The United States adopts **open data as official government policy**.

2013: The United Kingdom Cabinet Office publishes a policy paper proposing an **"open data charter"** for the G8 group of nations.

2015: **Open Data Charter** is formally adopted.

International organizations implement open data:

2010: World Bank announces **open access to their statistical databases** in advance of a **new access to information policy**.

2013: The **UN Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda** called for a data revolution to ensure accountability and monitor delivery, taking advantage of new technologies and access to open data for all people.

2017: Open data is included as a central part of the **Cape Town Action Plan for the Sustainable Development Goals**.

2020: **Data Strategy of the Secretary-General for Action by Everyone, Everywhere** identifies ways for sharing of data and statistics and strengthening UN role as a global data hub.

6. WHAT ARE THE PRECONDITIONS FOR MAXIMIZING THE BENEFITS OF OPEN DATA?

The GovLab's *Periodic Table of Open Data Elements* describes the enabling conditions and disabling factors that contribute to the success or failure of open data initiatives. Each context faces unique challenges, but current research and practice identifies five central issues that will either enable or disrupt the success of open data projects across different countries: Problem and demand definition, capacity and culture, partnerships, risks, and governance.

The GovLab. (2022). *Periodic Table of Open Data's Impact Factors*.

The World Bank's *Open Data Readiness Assessment* helps countries plan what actions governments should take in establishing an open data program. As a diagnostic and planning tool, it covers eight dimensions that it considers essential. These include senior leadership; policy/legal frameworks; institutional structures, responsibilities, and capabilities within government; government data management policies, procedures, and data availability; demand for open data; civic engagement and capabilities for open data; funding an open data program; and national technology and skills infrastructure.

World Bank. n.d. *Open Data Readiness Assessment (ODRA)*.

Overcoming Data Graveyards in Official Statistics provides an overview of institutional barriers that could limit the use of official statistics and thereby impede maximizing the benefits of open data. The paper focuses on “three core capacity challenges: i) the capacity of public sector entities to govern their data in a way that promotes its use while providing safeguards to prevent its misuse; ii) the capacity of official statistics producers to develop the products and services that meet user needs; and iii) tailoring the dissemination of official statistics in ways that correspond to the data literacy of their user communities.” Through the identification of these barriers, it should be possible to move from “the lack of data use in decision-making and general lack of trust and awareness of NSS data towards more effective and sustained use of data for decision-making.”

Open Data Watch & SDSN TRenDS. (2022). *Overcoming Data Graveyards in Official Statistics: Catalyzing Uptake and Use*.

Solutions to Close Gender Data Gaps includes a discussion of the enabling environment necessary for the implementation of solutions to close gender data gaps, but this also applies to open data. These solutions are only possible with governance processes that establish legal and strategic conditions and financial resources. Data producers must also create a culture of data use in decision-making and evidence-based policy formulation to catalyze data use and uptake.

Open Data Watch & Data2X. (2022). *Solutions to Close Gender Data Gaps*.

7. WHAT ARE THE ROLES OF DIFFERENT SECTORS OF SOCIETY WITH RESPECT TO OPEN DATA?

A wide range of stakeholders are involved in supporting open data in various capacities. From ensuring the publication of accessible open data, advocating for improvements, to promoting its use, these organizations serve many different functions. Many organizations serve multiple roles. Key organizations that support open data as a core goal are: United Nations Statistics Division (UNSD), Open Government Partnership (OGP), Open Data Charter, ODW, Open Data Institute (ODI), GovLab, and others.

Many stakeholders are both producers and users of open data. Local and national governments both collect official statistics and use them for evidence-based policies. Civil society organizations can facilitate engagement with respondents, use data to inform

efforts, and connect others to data as intermediaries. International organizations collect and publish data, fund data collection, and use data to inform efforts. They also serve as standard setters. Academia and scientific community conduct in-depth research using data they have collected or repurpose existing data. They can also help inform standards and develop methodologies. The private sector often collects massive amounts of big data and uses these same data as well as other sources to inform business decisions. The press or media tends to largely serve as a data intermediary, connecting citizens to information through data journalism. And philanthropic donors may fund data activities along the entire data value chain.

Open Data Watch. n.d. "Stakeholder Mapping." (Unpublished research)

8. WHAT ARE SOME OF THE BEST USE EXAMPLES OF OPEN DATA?

The GovLab project on Open Data in Developing Economies has compiled an inventory of the use of open data in developing economy case studies.

The GovLab. n.d. "Appendix A." Open Data in Developing Economies Appendices.

Humanitarian Open Street Map has been used in the response to the earthquakes in Haiti and Nepal, and the Typhoon in the Philippines to provide first responders with the open data they needed to find and rescue survivors. The Red Cross has coordinated its use in relation to the super typhoon Haiyan to help improve the situation in the Philippines. Volunteers can trace buildings and key infrastructure from satellite imagery. This helps inform humanitarian assistance on the ground.

Meyer, Robinson. (2013). Humanitarian Open Street Map. The Atlantic.

Open data, accessed via computers or through text messages on cell phones, can provide farmers with the most up-to-date information on prices and data on weather patterns that could affect their harvest. This overall project has worked to create sustainable employment for rural youth through active engagement in agribusiness for wealth creation and poverty reduction.

The Technical Centre for Agricultural and Rural Cooperation (CTA), Ustadi. (2020). Youth Economic Empowerment Through Agribusiness in Kenya (Vijabiz).

Nepal is currently focusing on building transparent and accountable public institutions following a period of disruptive civil war. By 2013-14, foreign aid represented 22 percent of the national budget and financed most development spending. NGOs, journalists, and civil society have demanded more comprehensive, timely and detailed information on aid flows, particularly geographic information, to show where money is being directed.

The Open Data Institute. "Effective Aid in Nepal." Open Data Handbook.

Open data can be used to advocate for more equitable public health outcomes: Following a lead poisoning epidemic in the Zamfara State of Northern Nigeria — the result of the local artisanal gold mining operations — a local non-profit organization, Follow The Money, took immediate action against their corrupt local government who they presumed was not distributing aid funds fairly. They launched the #SaveBagega initiative that relied heavily on visualizations and reports based on public data — some of which was open data and some of which they opened up through their campaign efforts — showing clearly where the previously released disaster relief funds should have gone. A global media outcry ensued and in January 2013, the voices of the people of Bagega had reached about 1 million people. Their story had been told by about 50 media organizations. By the end of January 2013, the federal government of Nigeria released about \$5.3 million for the cleanup of Bagega from the Ecological Funds, through the Ministry of Finance.

Ghosh, Rahul. "Extractives Remediation and Public Health." *Open Data Handbook*.

9. WHERE DO WE SEE MORE AND WHERE DO WE SEE LESS PROGRESS ON OPEN DATA AND WHY?

Arturo Rivera Perez and Cecilia Emilsson have discussed the findings and key policy messages of the 2019 OECD Open, Useful and Re-usable data (OURdata) Index. A growing number of OECD countries have scaled up the adoption of 'open by default' approaches by including formal requirements in open data strategies, laws, regulations, and other instruments. Governments are increasingly enabling their open government data portals as communication and feedback tools. Formal open government data requirements are essential but insufficient to ensure the release of, and access to, re-usable datasets on portals of good quality and which respond to a specific purpose or demand. OECD countries are growing more aware that availability of valuable open data is more relevant than quantity. This led to an acceleration of initiatives guiding the standardization and production of good quality data from earlier stages of the data value cycle (data generation). Across OECD countries, open data is becoming more intertwined with the development of government-wide data governance frameworks (e.g., national data strategies) and data management capacities, aimed to build an overall environment within the public sector that enables and incentivizes data re-use.

Rivera Perez, J. A., & Emilsson, C. (2020). *OECD Open, Useful and Re-usable data (OURdata) Index: 2019*.

The latest report of the Open Data Inventory (ODIN) highlights how official statistics have become more open in recent years. In addition to increases in coverage, more countries are publishing data in open formats, disseminating data through data portals, publishing more metadata, and adopting an open license or terms of use.

Open Data Watch. (2021). *Open Data Inventory Annual Report 2020/21*.

Open Data Watch has reviewed selected open data indexes and assessment tools that show where countries are making progress in terms of open data and development data in general. In addition to discussing each of the indexes and tools and their differences, this paper also highlights recent efforts to improve the measurement and monitoring of open data use. Measuring the impacts and use of open data is critical to making a case for increased investments.

Open Data Watch. (2022). Better Measurement and Monitoring of Data for Development.

10. HOW CAN I OR MY ORGANIZATION CONTRIBUTE TO THE OPEN DATA MOVEMENT?

The Open Data Handbook describes what data publishers can do once they have made their data publicly available. It emphasizes the importance of dissemination and connecting it to those that need it. A publisher can hold events to help publicize data and bring people together. They can also support hackathons by giving programmers a short time frame to develop applications using the data.

Open Knowledge Foundation. n.d. "So I've Opened Up Some Data, Now What?" Open Data Handbook.

To ensure that open data achieve use and impacts, publication is not enough. A transformation process is needed to overcome Institutional barriers that limit the use of official statistics and increase the capacity of data producers and public sector entities to i) govern their data in a way that promotes their use while providing safeguards to prevent misuse; ii) develop the products and services that meet user needs; and iii) tailor the dissemination of official statistics to match the data literacy of their user communities. This can be accomplished through a perpetual process of feedback and reflection to improve capacities in a way that meets user needs.

Open Data Watch & SDSN TRenDS. (2022). Overcoming Data Graveyards in Official Statistics: Catalyzing Uptake and Use.



Open Data Watch is an international, non-profit organization working at the intersection of open data and official statistics. Our work provides assistance to organizations that produce and manage official statistics focusing on how to use open data to meet Sustainable Development Goals. Concentrating efforts in three areas — policy advice, data support, and monitoring — Open Data Watch seeks to make development data better and more accessible for increased use and impact.

Learn more: www.opendatawatch.com
Sign up for our newsletter: bit.ly/StayingUpToDataODW