

# Open data for official statistics: History, principles, and implementation

Shaïda Badiee\*, Jamison Crowell, Lorenz Noe, Amelia Pittman, Caleb Rudow and Eric Swanson  
*Open Data Watch, 1110 Vermont Avenue (Suite 500), Washington, DC, USA*

**Abstract.** For data that are collected and managed by national statistical offices to reach their full potential and benefit to society, they must be made available to the public as open data. In the simplest terms, open data are data that can be freely used, modified, and shared by anyone for any purpose. This paper reviews the development of standards for the production and dissemination of open data. It discusses the implementation of these standards in national statistical systems and reviews tool kits, readiness assessments, and maturity models that are available to guide national statistical offices in the adoption of open data. The demand for open data has created challenges for official statistics, but it has also raised the profile of the statistical office and points to a new and expanded role as data brokers and data stewards. The paper concludes with a discussion of how open data in official statistics can be used to improve governance.

Keywords: Open data, licensing, microdata, privacy, governance, indexes

## 1. Introduction

The much-celebrated data revolution is premised on the possibility of open access to all forms of data, prominent among which are the data and statistics produced by the government agencies that constitute the national statistical system (NSS) [26]. The claims made on behalf of open data include greater transparency and increased efficiency of government functions, and the possibility of generating new products and private profits through innovative use of government data. A more fundamental claim is based on the right of the public to have access to information produced with public resources [99].

For data that are collected and managed by national statistical offices to reach their full potential and benefit to society, they must be made available to the public as open data. In the simplest terms, open data are data that can be freely used, modified, and shared by anyone for any purpose [72]. While this definition is straightforward, the implementation of open data can still pose technical and organizational challenges for statistical offices. This paper reviews the development

of standards for the production and dissemination of open data. It discusses the implementation of these standards in national statistical systems and reviews tool kits, readiness assessments, and maturity models that are available to guide national statistical offices (NSOs) in the adoption of open data.

Because data users frequently want to combine data from multiple sources, including privately-held databases, open data has increased the demand for data interoperability and the use of standard classification systems. There is also growing demand for access to more finely disaggregated data and microdata that can, under certain circumstances, reveal the identity of individuals or other entities. This paper addresses how to balance privacy issues with the public's right to open microdata and suggests strategies for achieving this balance. The demand for open data has created challenges for official statistics, but it has also raised the profile of the statistical office and points to a new and expanded role as data brokers and data stewards. The paper concludes with a discussion of how open data in official statistics can be used to improve governance.

This paper will focus specifically on open data for official statistics, although many of the concepts and principles are applicable to data produced by other functions of government, academic researchers, civil society organizations, and even private entities. Official statis-

---

\*Corresponding author: Shaïda Badiee, Open Data Watch, 1110 Vermont Avenue (Suite 500), Washington, DC 20005, USA. Tel.: +1 301 320 3020; E-mail: info@opendatawatch.com.

tics are defined by the International Association for Official Statistics as statistics made through activities that are normally carried out by “official” bodies belonging to governmental administrations [29]. These datasets produced by governments – typically by the national statistical office – as a part of their official function include some of the most important data that decision makers need to create policies, evaluate programs, and allocate resources.

### 1.1. History of the open data movement

The open data movement has its roots in the open-source, open science, and government transparency and accountability movements. The first use of the term open data comes from *On the Full and Open Exchange of Scientific Data* [47], which called for making environmental data available to the public so that scientists could study the global environment that transcends borders. The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities called for “a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose” [39].

The Fundamental Principles of Official Statistics (FPOS) set out the professional and scientific standards for NSOs and provides the foundation for the open data for official statistics movement. The United Nations Statistical Commission in 1994 adopted the FPOS, which had been adopted by the European Conference of Statisticians in 1992. The first of the ten principles anchors the need for statistical information in democratic governance and recognizes the citizens’ entitlement to public information:

Official statistics provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honor citizens’ entitlement to public information [99].

The commentary on the first principle anticipates the subsequent discussion of open data as requiring dissemination of data in usable formats and with suitable metadata:

[M]aking information available on an impartial basis requires dissemination activities, which provide information in the form useful for the users, and release policies, which provide equal opportunity of access. Sound statistical principles need to be followed when presenting statistics so that they are easy to understand and impartially reported [102].

Open data was also discussed in the open-source and computer science communities, which saw benefits to sharing code and data for reuse. In December of 2007, two prominent thinkers on computer science and the internet, Lawrence Lessig and Tim O’Reilly, organized a meeting in Sebastopol, California, to make the case for open government data. The preamble to the meeting’s list of eight principles articulated a fundamental claim about open data: “Information becomes more valuable as it is shared, less valuable as it is hoarded. Open data promotes increased civil discourse, improved public welfare, and a more efficient use of public resources. . . . By embracing the eight principles, governments of the world can become more effective, transparent, and relevant to our lives” [83]. The need for data to better understand and solve global problems has become a common theme in support of open data. Trevor Manuel, then Minister of Finance of South Africa, noting the need for good data to guide development policies, quoted Peter Drucker: “What can’t be measured can’t be managed” [38].

In 2009 the United States adopted open data as official government policy [49], creating a new repository for United States government data of all types. A year later the United Kingdom followed suit, announcing a “One-stop shop for Government data” [100]. In 2013 the United Kingdom Cabinet Office published a policy paper proposing an “open data charter” for the G8 group of nations [97] that was adopted in June of that year.

With national governments working to implement open data, there was renewed pressure on international organizations to do the same. One of the first movers was the World Bank, which announced open access to their statistical databases in April of 2010 in advance of a new access to information policy taking effect that July [110,111]. The World Bank’s adoption of open data was spurred by a hero of the open data movement, Hans Rosling, whose advocacy and work to make data applicable, interesting, and useful to policy makers and citizens alike, provided a huge service to the community [33]. Opening the World Bank’s database, particularly the World Development Indicators, which contained data from many of the United Nations’ special-

ized agencies, would help to normalize open data in the international realm.

The G8 Open Data Charter provided the model for the seven principles of the International Open Data Charter [57]. Open data was included as a central part of the Cape Town Action Plan for the Sustainable Development Goals in 2017, which further cemented open data's role in international governance [22]. To enshrine the concepts of open data for official statistics, a working group was established to add open data concepts to the Fundamental Principles of Official Statistics at the 50<sup>th</sup> meeting of the United Nations Statistical Commission [101]. That work is ongoing.

### 1.2. *Benefits of open data*

The case for open data has often been stated in political or ethical terms: citizens 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.

The economic case for open data is rooted in the theory of public goods [86,90]. Public goods have two attributes: they are non-rivalrous, meaning they are not reduced by use; and they are non-excludable, so they are available to everyone. The classic examples of a public good are national defense or a lighthouse whose benefits are shared by all without diminishing anyone's use. Other examples are air pollution – a public bad – and over-the-air television broadcasts. In all these cases, it would be inefficient or impossible to charge individuals for the use of the good.

Information produced by governments is clearly non-rivalrous: one person's knowledge of census statistics or use of meteorological information does not reduce the amount of information available to others. However, unlike national defense, access to information can be made exclusive by law or technical means, thus limiting access, and potentially allowing the producer to discriminate among users. At this point, the theory of public goods confronts the general principle of economic efficiency: goods – including intangible goods like statistical information – should be priced at their marginal – not average or total – cost of production. Information once it has been created costs very little to reproduce and disseminate. Indeed, the hallmark of the data revolution is that digital technology and global networks have reduced the marginal cost of delivering data to near zero [23,26]. For NSOs and other public

agencies that must bear the cost of producing statistics, this presents a conundrum: the efficient price for open data will not cover their cost. Budgets for the statistical agency or for the publications department may include revenues from the sale of paper publications and physical media such as CD-ROMs, often at significant mark-ups. With the advent of internet dissemination, the need for these modes of publication have largely disappeared and the shortfall in revenue must be made up in other ways. Governments could still charge for access to their information, but the welfare gains to the public can more than justify a loss of revenues.

Economists have quantified the value generated from opening datasets. One of the early examples is Pollock [81] who estimated the value of providing free access to public sector information that had been previously sold to be GBP 1.6–6 billion, 4–15 times the forgone sales revenue in the United Kingdom. One of the biggest examples of the economic impact of open data is the United States opening GPS data for civilian use, which is estimated to have created around USD 100 billion in economic value [121]. Research on the opening of Landsat satellite data and weather data from the National Oceanic and Atmospheric Administration (NOAA) in the United States points to similar economic and societal benefits [120]. Annual savings from the open Landsat data for non-governmental organizations, the federal government, and the private sector are estimated at between USD 350 and 436 million per year [34]. Beyond the direct financial benefits, the opening of data can produce a wave of innovation and new products and services as is evidenced by the many apps and products that have been developed and built from the opening of GPS data.

There are other benefits, besides financial, in international development to opening data. Open data has been critical to saving lives in natural disasters and emergency situations. Humanitarian Open Street Map has been used for 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 [40]. 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 [25]. Open data is a central part of the strategy for achieving the Sustainable Development Goals (SDGs) and is a critical part of the effective allocation of resources to solve international development challenges.

The case for open data is usually based on the economic and social benefits and contribution to good gov-

ernance described above, but there are also benefits for NSOs that take the lead on open data. Opening official statistics can increase their use and raise the profile of the NSO. Opening data can help the NSO to start a virtuous cycle of use in which, through the provision of high-quality data, the NSO enables decision makers to increase their impact, and they, in turn, provide more funding for the NSO [98]. In more simplistic terms, open data gives NSOs the opportunity to prove their inherent use to the public and to people in positions of power who may control funding.

### 1.3. Defining open data

In 2005 Rufus Pollock and the Open Knowledge Foundation [80] proposed a comprehensive definition of open data, built on the foundations of the Open-source Definition [72]. Pollock's proposal distinguished requirements for open access, which he called "social openness," from open licensing. Social openness meant: "The work shall be available as a whole and at no more than a reasonable reproduction cost, preferably downloading via the Internet without charge. The work must also be available in a convenient and modifiable form." The subsequent 10 principles all specify the terms of an open license. Originally titled Open Definition version 1.0, it would undergo multiple revisions before arriving at version 2.1 in 2015 [82], but the core distinction between technical openness and legal (or licensed) openness is characteristic of all subsequent definitions of open data [2].

In the introduction to the Open Definition 2.1, the Open Knowledge Foundation defines open as: "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)." And defines open data as: "Open data and content can be freely used, modified, and shared by anyone for any purpose" [72]. In its current form the Open Definition 2.1. has four elements or principles that define an open work:

#### 1.1 Open License or Status

The work must be in the public domain or provided under an open license.

#### 1.2 Access

The work must be provided as a whole and at no more than a reasonable one-time reproduction cost, and should be downloadable via the Internet without charge.

#### 1.3 Machine Readability

The work must be provided in a form readily processable by a computer and where the individual

elements of the work can be easily accessed and modified.

#### 1.4 Open Format

The work must be provided in an open format. An open format is one which places no restrictions, monetary or otherwise, upon its use and can be fully processed with at least one free/libre/open-source software tool.

The second part of the Open Definition presents nine criteria for an open data license. Many data publishers have adopted a Creative Commons license that satisfies these criteria. However, licensing terms must also be consistent with local and national laws governing copyright and intellectual property and the contractual authority of the national statistical office, so many variations are possible. Open data licenses and the Open Definition criteria are discussed in more detail below.

A second important document defining open data is the International Open Data Charter's Principles of Open Data [56]. The Open Data Charter is a set of principles and best practices for how and why government data, including official statistics, should be opened. The six principles and their accompanying commentary are described as "aspirational norms" for the dissemination and use of open data:

1. Open by Default
2. Timely and Comprehensive
3. Accessible and Usable
4. Comparable and Interoperable
5. For Improved Governance and Citizen Engagement
6. For Inclusive Development and Innovation

The elaboration of these principles includes useful advice for implementation of open data. The first principle has caused the most debate. "Open by default" is generally understood to mean that data (and other government information) should be open except where there are specific reasons for limiting access. These reasons include the protection of security, privacy, confidentiality, and intellectual property. Thus "open by default" requires a "negative list" of data that are not open; all other data may be assumed to be open. Put simply, "open by default" means that producers, users, and policymakers should presume that data will be published as open, unless there is a justifiable reason for it not to be.

The International Open Data Charter has been adopted by 74 national and local governments and endorsed by 55 organizations and non-state actors. Taken together the Open Definition and Open Data Charter

provide a working definition of open data that has been well adapted to the needs of many official statistical agencies.

The open science community has embraced a similar set of guidelines for the sharing of scientific research materials: the Findable, Accessible, Interoperable, Reusable (FAIR) principles. The FAIR principles for data were formally outlined in 2016 as a part of a movement to improve the infrastructure supporting the reuse of scholarly data [107]. Mons, Barend et al. [42] make the point, equally applicable to the Open Data Charter, that these are guiding principles, not standards and their implementation may be context specific. They argue that data should not be “Open by default” but “Accessible under well-defined conditions,” as they take into account some of the reasons noted above for limiting public access to certain data.

## 2. Implementing open data in national statistical systems

Implementation of open data in national statistical systems is where principles and good intentions meet operational realities. Open data is an activity, not just a one-time commitment. Successful implementation of an open data program depends on sustained political and financial commitment and goal-oriented management of data content, delivery systems, and user engagement. For some statistical agencies, moving to open data may be a small step, requiring only the adoption of an open license or providing a new download format. But for others it will be a larger challenge, requiring new training, new equipment, and new sources of funding. Careful preparation and self-assessment can ease the transition from a closed model of data dissemination to an open one. The following section will dive into common implementation steps for open data for relevant entities within governments, such as NSOs, including pitfalls and country examples.

### 2.1. Planning for open data

Planning for open data can take place through various mechanisms. Developing a National Strategy for the Development of Statistics (NSDS) is how some low- and lower-middle-income countries have planned for including open data in their official statistics [76]. PARIS21 recommends incorporating open data into each phase of the development of an NSDS [77]. As of May 2019, nearly sixty percent of all International De-

velopment Association (IDA) borrower countries were implementing their NSDS, although not all included open data as part of their strategy [78].

The World Bank offers the Open Data Readiness Assessment (ODRA) as a diagnostic and planning tool to help statistical offices and their senior management prepare for the adoption of an open data program [116]. Statistical agencies that have already taken steps to implement open data may also find the Open Data Institute’s Open Data Maturity Model a useful guide for maintaining improvements to their systems and organization [59].

An ODRA can be conducted by internal staff but more often an outside expert works with a client team. The process described by the ODRA User’s Guide is intended to be carried out over a period of nine weeks and involves eight dimensions [113]:

1. Senior Leadership – What is the organizational commitment and willingness to adopt open data? Is there high-level political support for open government or open data and does human capital support exist in the form of open data champions or specific officers tasked with implementing open data?
2. Policy/Legal Framework – What is the legal framework within which open data will operate, as demonstrated by existing laws concerning freedom of information, privacy, and licensing?
3. Institutional structure, responsibilities, and capabilities within government – What is the organizational setup within which open data can be implemented on a technical and bureaucratic basis, with an emphasis on open data literacy and performance assessment for service delivery?
4. Government Data Management Policies and Procedures – What are the systems for data management: data inventories, digitization efforts, and data interoperability standards across government, among others? Are key datasets available to be published as open data?
5. Demand for Open Data – What are the “pull” forces for open data from outside of government, as evidenced by civil society users of official statistics and government practices to involve citizens in policy decisions?
6. Civic Engagement and Capabilities for Open Data – Does the statistical office engage with outside stakeholders on sourcing feedback and joint data efforts? What is the environment for data and technology in the country, as evidenced by technical university graduates and app development, for example?

7. Funding an Open Data Program – Is the government willing and able to fund open data programs based on existing and planned funding for open data, as well as funding for data and technology more broadly.
8. National Technology and Skills Infrastructure – This component examines the availability and robustness of national technology infrastructure and the context within which open data applications and systems can be developed, for example, by looking at the existing web presence of government, the availability of e-services, and country internet penetration.

All eight components are important. They complement one another, and weakness in one can compromise others. When Vietnam completed an ODRA in 2018, the assessors found that while demand for open data was present, the legal framework did not provide a supportive environment for open data implementation [115]. In the years since, Vietnam has put in place new digital data management, sharing, and open data policies [35] and recently launched its open data portal [41].

The cost of implementing open data depends on existing capacity, both human and technical, and the scale of the open data program. The Open Data Institute [58] provides advice on planning and budgeting an open data initiative. Whether the plan is to develop a separate portal for official statistics or to implement a whole of government open data program, implementing open data requires staff training, software and hardware acquisition, and communication strategies, all of which require funding, again underscoring the importance of assessing the political and financial support a government is able to offer for open data implementation.

## 2.2. Implementing open data

Having assessed and planned for open data within the NSS, implementing agencies must put in place suitable policies to “aid both data consumers and data producers by clearly outlining the standards, processes and requirements for offering and acquiring public information” [117]. They can guide very high-level processes, such as the EU Implementation of the G8 Open Data Charter [12], or describe detailed activities for implementing open data as in the case of Ireland’s Open Data Strategy 2017–2022 [19].

Open data initiatives will need a champion to lead the initial foray into open data assessments and represent the interests of open data in budgetary and political discussions. The World Bank recommends including the

eventual implementer in the ODRA process, which in turn will require a pre-ODRA scoping process to ensure any resulting recommendations can be implemented by the members of the ODRA team [114].

Technical partnerships between the open data initiative and the responsible IT infrastructure team, either in-house or hired from outside, are crucial to open data implementation. Adoption of international standards can help countries struggling with the complexity of making their data open and interoperable. The Generic Statistical Business Process Model (GSBPM), while not explicitly focused on open data, nevertheless can give open data initiatives the structure they need to organize their program within the NSO and across line ministries. It is already used by 39 NSOs. Open data and further guidelines on interoperability can be added to the GSBPM, as described by the Friends of the Chair Working Group on Open Data [109]. The World Bank also makes a series of recommendations for implementing open data standards at a technical level, such as improving technical documentation and ensuring that public APIs and endpoints are interoperable [112].

There are many choices that must be made to implement an open data program: about the data to be included, the format of their publication, the user interface and tools provided, and the terms of use. There is no single standard for open data and particularly in the case of microdata, there may be reasons not to make data fully open. The following section describes some of the practical choices that satisfy the core principles of open data.

### 2.2.1. Machine readability

Machine-readable data are structured data in formats that can be read and processed by a computer. These include Extensible Markup Language (XML), JavaScript Object Notation (JSON), comma-separated values (CSV), and Microsoft Excel’s Open XML Spreadsheet (XLSX). When data are made available in formats that are not machine readable, users cannot easily access and modify the data, which severely restricts their use. Datasets, in particular very large datasets, on their own convey little information to a human. Only when those data are processed in some way – visualized, analyzed, or summarized – do they become informative or useful [96].

In many cases, countries only publish statistics through annual statistical yearbooks or other PDF publications. These publications can be helpful to users, as the text in conjunction with the tables gives context and explanation to the figures, which helps less

technical users understand the data. However, machine readability should be a high priority. The tables within these reports should be extracted and made available in a separate machine-readable file as well.

Many national statistical offices utilize data portals to provide machine-readable files, such as Moldova, Slovenia, Taiwan, Bosnia and Herzegovina, and many other countries who have used the PxWeb system provided free by the Swedish government to build their data portals [88]. However, many countries without data portals have successfully provided users both PDF and XLSX versions of the data published through annual publications, such as Rwanda's 2019 Statistical Yearbook.

### 2.2.2. Non-proprietary formats

To meet standards of openness, statistics should be published in multiple file formats to meet different users' needs. At least one of these file formats should be machine readable and non-proprietary.

Non-proprietary formats allow users to access data without requiring the use of a costly, proprietary software that may prevent some users from accessing the data. By definition, proprietary formats restrict either the ability to use the information or the ability to share the information [67]. Users typically need to pay to use the software, cannot read or modify the source code, and cannot copy the software or re-sell it as part of their own product [70]. On the contrary, non-proprietary data are available in formats that no entity has exclusive control over. The most common non-proprietary formats used by statistical offices are PDF, XLSX, and CSV. Many countries still publish data in XLS format, which is proprietary. Although XLS files can be opened with some open-source software, such as OpenOffice and LibreOffice, the format is based on BIFF (Binary Interchange File Format), whose use is restricted by various licenses.

### 2.2.3. Metadata

Datasets must be published with ample metadata to aid both discoverability and usability of the data [71]. Various metadata standards are used for statistical data, including Dublin Core, Data Catalogue Vocabulary (DCAT), and RDF Data Cube Vocabulary. The precise standard used is less important than ensuring that the standard is used across all datasets. Core pieces of information that should always be included are the source of data; definitions of indicators and concepts; publication dates; and contact information if users have questions. Datasets licensed with a requirement of attribu-

tion should include a preferred form of citation. Metadata should also be published in a centralized location close to the dataset. Serbia's Open Data Portal is a good example of how metadata should be published [87]. When navigating to a particular dataset, users can click on the "M" symbol to be redirected to extensive reference metadata from the same location used to download the dataset.

### 2.2.4. Download options

Users should have the ability to select the data they want or to download the whole data set. Bulk downloads are a key component of the Open Definition, which requires data to be "provided as a whole. . . and downloadable via the internet" [73]. Bulk downloads increase the accessibility of data, enabling users to download datasets efficiently. Bulk downloads should be implemented to the greatest extent possible and most appropriate for its users. For instance, if internet download speeds in the country cannot handle bulk downloads of an entire portal, bulk downloads for each dataset, publication, or category of data would be more appropriate.

In contrast to bulk downloads, Application Programming Interfaces (APIs) and custom download options allow users to download a subset of a larger dataset. This feature also increases accessibility, enabling users to efficiently extract specific data points for use, whether that be data for a particular year or geographic area.

### 2.2.5. Standardization and interoperability

Data standardization allows for collaborative research, large-scale analytics, and sharing of methodologies [50]. Standards are needed to provide a basis for assessing data quality, for comparing and cross-validating datasets, and making them interoperable [92]. To improve data standardization, the United States government established a National Information Exchange Model (NIEM). NIEM specifies agreed-upon terms, definitions, relationships, and formats – independent of how information is stored in individual systems [46]. With consistent terms, definitions, and structures, federal agencies can more easily integrate data from various entities.

Interoperability is the ability to join up data from various sources in a standardized and contextualized way [89]. For example, COVID-19 death counts show the age and sex distribution of the deceased but often lack other information about the deceased. After connecting death statistics to the 2018 US Census Small Area Income and Poverty Estimates Program, Adhikari et al. found that the excess burden of both infections and deaths was experienced by more impoverished and more diverse areas [1].

### 2.2.6. Data licensing

Adopting an open license is a core component of the definition of open data. For a license to be categorized as open, it should allow others to use, adapt, and distribute data, commercially or non-commercially, with the only limitation being attribution to the original source. Many countries neglect adopting or publishing data licenses. This may be due to the incorrect assumption that publishing data online implies authorization of public use. However, there can be widely different interpretations of what authorized use looks like to governments and users. An open data license is an opportunity for governments to encourage public use of their data by specifically addressing how people can use data, how they should attribute data, and what types of use, if any, are prohibited. These specifics encourage use and reuse of data by alleviating user concerns about legal ramifications of unapproved use. Data licensing – the point at which open data principles must adapt to national legal systems – is critical to the success of open data initiatives and is discussed in further detail below.

### 2.2.7. Barriers to access

The preceding describes choices that make data open. It is equally important to avoid choices that restrict access. Requiring payment for data is not compatible with the principles of open data. Users must not be required to purchase data to obtain access. However, payments for data visualizations, analytical services, or other value-added services are permissible, so long as the data are made available at no cost.

Requiring users to provide information about themselves to obtain access to data is also in conflict with openness. Some data portals offer users the ability to create accounts through registration to save datasets or create visualizations. This is acceptable, so long as users can still access and download datasets without registration. In other cases, data portal administrators ask users questions before granting access to data to better understand user's needs. However, if these questions are not voluntary, it is a violation of data openness standards. A better option is to allow users to voluntarily provide feedback.

## 2.3. Maintaining open data

To sustain an open data program governments and statistical agencies should foster an open data culture that ensures open data remains a priority through changing administrations [122]. One way to conceptualize the adoption of an open data culture and its benefits is

through the data value chain, which describes the progression of statistics from raw data to the publication of reliable statistics and their uptake to their ultimate role in providing evidence for policies that have an impact on peoples' lives [64]. In this framework, adopting an open data culture can help move NSOs farther along the data value chain to contribute to both uptake and impact of their data [109].

Self-assessment is an important practice for maintaining an open data program. The Open Data Maturity Model allows an organization to assess how well they publish and consume open data, identifying actions for improvement [59]. It was developed by the Open Data Institute and the United Kingdom's Department for Environment, Food and Rural Affairs for a public sector audience, but it can be applied to a wide range of organizations publishing or consuming open data. Organizations can use the model to assess operational and strategic activities around open data and compare themselves against others to highlight strengths and weaknesses. Through this, they can identify potential areas for improvement, adopt best practices, and improve processes.

Organizations conduct a self-assessment using the Open Data Pathway, an online app that facilitates the mapping of open data practices [60]. The model is based on five themes, each representing a broad area of activity:

1. Data Management Processes – This component examines the key business processes that underpin data management and publication including quality control, publication workflows, and adoption of technical standards.
2. Knowledge and Skills – This component examines the steps required to create a culture of open data within an organization by identifying the knowledge sharing, training and learning required to embed an understanding of the benefits of open data.
3. Customer Support and Engagement – This component examines the need for an organization to engage with both their data sources and their data re-users to provide sufficient support and feedback to make open data successful.
4. Investment and Financial Performance – This component examines the need for organizations to have insight into the value of their datasets and the appropriate budgetary and financial oversight required to support their publication. In terms of data consumption, organizations will need to understand the costs and value associated with their re-use of third-party datasets.



5. Strategic Oversight – This component examines the need for an organization to have a clear strategy around data sharing and re-use, and an identified leadership with responsibility and capacity to deliver that strategy.

Organizations can identify their own level of maturity on a scale of 1 to 5 for each of the themes, which charts an organization’s move from “ad-hoc uncontrolled processes to those that are repeatable, standardized and well-managed; [a move from] a reactive to a proactive approach within a particular area of study; [and a move from] isolated expertise [on open data. . .] through to wider organizational support” [59]. The maturity model has been used, among others, by the Queensland Department of Transport and Main Roads in Australia to meet its improvement targets for open data.

The insights derived from self-assessment tools and other review processes can be used to work towards an open data culture alongside monitoring the legal and policy environment of a statistical office. Policies shape behavior, so policies need to evolve with the open data program. For example, Ghana recently replaced its 35-year-old mandate with one that now gives the NSO greater control over data gathering, setting standards across the national statistical system, and coordinating and cooperating with international partners to further increase capacity [16].

Leadership and internal capacity are vital for launching an open data initiative and are equally important for maintaining open data. New positions such as a Chief Data Officer [93] can enshrine the importance of open data in the management structure. Though data literacy itself may be high in many of the agencies charged with producing data, open data literacy may not. Including open data in internal trainings and in onboarding materials can weave open data into the daily operations of data producing and publishing entities.

Engaging with civil society and other partners outside government is key to creating open data feedback loops from champions outside the government, as well as from within. The World Bank in its review of open data practices suggests that “[E]xternal stakeholders, unfettered by the constraints of official practice, can sometimes be powerful allies in lobbying for action, and it is the use of data by users that can best demonstrate the value of an Open Data initiative” [114]. There are a multitude of tools available to increase the accessibility of open data, such as those listed above, and they should be implemented in the course of adopting open data practices in order to increase awareness of open data. Public engagements such as hackathons, brief-

ings, press releases, blogs, and digital feedback mechanisms are also useful tools for establishing contacts with outside partners and receiving recommendations to improve open data practices. The United States Government, for example, publishes guidance for hosting open data engagement events that can be adapted to other countries’ situations [95].

Monitoring and measuring data use in conjunction with a data dissemination program can create a positive feedback loop that builds support for open data and the work of the NSO. With metrics on data use derived from website tracking programs such as Google Analytics, NSOs can make the case that their open data portals are more than just fancy websites: they are tools that are used by the public they serve. These metrics help NSOs to better understand the public’s data demands. Used to evaluate data dissemination efforts, they can point to potential gaps in dissemination strategies. Open Data Watch’s *Measuring Data Use* report [63] demonstrates some of the tools available and provides guidance for NSOs on monitoring data use on their sites. Monitoring and measuring data use can, in turn, improve data use and dissemination and strengthen a country’s open data culture by providing more value and use cases for open data.

### 3. Open data licensing

Data cannot be open if they are not accompanied by a license or a clear statement that they are in the public domain. Despite its importance, licensing is often disregarded when publishing government data. Perhaps governments or their statistical offices assume that data they publish are recognized as being open or they may not wish to commit to fully open data. Licensing is the first of four elements of the Open Definition. Even if the other three elements are satisfied, users without a license may be hesitant to use or share data.

In its most basic form, a license is simply a permit to do something. If a person or an organization owns something – whether tangible or intangible – they may license others to use it, and they may attach conditions to that license. An open data license in its simplest form waives copyright and other rights to the data and gives permission to use the data for any legal purpose. Even when data are not subject to copyright, other rights may exist under national law that should be waived by an open data license. Many statistical offices have adopted standard forms of licenses for their published data that satisfy the core principles of open data, but for others national laws or policies may require a bespoke license.

### 3.1. Copyright and open data

Copyright is a legal right to reproduce original works, license them for reuse, or otherwise assign a right over them to another person or entity. In many jurisdictions, the idea that data are copyrightable is highly contentious. Data are often interpreted in law as being mere facts, not original works. The proportion of girls in a school grade for instance is a fact, not an original work, but assembling and documenting a database on girls' education might be considered a creative work worthy of copyright; even in this instance, it would be the structure of the database itself that might receive *sui generis* protection and not the underlying data themselves. However, some countries consider all work by governments to be in the public domain and not subject to copyright under any circumstances.

If data are not copyrightable, then no license is necessary, but an explicit statement that the data are license-free should be included. Determining whether data are copyrightable can be complicated. Copyright laws vary greatly by country. Although there are international treaties between some countries that agree to uphold each other's copyright laws, there is no international copyright law [6]. The oldest such treaty, the Berne Convention for the Protection of Literary and Artistic works, outlines standards for determining whether a work is eligible for copyright, but they are not legally enforceable [108].

In the United States and many European countries, data are considered "facts" not covered by copyright law because they do not meet the threshold of originality, a criterion of the Berne Convention. In these countries, there are no legal barriers to releasing the data into the public domain license-free, especially if the data are owned by the government. However, in other countries not all data may be considered facts, or the data may be considered facts but "may be eligible for other appropriate or applicable forms of intellectual property protection, for example by virtue of *sui generis* database rights, depending on its jurisdiction of collection or creation" [10]. When possible, data producers should waive any of the following rights that may apply.

Moral rights, unlike other types of protection, are designed to protect the reputation of the creator rather than the economic value of a product. The specific aspects of moral rights vary by country, but generally include the right to attribution, how the work is displayed, and controlling how the work is altered or adapted [84]. In many cases, moral rights are not applicable to data. In the United Kingdom for instance, moral rights are only

applicable to "literary, dramatic, musical and artistic works and film, as well as some performances" [27]. However, in other countries moral rights may apply to statistical products, even those created as government works. For example, article 5 of Oman's Open Government License states "You are not permitted to take any act of compromising the data that would harm the honor and reputation of the data provider;" a provision added to protect the creator's moral rights [52]. Taiwan's Open Government Data License, Version 1.0 also protects moral rights, stating in Section 3.1, "By utilizing the Open Data provided under the License, User... shall make the reasonable efforts with respect to moral right protection of the third parties involved" [45]. Many open data licenses address moral rights by waiving them, but in most countries where moral rights exist, they cannot be waived. The most problematic aspect of moral rights as they pertain to statistical data is the protection of a data creator's reputation. If data are reused in a way that counters the producer's findings or a user criticizes the quality of the data, would this be considered damaging to the producer's reputation? If moral rights cannot be waived and they are applied to statistical data according to a country's laws, specific language should be included in the data license as to what would be considered an infringement of these rights.

In some countries, even when the data are not copyrightable, the databases that hold the data are, particularly when the design of the database or arrangement of data is an intellectual creation [53]. Database rights also protect the creator's rights in cases when its creation involves originality or complex design. In most cases, these rights do not apply to the content of the database but may, in certain cases, apply to items such as metadata to protect a creator's intellectual property rights [32]. If the creator's rights do apply to the content of a statistical database or its metadata, these rights should be waived in an open license.

In contrast to database rights, *sui generis* database rights may apply to databases and their content that qualify for copyright protection not due to originality in creation, but because the investment in obtaining and presenting the data was substantial [14]. This right is most notably recognized in the European Union, United Kingdom, and Russia. In 2017 a public consultation was held by the European Commission [13] to review both database rights and *sui generis* rights, but no changes to either resulted. These rights can be waived by adopting an open license.

### 3.2. Criteria for an open data license

While recognizing that the legal debate continues

about whether data can or even need to be licensed, an open license or explicit statement that data are license-free is an important mark of open data. A data license is an opportunity for data producers to encourage public use of their data by specifically addressing how people can use data, the form of attribution requested, and what types of use, if any, are prohibited. Even where data are “open by default,” licensing affirms policies pertaining to authorized use, alleviating user concerns about legal consequences of unapproved use. For example, a journalist accessing freely published official data might fear prosecution for using the data to criticize a government program. This fear has increased in recent years with a rise in “fake news” laws – often used to stifle reporting that government officials disagree with. Between 2016 and 2019, 65 journalists were imprisoned for publishing “fake news” worldwide [4]. A clear and unambiguous license can therefore confer a degree of legal certainty and protection for users that in turn gives them confidence in their use of the data.

The Open Definition lists nine criteria that a license must have to be considered fully open along with certain conditions that can be added to a license without affecting its openness [73]. An open license must irrevocably permit the following:

1. The license must allow free use of the licensed work.
2. The license must allow redistribution of the licensed work, including sale, whether on its own or as part of a collection made from works from different sources.
3. The license must allow the creation of derivatives of the licensed work and allow the distribution of such derivatives under the same terms of the original licensed work.
4. The license must allow any part of the work to be freely used, distributed, or modified separately from any other part of the work or from any collection of works in which it was originally distributed. All parties who receive any distribution of any part of a work within the terms of the original license should have the same rights as those that are granted in conjunction with the original work.
5. The license must allow the licensed work to be distributed along with other distinct works without placing restrictions on these other works.
6. The license must not discriminate against any person or group.
7. The rights attached to the work must apply to all to whom it is redistributed without the need to agree to any additional legal terms.

8. The license must allow use, redistribution, modification, and compilation for any purpose. The license must not restrict anyone from making use of the work in a specific field of endeavor.
9. The license must not impose any fee arrangement, royalty, or other compensation or monetary remuneration as part of its conditions.

The Open Definition also lists seven conditions or limitations that may be included in an open license. The most common is requiring attribution to the original source. The license may also include an integrity condition, requiring that modified versions carry a different name or version number. Many organizations specify that the user may make no claim of endorsement by the original source. While these provisions are discretionary, they may increase confidence in the data and prevent confusion about their provenance. For example, an attribution clause in a license means that data that might be reused several times over will always be traceable back to its original source. That may help the ultimate user determine whether the data are trustworthy.

Open licenses may include provisions, such as share-alike conditions, that require derivative works to be licensed under the same terms as the original. This condition does not alter the license’s openness, though it has sparked controversy in the open data community. It has been argued that share-alike licensing imposes a restriction on data use that conflict with the core principle of open data [106]. While share-alike licensing is intended to ensure the further dissemination of open data, it may have the unintended effect of inhibiting creative use of the data. For example, problems may arise when combining open data from several sources: if just one dataset has a share-alike condition, it can make it very difficult to pull together the various open data streams and re-publish them under a separate open data license.

A licensor can create a bespoke license that meets the criteria of the Open Definition. In practice, many data publishers choose a standard open license. Table 1 lists the most common open data licenses used for statistical data. Creative Commons is a non-governmental organization that has developed a set of licenses to facilitate the sharing of knowledge [7]. Open Data Commons is a project of the Open Knowledge Foundation [74]. Open Data Commons licenses are used less frequently for statistical data. In general, it is advised to not license data that do not fall within the domain of a country’s copyright law, unless there is some uncertainty that may warrant applying a Creative Commons CC0 or Public

Table 1  
Common data licenses used for official statistics

| Openness level | Data licenses  |
|----------------|--|
| Open access    | Creative Commons 0 (CC0): This is the CC's public domain dedication. By using this license, the dataset owner surrenders all rights to the public domain (unless moral rights cannot be waived).<br>Open Data Commons Public Domain Dedication and License (PDDL): This is the ODCs public domain dedication which serves the same purpose as CC0.<br>Creative Commons BY 4.0: This license allows users to use and reuse data for commercial and noncommercial purposes so long as users give attribution to the original source. |
| Limited access | Creative Commons BY SA 4.0: The "Share-Alike" license allows users to use and reuse data for commercial and noncommercial purposes so long as they cite the source and distribute the data and any derivatives under the same license.   |
| Closed access  | Creative Commons BY NC 4.0: This license allows users to use and reuse data for noncommercial purposes only so long as users give attribution.   |

Domain Dedication and License (PDDL) [91]. Otherwise, a statement that the data are license-free or in the public domain is sufficient.

The CC BY (attribution) license is frequently used to ensure that the original source of the data is identified in subsequent products. The Share-Alike license imposes an additional restriction by requiring that the same licensing terms be used for any derivative products. It is still considered an open license. However, the CC BY NC license, which is restricted to non-commercial use of data, fails to meet the standard for a fully open license. Creative Commons licenses have changed over time. In version 4.0, for example, moral rights are waived to the limited extent necessary to exercise the licensed rights. An extensive discussion of the terms of the licenses and the specific rights granted by them can be found on the Creative Commons website [8].

### 3.3. Restrictions on the misuse of statistics

As noted above, Creative Commons licenses waive the moral rights of the licensor to the extent possible under local law. But clauses prohibiting the misuse of statistics, or similar conditions, are sometimes included in bespoke data licenses, often at the expense of openness. They are problematic because the term "misuse" is generally undefined in these licenses, creating an unknown and unknowable liability for the data user.

Principle 4 of the Fundamental Principles of Official Statistics states that "The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics," which has been interpreted by some statistical offices as a mandate to restrict data use by aggressively pursuing broadly defined "misuse" of statistics. The Implementation Guidelines from the UN Statistical Commission recommend principle 4 be achieved through clear language in national and legal frameworks [103]. The specific examples of misuse dis-

cussed by the FPOS Implementation Guidelines include "overgeneralization," "misreporting or misunderstanding of estimated error," "false causality," and "data manipulation." When data are open, such misuse is more likely to be discovered, and as suggested by principle 4, the best response to misuse is to draw attention to it and criticize the perpetrators.

The Open Data Inventory (ODIN) reported in 2017 that 31 countries had data licenses (or terms of use that included data use provisions) that prohibited misuse, but used vague language, such as "misleading use is prohibited," without specific examples to describe misuse [63]. Many government agencies, in absence of a data license, use the terms of use of their websites to address data use provisions that would typically be included in a data license. Early versions of some countries' licenses, such as version 1 of United Kingdom's Open Government License, included similarly vague language about misuse, but this language has since been removed from later versions [44]. In later versions of their open government licenses, most countries have generally removed problematic language, but some that used the UK's license as a model have retained the original language prohibiting misuse.

To address instances of misuse, as defined by the FPOS Implementation Guide, such as claims of false causality and overgeneralization, many countries include a non-endorsement clause and disclaimer. For example, the Open License Agreement for the Central Statistics Office of the Commonwealth of Dominica states, "This license does not grant you the right to use the data in a way that suggests an official status or endorsement of you or your use of the data" and that they "will not be held responsible for damages resulting from its use or interpretation" [5]. These types of clauses do not necessarily prevent misuse but release the data producer from any responsibility for the misuse. Other countries have gone further, giving recommendations to data users that may help prevent misuse, such as Senegal's National Agency for Statistics and Demography,

which suggests that users not only provide attribution, but take into account the indicator's relevance before using, do not modify the specific figures, and provide explanatory information about the indicator in order to prevent misunderstanding [43].

#### 4. Special issues for open microdata

Microdata are data on the characteristics of members of a population, such as individuals, households, or establishments, collected by a census, survey, or experiment [104]. Because microdata contain information on the individual unit of analysis (whether individual, household, or other unit), they provide a richer dataset than macrodata and can be used to perform analyses across multiple dimensions, such as sex, age, location, or ethnicity. They are essential for validating previous analyses, testing new hypotheses, and measuring the impact of development policies and programs. The full benefits of microdata, however, will only be realized if they are made open and available to the public for use while respecting the obligation of statistical offices to uphold principle 6 of the Fundamental Principles of Official Statistics: "Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes."

The case for open microdata has been supported by many international agencies, including the OECD, citing open microdata as a crucial element in open government and democratic societies and as a driver of economic and social benefits through innovation and new uses of data [51]. The International Household Survey Network (IHSN) echoes these sentiments, adding enhanced credibility of national statistical offices and increased funding opportunities as other potential outcomes of open microdata [30].

##### 4.1. Statistical disclosure control

Despite international agreement on the importance of opening microdata, there is no straightforward definition of open microdata analogous to the Open Definition that applies to aggregate indicators or macrodata and to other non-personal government records. The Open Definition provides the basis for assessments of open macrodata (such as the Open Data Inventory and the Global Data Barometer), but there are no similar assessments of the openness of microdata. The lack of standards for open microdata is due in part to uncer-

tainty inherent in statistical disclosure control (SDC). The International Household Survey Network explains the problem this way:

SDC seeks to treat and alter the data so that the data can be published or released without revealing the confidential information it contains, while, at the same time, limit information loss due to the anonymization of the data. . . SDC is characterized by the trade-off between risk of disclosure and utility of the data for end users. The risk – utility scale extends between two extremes; (i) no data is released (zero risk of disclosure) and thus users gain no utility from the data, to (ii) data is released without any treatment, and thus with maximum risk of disclosure, but also maximum utility to the user (i.e., no information loss). The goal of a well-implemented SDC process is to find the optimal point where utility for end users is maximized at an acceptable level of risk [31].

To maintain the privacy of survey respondents, the right level of SDC needs to be found, but microdata from different surveys, geographical regions, and countries may have separate SDC requirements based on the sensitivity of the information, the presence of outliers in the data that may be de-anonymized, and a host of other factors. At the bare minimum, personally identifiable information such as names, email, addresses, and phone numbers should be removed from the data before making them public. A review of SDC practices and procedures is beyond the scope of this paper, but guidance on best practices for SDC is provided by the International Household Survey Network's SDC Practice Guide [30].

Because disclosure may cause harm to individuals or establishments, it may be necessary to exclude certain uses of the data or users from access to portions of the data. In many instances, the data have been collected with an explicit promise of confidentiality and that commitment may be legally enforceable. In some instances, data sharing, or providing access to data to a limited number of people, might be more suitable than open release of a dataset. Data sharing agreements are often used when data are particularly sensitive because of personal privacy, intellectual property, or other reasons. Microdata sharing differs from open data because the data are not made open and public to all but to a select group of people or organizations.

The World Bank's Microdata Catalog [118] defines five categories of access for surveys and censuses. (See Table 2). The same convention has been adopted by many national data archives. Open access is the only

Table 2  
World Bank Microdata Catalog access categories<sup>1</sup>

| Access category         | Description  |
|-------------------------|--|
| Open access             | Microdata are provided under terms of use that permit use or redistribution for any purpose whether commercial or non-commercial. No login is required to obtain data. These datasets substantially meet the stands of the Open Definition.  |
| Direct access           | Microdata are made available, but only to registered and unregistered users for statistical and scientific research purposes. Use is restricted solely to the user and may not be redistributed or sold to other individuals, institutions, or organizations without prior written agreement. Citation is required.  |
| Public use files (PUFs) | Microdata are made available online, subject to certain conditions. These data are made easily accessible because the risk of identifying individual respondents or data providers is considered to be low. Terms of use are the same as Direct Access, but users are required to register before obtaining access to the datasets.                                    |
| Licensed files          | Microdata are only available to certain users. Access is granted to authenticated users who have received authorization to access them after submitting a documented application and signing an agreement that specifies the purpose for which the data will be used. The users must be acting on behalf of an organization, who must take responsibility for the use. |
| No Access               | Microdata are not available or have no defined policy.   |

<sup>1</sup> Adapted from <https://microdata.worldbank.org/index.php/terms-of-use> and [https://dimewiki.worldbank.org/wiki/Microdata\\_Catalog#Access\\_conditions](https://dimewiki.worldbank.org/wiki/Microdata_Catalog#Access_conditions).

category that is indisputably open. Other categories imply restrictions that would disqualify them under the Open Definition [73] or by the standards applied by the Open Data Inventory to macrodata [65]. Another mechanism for protecting data confidentiality is the creation of data enclaves: physically controlled environments in which researchers may submit data requests but are not able to access the data directly. The National Opinion Research Center (NORC) at the University of Chicago, for example, offers a secure environment where public and private organizations can store data and provide controlled access to authorized users [48].

#### 4.2. Choosing a microdata license

Licenses for microdata can vary for a number of reasons, but the main determining factor is usually privacy concerns and statistical disclosure control. Different types of statistical data have different levels of disclosure risk and the level of risk determines the degree of openness the license will offer. Licenses, like data, are not open or closed, but rather exist on a spectrum. Data producers should strike a balance by pursuing the maximum level of openness appropriate to counter disclosure risk. Figure 1 shows the disclosure risks of various statistical outputs, along with the suggested access levels (or openness) of their applicable data licenses.

The movement to open microdata is at a similar place that the open data movement was years ago. Important work is being done to develop standards, document the availability of the microdata sets, identify data gaps, and advocate for better funding. Much of this work is expected to be led by the Inter-Secretariat Working Group on Household Surveys and its member organi-

zations [109]. As these data play an important role in many different policy and government decisions, the process of overcoming these challenges and making these data open and available, while protecting the privacy of the survey respondents, will be rewarded with increased use of this data leading to large societal benefits.

#### 5. Assessing data openness: Advocacy and rating systems

To advocate for improved availability and use of open data in national statistical systems, there must be unbiased assessments of open data to identify opportunities for improvement. These assessments are typically performed by NGOs, as they are better able to remain unbiased and stay above any political sensitivities that might limit large multilateral organizations from performing these assessments [3]. The results from these assessments can be used to: track international progress on open data to better address data use barriers and capacity issues; incentivize improvements in open data progress; and help provide NSOs with actionable insight on how they can make their data more open.

The assessments reviewed in this section are: Global Data Barometer, OECD OURdata Index, Open Data Inventory, and the European Open Data Maturity Assessment. A few self-assessments were noted in the section on the implementation of open data in national statistical systems but the assessments in this section are largely performed by outside assessors, although some of them request feedback or input from the organizations they are studying. These assessments can be further distinguished from the self-assessments as they

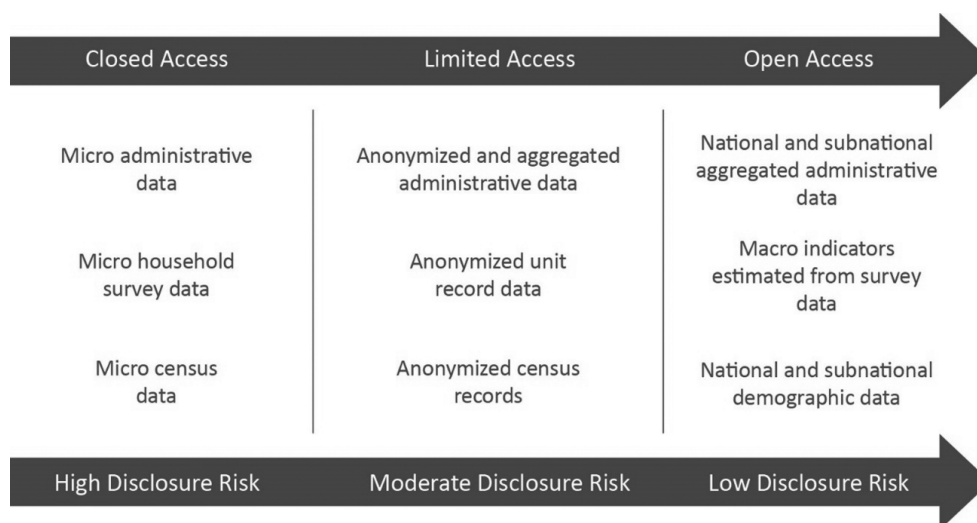


Fig. 1. Disclosure risk and access levels of different statistical outputs.

are designed to be shared with the public, as well as used as internal tools for improvement. Governments, academics, or others interested in the findings from these indexes may choose to review multiple indexes to get a bigger picture view of open data in the region or country and to identify opportunities for improvement.

A variety of options are available for assessing data openness. The sections below provide a more in-depth overview of the major assessments of open data, what they cover, and how they might be used. The Global Open Data Inventory (GODI) was produced by the Open Knowledge Foundation [69] but an in-depth review of the assessment will not be provided in this section as it is no longer an active assessment.

### 5.1. Global Data Barometer

The Global Data Barometer (GDB), a research project underway, aims to track data governance, use, and availability across over 100 countries using country-level expert surveys combined with secondary data. About half of the thematic data categories covered in the GDB pertain to official statistics. The Global Data Barometer is a successor to the Open Data Barometer [17,119]. When completed it will provide an index of data governance, capabilities for data use, data availability, and impact. Through this research, the Global Data Barometer seeks to assess government, private sector, and civil society capacity to harness data for the public good. And it will highlight uses and abuses of data sharing, open data, and algorithmic decision-making.

Discussions for the creation of the Global Data Barometer began at the Open Government Partnership global summit in May 2019, and the study analysis and report are scheduled to be launched in late 2021. Prior to 2020, the Global Data Barometer was known as the Open Data Barometer [119], but its new name reflects a focus on more than just open data concepts to include data collaboratives, use of algorithms, and the evolving landscape of data and development.

The index will be based on a country-level expert survey, combined with secondary data to produce an overall index, sub-index, and individual scores. The primary peer-reviewed, expert survey will be conducted every two years. It will track data governance policies and practices in over 100 countries and across 8-15 thematic modules that address key data gaps. Each thematic module will be co-created with partners. Themes currently under consideration include: Gender and inclusion, Land, Telecoms, Energy, Transit, Climate, Crime and Justice, Agriculture, Construction and Infrastructure, Procurement, Fiscal Transparency, Education, Urban Development, Local Government, Extractives, and Geodata.

### 5.2. OECD OURdata Index

The OECD OURdata Index measures the availability, accessibility, and reuse of government data through an official survey sent to public officials in 32 OECD countries and one accession country (Colombia) [51]. The public results of this index do not allow for an in-depth analysis of data gaps or recommendations, but the

report offers many examples from OECD countries and best practices. The OURdata Index seeks to increase data flows, challenge data monopolies, and inform automated decision-making models based on emerging technologies such as AI. The pilot version first launched in 2015 with the latest set of results having been shared in 2019. It has been developed by the OECD Open and Innovative Government Division, within the Directorate for Public Governance.

The results of the OECD Open Government Data Survey sent to public sector officials are structured around three pillars: data availability, data accessibility, and government support for data reuse. For data availability, the index measures the scope of datasets available on open data portals, the extent to which governments promote open government data at the national level, and user involvement in data policy processes. The data accessibility pillar covers requirements for unrestricted access, such as terms of use and metadata, and the role of the ecosystem and portal in ensuring data quality. And under the government support for data reuse, the index measures how governments promote the reuse of government data within government and beyond.

### 5.3. *European Open Data Maturity Assessment*

The European Open Data Maturity Assessment covers the maturity of open data policies, portals, impacts, and quality through a questionnaire sent to national open data representatives in 32 EU28+ countries. The European Data Portal, a project of the European Commission, conducts this assessment as an annual exercise to provide the EU28+ countries with an overview of their maturity and progress [15]. The exercise assesses maturity across four dimensions, including policy, portal, impact, and quality. It categorizes countries into four groups: trend-setters, fast-trackers, followers, and beginners. These assessments have been conducted since the European Data Portal was launched in 2015 and document year-on-year progress since then.

Data for this research is collected through a questionnaire sent to the national open data representatives that work in collaboration with the European Commission and the Public Sector Information Expert Group. The questionnaire is structured to collect detailed metrics for each of the four dimensions. These dimensions and metrics were developed during a major methodology revision in 2018. The government survey completed by officials receives further validation and analysis from the European Data Portal team in cooperation with government officials. The 2019 version of the Open Data Maturity Assessment included 32 countries.

### 5.4. *Open Data Inventory*

The Open Data Inventory (ODIN) independently rates the coverage and openness of official statistics available from or linked to national statistical office websites in 180 countries. These assessments help identify gaps and help improve access to open data produced by national statistical offices (NSOs) [66]. The ODIN scores provide an indicator of how complete and open the data provided by the NSO are. The summary scores for social, economic, and environmental statistics and summary scores for coverage and openness provide a picture of the national statistical systems' strengths and weaknesses. Open Data Watch published the first round of ODIN assessments in 2015. The most recent version was launched for 2018/19 and included 178 countries. The next ODIN round is expected to be released in early 2021.

The statistics on NSO websites are assessed by trained researchers on ten elements of coverage and openness. Coverage scores are based on the availability of key indicators and appropriate disaggregations over time and for geographic subdivisions. The openness assessment is based on the first four principles of the Open Definition. It tests whether data available in 22 topical categories can be downloaded in machine-readable and non-proprietary formats; are accompanied by metadata; can be selected directly by users or through APIs or for bulk download; and whether the terms of use satisfy requirements of an open license.

The ODIN assessment process assumes NSOs to be at the center of national statistical system, taking charge of the dissemination of official statistics as open data. Even for decentralized national statistical systems, ODIN looks for NSO websites to provide links to other ministries' data sites. Feedback from government officials is solicited to engage NSOs in improvements to the coverage and openness of official statistics.

## 6. **Open data for improved governance**

From its beginning, the open data movement has sustained the conviction that through open data, “[G]overnments of the world can become more effective, transparent, and relevant to our lives.” [83]. The same conviction is at the heart of the first Fundamental Principles of Official Statistics: “Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, de-



mographic, social and environmental situation” [99]. In this section we examine some of the ways that open data has contributed to improved governance and how the governance of official statistics is evolving in the open data era.

Open data is a core part of the strategy of international initiatives to make the activities of governments and private firms more transparent. The International Aid Transparency Initiative (IATI) has developed an open standard that provides guidance on how to report on development activities and resource flows [28] and makes those data openly accessible. Open data on development finance allows decision makers to better direct resources and improve the efficacy of their development initiatives. Honig and Weaver argue that the IATI standard not only enables the study of resource allocation in international development but also creates incentives for more transparency in aid activities [24].

The Extractive Industries Transparency Initiative (EITI) has adopted a similar strategy. It encourages countries and companies engaged in oil, gas, and mineral extraction to adopt the EITI global standard for “open and accountable management” by publishing information “. . . from the point of extraction to how revenues make their way through the government, and how they benefit the public. By doing so, the EITI seeks to strengthen public and corporate governance, promote understanding of natural resource management, and provide the data to inform reforms for greater transparency and accountability” [11]. Other organizations that have created open data standards to improve governance are the Open Contracting Partnership [54], Open Corporates [55], and Open Ownership [75].

At the national level, open data provides a critical feedback loop for citizens to monitor and hold their governments accountable. The study of the use and impacts of open government have been largely focused on the United States and other high-income economies [85], but a number of case studies demonstrate how open data can help citizens hold government accountable and make it more effective. Open data on government spending is used to tackle corruption and hold government accountable, as two Ebola survivors did in 2018 by suing the government after an audit found that 30 percent of funds donated to fight the epidemic were unaccounted for [94]. Open data on government contracting has been an integral tool of organizations that work to monitor and assess the functioning of parliaments [37].

The COVID-19 pandemic has demonstrated the value of open data and the consequences of a lack of data. A

study from the Proceedings of the National Academy of Sciences found that the lack of precise real-time data in the United States in March may have prevented the country from containing the virus [79]. The world has seen an increase in demand for COVID-19 data and spikes in traffic to NSO websites. The United Kingdom’s Office for National Statistics, for example, saw an increase in website traffic by as much as 185 to 457 percent, depending on the metric used [105].

Open data brings out the value of existing data and increases their use, but with wider use, gaps in data coverage, availability, and timeliness surface, calling for innovative solutions. As NSOs and governments struggled to keep up with demand for COVID-19 data, other actors have stepped in to provide data on the spread of the virus and movement patterns. These unofficial data have come from a variety of sources, such as private cell phone records and from the internet of things. They have been used to perform contact tracing and to estimate the spread of the COVID-19. Unofficial sources may be useful for filling gaps in disaggregated data for the SDGs [36]. Mobile phone data can be used to supplement censuses and population surveys and provide more timely population data, which is especially important for low-income countries that may not have the capacity or resources for a decennial census [9].

To this end, some have advocated a revision of the role of the NSO to accommodate these new types of data and to help the public create value with them. This new role, often described as data stewardship, would shift NSOs’ mission beyond producing and disseminating their own statistics to coordinating between actors in the data ecosystem to find new sources and create value from data. The GovLab defines data stewards as: “organizational leaders or teams empowered to create public value by re-using their organization’s data (and data expertise); identifying opportunities for productive cross-sector collaboration and responding pro-actively to external requests for functional access to data, insights or expertise. They are active in both the public and private sector, promoting trust within and outside their organization” [20].

The case for data stewards is often based on enhanced data sharing between public and private organizations [21], but the role also includes validating unofficial data, increasing interoperability between datasets, and promoting the open dissemination of data. A data steward’s success should be measured by increased demand and use of statistics. While the role of data steward is new, it fills a clear need in the data revolution and is being actively discussed by national and international

statistical organizations. Data stewardship was included in discussions at the 51<sup>st</sup> session of the United Nations Statistical Commission [103].

As more actors and data providers join the data ecosystem, clear leadership from NSOs as data stewards will become increasingly important. This paper has reviewed the history of open data and the evolution of open data standards. The new role of data stewardship brings NSOs and open data for official statistics one step closer to the original goals of the open data movement to improve governance.

## Acknowledgments

This article has benefitted from discussions and edits from Tom Orrell, Data Ready, and Grant Cameron, Director of the Sustainable Development Solutions Network Thematic Data. The authors would also like to extend their gratitude to Deirdre Appel and Elettra Baldi, Open Data Watch, and Yu En Hsu for their help researching and editing the paper.

## References

- [1] Adhikari S, Pantaleo NP, Feldman JM, Ogedegbe O, Thorpe L, Troxel AB. Assessment of Community-Level Disparities in Coronavirus Disease 2019 (COVID-19) Infections and Deaths in Large US Metropolitan Areas. *JAMA Network Open*. 2020; 3(7). doi: 10.1001/jamanetworkopen.2020.16938.
- [2] Ayre L, Craner J. Open data: What it is and why you should care. *Public Library Quarterly*. 2017; 36: 173-184. doi: 10.1080/01616846.2017.1313045.
- [3] Belkindas M, Swanson E. International support for data openness and transparency. *Statistical Journal of the IAOS*. 2014; 30(2): 109-112. doi: 10.3233/SJI-140807.
- [4] Berger M. There's a worrying rise in journalists being arrested for 'fake news' around the world. *The Washington Post*. 12 December 2019. <https://www.washingtonpost.com/world/2019/12/12/theres-worrying-rise-journalists-being-arrested-fake-news-around-world/>.
- [5] Central Statistics Office of Dominica. n.d. Open License Agreement. Accessed 14 October 2020. <http://stats.gov.dm/open-licence-agreement/>.
- [6] Copyright Clearance Center. n.d. International Copyright. Accessed 14 October 2020. <https://www.copyright.com/learn/international-copyright/>.
- [7] Creative Commons. n.d. About CC. Accessed 14 October 2020. <https://creativecommons.org/faq/#what-is-creativecommons-and-what-do-you-do>.
- [8] Creative Commons. n.d. Considerations for licensors and licensees. Accessed 14 October 2020. [https://wiki.creativecommons.org/wiki/Considerations\\_for\\_licensors\\_and\\_licensees](https://wiki.creativecommons.org/wiki/Considerations_for_licensors_and_licensees).
- [9] Devile P, Linard C, Martin S, Gilbert M, Stevens F, Gaughan A, Blondel V, Tatem A. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*. Nov 2014; 111(45): 15888-15893; doi: 10.1073/pnas.1408439111.
- [10] Doldirina C, Eisenstadt A, Onsrud H, Uhler P. Legal Approaches for Open Access to Research Data. <https://osf.io/preprints/lawarxiv/n7gfa/download>.
- [11] EITI (Extractive Industries Transparency Initiative). (2019). "The EITI Standard 2019." <https://eiti.org/document/eiti-standard-2019>.
- [12] European Commission. (2013). EU implementation of the G8 Open Data Charter. <https://ec.europa.eu/digital-single-market/en/news/eu-implementation-g8-open-data-charter>.
- [13] European Commission. (2017). Public consultation on the database directive: application and impact. [https://ec.europa.eu/info/consultations/public-consultation-database-directive-application-and-impact-0\\_en#objective](https://ec.europa.eu/info/consultations/public-consultation-database-directive-application-and-impact-0_en#objective).
- [14] European Commission. (2018). Protection of databases. <https://ec.europa.eu/digital-single-market/en/protection-databases>.
- [15] European Commission. (2019). Open Data Maturity Report 2019. [https://www.europeandataportal.eu/sites/default/files/open\\_data\\_maturity\\_report\\_2019.pdf](https://www.europeandataportal.eu/sites/default/files/open_data_maturity_report_2019.pdf).
- [16] Ghana Statistical Service (GSS). n.d. "About GSS – Legal Mandate." Accessed 15 September 2020. <https://www.statsghana.gov.gh/aboutgss.php?category=MjkwMzA1NjI0LjE0MTU=/webstats/oq43q9p651>.
- [17] Global Data Barometer. n.d. About. Accessed 14 October 2020. <https://globaldatabarometer.org/about/>.
- [18] Global Data Barometer. n.d. Global Data Barometer. Accessed 14 October 2020. <https://globaldatabarometer.org/>.
- [19] Government of Ireland. (2017). "Open Data Strategy 2017–2022." Government Reform Unit, Department of Public Expenditure and Reform. [https://data.gov.ie/sites/default/files/files/Final%20Strategy%20online%20version\(1\).pdf](https://data.gov.ie/sites/default/files/files/Final%20Strategy%20online%20version(1).pdf).
- [20] GovLab. (2020). Wanted Data Stewards: (Re-)Defining the Roles and Responsibilities of Data Stewards for an Age of Data Collaboration. <http://www.thegovlab.org/static/files/publications/wanted-data-stewards.pdf>.
- [21] High-Level Expert Group on Business-to-Government Data Sharing. (2018). Towards a European strategy on business-to-government data sharing for the public interest. <https://www.euractiv.com/wp-content/uploads/sites/2/2020/02/B2GDataSharingExpertGroupReport-1.pdf>.
- [22] HLG-PCCB (High-level Group for Partnership, Coordination and Capacity-Building for Statistics for the 2030 Agenda for Sustainable Development). (2017). "Cape Town Global Action Plan for Sustainable Development Data." <https://unstats.un.org/sdgs/hlg/Cape-Town-Global-Action-Plan/>.
- [23] HLPEP (High-Level Panel of Eminent Persons on the Post-2015 Development Agenda). (2013). A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development. New York: United Nations. <https://sustainabledevelopment.un.org/content/documents/8932013-05%20-%20HLP%20Report%20-%20A%20New%20Global%20Partnership.pdf>.
- [24] Honig D, Weaver C. A race to the top? The aid transparency index and the social power of global performance indicators. *International Organization*. 2019; 1-32. doi: 10.1017/S0020818319000122.
- [25] ICT Update. (2017). The impact of open data on smallholder farmers. <http://ictupdate.cta.int/en/article/the-impact-of-open-data-on-smallholder-farmers-sid0e471cb3b-e005-4520-96d3-8c28e40d5f5b.html>.
- [26] IEAGS (Independent Expert Advisory Group Secretariat) (2014). A World That Counts: Mobilizing the Data Revolution for Sustainable Development. United Na-

- tions Secretary General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development. <https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>.
- [27] Intellectual Property Office. (2015). The rights granted by copyright. <https://www.gov.uk/guidance/the-rights-granted-by-copyright#:~:text=Moral%20rights%20protect%20those%20non,be%20sold%20or%20otherwise%20transferred>.
- [28] International Aid Transparency Initiative. n.d. International Aid Transparency Initiative. Accessed 14 October 2020. <https://iatistandard.org/en/>.
- [29] International Association for Official Statistics. Objectives. <http://www.iaos-isi.org/index.php/about-us/foundation-and-objectives>.
- [30] International Household Survey Network. n.d. "Benefits of microdata dissemination." Accessed 14 October 2020. <http://www.ihsn.org/dissemination-benefits>.
- [31] International Household Survey Network. n.d. Statistical Disclosure Control (SDC): An Introduction. Accessed 14 October 2020. [https://sdcpractice.readthedocs.io/en/latest/SDC\\_intro.html#foot18](https://sdcpractice.readthedocs.io/en/latest/SDC_intro.html#foot18).
- [32] Khayyat M. Open Data Licensing: More than meets the eye. *Information Polity*. 2014; 20. doi: 10.3233/IP-150357. [https://www.researchgate.net/publication/276274546\\_Open\\_Data\\_Licensing\\_More\\_than\\_meets\\_the\\_eye](https://www.researchgate.net/publication/276274546_Open_Data_Licensing_More_than_meets_the_eye).
- [33] Khokhar T. Hugs and databases: in memory of Hans Rosling. 2017. <https://blogs.worldbank.org/opendata/hugs-and-databases-memory-hans-rosling>.
- [34] Landsat Advisory Group. (2018). The value proposition for Landsat applications. NGAC Landsat Economic Value Paper – 2014 update. Reston, VA: National Geospatial Advisory Committee. <https://www.fgdc.gov/ngac/meetings/december-2014/ngac-landsat-economic-value-paper-2014-update.pdf>.
- [35] Lee C. (2020). New policy in Vietnam on digital data management, connection and sharing. <https://web.archive.org/web/20201002181156/https://vietnamtimes.org.vn/new-policy-in-vietnam-on-digital-data-management-connection-and-sharing-19547.html>.
- [36] Macfeely S, Nastav B. (2019). "You say you want a [data] revolution": A proposal to use unofficial statistics for the SDG Global Indicator Framework. <https://content.iospress.com/download/statistical-journal-of-the-iaos/sji180486?id=statistical-journal-of-the-iaos%2Fsjj180486>.
- [37] Mandelbaum A. (2011). Strengthening Parliamentary Accountability, Citizen Engagement and Access to Information: A Global Survey of Parliamentary Monitoring Organizations. <https://www.ndi.org/sites/default/files/governance-parliamentary-monitoring-organizations-survey-september-2011.pdf>.
- [38] Manuel T. (2006). Quoted in Schneider, Marguerite, Dasappa, Princelle, Khan, Neloufar, Khan, Azam. (2009). Measuring disability in censuses: The case of South Africa. *ALTER – European Journal of Disability Research/Revue Européenne de Recherche sur le Handicap*. 3. doi: 10.1016/j.alter.2009.04.002.
- [39] Max Planck Institute. (2003). Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. <https://openaccess.mpg.de/Berlin-Declaration>.
- [40] Meyer R. (2013). How Online Mapmakers Are Helping the Red Cross Save Lives in the Philippines. <https://www.theatlantic.com/technology/archive/2013/11/how-online-mapmakers-are-helping-the-red-cross-save-lives-in-the-philippines/281366/>.
- [41] Ministry of Information and Communications. (2020). National data portal inaugurated. <https://english.mic.gov.vn/Pages/TinTuc/144405/National-data-portal-inaugurated.html>.
- [42] Mons B, Cameron N, Velterop J, Dumontier M, da Silva Santos, LOB, Wilkinson M. (2017). Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud. <https://content.iospress.com/articles/information-services-and-use/isu824>.
- [43] National Agency for Statistics and Demography. (2014). Legal Notice. [http://www.ansd.sn/index.php?option=com\\_content&view=article&id=127&Itemid=273](http://www.ansd.sn/index.php?option=com_content&view=article&id=127&Itemid=273).
- [44] National Archives, The. n.d. Open Government License for public sector information. Accessed 14 October 2020. <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/1/>.
- [45] National Development Commission. n.d. Open Government Data License, version 1.0. Accessed 14 October 2020. <https://data.gov.tw/license#eng>.
- [46] National Information Exchange Model. (2019). "NIEM Model." National Information Exchange Model. March 13, 2019. <https://www.niem.gov/about-niem/niem-model>.
- [47] National Research Council. (1995). *On the Full and Open Exchange of Scientific Data*. Washington, DC: The National Academies Press. doi: 10.17226/18769.
- [48] NORC. N.d. "NORC Data Enclave." Accessed 15 October 2020. <https://www.norc.org/Research/Capabilities/Pages/data-enclave.aspx>.
- [49] Obama B. (2009). Transparency and Open Government. <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>.
- [50] Observational Health Data Sciences and Informatics. (2020). "Data Standardization." Observational Health Data Sciences and Informatics. <https://www.ohdsi.org/data-standardization/>.
- [51] OECD. 2020. OECD Open, Useful and Re-usable data (OURdata) Index: (2019). <http://www.oecd.org/digital-government/ourdata-index-policy-paper-2020.pdf>.
- [52] Omanuna. n.d. Open Government License – Oman. Accessed 14 October 2020. <https://www.oman.om/wps/portal/index/opendata/ogl!/ut/p/a1/>.
- [53] Open Aire. (2020). Sui Generis Database Right (SGDR). <https://www.openaire.eu/sui-generis-database-right-sgdr>.
- [54] Open Contracting Partnership. n.d. Open Contracting Partnership. Accessed 14 October 2020. <https://www.opencontracting.org/>.
- [55] Open Corporates. n.d. Open Corporates. Accessed 14 October 2020. <https://opencorporates.com/>.
- [56] Open Data Charter. n.d. "Principles." Accessed 14 October 2020. <https://opendatacharter.net/principles/>.
- [57] Open Data Charter. n.d. [internet]. Our History. Accessed 14 October 2020. <https://opendatacharter.net/our-history/>.
- [58] Open Data Institute. (2014). "How to plan and budget an open data initiative." <https://theodi.org/article/how-to-plan-and-budget-an-open-data-initiative/>.
- [59] Open Data Institute. (2015). Open Data Maturity Model. <https://theodi.org/article/open-data-maturity-model-2/>.
- [60] Open Data Institute. n.d. Open Data Pathway. Accessed 14 October 2020. <http://pathway.theodi.org/>.
- [61] Open Data Watch and Data2X. (2019). Bridging the Gap: Mapping Gender Data Availability in Africa. <https://opendatawatch.com/monitoring-reporting/bridging-gender-data-gaps-in-africa/>.
- [62] Open Data Watch, Data2X, and ECLAC. (2020). Bridging the Gap: Mapping Gender Data Availability in Latin America and the Caribbean. <https://opendatawatch.com/publications/>

- bridging-gender-data-gaps-in-latin-america-and-the-caribbean-technical-report/.
- [63] Open Data Watch. (2018). Measuring Data Use: An Analysis of Data Portal Web Traffic. <https://opendatawatch.com/publications/measuring-data-use/>.
- [64] Open Data Watch. (2018). The Data Value Chain: Moving from Production to Impact. <https://opendatawatch.com/reference/the-data-value-chain-executive-summary/>.
- [65] Open Data Watch. (2019). Open Data Inventory 2018-19 Annual Report. <https://opendatawatch.com/publications/open-data-inventory-2018-19-annual-report/>.
- [66] Open Data Watch. n.d. ODIN. Accessed 14 October 2020. <https://odin.opendatawatch.com/>.
- [67] Open Election Data Initiative. (2020). "Principle 6: Election Data Is Open When It Is in a Non-Proprietary Format – Section 2: Open Election Data Principles – Unleashing the Potential of Election Data – Open Election Data Initiative." Open Election Data. <https://openelectiondata.net/en/guide/principles/non-proprietary/>.
- [68] Open Knowledge Foundation. (2015). "Announcement – Open Definition 2.1." <https://blog.okfn.org/2015/11/10/announcement-open-definition-2-1/>.
- [69] Open Knowledge Foundation. (2017). The Global Open Data Index 2016/2017 – Advancing the State of Open Data Through Dialogue. <https://index.okfn.org/about/>.
- [70] Open Knowledge Foundation. (2020). "Proprietary." Open Data Handbook. <https://opendatahandbook.org/glossary/en/terms/proprietary/>.
- [71] Open Knowledge Foundation. (2020). "Metadata." Open Data Handbook. <https://opendatahandbook.org/glossary/en/terms/metadata/>.
- [72] Open Knowledge Foundation. n.d. History of the Open Definition. Accessed 14 October 2020. <http://opendefinition.org/history/>.
- [73] Open Knowledge Foundation. n.d. Open Definition 2.1. Accessed 14 October 2020. <https://opendefinition.org/od/2.1/en/>.
- [74] Open Knowledge Foundation. n.d. Open Data Commons. Accessed 14 October 2020. <https://opendatacommons.org/>.
- [75] Open Ownership. n.d. Open Ownership. Accessed 14 October 2020. <https://www.openownership.org/>.
- [76] PARIS21. (2017). National Strategies for the Development of Statistics. <https://paris21.org/national-strategy-development-statistics-nsds>.
- [77] PARIS21. (2017). NSDS Guidelines 2.3. Specific Issues – Open Data. <https://nsdsguidelines.paris21.org/node/530>.
- [78] PARIS21. (2019). "2019 NSDS Progress Report: National Strategies for the Development of Statistics." <https://paris21.org/sites/default/files/2019-05/NSDS%20Status%20Report%20May%202019.pdf>.
- [79] Perkins A, Cavany S, Moore S, Oidtman R, Lerch A, Poterek M. Estimating unobserved SARS-CoV-2 infections in the United States. *Proceedings of the National Academy of Sciences*. 2020(36): 22597-22602; doi: 10.1073/pnas.2005476117.
- [80] Pollock R. (2005). "RFC: Open Knowledge Definition v0.1." <https://lists-archive.okfn.org/pipermail/okfn-discuss/2005-August/005233.html>.
- [81] Pollock R. (2011). "Welfare gains from opening up Public Sector Information in the UK", University of Cambridge, undated. [https://rufuspollock.com/papers/psi\\_openness\\_gains.pdf](https://rufuspollock.com/papers/psi_openness_gains.pdf).
- [82] Pollock R. (2015). "Announcement – Open Definition 2.1." Accessed 14 October 2020. <https://blog.okfn.org/2015/11/10/announcement-open-definition-2-1/>.
- [83] Public.Resource.org. (2007). "Open Government Data Principles." [https://public.resource.org/8\\_principles.html](https://public.resource.org/8_principles.html).
- [84] Rosenblatt, Betsy. (1998). Moral Rights Basics. <https://cyber.harvard.edu/property/library/moralprimer.html>.
- [85] Safarov I, Meijer A, Grimmelikhuijsen S. Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*. 2017; 22: 1-24. doi: 10.3233/IP-160012.
- [86] Samuelson PA. The Pure Theory of Public Expenditure. *Review of Economics and Statistics*. 1954; 36(4): 387-89. doi: 10.2307/1925895.JSTOR 1925895.
- [87] Statistical Office of the Republic of Serbia. n.d. Open Data. Accessed 14 October 2020. <http://opendata.stat.gov.rs/odata/?id=en-us>.
- [88] Statistics Sweden. n.d. PxWeb-examples. Accessed 16 September 2020. <https://www.scb.se/en/services/statistical-programs-for-px-files/px-web/pxweb-examples/>.
- [89] Steele L, Orrell T. (2017). "The frontiers of data interoperability for sustainable development." Development Initiatives (2017). [https://www.publishwhatyoufund.org/wp-content/uploads/2017/11/JUDS\\_Report\\_Web\\_061117.pdf](https://www.publishwhatyoufund.org/wp-content/uploads/2017/11/JUDS_Report_Web_061117.pdf).
- [90] Stiglitz J. (1999). "Knowledge as a Global Public Good" in *Global Public Goods: International Cooperation in the 21<sup>st</sup> Century*. Ed. by Inge Kaul, Isabelle Grunbert, and Mar A Stern. United Nations Development Programme: New York.
- [91] Tauberer J, Mill E, Gray J, Higgins P, Weinberg M, Vollmer T. Open Government Data, best-practices language for making data "license-free." <https://theunitedstates.io/licensing/>.
- [92] The Center for Open Data Enterprise. (2018). "Data Standards." U.S. Open Data Toolkit. 14 April 2018. <https://www.usopendat toolkit.org/best-practices/data-standards>.
- [93] The Data Foundation. (2019). Future of Open Data: Maximizing the Impact of the OPEN Government Data Act. <https://www.datafoundation.org/future-of-open-data-maximizing-the-impact-of-the-open-government-data-act>.
- [94] The Guardian. (2018). "Ebola survivors sue government of Sierra Leone over missing millions." <https://www.theguardian.com/global-development/2018/jan/05/ebola-survivors-sue-sierra-leone-government-over-missing-ebola-millions>.
- [95] U.S. General Services Administration. n.d. "Open Data Engagement Guidance." [Data.gov](https://resources.data.gov/resources/odeg/). Accessed 15. September 2020. <https://resources.data.gov/resources/odeg/>.
- [96] U.S. General Services Administration. (2012). "A Primer on Machine Readability for Online Documents and Data." [Data.Gov](https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data). <https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data>.
- [97] UK Cabinet Office. (2013). Policy Paper - G8 Open Data Charter and Technical Annex. <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>.
- [98] UN Women, Open Data Watch, and Individual Depuration Measure. (2019). Navigating the Politics of Open Data. <https://opendatawatch.com/publications/navigating-the-politics-of-open-data>.
- [99] UNGA (United Nations General Assembly). (2014). Resolution adopted by the General Assembly on 29 January 2014, 68/261. Fundamental Principles of Official Statistics. <https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf>.
- [100] United Kingdom. (2010). Nuber10.gov.uk. "One-stop shop for Government data launched." <https://webarchive.nation>

- alarchives.gov.uk/20100202201223/http://www.number10.gov.uk/Page22218.
- [101] United Nations Statistics Division. (2018). Working Group on Open Data. <https://unstats.un.org/open-data/>.
- [102] UNSC (United Nations Statistical Commission). (2015). United Nations Fundamental Principles of Official Statistics Implementation guidelines. <https://unstats.un.org/unsd/statcom/doc15/BG-FPOS.pdf>.
- [103] UNSC (United Nations Statistical Commission). (2020). High Level Forum on Official Statistics: Data stewardship – a solution for official statistics’ predicament? – Side Event, 51st Statistical Commission (3–6 March 2020). <http://webtv.un.org/search/high-level-forum-on-official-statistics-data-stewardship-%C3%A2%E2%82%AC%E2%80%9C-a-solution-for-official-statistics%C3%A2%E2%82%AC%E2%84%A2-predicament-side-event-51st-statistical-commission-3-6-march-2020/6137789693001/?term=&lan=english>.
- [104] United States Bureau of the Census. (1998). “Survey Design and Statistical Methodology Metadata.” August 1998, Section 3.4.4, page 39. <https://stats.oecd.org/glossary/detail.asp?ID=1656>.
- [105] Wiley D. (2020). Twitter Post. 25 June 2020 1:41pm. <https://twitter.com/darrenwaters/status/1276208924149022721>.
- [106] Wiley D. (2007). “Noncommercial Isn’t the Problem, Share-Alike Is.” <https://opencontent.org/blog/archives/347>.
- [107] Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A, Blomberg N, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016; 3(1): 160018. doi: 10.1038/sdata.2016.18.
- [108] WIPO. n.d. Berne Convention for the Protection of Literary and Artistic Works. Accessed 14 October 2020. [https://www.wipo.int/treaties/en/text.jsp?file\\_id=283698](https://www.wipo.int/treaties/en/text.jsp?file_id=283698).
- [109] Working Group on Open Data. (2020). “Guidance on the implementation of open data in national statistical offices.” [https://unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3v-Guidance\\_OD-E.pdf](https://unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3v-Guidance_OD-E.pdf).
- [110] World Bank. (2010). “World Bank Group Opens Data to All.” <https://www.worldbank.org/en/news/press-release/2010/04/20/world-bank-group-opens-data-to-all#:~:text=WASHINGTON%2C%20April%2020%2C%202010%20%E2%80%94,people%20in%20the%20developing%20world>.
- [111] World Bank. (2010). “The World Bank policy on access to information.” <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/391361468161959342/the-world-bank-policy-on-access-to-information>.
- [112] World Bank. (2014). “Technical assessment of open data platforms for national statistical organisations (English).” Working Paper. <http://documents.worldbank.org/curated/en/744241468334210686/Technical-assessment-of-open-data-platforms-for-national-statistical-organisations>.
- [113] World Bank. (2015). Open Data Readiness Assessment Users’ Guide version 3.1. Open Government Data Working Group. [http://opendatatoolkit.worldbank.org/docs/odra/odra\\_v3.1\\_userguide-en.pdf](http://opendatatoolkit.worldbank.org/docs/odra/odra_v3.1_userguide-en.pdf).
- [114] World Bank. (2017). “World Bank Support for Open Data 2012–2017.” <http://opendatatoolkit.worldbank.org/docs/world-bank-open-data-support.pdf>.
- [115] World Bank. (2019). “Digital Government and Open Data Readiness Assessment – Prepared for the Government of the Socialist Republic of Vietnam.” <http://documents1.worldbank.org/curated/en/311651553511049630/pdf/Digital-Government-and-Open-Data-Readiness-Assessment.pdf>.
- [116] World Bank. n.d. Readiness Assessment Tool. Open Government Data Working Group. Accessed 8 September 2020. <http://opendatatoolkit.worldbank.org/en/odra.html>.
- [117] World Bank. n.d. “Open Data Policies.” Open Government Data Working Group. Accessed 17 September 2020. <http://opendatatoolkit.worldbank.org/en/starting.html#policies>.
- [118] World Bank. n.d. “Terms of Use.” Accessed 14 October 2020. <https://microdata.worldbank.org/index.php/terms-of-use>.
- [119] World Wide Web Foundation. (2017). The Open Data Barometer. <https://opendatabarometer.org/barometer/>.
- [120] Young A, Rogawski C, Verhuslt S. (2016). NOAA Open Data Portal: Creating a New Industry Through Access to Weather Data. <http://odimpact.org/files/case-studies-noaa.pdf>.
- [121] Young A, Rogawski C, Verhuslt S. n.d. United States Opening GPS Data for Civilian Use: Creating a Global Public Utility. <http://odimpact.org/case-united-states-opening-gps-data-for-civilian-use.html>.
- [122] Zuiderwijk A, Janssen M. Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*. 2014; 31(1), 17-29. doi: 10.1016/j.giq.2013.04.003.