GLOBAL REPORT

First edition

Global Data Barometer

Contents

Acknowledgements

Suggested citation: Global Data Barometer (2022). First Edition Report – Global Data Barometer. ILDA. DOI: https://doi.org/10.5281/zenodo.6488349

This report and the whole Global Data Barometer project have been possible thanks to the participation of a large network of collaborators. With thanks to the following (in alphabetical order):

Report authors: Tim Davies & Silvana Fumega with contributions from Jonathan Gray

Project direction: Silvana Fumega

Research and data coordination: Nicolas Grossman

Research and capacity building: David Zamora

Index design and calculation: Nicolas Grossman and David Zamora

Data process and analytical support: Ana Florido

Communication: Fiona Chawana with the key support of Catrina Cortes and Violeta Belver

Early project development: Katie Clancy, Veronica Cretu, Tim Davies, Amy Johnson, Oscar Montiel, Fernando Perini, Stephen Walker

Survey tool development: David Raznick (Open Data Services Co-operative)

Website visual tools: Socio Publico

Administrative host: ILDA

Financial and in-kind support: IDRC, Global Affairs Canada, and the William and Flora Hewlett Foundation (via Open Data for Development), the Data for Development Global Research Hub (D4D.net), and UNFPA, thematic and regional partners (see below)

Research Advisory Committee:

- Rachel Adams - Human Sciences Research Council, South Africa
- Ana Brandusescu - McGill University, Canada
- Michael Canares - Step Up Consulting, Philippines
- Craig Hammer - The World Bank
- Arturo Muente Kunigami - Interamerican Development Bank
- Leonida Mutuku - LDRI, Kenya
- Suzana Russell - University of the West Indies, Jamaica
- Fabrizio Scrollini - ILDA, Uruguay
- Stefaan Verhulst - The GovLab, United States of America
- Aimee Whitcroft - Independent Expert, New Zealand

Thematic partners: The Global Initiative for Fiscal Transparency, Land Portal, The Open Contracting Partnership, The Open Government Partnership, Open Ownership and Transparency International.

Regional partners:

- Access Info Europe
- Birzeit University
- ILDA
- Caribbean Open Institute
- Open Data China
- Open Data Kosovo
- InfoCulture
- Internet Society (HK Chapter)
- Institute for Development of Freedom of Information
- Local Development Research Institute
- Data for Development (D4D) Asia Hub
- The Pacific Community

Local researchers:

For Access Info Europe: Michal Škop, Niels Erik Kaaber Rasmussen, Open Knowledge Estonia, TIEKE Information Society Development Centre, Caroline Murgue, Elsa Perreau, Denis Parfenov, Openpolis, Liene Gatere, Rugile Trumpyte, Open State Foundation, Gustavo Magalhaes, Asedie, Elenor Weijmar, Javier Ruiz, Stephan Anguelov, Gong, Erasmia Tsipou & Sophocles P. Geroules, Christoph Schwaiger, Transparency International Romania, Lucia Cizmaziova, Rebecca Williams, Suk Kyoung Kim, Guy Zomer, Elsa Foucraut (TI chapter)

For the Caribbean Open Institute: Michael Yee Shui, Fitzgerald Yaw, Victor Gonzalez, Laurajan Obermuller, Michelle McLeod, Louise Mathurin-Serieux, Shurland George

For the Center for Continuing Education at Birzeit University: Hatem Ben Yacoub, Fatma Mosaad, Issa Al Mahasni, Raed Al Sahrif, Ahmad Qadi

For The Open Development Initiative (East-West Management Institute): Nikesh Balami, Vino Lucero, Ramathi Bandaranayake, Tong Khanh Linh

For IDFI Georgia: Arpine Mazhinyan, Igbal Safarov, Mikhail Doroshevich, David Eristavi, Sholpan Aitenova, Gulnura Toralieva, Elberel Tumenjargal, Livia Turcanu, Sandro Rochikashvili, Rustam Muhamedov, Nadiia Babynska, Hurriyat Khudoykulova

For ILDA: Virginia Brussa, Milenka Villegas Taguasi, Larissa Galdino de Magalhães Santos, Carlos David Carrasco Muro, Juan Pablo Marín Díaz, Jorge Umaña Cubillo, Eduardo Bejar, Iris Bertila Palma Recinos, Julio Roberto Herrera Toledo, Daniel Emilio Rodriguez Rivera, Aura Eréndira Martínez Oriol, David Riveros García, Aída Martínez Mórtola, Ana Isabel Fiafilio Rodriguez, Eliana Álvarez, Pablo Secchi (TI chapter)

For InfoCulture Russia: Olga Parkhimovich

For ISOC Chapter Hong Kong: Benjamin Zhou

For Local Development Research Institute (LDRI): Tsandzana Dércio, Kossi Amessinou, Oarabile Mudongo, Abzeta Koulsoum Ouedraogo, Poncelet O. Ileleji, Donatien Abel Gbessi Gbala, Mobile Web Ghana, Local Development Research Institute, Lamii Kpargoi, Logos Open Culture, Hatem Ben Yacoub, Tsandzana Dércio, Nashilongo Gervasius, Yusuf Suleiman, Richard Ndicunguye, Open Data pour Elles (Open Data for Her Network), Abdulai Kallon, Mbongeni Hlabano, Emmanuel Vitus, Asma Cherifi

For Open Data China: Feng Gao

For Open Data Kosovo: Dafina Olluri

For The Pacific Community: Ida Carapelle

Additional data review: Mailén García, Larissa Magalhães, Aura Martinez Oriol and Esbeidy Torres Hondal.

Disclaimer

The data used to produce this report has been produced with the support of the partner organizations. However, data and analysis does not necessarily represent the views, positions or opinions of those individual organizations, and any errors or omissions are the sole responsibility of the Global Data Barometer project. In particular, the views expressed herein do not necessarily represent those of IDRC or its Board of Governors. The data in used in this report belongs to the period May 2019 - May 2021.

Introduction

Data is a source of power. It can be exploited for private gain, and used to limit freedom, or it can be deployed for the public good: as a resource for tackling social challenges, enabling collaboration, driving innovation and improving accountability.

Over the last decade, data has risen to the top of national and global policy agendas: as nations seek to develop their economies, use data to address social challenges, and respond to citizen concerns about the uses and abuses of data. Yet, progress towards effective data governance, and to realising the public value of data, remains highly uneven across countries, regions and sectors. For example, while data protection laws are now widespread across the globe, many lack key redress mechanisms to allow individuals and communities to effectively exercise their data rights, and few comprehensively address emerging issues around location data, or algorithmic decision-making. In critical areas like climate action, significant data gaps can frustrate local action to protect ecosystems and respond to climate vulnerability. And when we look beyond the simple availability of datasets, to examine whether the data provided meets user needs, we find cases of data that's collected and shared, but that lacks key features or quality assurances that would allow it to properly power civic action, improved public services, and economic development.

This first edition of the Global Data Barometer provides evidence and insight into the development of national and global data ecosystems: offering critical comparisons, analysis and examples that can help drive top-down and bottom-up action to realize the opportunities of the 'data revolution'[1,2], while navigating its risks. In this sense, the Barometer aims not only to produce assessments of countries based on the state of their data, but also to assemble and support collective learning around what works, and about how to effectively intervene with and around data. In this report we summarize a number of key findings and highlight how the data gathered through the Barometer can be used in further exploratory work.

Barometer highlights

Shaping data for the public good is possible. But there is a long way to go. Nearly every benchmark set by our survey indicators was met *somewhere* in the world - showing that it is possible to both govern data well, and to make it available and used for the public good. However, no individual country scored over 70 out of 100, and the mean country score against the Barometer benchmark was 34.38 out of 100. This shows that every country has work to do updating policy, building capacity, sharing data, and promoting data use in order to make sure that data works as resource for sustainable development, and highlights the importance of continued focus on shaping data policy and practice to deliver the public good.

Open data agendas are alive, but not spreading. Applying open data criteria strictly, there has been little growth in the percentage of datasets that are fully machine-readable, openly licensed, freely available, and in bulk (10.63%) since the last global measure carried out by the Open Data Barometer in 2016. Although new national open data initiatives have launched since 2016, others have disappeared. However, where initiatives have been sustained, they are often better resourced and more embedded than they were in the past, and open data principles are also embedded in a number of sectoral initiatives.

Capacity gaps remain a drag on delivering value from data. While digital divides in terms of access may be narrowing, bringing more people into a datafied society, gaps in terms of government, private sector and civil society capability to create and use data for the public good remains significant. Limited

provision of training and capacity building in government risks undermining the production and governance of high-quality data, and there is pressing need to move from ad-hoc and small-scale training to building the data literacy of societies at scale.

Well-drafted frameworks deliver better data. We've looked at 7 different sectoral policies, and examined how far governance rules require data to be collected and shared in structured forms. Examining the implementation gap between rules and data availability highlights that when rules are explicit about data collection and sharing, data is more likely to be available, open, and to contain the data fields that matter.

Partnerships are powering data use. Collaboration between traditional civil society and civic technologists, or between journalists and private sector application providers are driving new uses of data to highlight corruption, promote public integrity, monitor environmental issues and shape policy debates.

By looking at data practices in specific sectors, this first edition of the Global Data Barometer contributes to work exploring, among other topics:

Climate action. The global response to COVID-19 demonstrated that new data infrastructures can be built rapidly, yet there are significant and pressing gaps in the availability of emissions, biodiversity, and climate vulnerability datasets. Data that should be available to support local action on combatting, and adapting to, climate change is often only available in aggregated and out-of-date forms. Our evidence has the potential to support participatory action on improving climate data ecosystems by helping communities identify and compare good practices, and gaps.

Political integrity. If countries that are already providing political integrity information online were to shift from paper-based processes to collecting structured data, they could unlock new approaches to accountability. Although a lack of interoperability among political integrity datasets remains a key problem in many countries, our data can be used to explore bright-spots and best-practice examples where digital-first disclosure systems are driving change. We also provide new baseline evidence on the prevalence of rules for disclosure of lobbying and highlight a lack of structured lobbying data available, enabling progress towards greater disclosure to be tracked in future.

Public finance and contracting. The relatively high levels of structured and open data publication detected by our survey for government budget and spending data, and for public procurement data, suggest a positive influence of global campaigns and capacity-building initiatives in promoting data publication and use. However, a close look at the available data also reveals that while data is increasingly available on the 'input' side of public investment (e.g. budget allocations, contracting tenders and awards, etc.), there is significant progress still to be made in tracking the 'output' side by providing joined-up data on the implementation of contracts, or the impacts of spending, particularly on issues of equity and sustainable development.

Regional analysis, and recommendations against each of the Global Data Barometer's four pillars, highlight areas for action, tailored to different country contexts. These include calls to:

Strengthen leadership and strategy to scale up and embed the skills, institutions, and freedoms required for data to be governed and used for the public good.

Develop robust data sharing frameworks including at the sub-national level, so that potential data abuses are limited, and positive re-use of data, whether from public, private or non-profit sectors, is enabled.

Deepen emphasis on equity and inclusion, recognizing that data governance, capability, availability and use all need to explicitly consider the needs of marginalized populations.

Increase the transparency of government data use and make the public more aware of when governments are collecting, sharing or using data. This can promote more accountable data practice and support greater collaboration across sectors in using data effectively.

In the regional analysis chapter, we identify potential strengths and weaknesses for each country, as well as flag where countries may offer potential bright-spot examples with potential for peer-learn. This chapter also offers valuable indications of how far open data policies and initiatives in each country have improved or moved backward since the last comparable measures in 2016, 2017 or 2020.

About the Barometer

The *Global* Data Barometer (hereafter, the Barometer) is a project of the Data for Development Network (D4D.net) and builds upon the Open Data Barometer (ODB) study, run by the Web Foundation and the Open Data for Development Network (OD4D) between 2013 and 2020 (with the final global edition completed in 2016, and the final regional edition, based in Latin America and the Caribbean, completed by ILDA in 2020). The new Barometer draws on primary data from a global expert survey carried out in mid-2021 and looks at evidence for the period May 1st 2019 - May 1st 2021. This is combined with secondary data from trusted sources to generate a range of metrics. The Barometer provides:

- **Rating not ranking** - primary indicators and scores are based on a 0 - 100 scale, where 100 is designed to measure 'best practice', defined against internationally agreed norms or frameworks. Few countries score 100 out of 100 on any indicator. While comparisons between countries can be used to look for relative strengths and weaknesses, the greater value in this model is in showing individual areas for improvement in each country.
- **Indicators *and* evidence** - primary indicators are made up of a number of structured sub-questions and are backed with qualitative evidence that can be used to gain a deeper understanding of each country's context.
- **Responsive assessment** - indicators have been designed to accommodate differing political systems and state structures (e.g. federal and non-federal systems), while reserving the highest scores for cases where governance rules, data capabilities, or available data provide coverage of the whole of a country's population.
- **A global network** - research has been carried out through a network of regional hubs, mainly by in-country researchers. Findings have then been cross-checked with a network of global thematic expert partners. This model contributes to global capacity building, creating a community of researchers and practitioners exploring data for the public good.
- **Actionable insights** - each Barometer indicator has been designed to measure features of governance, capability, availability and data use that are within the power of governments to address, and of national and international civil society to support and influence.
- **Open data** - all the scores and evidence gathered by the Barometer are published alongside this report as open data, supporting further research and analysis. We welcome further work to remix Barometer results.

The Global Data Barometer looks beyond the legal and technical conventions of open data to explore broader landscapes of data availability, governance, use and impact, aiming to put issues and communities at the centre. Rather than only focusing on how data is made available by governments, it also explores the social life and settings of data and how it has been put to work in relation to issues – including through "bright spot" examples, as well as citizens, journalists and activist groups who make and use data. The Barometer aims not only to produce assessments of countries based on the state of their data, but also to assemble and support collective learning around what works and how to effectively intervene with and around data.

Ultimately, we hope that this first edition of the Global Data Barometer will contribute not only to improving the governance, availability and use of data, but will also help to develop civil society capacity to democratically shape the "data revolution". Broadening participation around public data advocacy and policy-making around the world increases the chances of changing what counts and composing data infrastructures that are capable of making a difference.

About this report

This report provides an overview of Barometer findings. The Barometer includes 39 primary indicators, and over 500 sub-questions, covering 109 countries (delivering more than 60,000 data points in total). In this report, we select just a few of these to explore, providing a non-exhaustive overview of some of the topics that could be explored further using Barometer data.

- **Section 1** provides a short overview of the key concepts used in the Barometer and a short description of the methodology
- **Section 2** looks at the four key pillars of the Barometer (governance, capability, availability and use), and provides headlines from each.
- **Section 3** provides a regional analysis, drawing on insights from Barometer regional hubs to understand the unique context of each region and the relative strengths and weaknesses of countries.
- **Section 4** provides a short summary of learning from the first edition and highlights directions for future work

The full methodology, and details of how to access and work further with Barometer data, are contained in Appendices.

About the Data for Development Network (D4D.net)

D4D.net is a global research alliance that works to strengthen collaboration across a broad network of stakeholders by developing and mobilising the knowledge needed to advance the use of data to address critical development challenges across the Global South. D4D.net was launched in 2021 as an initiative of the International Development Research Centre (IDRC) and builds on the activities of the OD4D network and global partners interested in strengthening collaboration on open data, responsible AI, big data, privacy rights, intellectual property, cybersecurity, online surveillance, CRVS, and more.

Data for the public good

What does it mean to assess data for the public good?

A number of thinkers have discussed the idea that digital data should be treated *as* a public good[3]. In other words, use of a dataset by one person does not diminish the availability of that dataset to other users, hence the greatest value for data can be created by providing free and open access to data. For many datasets, value is also increased through scale and standardisation: network effects[4] mean that although having, for example, data from one country is useful, it can be more than twice as valuable to make connections between datasets from two or more countries. This is one of the reasons why multinational firms with vast data resources have become so powerful. And it is one of the reasons why developing open data as a public good often involves an effort to create new, interoperable, public data infrastructures that can join the dots between datasets from different countries and regions. The idea of data as a public good was central to the Open Data Barometer (our predecessor study).

However, there are exceptions to the idea that all data should be 'open by default', such as when the data is subject to legitimate privacy interests, or when there is significant potential for harm to result from data use. The data spectrum[5] is a widely used tool that highlights that while some datasets might be best thought of **as** a public good and provided openly, other datasets may be better managed as shared club goods, commons, or private property, and other datasets (e.g. sensitive personal data) should be managed as closed resources, carefully protected and only made available under strictly controlled arrangements.

In the Global Data Barometer, when we talk about data **for** the public good, we are ultimately seeking to ask two related questions:

- Is data of all forms (closed, shared and open) managed in ways that promote the public good?
- Is relevant data being collected, shared and opened to support particular wider public goods (health, education, sustainable development, justice etc.)?

Importantly, we recognize that **the** public good is a contested concept. There are many publics, many different visions of how society should be organized, and there are many views on the goals we should individually and collectively work towards. In the Barometer, the Sustainable Development Goals (SDGs), agreed through a broad international process, provide a common point of reference for identifying a set of particular public goods that data might help deliver, and that we can provide some global assessment against: from good health for all, to climate action, and to just and strong institutions.

Alongside the SDGs, we draw on other widely adopted international agreements and norms to guide the metrics that we have used. We also make all our underlying data available to allow for deeper, and/or alternative, analysis. The Barometer contains, among other data points, details on: the particular features of data governance rules; the kinds of stakeholders targeted by capacity building interventions; the particular properties of published datasets; and examples of data use and impacts. This reflects the need to go beyond simply asking about the online availability of certain datasets, to ask: whether the provision of data is built on solid foundations; whether data has the features required for it to be used in solving particular public problems; and whether communities have widely distributed capabilities to work with data.

The motivation for exploring these questions can be found in four assumptions embedded in the Barometer about what it means to work towards data for the public good:

- The collection and sharing of data should be governed by legitimate public rules that respect fundamental rights;
- Countries and communities need broad-based and widely distributed capabilities for data sharing, publication and use in order for data to be governed and used domestically for the public good;
- Particular high-priority datasets need to be available, accessible, open and, in many cases, interoperable for re-use to support widespread data use for particular public goods;
- There is a positive feedback loop between data use and data supply that can contribute to thriving national and sectoral ecosystems of data for the public good.

Of particular note, in the Barometer we ask at a number of points whether the frameworks governing data availability and use are clearly set out, and whether they have the force of law. In other words, is it theoretically possible for citizens to hold powerful actors to account, and in states with a degree of public participation in law making, are the rules themselves subject to some form of citizen control or oversight? Of course, states vary substantially in how far the rule of law operates, the extent to which laws reflect public or private interests, and the extent to which all peoples in their jurisdiction are afforded rights of citizenship. Individuals may also increasingly have access to novel data governance arrangements which are less directly reliant on state jurisdictions[6,7]. In this first edition of the Barometer, we don't yet fully take all these elements into account.

You will notice also throughout this first edition that we focus primarily on the availability and use of **public data** for the **public good**. In other words, although we include indicators that address the governance of private data (e.g. presence of data protection laws and institutions) and have asked researchers to tell us, when government is not providing data, whether it is available from other sources, including the private sector, we do not measure the **effectiveness** of governance mechanisms in protecting citizens in a given country from harms from private sector uses or abuses of data, nor do we capture comparative cross-country evidence on the extent to which private sector data is being used to deliver the public good in particular contexts. This omission does not mean these issues are not important, rather, they were more difficult to address throughout current methodology of country-level research.

Fundamentally, our approach to the public good recognizes that the construction of public good is an ongoing, unfinished and contested process. A greater role for data in society will not inevitably lead to better societies: data policies and practices need to be intentional if they are to deliver change, and it is ultimately the focus and progress of this intentional work that the Barometer measures.

Beyond data for good?

There are a number of 'data for good' (or increasingly, with emphasis on the data analysis methods used, 'artificial intelligence for good') projects around the world. These look to make use of public or private datasets towards some (often loosely defined) social good goal. Discussions of 'data for good' are generally restricted to exploring whether particular social good uses of a dataset were successful against their stated aims.

However, a dataset that is used 'for good' in one instance, may also be used 'for bad' in others, or may have been collected, managed or more widely used in ways that act against the public good. For this reason, a data for the public good lens seeks to take a broader look at concepts of data governance, capability, availability and use.

Methodology

Building the Barometer

The Global Data Barometer (GDB) was born out of previous work on the Open Data Barometer (ODB)[8], *"a global measure of how governments are publishing and using open data for accountability, innovation and social impact"*. This was in turn prompted by the crowd-sourced Open Data Index[9], which sought to provide *"the most comprehensive snapshot available of the state of open government data publication"*. The Open Data Index started with a list of ten areas of data "which most governments could reasonably be expected to collect", and asked seven questions about the availability of data in these areas. The Open Data Barometer added further questions about readiness, implementation and impact in order to appraise the state of open data around the world. Four full editions of the Open Data Barometer were produced by the World Wide Web Foundation between 2013 and 2016, with a smaller 30-country 'Leaders edition' published in 2018. A separate edition of the Open Data Barometer for Latin America and the Caribbean[10] was produced in 2020/21 by ILDA.

In mid-2019, discussion began with the Open Data for Development (OD4D) network and Open Government Partnership (OGP) research team about 'rebooting' the ODB: in particular to provide updated data on the availability of certain key datasets. Through an initial scoping process, informed by the conclusions of The State of Open Data: Histories and Horizons book[11] and interviews with past users of the ODB, we identified the need for a broader framework that reflects current data debates: looking not only at open data, but also at data sharing and the governance of private data. In early 2020, we brought together members of the OD4D network and invited experts to shape the design of a new study and survey instrument. This led to a framing around 'data for the public good', and the choice to move from indicators scored against on a guided 0 - 10 scale to using indicator scores based on detailed and discrete sub-questions that generate a score from 0 - 100. Participants at the design workshop asked that each data-point in the study be based upon existing normative frameworks, international agreements, standards or evidence, and that the study design be better able to surface 'bright spots' and encourage peer-learning, as opposed to placing emphasis on country rankings. By clustering prospective indicators, the workshop developed the four pillar structure of the GDB: governance, capability, availability, and use and impact.

Governance	Capabilities	Availability	Use and Impact
			
Covering data management; data protection; data sharing and open data, as well as looking at how data is addressed within sectoral regulations.	Covering foundations (e.g. connectivity; education), government data capability and institutions, and capability within private sector & civil society.	Addressing data existence, features (quality), openness, and coverage (extent) across a number of thematic areas.	Preliminary analysis based on example use-cases.

With funding support from Canada's International Development Research Centre (IDRC) and the new Data for Development Network (D4D.net), a small Barometer team started work in mid-2020 to more fully conceptualize the new study, exploring a wide range of data themes. Working with thematic and regional partners, draft indicators and a researcher handbook were published in early 2021. In response to project disruption as a result of the COVID-19 pandemic, and early testing that revealed a significant potential trade-off between question quantity and answer quality, the first edition of the Barometer was reduced in scope to focus on a limited number of thematic indicators.

Indicators for the first edition were prioritized based on the presence of partnerships that could support data review, analysis and re-use, and to fill critical data gaps in areas of global importance. We chose to focus in particular on the intersection of data with long-standing issues of accountability, power and money, and to include a focus on urgent global issues including the climate emergency and the COVID-19 pandemic. The intent of the Barometer is to expand thematic coverage year-on-year with managed change over time in the sample of thematic datasets, capabilities, governance rules and use-cases explored: balancing longitudinal comparability with responsiveness to emerging data issues. A research advisory committee reviewed the final design of the Barometer's indicator framework, providing critical feedback that was used to strengthen the study.

From May 2021 until late 2021, field work took place in 109 countries, managed through a network of regional hubs. An expert researcher for each country completed an in-depth survey with responses going through regional and global reviews. Preliminary data was shared with thematic partners for additional validation with responses cross-checked, outliers reviewed, and final validation checks carried out by the Barometer team. In parallel, secondary data sources were accessed, reviewed and normalized. (Note that while the data used to produce this report has been produced with the support of the partner organizations, data and analysis does not necessarily represent the views, positions or opinions of those individual organizations, and any errors or omissions are the sole responsibility of the Global Data Barometer project.)

Once initial data was available, regional and thematic partners used dashboards from the Barometer team, and their own data analysis, to identify key themes and messages. These have informed the drafting of this global report. The regional chapter presents reports from each research hub.

Throughout this process we have critically reflected upon the Barometer methodology, identifying particular research challenges to be addressed in future iterations of the project. The release of this report, and the accompanying presentation and open datasets available from the Barometer website, marks the end of the first phase of building the Barometer. Yet, there is much more work to be done. The final section of this report outlines some of the areas for future development, and some of the learning to take forward into the next stages of this work.

Countries and governments

Barometer data has been collected at the country level. To reflect an inclusive and citizen-centric perspective on the ultimate usability and governance of data, while also accommodating federal systems, indicators have been designed to give the highest scores when rules, interventions or data has nationwide coverage with no significant omissions. However, indicators are also sensitive to identifying good practice at a sub-national level, even when this is not reflected nationwide.

In short, while it should be possible, for example, to identify and explore examples of good practice from a country with high quality land ownership data available for a major city or sub-national state, the highest scores should be reserved for countries that provide land ownership data meeting the needs of all citizens, whether urban or rural, or regardless of the kind of tenure they are interested in.

The Barometer has also been designed to focus on issues that are within the power of governments to affect. Governments have a range of tools at their disposal, from policy-making and legislation, to providing incentives or building shared infrastructures for data accessibility and use.

Survey methodology

Years covered by survey	Country Surveys Completed	Regional Hubs	Researchers	Thematic Partners	Secondary indicators
2	109	12	113	6	14

Primary Indicators	Primary Variables	Data Points Collected (including supporting data)	Unique URLs reviewed	Words of justification
39	607	107,389	17799	581040

At the heart of Barometer data collection is our expert survey. Each regional hub recruited and trained country researchers, who carried out initial data collection, following a detailed research handbook (available online at https://handbook.globaldatabarometer.org/2021/). Survey responses were then reviewed by regional hubs and other national researchers with comments sent back to researchers to carry out additional evidence collection and checking. A further round of review and researcher-led updates took place with input from the global team and from thematic partners, comparing responses to specific questions across countries. During a final validation phase, further corrections and updates were made by the global team, drawing on evidence from country researchers and thematic partners. In several countries data coming from the government survey (a shorter version of the expert survey) was also taken into account during the review process.

Most expert survey indicators are based on a common pattern, consisting of three subsections:

- **Existence** - assessing whether there is evidence that a governance framework, capability, type of data, or data use exists in the country, and the nature of that existence.
- **Elements** - generally split into two parts to assess:
    - **Quality related features** of a law, dataset or capability. Wherever possible, the selection of features was based on both widely agreed international norms, and on clear use-cases.
    - **(Open) data related features** of a law or dataset, using a common set of sub-questions to assess issues such as references to/presence of structured data and licenses.
- **Extent** - assessing whether the governance, capability or data evaluated is applicable, relevant or useful across the whole country and for all citizens, or whether it has limitations. In the case of data use indicators, this looks at evidence that use is leading to impacts.

Researchers provided a written justification and sources for each indicator, and many sub-questions invited additional supporting information, such as:

- The URL to specific laws, policies, or dataset distributions
- The file format in which data is available
- The license under which a dataset is published
- The most recent update date of a dataset

The written justifications, and supporting data are all contained in the Barometer's open dataset, linked to question responses, as a resource for future research.

The majority of questions in the **element** subsection could be scored on a scale of 'No', 'Partially' or 'Yes'. The 'Partially' option was introduced to respond to variation between countries that may mean, for example, that a particular dataset feature does not exist in the way described by the question, but is available in a comparable form, or a form with a notable limitation, or to capture cases where a sub-question asks whether a list of components are present in a dataset, and only some were found. In selected indicators, specific guidance on when to use the partial response was provided to researchers. Most 'partial' answers triggered a supplementary question asking for detailed justifications. Partial sub-question answers receive 50% of the score that a 'Yes' answer receives.

These element sub-question scores sum up 100 points, meaning that, for example, a dataset with all the quality and openness features, will start the assessment with a score of 100. This score is then reduced proportionally if existence or extent sub-questions indicate that there are factors that weaken the availability, enforceability or scope of the relevant datasets, frameworks or interventions. For example, there may be a strong governance framework for data protection, but that only applies in a particular sector such as health, or that only applies in one of the states of a federal system.

To achieve this, questions in the **existence** and **extent** subsections work as multipliers (with the exception of the governance indicator 'To what extent do relevant laws, regulations, policies, and guidance require that data collection and publication be accessible to people with disabilities?', where only extent and not existence is used as a multiplier). For example, if the existence and nature of a framework of a dataset meets the highest bar set by the Barometer (e.g. has the force of law, or is released by the government), then the multiplier value is 1. If a dataset is available, but not as a result of government actions, the multiplier will be 0.9. This makes the maximum score available for datasets not provided from government 90 points overall (100 * 0.9). The same thing occurs with Extent subsection: a capability, framework or data availability with a wide coverage will keep all points earned at the elements subsection, while an isolated example will be affected by the multiplier reducing its score.

A full table of multipliers is provided in the appendix.

Secondary data

In addition to 39 primary indicators, pillar and module scores also draw upon 14 secondary indicators. These are taken from carefully reviewed external sources and have been each transformed onto a 0 - 100 scale, with missing values imputed where appropriate. A full list of secondary indicators is included in the methodology appendix.

Scoring and calculation

Individual primary and secondary indicator scores each belong to both a pillar and a module and are used to calculate:

- **An overall score** representing the performance of a country across the four pillars.
- **Pillar scores** for the governance, capability, availability and use and impact pillars.
- **Module scores** for each of the thematic modules (Company Information, Land, Political Integrity, Public Finance, Public Procurement, Climate Action, Health & Covid-19)

Each indicator is first assigned a weight based on the indicator type with primary indicators weighted higher than secondary, and governance, capability and availability indicators weighted higher than use and impact. Secondary indicator weights are also differentiated based on the type of secondary source, with secondary index variables weighted higher than single metrics from a secondary source, and those higher than dichotomous secondary variables. A full breakdown of the weighting is provided in the methodology appendix. Weights are scaled to place each pillar or module score on a 0 - 100 scale, and each indicator is then multiplied by its weight and the results summed.

In-line with the design of the Barometer as a rating, rather than ranking instrument, we do not carry out any min-max scaling of results, meaning that a country would only score 100 out of 100 on a given pillar if all the input indicators also score 100 out of 100. The gap between a country score and 100 on any pillar or module score therefore represents the gap between current performance, and a normative ideal that the Barometer indicators represent.

However, the Barometer norms are not designed to be unattainable. If we take the maximum score given on each indicator and construct an imaginary country that combines the best performance found across each of the countries in the Barometer, it would score 95.92, proving that virtually all the benchmarks set by the Barometer are, in theory, attainable today.

Learning and limitations

In seeking to provide a broad view of data for the public good, across more than 100 different countries, the Barometer has an ambitious goal. While this report and the dataset released alongside it, move us closer to delivering on that goal, it is important to note some limitations of both the Barometer methodology and the data gathered in this first edition in particular:

- We were not able to include as many data governance or capability indicators as we had initially planned. In particular, we have more limited coverage of artificial intelligence uses of data and only initial insights into data sharing frameworks and capabilities. However, the GDB is an innovative measurement tool that offers unique perspectives on the use of data worldwide. As such, this tool will need to be further refined and evaluated to increase its accuracy, and potentially include new indicators and themes.

- Few secondary sources we identified offered gender disaggregation, and our expert survey method offers only limited opportunities to gather robust evidence on the extent to which data revolutions are gender-balanced or have significantly gendered impacts. While some indicators include sub-questions on gender and inclusion, we have not always been able to include these in indicator scores. Additional efforts will be added to future editions to provide many other data points around inclusion, in general, and gender equality, in particular.

- Certain survey questions were less successful at generating robust comparable data. There have been multiple rounds of review, however, it remains possible there are both false-positives and false-negatives in the Barometer data when examining whether certain governance rules, datasets or uses exist, and assessing the elements or features they have. Therefore, we are providing opportunities for users to contact the Barometer and express their opinions on particular answers, if needed.

- We were only able to include a small number of use and impact indicators in this edition. These indicators also generated some particular data collection challenges as the availability of evidence on data use and impact varies significantly between countries, not always in proportion to the actual levels of data use. Therefore, the use and impact pillar has limited weight in final score calculations to reflect this - and cross-country comparison of use and impact scores should be treated with caution. However, the qualitative data from these indicators offers valuable insights into data use worldwide.

Throughout this report, specific methodological limitations are noted where relevant. Like many of the datasets surveyed by the Barometer, our data should be approached critically and as one tool among others for constructing knowledge and action.

Use the data

All the data and evidence collected by the Global Data Barometer is available as open data for further analysis. You can:

- **Read the research handbook** to see how each primary indicator was defined and review the guidance given to researchers.
- **Access overall scores** by pillar, module, indicator or country.
- **Explore the supporting data** to examine justifications, examples and sub-question

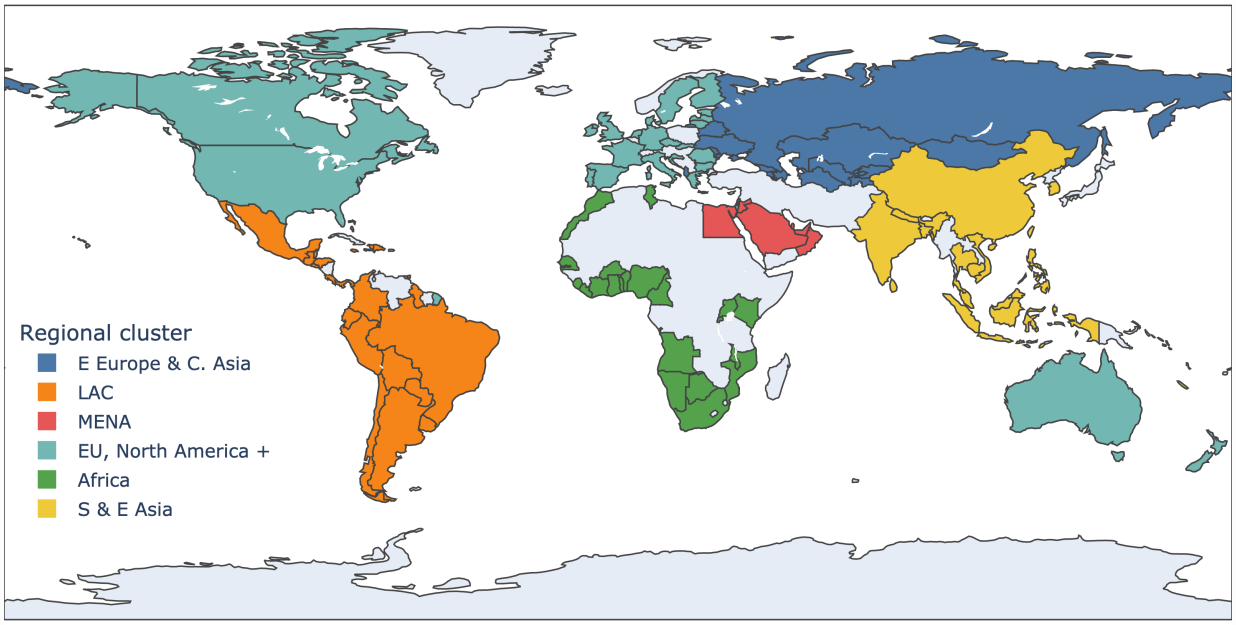
responses.

Find out how to access Barometer datasets at https://www.globaldatabarometer.org

Regional clusters

In this report, we use six regional clusters to present data. The regions used are based on both geographic proximity, and how countries were addressed by our regional research hubs. Each research hub was invited to select the countries they would focus on based on regional priorities and practical considerations. The full list of countries in each region can be found in Section 3, along with details of the hubs working on that region. The regions we use in the report are:

- Africa
- Eastern Europe and Central Asia
- European Union, United Kingdom, North America, Israel, Australia and New Zealand (sometimes referenced as 'EU, North America+')
- Latin America and the Caribbean
- Middle East and North Africa (MENA)
- South and East Asia



Regional groupings: Map showing assignment of countries to regional clusters/groups.

Pillars

- Governance
- Capabilities
- Availability
- Use and impact

The Global Data Barometer is based around four pillars, each representing a different aspect of data for the public good:

- **Governance** involves making sure that the collection and sharing of data is governed by legitimate public rules that respect fundamental rights. This covers rules that restrict access to, and use of, certain kinds of data **and** rules that promote availability and re-use of other kinds of data.
- **Capabilities** relate to the resources (connectivity, skills, institutions, training, etc.) and the opportunities to use them (political freedoms, supportive environment etc.), that support collection, management, sharing, and use of data in ways that can contribute to sustainable development.
- **Availability** surveys the presence, openness and key features of selected datasets in order to understand how far each country is making key datasets accessible in structured online forms that are fit-for-purpose for public good use-cases.
- **Use & impact** looks for evidence of particular cases of data use and explores which stakeholder groups are making use of data in each country.

The following sections present key data from each of these pillars, drawing on both the quantitative Barometer indicators, and qualitative evidence.

Governance

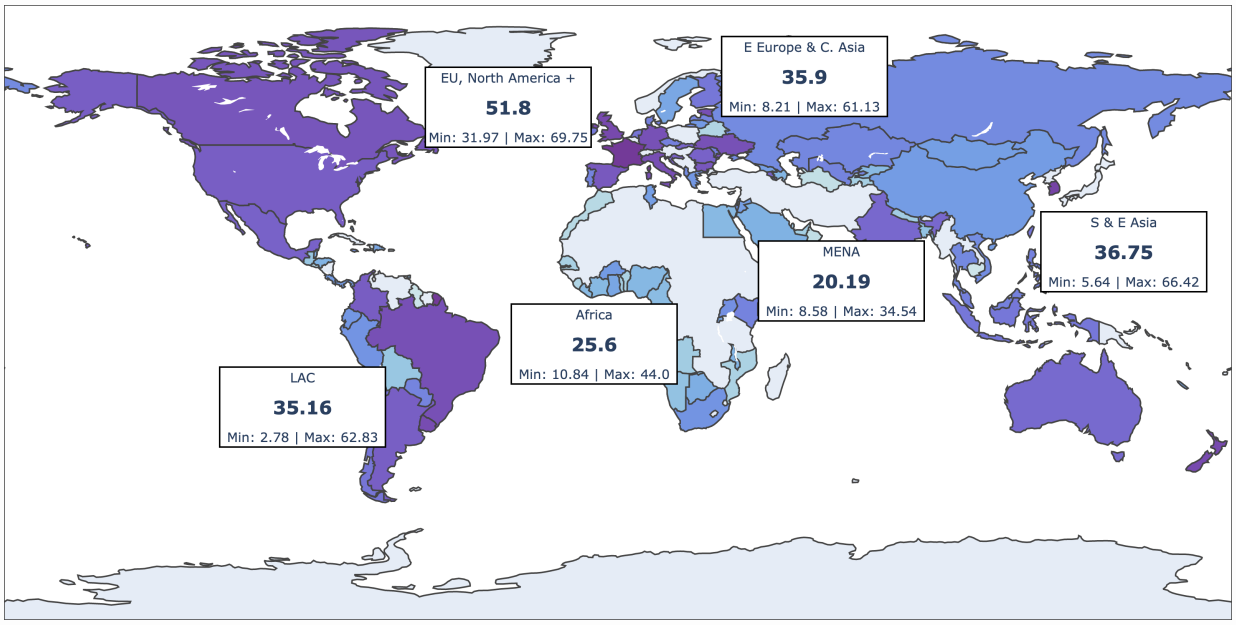
Governments can shape the collection, availability and use of data in a number of key ways, including by setting rules, adopting norms, following good practices, and establishing or engaging with oversight institutions and networks.

Governing data for the public good combines good data management with the provision of robust policies and frameworks that both protect the data rights of individuals and communities, and that make non-sensitive data widely available for re-use. Data governance for the public good should explicitly address issues of inclusion, ensuring that the production and use of data narrows, rather than widens, social inequality.

In this first edition of the Barometer, we have looked specifically at how far data governance is based on binding rules. In other words, are there laws, regulations, policies or guidance that are enforceable, and that, assuming the functioning rule of law, provide the foundations for a consistent and stable approach to how data-related issues will be managed.

Summary

- Countries increasingly have laws for protection of personal data. 98 out of the 109 countries surveyed by the Barometer have some form of framework, although in 13 countries these lack the force of law, and in 12 countries protections are limited to particular sectors, lacking full coverage of both public and private sector data use. The majority of frameworks embed clear principles of choice or consent, and rights to access and correct data. However, fewer cover breach notification, and there is evidence of a need to modernize many frameworks so that they better address particular risks around location data and the algorithmic use of personal data.
- Data management and open data policies are increasingly well-established, although far fewer countries have robust data frameworks to govern data-sharing, potentially creating gaps when it comes to both the facilitation and regulation of the exchange of sensitive data between government and other sectors, and limiting opportunities to secure the trustworthy use of data for the public good. 30 countries now have legally binding open data policies, although many countries are yet to adequately address the standardisation and interoperability of published open data.
- Just over half of the sectoral laws and frameworks we identified addressed the collection and publication of data within binding rules. Public finance, right to information (RTI) performance and asset declaration rules were the most likely to specify that data collected should be published as structured open data. Although many frameworks address data quality through establishing oversight institutions, less than half of frameworks set out verification processes, and just 36.69% explicitly support collection of structured data. There is a clear link between laws specifying that data should be published and data being made available, although the ‘implementation gap’ between requirements to publish and publication varies by sector.



Governance pillar regional scores: Our EU, North America+ regional grouping achieves the highest scores on the governance pillar. Countries in Middle East and North Africa have the lowest average score.

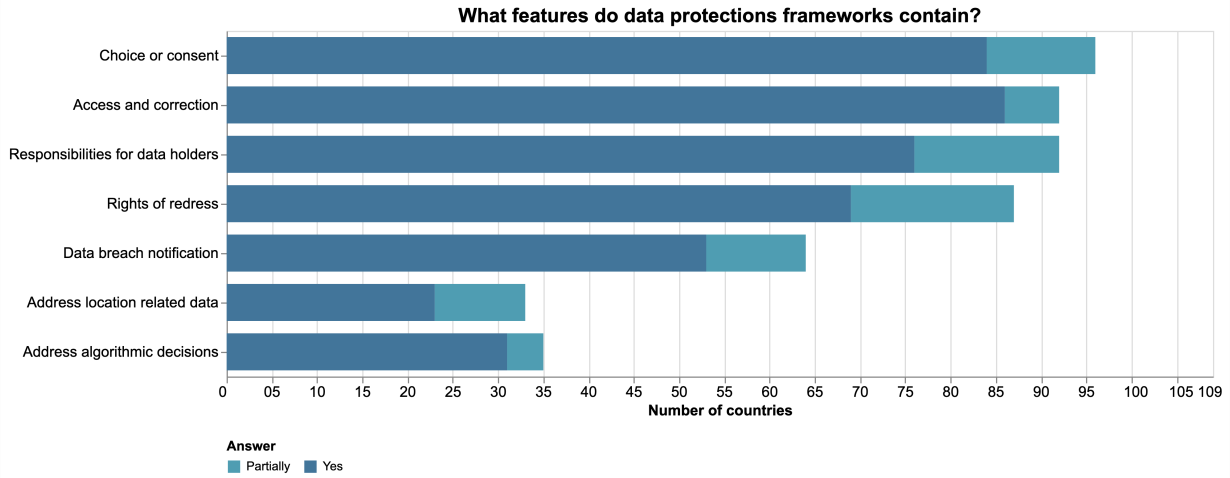
A note on 'frameworks'

In the Barometer we often abbreviate "relevant laws, regulations, policies, and guidance" as "framework." This is because the basis for governing, collecting or publishing data is often distributed across multiple laws, regulations, policies, and guidance documents. For example, one law may empower an agency to collect data, another regulation or memorandum may specify that data should be provided in a structured form, and another law may mandate that when data is provided, it should be under open license. "Framework" is used to represent the collection of relevant laws, regulations, policies, and guidance, it does not imply that a government itself necessarily presents or understands these as a unified framework.

How well developed are data protection frameworks?

The steady growth in the global coverage of data protection laws has been widely studied[12,13], and Barometer evidence reflects the trend with 98 (89.9%) of the countries surveyed having some form of data protection framework. However, of these, a number are limited to particular sectors, such as financial services, or apply only to central government data collection, providing no protection for citizens against data abuses by the private sector. In other cases, the public sector is exempted from data protection requirements. In Latin America and the Caribbean, the Middle East and North Africa, South and East Asia and Africa there are still countries yet to establish any form of data protection framework, or where frameworks need significant strengthening.

Robust data protection frameworks should be seen as a pre-requisite for work that promotes the re-use of data that may be ultimately derived from, or linked to, data about individuals, although they are only one part of a robust data governance regime, which must also consider how to regulate other forms of data-related harm. Recent developments in global standards for data protection frameworks have placed particular emphasis on improving breach notification, recognizing particular sensitivities of location data, and addressing the use of data within artificial intelligence applications and algorithms[14,15].



What features do data protection frameworks contain?: The majority of data protection frameworks cover choice and consent, access and correction, responsibility for data holders and rights of redress. Fewer cover data breach notification, and a minority explicitly address location related data, and use of data in algorithmic decision making.

Barometer data shows that, while certain aspects of data protection regulation are widespread, 45 of the countries with frameworks (45.9 %) appear to lack robust provisions for data breach notifications, and 29 (29.6%) have limited right of redress in cases of harm arising from abuses of data. Just 23 (23.5%) available frameworks robustly address location data, with marginally more (31 / 31.6%) addressing algorithmic decision-making.

Notably, in 21.4% of the countries with frameworks, researchers reported at least some degree of limitations being placed on the operation of data protection frameworks in the context of the COVID-19 pandemic, ranging from regulators announcing they would take a lighter-touch approach to reporting and enforcement for organizations involved in pandemic response, through to suspension of consent requirements for health-related data sharing, or making allowances to enable mobile phone records for monitoring population movements. While in some cases, this involved using provisions already in place for emergency situations, campaigners have expressed concern that some data protection processes may have been weakened during pandemic response, highlighting a need to both track how far exemptions or changes stay in force after the pandemic and to explore whether regulators return to review any practices that might have developed during a period of softer regulation.

It is also worth noting that there are increasing concerns that individual rights-based data protection frameworks only cover some of the potential harms of data collection and use that need to be managed. The qualitative data collected by the Barometer offers some insight into how national data protection systems are functioning, but future work will need to address a wider range of modalities for ensuring data is not used to threaten the rights of both individuals and communities.

How well developed are data management, sharing and access frameworks?

Selected governance pillar indicator scores

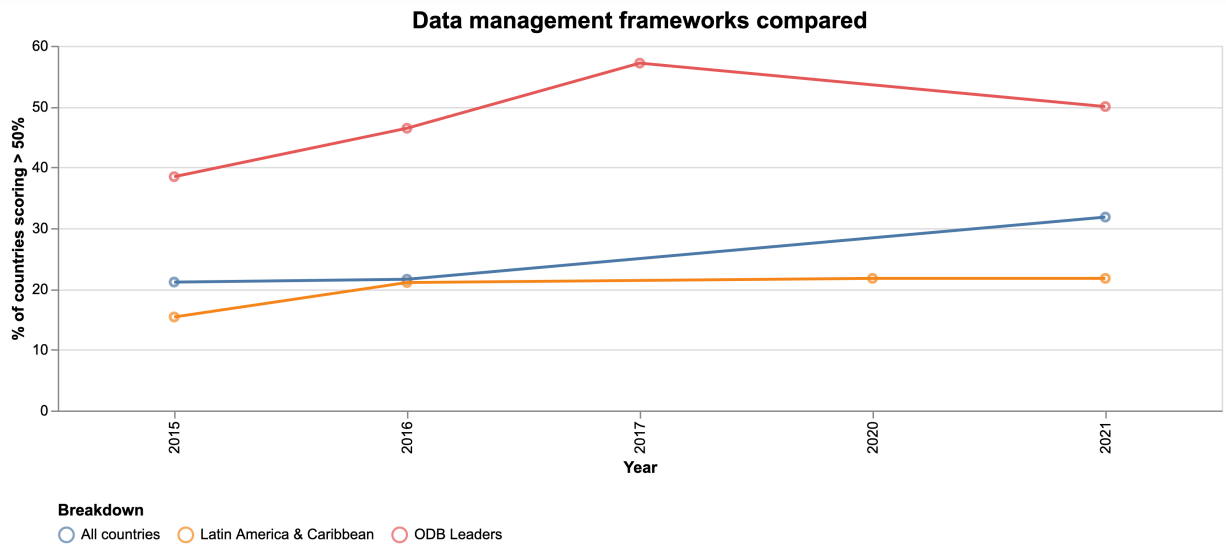
Countries with data management frameworks	Countries with data sharing frameworks	Countries with open data policies
71.6%	62.4%	67.9%
Mean score: 30.28 out of 10	Mean score: 24.05 out of 10	Mean score: 37.69 out of 10

The value of data for the public good, regardless of whether it is open data or not, is increased when data is more easily discoverable, when data comes with clear documentation, when data quality has been assured, when appropriate technical standards are used, and when user feedback is sought to improve data management. Governments may promote consistent and high quality approaches to data management through a variety of routes, including national data strategies, data management guidance

and data management standards. Clear frameworks for data management and sharing can also protect against harms of data misuse by ensuring data is handled according to transparent processes, and that there are lines of accountability around data use.

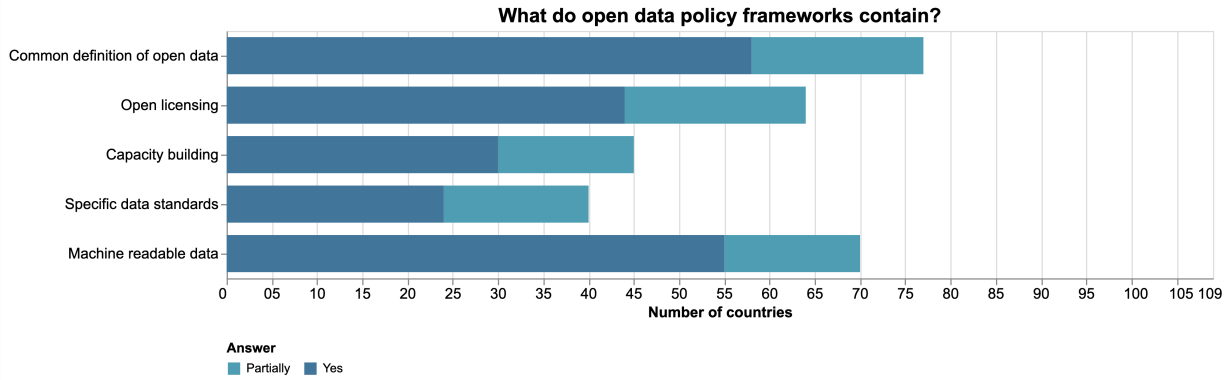
The Barometer asked researchers to look for the presence of data management frameworks and to explore their features. We found 36 (33%) countries with binding data management frameworks and an additional 42 (38.5%) with less-than-binding frameworks (for example, guidance or strategies). Of these, just 8 (10.3%) had evidence of robust documented mechanisms to solicit or integrate feedback from external users to improve data quality, although 49 (62.8%) referenced minimum standards for meta-data.

Overall, 29 countries scored greater than 50 out of 100 for the quality and coverage of their data management frameworks. As our data management indicator has been designed to be broadly comparable with an earlier Open Data Barometer indicator, used since 2015, we can track how the quality of data management frameworks has changed over time. As the chart below reveals, looking only at countries included in both the ODB and GDB that have scored above 50% on this indicator in any given year, there is evidence of steady progress with a rise from 21% of countries reaching this threshold in 2016 to 31% in 2021. The chart also compares scores for just the 28 countries included in the 2017 ODB Leaders Edition (showing a slight fall in score) and for up to 23 countries in Latin America and the Caribbean region (no recent growth) for which further data was collected in 2020. Adding these comparisons suggests, in particular, that a modest global trend towards stronger data management is coming from countries beyond the Open Data Charter signatories covered by the ODB Leaders Edition.



Comparison of Open Data Barometer and Global Data Barometer data management scores. There has been modest global progress towards stronger data management frameworks from 2015 - 2021, although this is less evident in the countries covered by the Open Data Barometer 'Leaders Edition'. Progress towards improved data management in Latin America appears to have stalled. Note: differences in methodology between the ODB and GDB may explain some of the variation shown above.

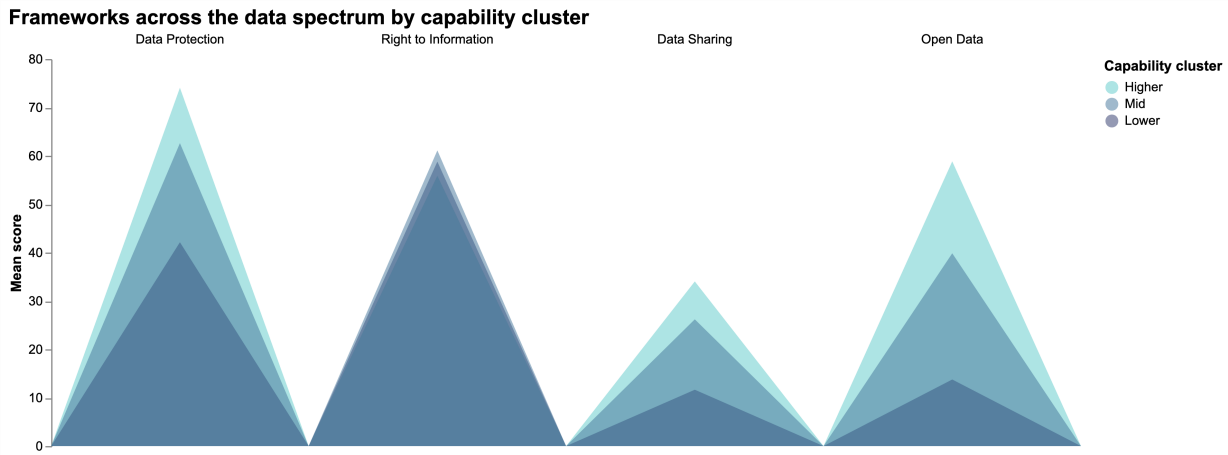
When it comes to open data policies, we found 74 countries with some form of policy in place, 30 of which had legally enforceable policies. 91.9% of policies provide a common definition of open data with 83.8% requiring machine readable data and 78.4% promoting open licensing without any restrictions beyond attribution and share-alike. However, just 47.3% address common standards for data, suggesting an opportunity for future work to focus more on standardisation and interoperability of published open data.



What do open data policy frameworks contain? Open Data policies are more likely to cover data formats than they are to address licensing requirements or capacity building amongst officials.

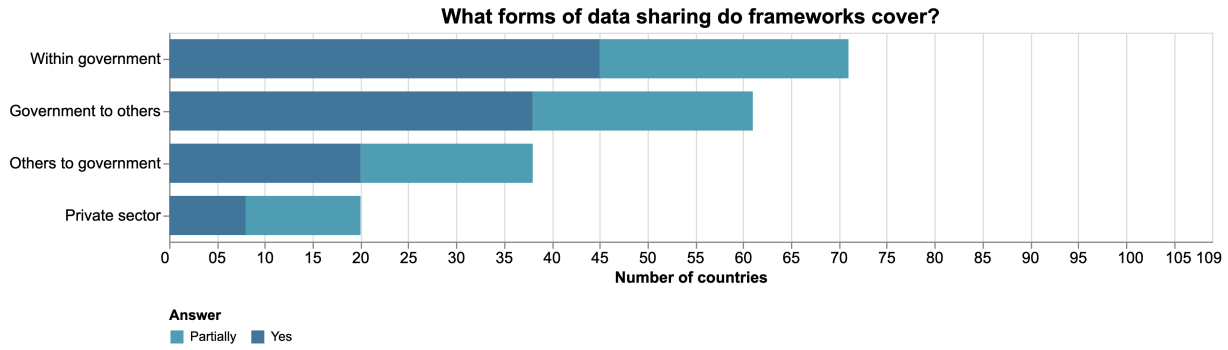
Protection and access across the data spectrum

Realising the value of data for the public good involves finding the right approach for each dataset that can maximize access and re-use, while respecting individual and collective data rights. The Open Data Institute has developed the Data Spectrum[5] as a tool for thinking about whether data should be closed (kept securely), shared (provided to trusted third-parties, often on the basis of formal agreements), or open (available for anyone to access and re-use without restriction).



The data spectrum explored: Higher and mid capability countries are making progress in developing well-resourced and robust open data and data sharing frameworks, while low capability countries have made less progress here. There is limited difference on account of digital capability between the quality of country right to information frameworks, while countries with high capability are more likely to have strong data protection frameworks. Visualization based on the Open Data Spectrum developed by the Open Data Institute.

By dividing Barometer countries into three capacity clusters (see next chapter), we can explore the relative maturity of governance frameworks for each part of the data spectrum. As the chart above shows, there is relatively little variation between countries on basis of their capabilities when it comes to the presence of robust right to information frameworks, and both high and mid-capability countries have reasonably strong open data frameworks. However, across all clusters, the governance of data sharing is less well developed (at least when it comes to being placed on a legal footing), and low-capability countries lag significantly in terms of data protection, data sharing and open data frameworks. With a significant focus on data sharing-based data for good initiatives, including in low-capability countries, this may be a particular cause for concern and highlights at area in need of significant development.



What do data sharing frameworks cover? Data sharing frameworks are more than three times as likely to regulate data sharing within government as they are to regulate private sector data sharing arrangements.

Looking deeper into the kinds of data sharing covered by the frameworks that do exist, we find the majority (92.6% of the 68 frameworks identified) govern data sharing within government, with 79.4% covering how government should share data with other sectors, and 51.5% addressing how other sectors should share data with government. Just 16.2% explicitly address artificial intelligence uses of data, and only 26.5% look at data sharing within the private sector. Increased governance of private sector data sharing has been on the agenda of the European Union through the concept of 'Data Spaces' introduced in the European Data Strategy[16], but this appears not to have fed through as yet into national frameworks. In a number of cases, where researchers were unable to locate data sharing frameworks, they looked at data sharing provisions within data protection legislation. In future editions of the Barometer, we will look to strengthen the definition of data sharing frameworks so that these cases would not be counted, as they generally do not demonstrate a focus on mechanisms that specifically govern data sharing. If anything, we anticipate this has led to a marginal over-counting of data sharing frameworks in this edition.

Overall, the qualitative evidence collected by the Barometer survey revealed many different approaches around the world to the governance of data sharing, including approaches focused on setting rules, providing platforms, promoting interoperability, creating new government powers, and providing guidance to government and industry. Much practice remains at the level of policy, rather than binding legal rules.

How far does governance take accessibility and inclusion into account?

Two of the key barriers to the inclusiveness of data policy observed early in the rise of the open data movement were a tendency for data and portals to be provided in only one language, even in countries where many languages are spoken, and for data platforms to be designed with little attention paid to accessibility features, such as compatibility with assistive technologies like screen readers, or through implementing universal design principles when creating data-related websites and tools.

The Barometer included two indicators designed to identify the extent to which countries have governance frameworks to promote inclusive accessibility of data. One of these, looking at language, has been excluded from scoring, as we were not able to secure reliable enough results from our survey question. The other, on accessibility, reveals that while 66 countries have some form of law, regulations, policy or guidance that requires data collection and publication be accessible to people with disabilities, this is derived in most cases from general accessibility rules with just 17 countries having a dedicated accessibility framework or specific provisions relating to data.

Regionally, we see significant variation in accessibility frameworks with the lowest scores in the Middle East and North Africa region and the next-lowest mean score coming in Africa.

Mean scores on accessibility indicator in the governance pillar by region

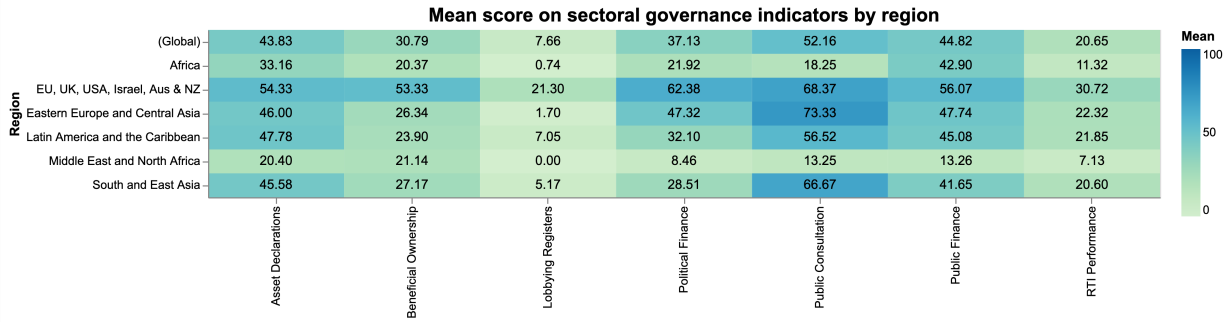
Africa	Eastern Europe and Central Asia	European Union, North America, Israel, Australia and New Zealand	Latin America and the Caribbean	Middle East and North Africa	South and East Asia
19.11	22.33	58.97	30.4	5.06	30.78

In this first edition of the Barometer, we were not able to explore the extent to which individual governance frameworks for data address inclusion, although this is an area for future work, that will build on groundwork laid through a number of wider inclusion focused questions in this edition, including a contextual question asked to identify particularly relevant patterns of exclusion and marginalization in each country.

Are sectoral laws, policies or guidance 'data aware'?

Almost half of the score assigned to the governance pillar of the Barometer is derived from sectoral governance indicators. In these indicators, we explore the extent to which particular thematic laws and frameworks are 'data aware'. That is, do laws or policies pay attention to the fact that governmental processes, such as setting national budgets, carrying out public consultations, or putting in place mechanisms to improve political integrity, all generate and rely upon data. Early efforts to open up government data often relied on administrative decisions to publish datasets. Decisions that could be revoked at any time. When collection and publication of data is placed on a firm footing, both civil society, and businesses, are more likely to be able to rely upon it and build processes that make use of the data. And when data collection and publication is addressed explicitly within law and regulations, it can gain greater legitimacy, being subject to greater public scrutiny.

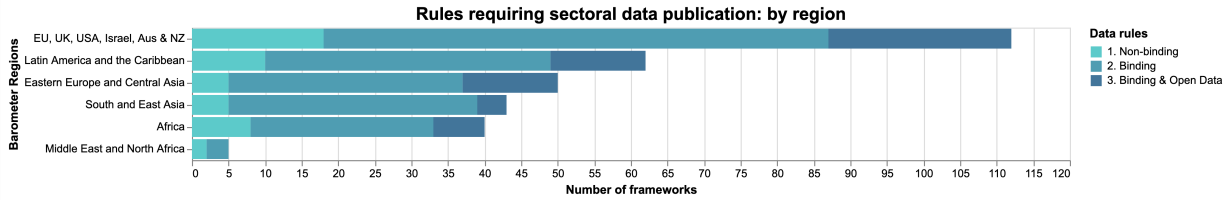
Each sectoral governance indicator is paired with a data availability indicator to support research into the relationship between data collection and publication rules and the data being shared or open. Within each indicator, we look at particular features that are important for the quality of data that might be produced, including general features across all kinds of data (such as providing structured data, timely updates, and having quality-assurance or verification processes), and topic-specific features (such as interoperable identifiers, or disaggregation by important variables).



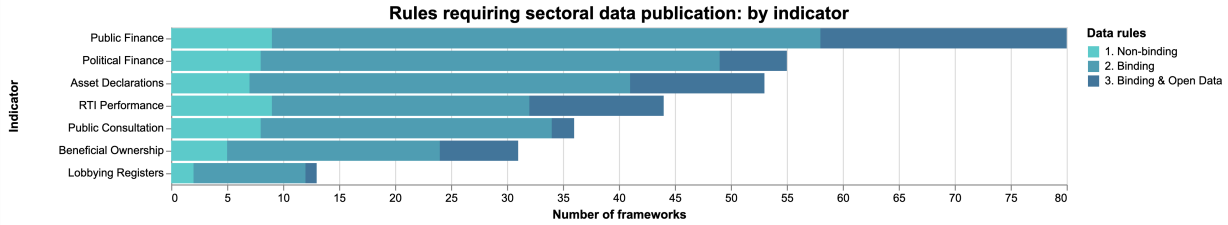
Overall scores on sectoral governance indicators by region. The highest mean scores, indicating laws or frameworks exist more widely, and have more of the relevant features, were found for 'asset declarations'. The lowest scores were found for lobbying registers.

The chart above shows the mean scores on each of these indicators, disaggregated by region. 186 out of 507 (36.69%) operational governance frameworks identified by our survey support the collection of structured data, and 246 (48.52%) require some form of verification process to assure the quality of data. Recognizing the importance of institutions in enacting governance rules, 358 instances of the surveyed frameworks (70.61%) empower an agency or official to oversee elements of data collection and publication.

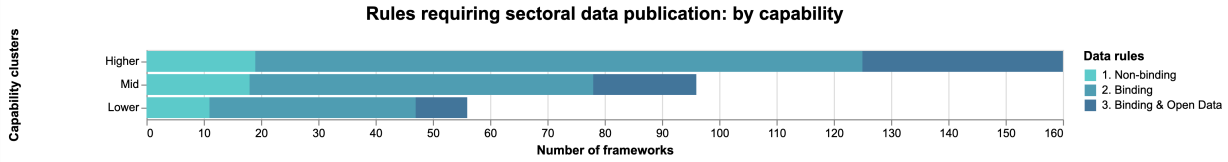
When it comes specifically to the publication of open data, we find no mention of publishing data in some 195 (38.46%) of the operational laws, policies, regulations or guidance, with 48 (9.47%) outlining some requirements to publish data in non-binding policy or guidance, 202 (39.84%) including this in binding policy, regulation or laws, and 62 (12.23%) providing the strongest forms of requirement for open data publication. Of these, 25 are found in the European Union, United Kingdom, North America, Israel, Australia and New Zealand, and 35 among higher capacity countries with laws on public finance accounting for 22 of the strongest open data requirements.



Sectoral rules requiring data publication by region: All regions, except MENA, have some binding requirements to publish open data as part of sectoral governance frameworks



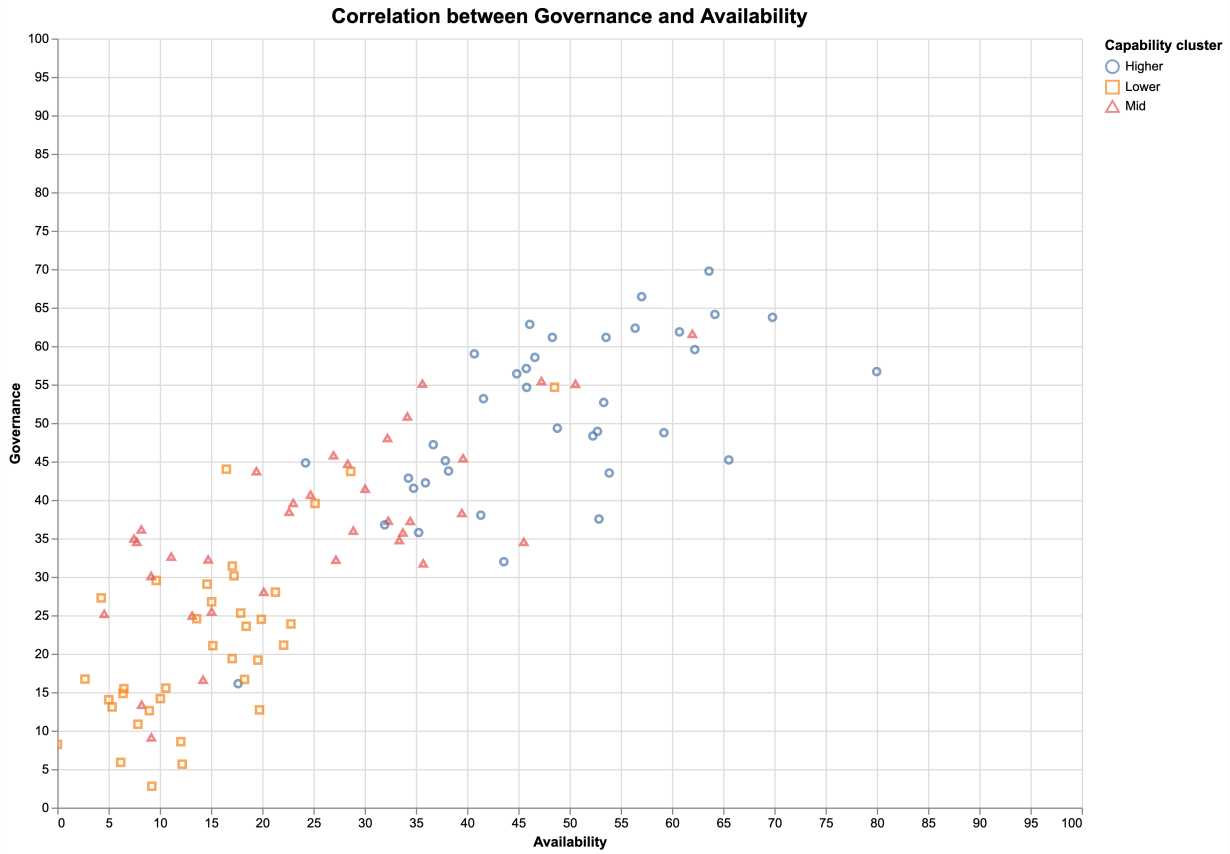
Sectoral rule requiring data publication by topic: Public finance frameworks are the most likely to have binding requirements for the publication of both data and open data



Sectoral rules requiring data publication by capability cluster: The majority of frameworks with binding requirements to publish open data exist in higher capability countries

Governance & Implementation

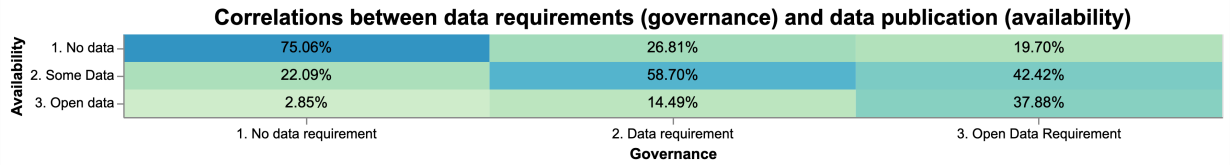
We can use Barometer variables to explore the connection between the quality of data governance and the quality of data availability. At a high level, the chart below shows a scatter plot of the governance pillar scores against the availability pillar, indicating a broad correlation. However, we can see that this relationship can, to a degree, also be explained by country capability.



Correlation of governance and accountability pillars: There is a positive relationship between governance and availability scores

We can also look at the level of paired data governance and data availability indicators to develop the three-by-three matrix below, which explores the extent to which there are cases where, for each country and dataset pair (109 x 7 = 763 cases in all), there are frameworks or rules that require data to be collected and published, and whether that data is then found to be available or not.

As the matrix below shows, in 75.06% of cases where there is no framework providing a requirement that data be collected or published, there is also no data available. When the publication of data is required by governance frameworks, in 58.7% of cases data is available in some form, and in 14.49% the data meets the open definition. This still leaves an implementation gap in 26.81% of cases, where governance frameworks require publication of data, but no data could be found by our survey. The implementation gap is larger when it comes to open data, where only 37.88% of binding requirements to publish open data appear to result in open data being available, although some data, albeit falling short of the open definition, is available in a further 42.42% of these cases.



Comparison of data governance requirements to data availability: Using all governance and availability indicators pairs shows stronger data requirements drive data greater data availability. Percentages are given by column.

The implementation gap also varies between data categories. For public finance data, there is a significant quantity of open data available even when formal rules were not found that require this, whereas for the interest and asset disclosures of politicians, strong requirements for open data in governing frameworks are not reflected yet in the availability of structured and open data.

How can countries improve data governance in future?

For each country, the Global Data Barometer website contains a profile that shows the breakdown of individual scores against governance indicators. This can be used to identify particular areas for improvement. Drawing on the quantitative and qualitative evidence captured for this governance pillar, we highlight the following common areas for action.

Review and refresh foundational frameworks

There are 16 countries covered by the Barometer that scored below 5 on our data protection indicator. Given the foundational importance of data protection frameworks to manage risks of data misuse while supporting public good data use, this suggests a priority need to either create, implement or strengthen data protection frameworks in these countries. However, **all** countries need to keep their data protection and governance frameworks under regular review, ensuring there are processes to review the effectiveness of mechanisms designed to present abuses of data, and responding to the changing landscape of data risks and opportunities. In particular, care must be taken to ensure the strengthening of data protection rules does not undermine legitimate public interests in accountability data.

In many countries, open data and data management frameworks also need strengthening, addressing data standardization and interoperability in particular.

If data is to be use towards an inclusive model of the public good, countries also need to focus on frameworks for making sure data and data-related platforms are accessible to people with disabilities.

Establish and iterate on clear frameworks for data sharing

The next decade is likely to see increased voluntary and mandated data sharing arrangements between businesses in industry sectors, between business and government, and in supporft of data collaborative arrangements oriented towards addressing humanitarian and development challenges. Without clear frameworks that facilitate and govern such arrangements, there are risks that positive uses of data will be missed, and that abuses of data will proceed unchecked.

Countries need to identify appropriate models to govern data sharing involving government and data sharing across the wider economy. This is likely to require a combination of broad national consultation to develop legitimate approaches, as well as international engagement that can help align approaches to data sharing involving international data flows.

Address data collection and publication in sectoral legislation and policy

The Barometer has explored a number of specific sectors where countries could establish or strengthen rules requiring the collection and publication of structured data, including:

- Lobbying registers;
- Data on the performance of right to information frameworks; and
- Beneficial ownership registers.

However, these are just a few examples of the kinds of sectoral legal frameworks countries may look to establish. The underlying point explored by the Barometer is that whenever legislation or policy is likely to involve the creation of data, there should be explicit attention paid to the rules that govern how the data should be provided, including addressing how data will be verified, privacy protected, and non-private data made available under open licenses.

Future editions of the Barometer will cover other sectoral governance rules. Governments should look to identify how good data management and open data practice can be embedded in any ongoing sectoral reforms, and civil society organizations should develop their capacity to scrutinize the impact of each reform on the production and sharing of data.

Capabilities

To realize the benefits of data for the public good, countries need a wide range of resources from broadly accessible foundations of Internet access and data infrastructures to basic digital skills across a population, institutions supporting the realization of data rights and good data management, and availability of advanced analytical tools and skills in government, private sector and civil society. In the context of the Barometer, capability involves not just the presence of digital resources, and access to data-literacy, but also involves the freedoms and opportunities for skills and resources to be put to use in service of wider social goals.

The metrics in this pillar of the Barometer reflect both the background conditions in a country for data to be used for the public good, and particular points of intervention that may be available to governments, or possible for external partners to support, that can contribute to an environment in which data is governed, made available, and used, to address social challenges and realize social goals.

Summary

- Digital divides in terms of Internet connectivity are narrowing. However, without efforts to narrow data literacy gaps, ensure effective institutions that can regulate data, and support broad based engagement with data, there is an ongoing risk that a greater number of people will have their data captured and used for private gain without having the ability to make use of data in order to advance their own interests and the wider public good.
- Many of the datasets with the greater relevance to the daily life of communities are often the responsibility of local governments. However, just 22% of countries appear to have evidence of sustained and institutionalized capability to manage data effectively at the sub-national level, with few countries showing evidence of widespread local policies on open data (16 / 14.68%) and data sharing (16 / 14.68%). However, 47 countries have bright-spots among states or municipalities demonstrating reasonable sub-national data capabilities, highlighting opportunities for peer-to-peer learning between states, cities and municipalities, as well as highlighting the potential returns on investments in local capability.
- The level of government-provided training to develop civil servants' data literacy and data skills remains low. Just 23 countries had evidence of widespread and regular training for civil servants on data matters as part of a planned and sustainable strategy, although the majority of countries have at least some training provision. There are significant opportunities to expand the range of topics covered by data-related training and the reach of training provision across government agencies.
- There is substantial national and regional variation in capabilities. The capability pillar of the Barometer has the greatest range between the highest and lowest scores, and the highest observed maximum pillar score (91.2). In Latin America and the Caribbean, for example, the opportunities created by higher human capital and political freedoms appear held back by lower levels of digital skill among the population as a whole, while in the Middle East and North Africa, business and government capacity to use data are strong, but civil society freedoms remain weak.

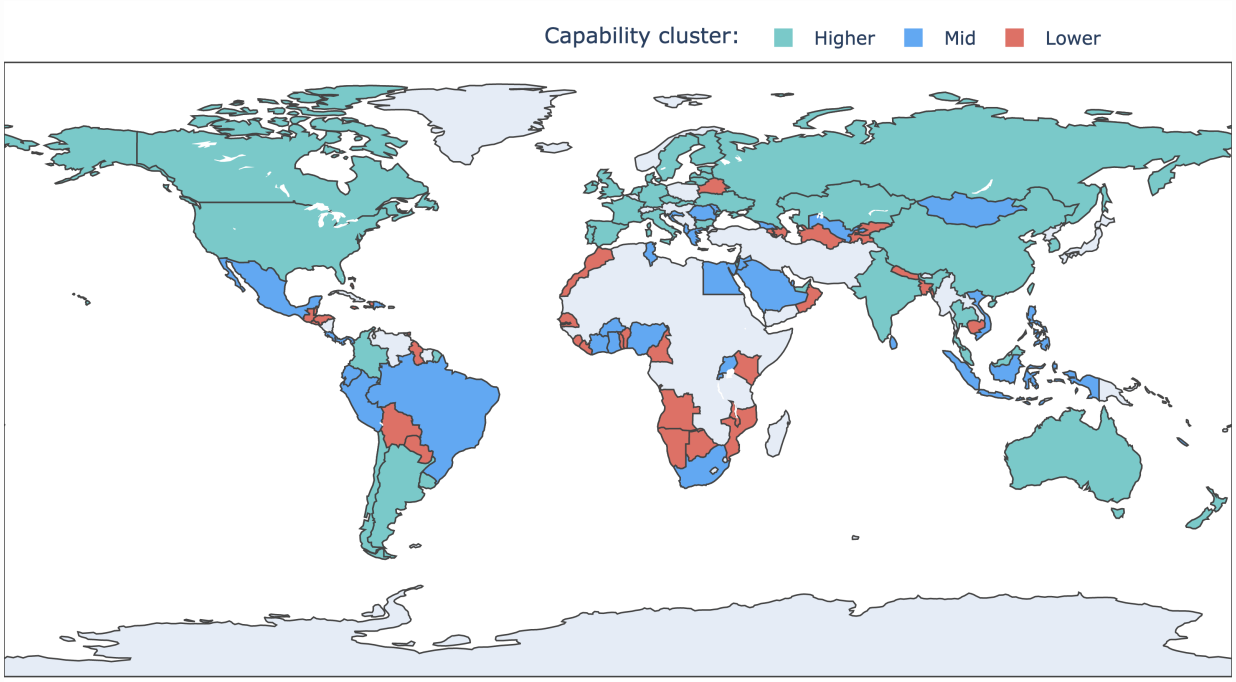
Methodological note

The capability component of the Barometer is built through a combination of ten secondary data sources, and four primary indicators. The primary indicators seek to fill particular gaps in knowledge about the presence of government-backed capacity building initiatives for civil servants, around open data, and data re-use, as well as data management capability at the sub-national level. Because our primary survey methodology relies upon public evidence of capacity building, training or support for data re-use, it is possible that it has not identified undocumented cases of training provision or capacity building taking place inside government institutions.

In putting together this capability component, we encountered a number of other significant data gaps that have limited our ability to generate a robust comparative view of the data-use capabilities of civil society, media, academia and government. The framework of the World Bank Statistical Performance Indicators[17] describes some of the key data issues. We also identified limitations in existing measures of general digital skills, including the limited country coverage and comparability of reporting against the data-intensive components of SDG 4.4.1. In a number of cases, such as assessing the data capabilities of media or civil society, we were not able to identify a primary indicator that could generate sufficiently robust data within the constraints of our survey method. As a result of these limitations, the capabilities component in this first edition of the Barometer is significantly weighted towards assessing government capability, and work will be needed in future editions of the Barometer to develop a more holistic set of indicators.

How does capability vary across the globe? Higher, mid and lower capability countries

Taking the overall Barometer capability component score (a weighted average of capability indicators), we divide countries into three equally sized groups, labeled here as 'lower', 'mid' and 'higher' capability countries. We use these clusters throughout the report in order to identify potential actions for countries based upon their relative starting points and levels of capability. The data collected for this component can also be used to explore relationships between data governance, capability, availability and use, exploring questions around the kinds of intervention that best promote data for the public good.



Country assignment to capability clusters. Clusters are generated at 33% and 66% percentile cuts in the weighted capability component score, giving three equally sized groups.

Countries in the higher capability cluster	Countries in the mid capability cluster	Countries in the lower capability cluster
Argentina,	Albania,	Angola,
Australia,	Brazil,	Armenia,
Bulgaria,	Burkina Faso,	Azerbaijan,
Canada,	Costa Rica,	Bahamas,
Chile,	Croatia,	Bahrain,
China,	Côte d'Ivoire,	Bangladesh,
Colombia,	Dominican Republic,	Belarus,
Czechia,	Ecuador,	Belize,
Denmark,	Egypt,	Benin,
Estonia,	Georgia,	Bolivia,
Finland,	Ghana,	Botswana,
France,	Greece,	Cambodia,
Germany,	Indonesia,	Cameroon,
Hong Kong,	Jordan,	El Salvador,
India,	Kosovo,	Gambia,
Ireland,	Malta,	Guatemala,
Israel,	Mexico,	Guyana,
Italy,	Moldova,	Haiti,
Kazakhstan,	Mongolia,	Honduras,
Latvia,	New Caledonia,	Jamaica,
Lithuania,	Nigeria,	Kenya,
Malaysia,	Panama,	Kyrgyz Republic,
Netherlands,	Peru,	Liberia,
New Zealand,	Philippines,	Malawi,
Portugal,	Qatar,	Morocco,
Republic of Korea,	Romania,	Mozambique,
Russia,	Rwanda,	Namibia,
Slovakia,	Saint Lucia,	Nepal,
Spain,	Saudi Arabia,	Oman,
Sweden,	South Africa,	Paraguay,
Taiwan,	Sri Lanka,	Senegal,
Thailand,	State of Palestine,	Sierra Leone,
UAE,	Tunisia,	Tajikistan,
USA,	Uganda,	Togo,
Ukraine,	Uzbekistan,	Trinidad and Tobago,
United Kingdom,	Viet Nam	Turkmenistan
Uruguay		

The highest capability countries demonstrate high levels of affordable Internet connectivity and human capital (measured by the UN E-Government Survey's Human Capital Index), as well as clear evidence of government support for digital and data practice and businesses making use of digital tools in their operations. While most countries in the higher capability cluster score highly on measures of political freedom, the cluster also includes a number of outliers (China, Kazakhstan, Russian Federation and United Arab Emirates) with lower levels of political freedom according to Freedom House rankings. Even within the high capability cluster, there is significant variation in how far government is providing training to develop civil servant training and skills, and the extent to which governments are providing

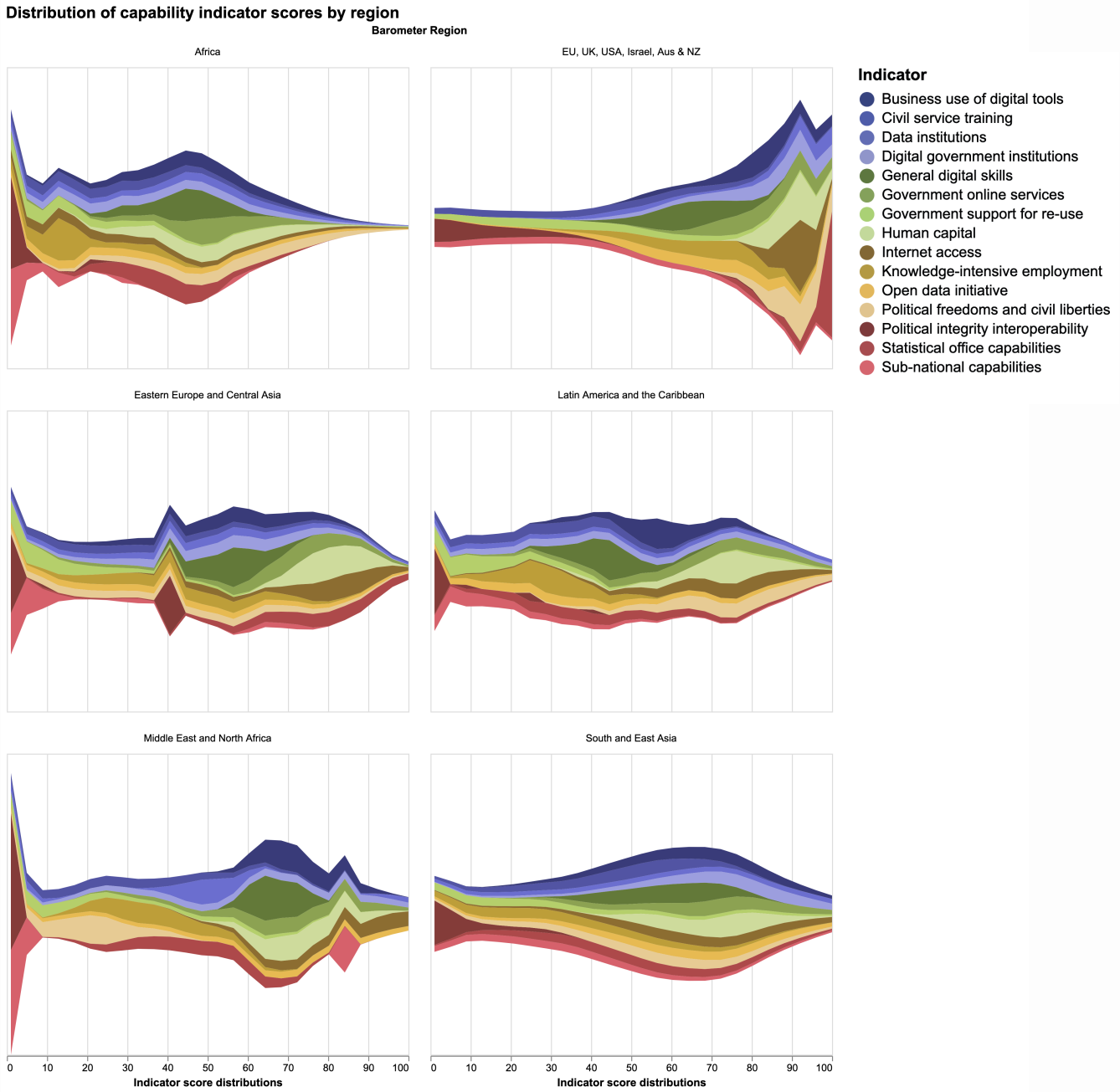
support for data re-use. Even high capability countries are mostly yet to develop robust interoperability infrastructures for public data as evidenced by low scores on the political integrity interoperability indicator.

In moderate capability countries, government is less likely to have adopted robust standards and practices in the production of official statistics, and countries score lower on the UN Government Online Services Index. However, 25 countries in this cluster still have some form of Open Data Initiative, although they are considerably less likely to have been assessed in our survey as having effective data management capability at a city, regional or local government level. Governments in this cluster are also less likely to be providing current support for data re-use than those in the higher capability cluster.

The lower capability cluster contains many countries where data institutions are currently lacking or weak, human capital scores are lower, fewer businesses currently use digital tools, and fewer people are employed in knowledge intensive industries. Internet access is also often more expensive and less extensive in this cluster. These countries may face particular challenges in providing and using data for the public good. However, during the study period, 11 countries in this cluster had an Open Data Initiative, and a number had evidence of capacity building, particularly around statistical data.

Which factors shape capability in each region?

The density chart below shows the distribution of unweighted scores for each primary and secondary Barometer capability indicator by region. A wider band to the left of each chart would show that scores on that indicator for that region skew towards the lower end, while wider bands to the right shows that scores skew higher. If the band of color for a particular indicator is relatively even across the chart, it indicates a spread of values for countries in the region.



Density chart showing the distribution of capability indicators by region: both primary and secondary indicators are shown, providing a view into relative strengths and weaknesses of each region, and how each variable is distributed.

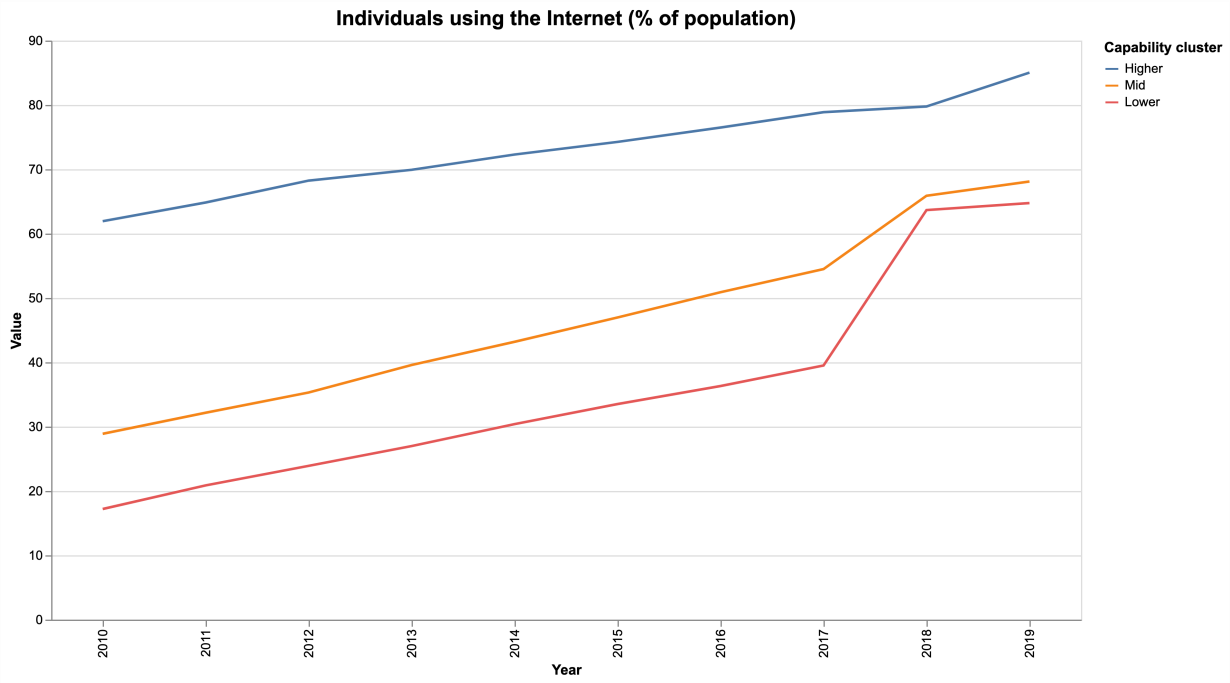
Qualitiative evidence collected by the Barometer survey suggests that the COVID-19 pandemic has acted as a powerful test of government, civil society and private sector data capabilities. In a number of cases, researchers report evidence that sub-national governments in particular have demonstrated the ability to respond rapidly to the data demands of the pandemic, rapidly taking on responsibility for creating and maintaining new data systems. However, in other cases, the pandemic has shown up weaknesses in the ability of state or civil society to use data for crisis response, or to provide scrutiny of emergency decision making.

Sub-national capabilities do not feature prominently in the chart above with all regions seeing both low levels of sub-national capability and broad distribution of scores with respect to the extent of robust sub-national practices around open data, data management and civil service capacity building, and none of our chosen secondary variables allowing us to explore whether the capabilities identified are concentrated in urban centers, or whether they are more evenly distributed across a country. However, qualitative evidence highlights the importance of a balance between national and sub-national capability to support inclusiveness and innovation. Researchers report, for example, on the situation in Spain, where strong data management capacity in the 17 autonomous regions is not always matched by

infrastructures or processes to integrate data at the national level, leading to strong data resources for engagement at the regional level, but making it arduous for re-users to obtain an overall national picture on key issues. For populations living near country borders, involved in trade, or working on cross-border issues, international data interoperability, and shared practice across countries also becomes important.

How are data capabilities developing?

The Barometer addresses capabilities through four dimensions: foundations (covering basic internet access, digital skills and human capital); government capabilities (to govern, produce and share data); private sector capability (to use data to generate economic value); and civil society and media capability (including the political freedoms that support broad ranging use of data for the public good, and civil society oversight of data governance).



Percent of population with Internet Access (Source: ITU) by capability cluster over time: showing convergence towards high levels of Internet access.

Over the last decade, Internet access has rapidly expanded across the globe with connectivity metrics rising much faster than those for digital skills or human capital. Even in mid and lower capability countries, the likelihood that any individual will have aspects of their life digitized and will at least have an access point where they could engage with digital services, has increased sharply. Rising connectivity provides opportunities and threats. Key to managing these is a balance of governmental capability, political freedoms to support checks-and-balances on government, and capability and engagement from other parties, including academia, the private sector, civil society and media.

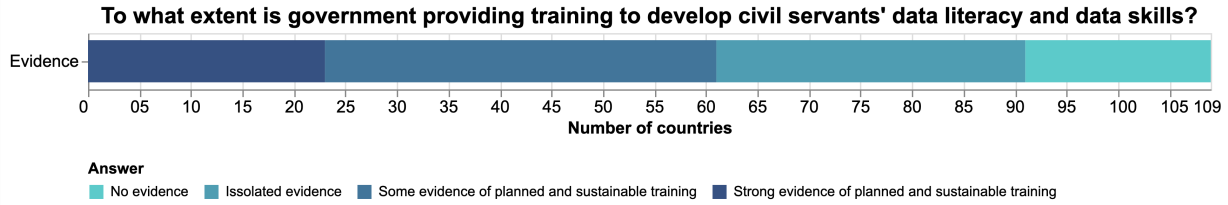
How are government data capabilities developing?

Mean scores by cluster on selected secondary capability indicators

-	Government Online Services Index	Digital Government Strategy	Data Institutions	Standards in Statistics
Higher	84	83.78	81.76	86.08
Mid	65.59	54.95	53.47	55.28
Lower	49.64	36.57	22.92	41.81

National governments have a critical role in data governance and stewardship, including a role in shaping data strategies, establishing and funding key governance institutions such as data protection authorities, providing digital services for data collection and access, and setting and adopting clear data standards. The secondary indicators in the Barometer show that lower capability countries face particular challenges when it comes to the presence of institutions in charge of data governance, management and data protection, and in having both key infrastructure (government cloud platforms), and strategies (including technology and interoperability strategies).

The Barometer explores the extent to which there is evidence that governments are investing in capacity building through providing training programmes to civil servants. Although over 90 countries have evidence of training being available, under 25% of countries appear to have widespread ongoing and sustainable capacity building programmes for public servants on data literacy and data skills.



Government provision of data literacy and skills training: Although over 90 countries have evidence of training being available, under 25% of countries appear to have widespread ongoing and sustainable capacity building programmes for public servants on data literacy and data skills.

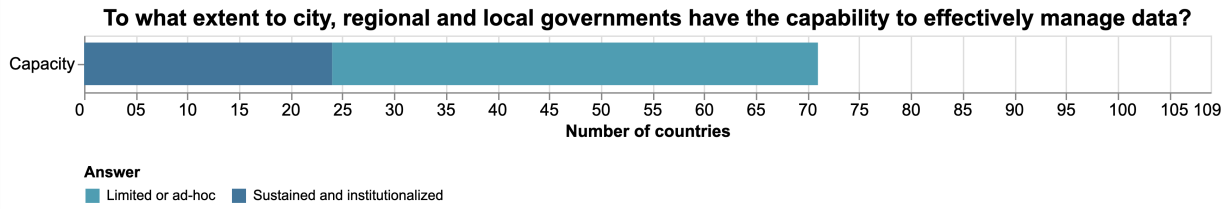
Notably, training availability is particularly limited at the sub-national level. Even where countries do have training programmes, reported figures on the number of civil servants trained suggest only a very small percentage of official are gaining access to targeted data-related skills training. For example, in Bulgaria, researchers note that, according to evidence from the Institute of Public Administration, “in the 2019-2021 period, all the [data related] trainings gathered 68 groups and a total of 1102 trainees”. This is less than two people for each of the 570+ executive or local government bodies in the country. A similar picture is observed in the UK, where the UK Data Science Campus reports having trained 681 analysts in data science tools as of April 2021. Researchers from the Gambia noted that, while government-provided data-related training is rare, technical staff from national government Ministries often access skills through degree programmes, generally studying abroad. Overall, across countries and regions, there is a sense that the actual, or potential, demand for capacity building currently far outstrips supply.

Scaling up data-related capacity building initiative requires leadership and strategy. The Republic of Korea is one of a few countries with evidence of a systematic focus on training through Article 25 (Education and Training about Public Data) of the Act on Promotion of the Provision and Use of Public Data, which requires the Minister of the Interior and Safety to formulate education and training policies for officials related to public data. By contrast, the researcher for Sweden noted that national data strategies contain no mention of data literacy and data skills training, and a mapping of introductory courses in 193 government authorities found no evidence of education related to digital literacy and skills.

For lower capability countries, donors have often been a key catalyst of capacity building. For example, researchers from Kosovo report that extensive training inputs were delivered when the country’s open data initiative was first launched, often drawing on NGO support to deliver the training. However, that training has not been sustained or systematized over time.

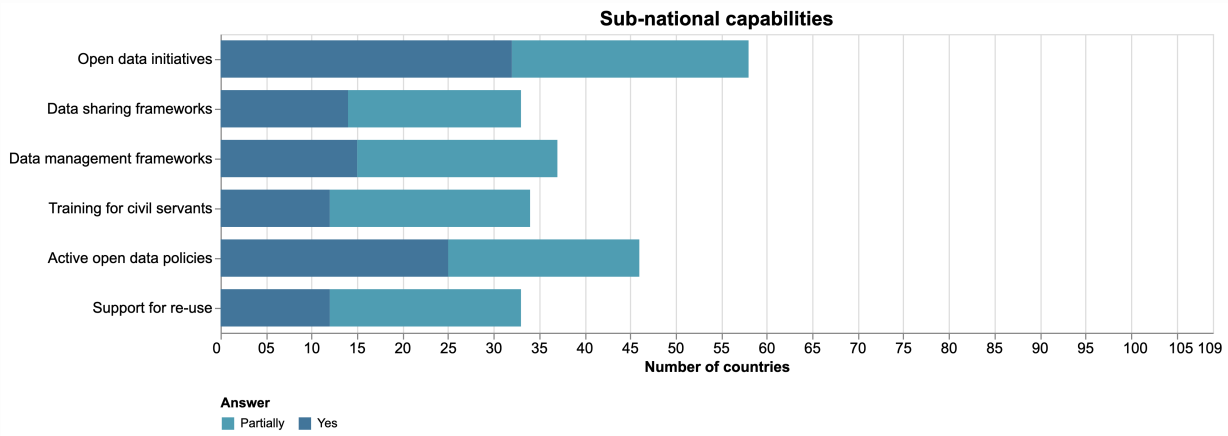
How are sub-national capabilities developing?

Although 71.6% of countries covered by the Barometer have some form of national data management policy or law, we only found evidence of sustained and institutionalized sub-national data management capacity in just 22% of countries.



Sub-national capabilities: Less than 25% of countries can demonstrate sustained and institutionalized data management capability at the sub-national level

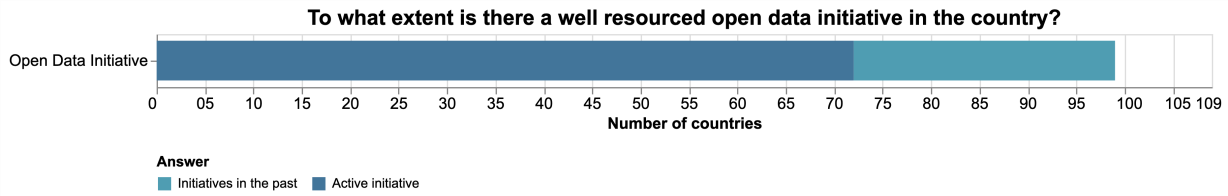
The sub-national capabilities identified by our survey centred on the presence of open data policies and initiatives with less evidence that sub-national governments are, as a matter of course, providing training to civil servants, or establishing clear local data sharing or data management frameworks. This is a notable gap, particularly given increasing interest in the potential role of private sector data as an input into city and state-level decision making. As recent work by[18] highlights, local governments face significant barriers and power imbalances when negotiating access to private sector data, and without effort to develop sub-national capacity and coordination for engaging in business-to-government data sharing, there are risks that local government will not be able to secure the greatest public good from data sharing arrangements.



Focus areas of sub-national capability: The sub-national capabilities identified by our survey centred on the presence of open data policies and initiatives, with less evidence that sub-national governments have clear data sharing or data management frameworks.

How are open data initiatives developing?

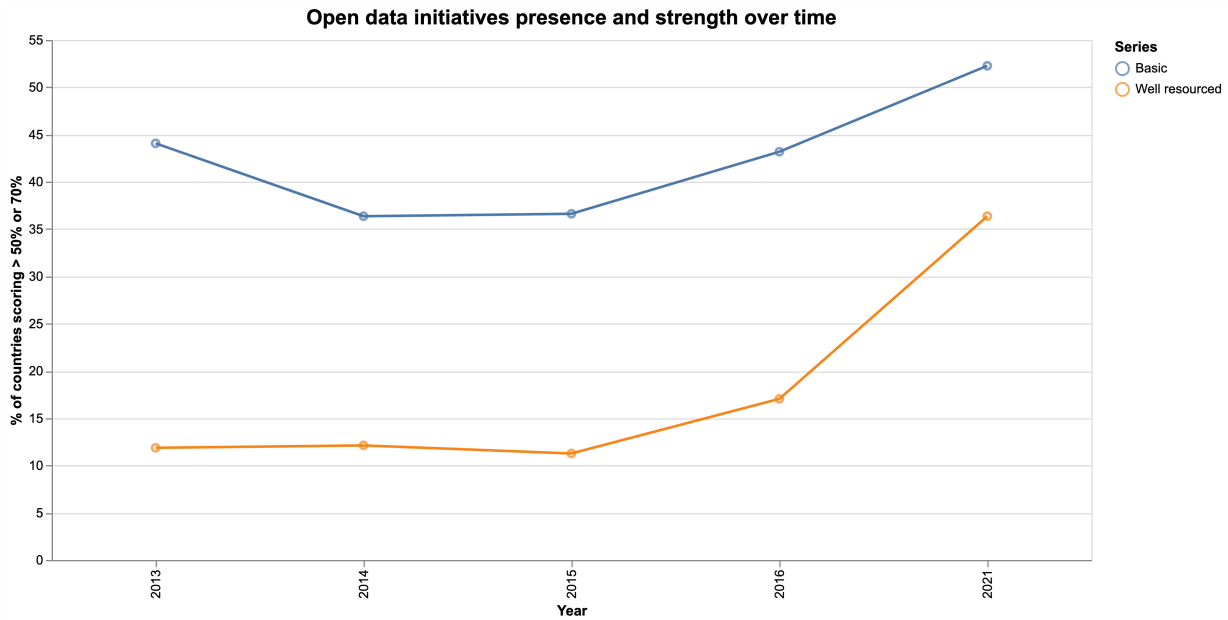
Alongside the 58 countries with some form of local open data initiative, almost 100 countries have, at some point, launched a national-level open data initiative. However, our survey found that only 72 of these remain active, suggesting that in recent years a number of open data initiatives have been cancelled or become dormant.



Open data initiatives: Nearly 100 countries have launched open data initiatives, although only 72 remain active

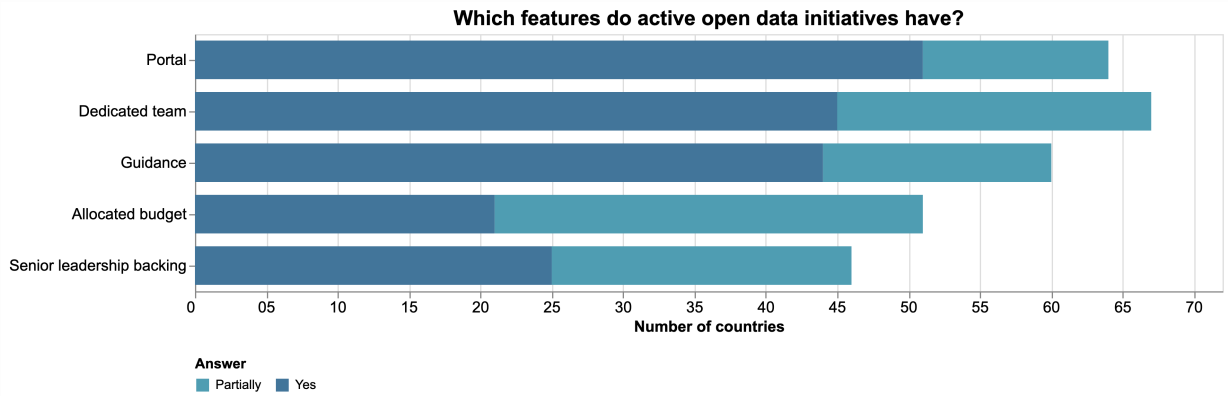
The Open Data Barometer has tracked the presence of open data initiatives since 2013. Taking countries covered by both the Global Data Barometer and the ODB, we can make a rough comparison of the presence of active open data initiatives (countries scoring more than 5 out of 10 on the comparable Open Data Initiative indicator from the two studies) and the percentage of initiatives that appear to be well-resourced (countries scoring more than 7 out of 10). As the chart below shows, the

number of active initiatives has only seem modest growth since 2016, and a number of countries with leading initiatives in 2016, such as Mexico and the United Kingdom, have seen significant reductions in their indicator score as open data activities have been de-prioritized or neglected.



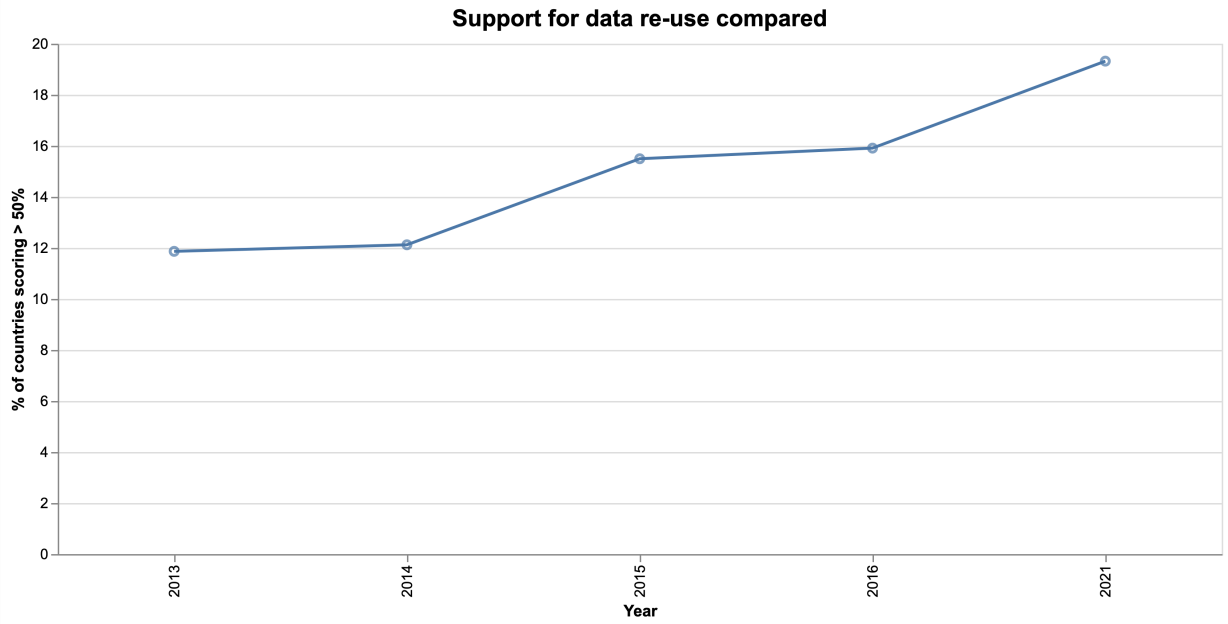
Strength of open data initiatives: Over 50% of countries covered by both the ODB and GDB now have active open data initiatives, representing modest growth on 2016. At the same time, the number of well-resourced initiatives has grown substantially since 2016.

Overall, the number of countries with evidence of well-resourced initiatives has grown, suggesting that where open data initiatives have remained active, they have become increasingly institutionalized with dedicated teams and technical infrastructure in place. A look at the sub-elements of the Global Data Barometer's open data initiative question reveals that a lack of allocated budget and limited senior leadership backing are pressing weaknesses for many initiatives, potentially representing the challenges of securing attention for open data activities when policy attention has moved to other related agendas such as artificial intelligence and data governance.



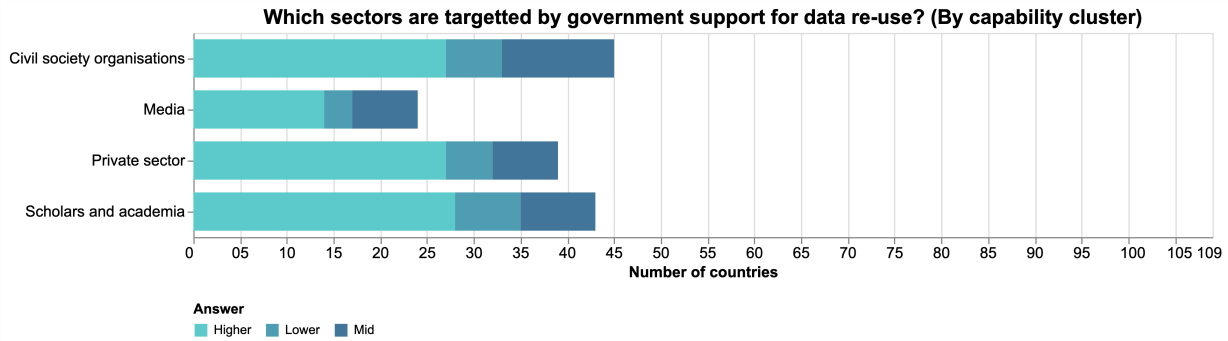
Features of active open data initiatives: Active open data initiatives generally have dedicated teams and technical infrastructure in the form of portals, but are often lacking allocated budget, and increasingly lack clear backing from senior government leaders

By comparing Global Data Barometer and Open Data Barometer indicators on government support for data re-use, we can also see a positive trend with a steady growth in the number of countries scoring above 5 on this indicator over time, albeit still a minority of countries reaching this threshold.



Support for data re-use: There has been modest growth in the number of countries providing support for data re-use over time

Overall, less than 50% of governments provide regular support for data re-use with the greatest support provided to civil society and scholars and the least support available for media.



Sectors provided with support (by capability cluster): Less than 50% of government provide regular support for data re-use, with the greatest support provided to civil society and scholars, and the least support available for media

What wider capabilities do countries have or need?

A countries ultimate capability to use data for the public good is not determined only, or even in the majority, by the direct technical capacity or interventions of government. The presence of technical skills, and the freedom to deploy those skills, in the wider population are key ingredients of data use for the public good.

There is limited globally comparable data on the prevalence of the kinds of digital skills that are required to govern and use data for the public good (and indeed, there are research gaps in identifying the full range of skills that should be tracked). We've turned to a number of secondary indicators in this component of the Barometer, though conscious that these do not offer us disaggregation by key dimensions of gender or marginalized status. To secure public good outcomes from data requires not only that countries have skilled citizens with the freedom to scrutinize or use data effectively, but that those skills are broadly distributed across age, gender, ethnicity, geography and social group.

Mean scores by cluster on selected secondary capability indicators

-	Business Use of Digital Tools	Knowledge-intensive employment	Human Capital	Political Freedoms and Civil Liberty	Digital skills in population
Higher	79.92	60.89	86.93	78.16	64.36
Mid	58.96	34.36	71.21	58.74	51.71

-	Business Use of Digital Tools	Knowledge-intensive employment	Human Capital	Political Freedoms and Civil Liberty	Digital skills in population
Lower	49.38	21.86	61.02	49.14	45.94

How can countries improve their data capabilities?

The Global Data Barometer website contains country profiles that show scores for each indicator and a breakdown of how these scores were generated. These can indicate particular areas of improvement for each country to focus on, and looking at the profiles of other countries, can highlight examples to learn from. Drawing on the quantitative and qualitative evidence captured in the capabilities pillar, we've identified three cross-cutting areas for action:

Sustaining and institutionalizing capacity building

Lower capability countries need to move from one-off or pilot capacity-building interventions to more sustained delivery of training and support around data collection, governance, provision and use. This involves programmes in the public and the private sector, as well as supporting the institutionalization of independent capacity building support for media and civil society. Capacity building needs to address both technical skills and critical social science and humanities skills for working with data. For low capability countries to improve data governance and management will require a focus on strengthening the national institutions that oversee data protection, open data, and data management, particularly developing capability around the use of common data standards.

Donor strategies need to focus on mainstreaming data-related capacity building, rather than providing short-term externally delivered programmes. This may bring slower but more sustained returns.

Particular attention should be paid to engaging marginalized groups as partners in the delivery of capacity building interventions to promote increased understanding of how data affects different populations.

Developing strategy and leadership

Mid-capability countries need to focus on the leadership and strategy that will deliver increased capability. There should be high-level political leadership not only for developing the role of data in the economy, but also for ensuring the country has skills, institutions, and freedoms required for data to be governed and used for the public good. Training and capacity building strategies should establish clear targets for increasing civil service data skills and for making wider training available to other sectors through established institutions.

A number of mid- and higher-capability countries have reasonably high levels of technical capacity in government and business, yet have a more constrained political environment, where it is harder for independent civil society to develop and support programmes on public good data governance and use. In these countries, efforts are needed to increase the space for independent and critical action with data and to develop models of data-enabled problem solving that allow government, academia, private sector and independent actors to work together in trusted ways.

Scaling and embedding capability

For higher capability countries the key challenge to meet is in scaling and embedding capacity building activities, particularly at the sub-national level. Ambitious targets should be set for training delivery and for the inclusion of technical and critical data skills in professional development frameworks. Renewed national leadership is needed for open data activities and to ensure countries can secure balanced public good outcomes from data sharing initiatives with the private sector.

In higher capability countries, innovations around data for the public good often originate at the sub-national level, with cities, states or municipalities that are 'outliers' when compared to the national picture. Where this is the case, countries should support efforts to share learning from these local leaders and should address national standardisation and interoperability frameworks to avoid the creation or deepening of significant data-divides within the country.

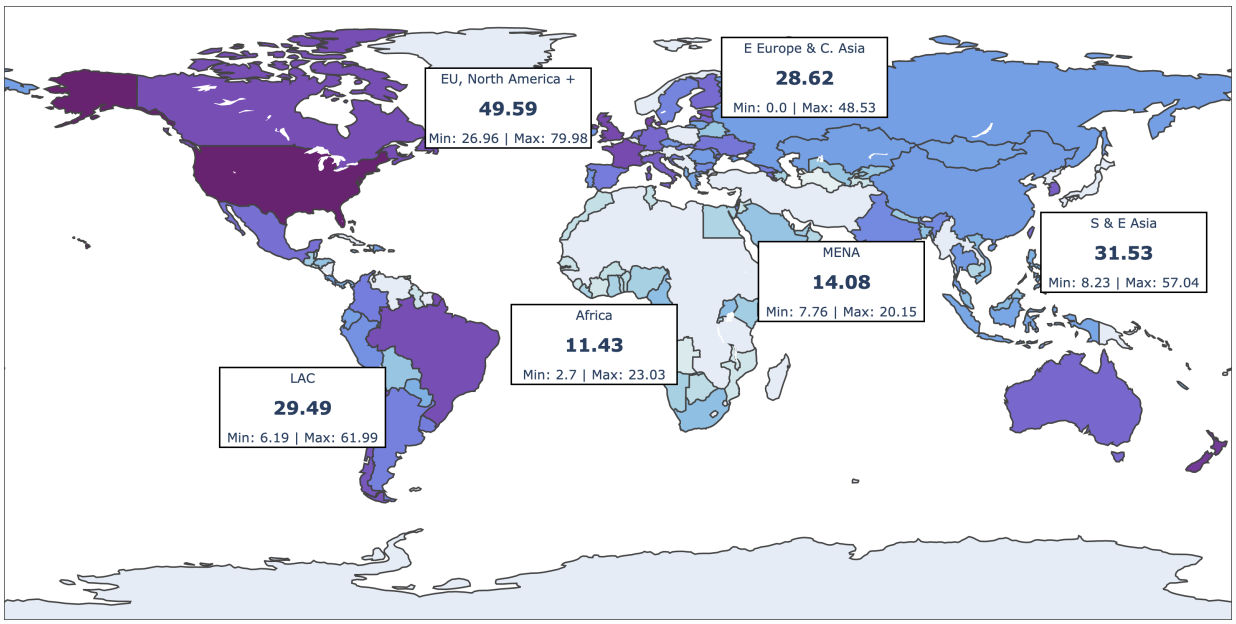
Availability

Not all data is created equal. Whether or not a country is able to realize the benefits of data for the public good will depend on whether key datasets relevant to solving social, political and environmental challenges are made widely available. In this first edition of the Barometer, we assess the availability of 19 datasets, selected based on their potential to address key issues such as climate change, public health, political integrity, and land rights.

Summary

- On a strict definition, the proportion of datasets published as open data (free of charge, in bulk and machine-readable forms, and openly licensed) has remained relatively stagnant over the last decade with 10.63% of the datasets surveyed meeting the open definition. This is only marginally above the high of 10% recorded in the 2015 Open Data Barometer. However, adopting a more flexible approach to assessment, we find a total of 17% datasets that either meet, or come close to, being provided as open data.
- The COVID-19 pandemic has demonstrated the importance of global data sharing and the capacity of the global community to establish shared data infrastructures, both through official measures, and through the collaborative efforts of independent researchers, citizen scientists and data activists. Drawing on secondary metrics, we find that data is available on COVID tests and positivity for 84.4% of countries and statistical data on health system capacity is avaialble for 86.2%. However, real-time healthcare system capacity metrics are available in just 46.8% of countries, and key data points that might be used to assess issues of equity in health system access and COVID vaccination roll-out are frequently missing even when data is available.
- Significant gaps also exist in the availability of key datasets to support national responses to the climate emergency. Although aggregate statistics are widely reported into international systems, detailed and disaggregated data on emissions, biodiversity, and climate vulnerability is rarely available as open data for use in-country.
- Where there have been concerted global initiatives, data availability and quality appear to be relatively high. For example, there have been significant global efforts over the last decade through fora including the Open Government Partnership to promote spend and procurement transparency and to build data capacity in this area. Our survey finds that 106 countries have at least some budget and spending information online and 100 have at least one example of public procurement data provided (though often from particular agencies, states or municipalities, rather than integrated national datasets).
- In areas of emerging focus, such as beneficial ownership, political integrity, and land tenure, the Barometer offers a baseline measure of data availability and openness, finding less than 50% of countries with data available against these indicators on average and less than 10% with open datasets. Future editions of the Barometer will be able to monitor progress against these baselines.
- Headline figures on data availability and openness hide significant regional variations. For example, more than half of the open datasets identified by the Barometer came from the European Union, UK, North America, Israel, Australia and New Zealand; whereas, in Africa, less than 40% of datasets sought were available online in any form.

- There is relatively little evidence that when key datasets are not available from government other data providers offer adequate substitutes. In just 84 of the 829 cases when governments were not providing surveyed datasets could researchers locate data from an alternative source.



Availability pillar regional scores: The highest scores are seen in the EU, North America+ regional grouping, with the lowest scores in Africa.

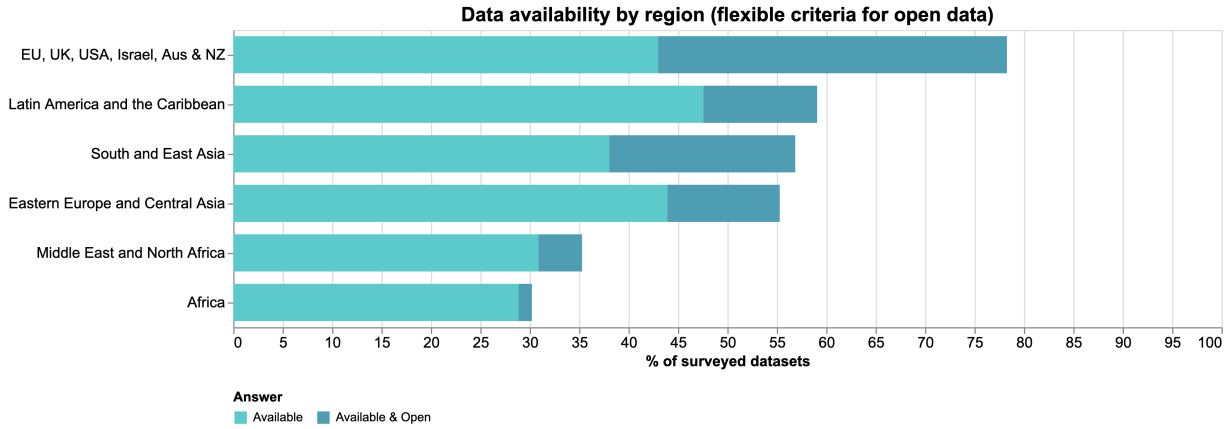
How open is the data covered by the Barometer?

According to the Open Definition[19], a dataset *"is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness."*. In practice, this translates into the data being available in a digital, machine-readable and non-proprietary form, either free or at no more than the cost of reproduction, and under explicit terms that permit re-use (rather than restrictive copyright terms for example). The central idea of the Open Definition is that when technical or legal barriers (either explicit, or as a result of uncertainty) to re-use are removed, data can be put to much wider use.

Barometer data provides us with two ways of calculating the total number of open datasets available across our sample. A strict approach, counting only those datasets that are available from government, and that researchers have assessed as having robust machine readability, bulk access, licensing, and being available at no cost, reveals 197 datasets (10.63%) meet the open definition. This represents modest progress on the figure of 7% of datasets fully open recorded by the Open Data Barometer in 2016, although matches the number found in 2014 and 2015 (differences in sample of both countries and datasets across studies mean these comparisons give a rough indication only).

If we adopt a more flexible assessment model and include data which has minor weaknesses when assessed against the open definition (for example, a license is not explicitly stated next to the dataset, but can be identified with extra research, or the data can be converted to a machine-readable form, but is not provided at source in the ideal structures or formats); in other words, we include datasets where Barometer researchers answers 'partially' in response to the questions on machine readability, bulk data, license and cost), we find a total of 315 datasets (17%) that could be considered to be, in effect, open data.

This headline picture hides significant variation between regions and sectors. The chart below shows the proportion of datasets our researchers found to be both online, and available as open data (using flexible criteria, to include cases where conditions such as bulk access or licensing are partially met) by region.

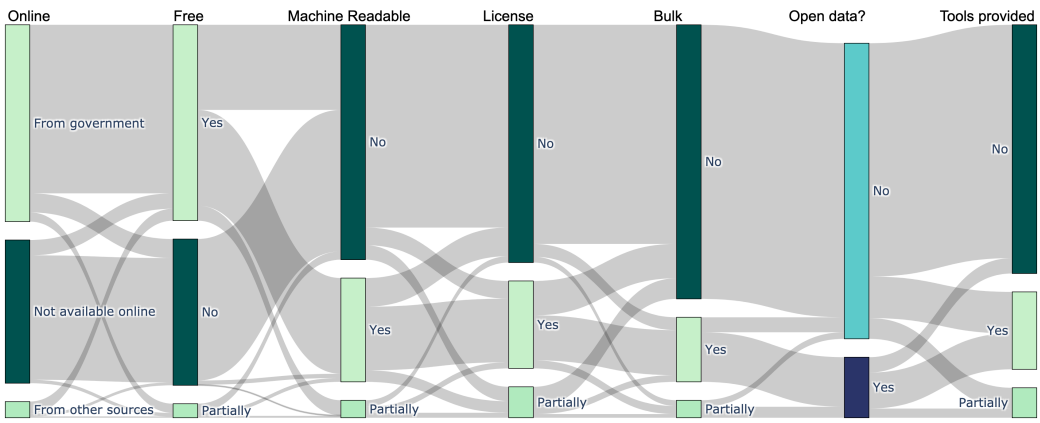


Data availability be region: Variation between the number of open datasets across regions is far greater than the variation between the availability of data in any form, highlighting a double data divide - firstly in access to any data, then in terms of access for data re-use.

Overall, more than half of the 315 open datasets identified by the Barometer are found in the 'European Union, UK, North America, Israel, Australia and New Zealand' region, whereas just 11 of the 374 datasets checked in the Africa region were judged to be open data.

In 2016, the Open Data Barometer concluded that the main factor holding back the spread of open data was the lack of open licensing[20]. As the diagram below shows, in 2021, the biggest limiting factor appears to be the lack of bulk data provision.

The Global Data Barometer availability assessments also include a variable asking whether there are 'accessible and open official tools available to help users explore a dataset'. Although provision of bulk data is important to support innovative data re-use, in many cases, users also want to be able to directly access facts or insights from a dataset without necessarily downloading a full or complex dataset. We found that in 46.4% of cases where data is available online, and in the majority of cases where data was published as open data, some form of online tool was available to explore it. The Barometer dataset contains links to all the tools identified, providing a future opportunity to explore the kinds of interfaces that might be created to increase the range of users who can benefit from provision of structured and open data.



Barriers to open data availability: The Sankey diagram shows the aggregate assessment of all datasets against the open definition criteria (Online, Free, Machine Readable, Open License, Bulk data) as well as the presence of accessible tools to explore available data. Partial responses are shown separately, but are counted towards meeting the open definition criteria (i.e. using a flexible approach to the open data definition).

Are other organizations providing data when government does not?

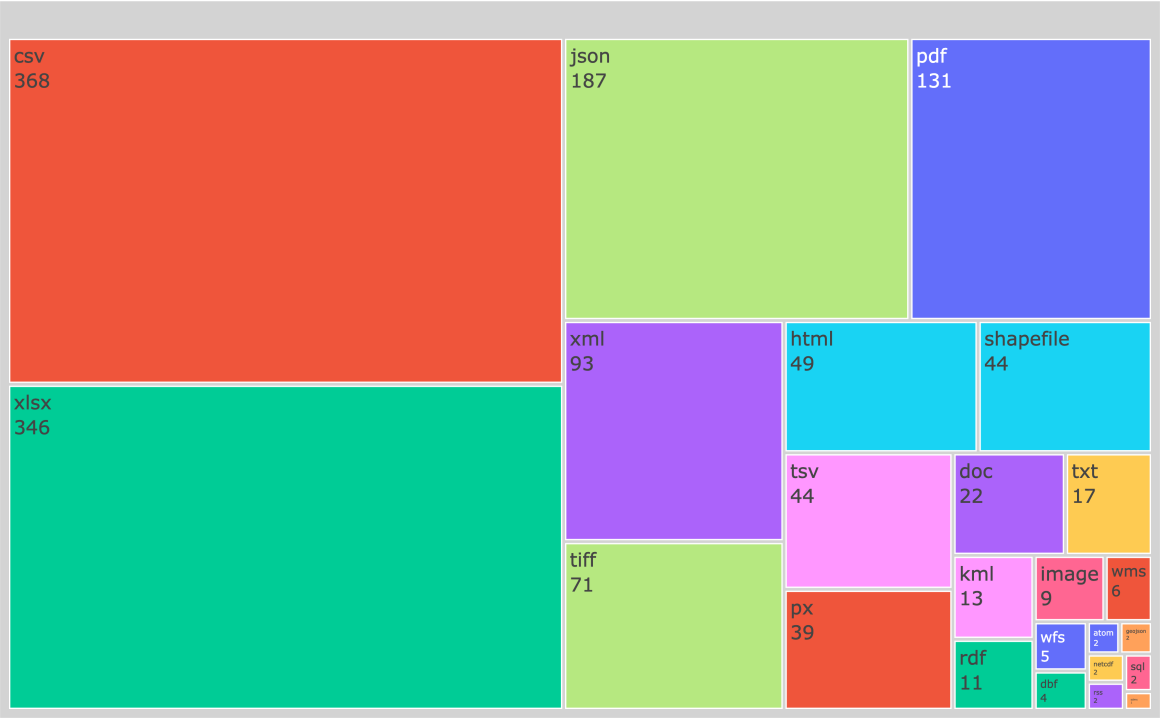
For each dataset surveyed, if researchers could not locate a data source made available by government, or because of action by government (e.g. a government mandate that a third party should publish the dataset), we asked them to identify if relevant data was available from another source. In just 84 of the 829 cases when governments were not providing data was an alternative source identified. These alternative sources included academic platforms, civil society websites, and data platforms provided by donors or multilateral organizations such as the World Bank or World Health Organization. In a small number of cases, particularly in relation to company information, private sector data providers were identified, but generally only offering paid-for, rather than open access to, data.

Overall, we see little evidence that alternative data sources are currently offering effective substitutes when there are gaps in government-provided data, although for lower capacity countries, international organizations and global civil-society or research-led platforms can provide useful data hosting and publication environments. For datasets provided through these alternative platforms, on average 45% of the desirable fields or features we looked for were provided, in comparison to 58% of the features when data was provided by, or because of, government action.

Is the available data fit for purpose?

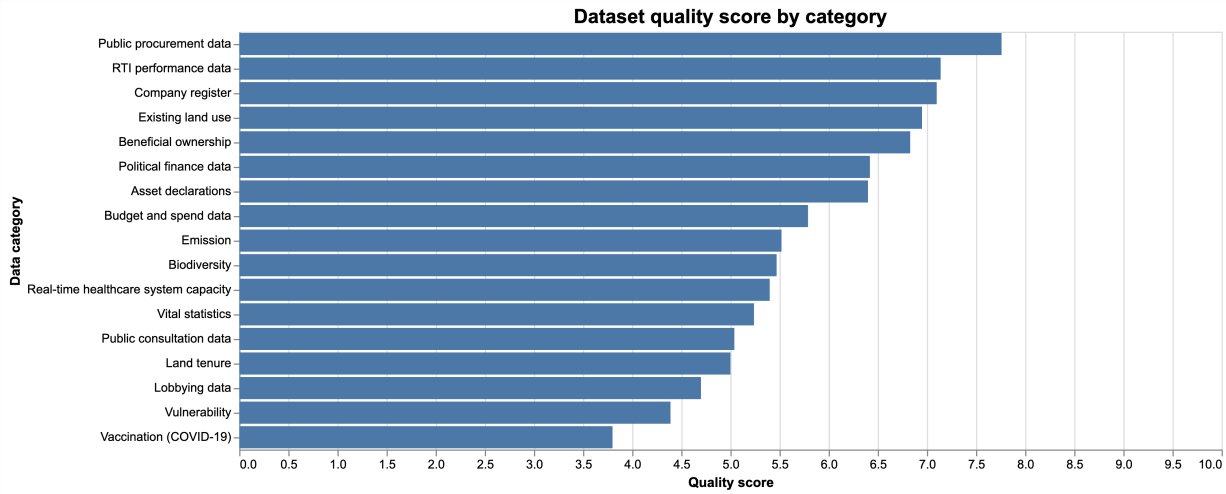
While the true test of whether data is fit for purpose is whether the data is being successfully put to use to solve a range of social challenges, the Barometer offers a number of indicators that capture aspects of data quality. We find that 73.78% of online datasets were assessed as reasonably 'timely and updated' and 61.34% of online datasets provided some degree of historical data that would allow users to track change over time (for example, records of past land use, or information about the previous as well as this year's budget). That still leaves almost 40% of datasets where only current data is being published, potentially creating challenges for a number of accountability or analytical use-cases.

The file format(s) a dataset is provided in can impact on its usability. Researchers provided free-text reports of the main file formats used by machine-readable datasets, which we have recoded to discover a narrow preference across published data for the non-proprietary CSV format and growing evidence of the use of other structured formats like json and XML.



Treemap of file extensions detected in free-text reports on file formats used by datasets.: CSV is the most commonly encountered file format, followed by Excel (xlsx) files, json, pdf and xml

Each primary availability indicator in the Barometer includes a number of sub-questions designed to check for the availability of certain dataset fields or features that are important to public good use-cases for the data, and which it is reasonable to expect that data publishers should provide. Although these field and feature lists are tailored to each data category, and some may be trickier to deliver than others, by looking at the mean weighted score on these sub-questions for each, we can identify the datasets most likely to be fit for purpose.



Mean dataset quality score by category. Procurement data scores highest for the presence of key fields and features in published data, while health and climate datasets are less likely to contain the full range of features checked by the Barometer survey.

Procurement data tops this table, although as noted below, many of the datasets assessed for this indicator are sub-national or single agency examples, which may act as positive outliers, and which are not always representative of the quality of data available across the country as a whole. Notably, health and climate datasets are among the most likely to have significant quality gaps when measured on the Barometer indicators.

A number of the dataset fields and features we checked for can, if present, be used to specifically focus on patterns of inclusion and exclusion. This includes data fields relating to gender, disability, or status within a marginalized group or indigenous populations. Comparing the number of times these fields were identified to the average number of times non-inclusion related fields were identified, we find that inclusion related fields are almost 50% less likely to be provided than other kinds of field. In other words, there are significant gaps to be addressed in the extent to which available data supports action on inclusion and challenging patterns of discrimination or marginalization.

Which datasets are available?

In this first edition of the Barometer, we looked at the availability of 19 datasets, organized under a number of thematic modules. This section presents a number of selected observations for each module, along with summary table entries for each dataset.

Methodological note

Summary tables present four values for each dataset covered by our primary survey.

- **Available online** represents the number of countries in which some data meeting the indicator definition was found online. This may include data only available at a sub-national level, or data with significant limitations (such as being presented in non machine-readable forms, or on websites requiring registration or payment)
- **Open data** represents the number of countries with at least some data meeting the open definition. We adopt our flexible definition of open data, including countries where researchers responded 'partially' to one or more of the open data assessment questions.
- **The quality score** represents a weighted average of how many desirable fields or features for the dataset were found to be available in online data. Note that the quality criteria are different for each data category.
- **The openness score** represents a weighted average of the standard openness questions asked for each dataset. These include questions about the presence of timely and historical data, as well as the availability of accessible tools for exploring data. The criteria used are common across each data category.

Health and COVID-19

The COVID-19 pandemic has underscored the importance of reliable, accessible, and trusted health data to enable coordinated action. While the governance of health data is addressed by indicators on the presence of data protection frameworks, in the availability pillar of the Barometer, we have looked for evidence that data is available to support both long-term health planning and to enable rapid-response specifically in the context of the COVID-19 pandemic.

Dataset	Available online	Open data	Mean quality score	Mean openness score
Vaccination (COVID-19)	95	31	3.8	6.37
Healthcare system capacity*	94	n/a	n/a	n/a
Testing data (COVID-19)*	92	n/a	n/a	n/a
Vital statistics	85	29	5.24	6.43
Real-time healthcare system capacity	51	17	5.4	5.83

* Indicators based on secondary data that doesn't enable us to calculate quality or openness scores.

Although backward-looking statistical data on health systems, and aggregated data on COVID-19 testing, are widely available, real-time data on healthcare system capacity and data-points that could support action on health inequality are much less likely to be provided online or in open formats.

At the time of our data collection, the Our World in Data project had managed to bring together data on COVID testing from over 100 countries, 92 of which are covered by the Barometer. Although they have been able to make the data, harvested from national Ministry of Health websites, or regional sources such as a the Africa Centres for Disease Control and Prevention, broadly comparable, documentation of the different data sources[21] highlights the significant challenges involved in locating data and finding documentation that can explain whether test figures relate to individuals or samples and which kinds of tests (e.g. PCR or antigen) are covered. Bringing together data from so many countries in such a short time is a remarkable achievement, but it also remains more complex than it should be with significant scope to strengthen data publication practices at source, including through more machine-readable publication and greater provision and adoption of standards for data publication[22].

The same activities, such as monitoring availability of regular and intensive care hospital beds, can, depending on the data infrastructures in place in a country, be used to generate both aggregate statistical data and to provide real-time information about health service capacity. In some countries, this real-time data is vital to allow individuals to know where they can access care, and at critical times, it may also be of significant-value to journalists or civil society organizations to scrutinize healthcare provision, performance and equity. In a number of countries, however, real-time data is either not collected, or only shared with government agencies or medical professionals. Drawing on data from the Open Data Inventory (ODIN), we find that 86.2% of Barometer countries provide statistical data on healthcare system capacity, but only 46.8% have real-time, or near real-time, data available, and of these, just 33 countries provide facility (e.g. hospital or clinic) level data, and only 35 have information on bed availability. In a number of countries, real-time data has been made available for the first time in the context of COVID-19, raising questions about whether states will cease this publication in future, or put the provision of healthcare capacity information on a more stable footing.

The quality sub-questions for our survey indicator on vaccination rates paid particular attention to equity of vaccine distribution, asking whether available data was disaggregated by age, gender, geography, disability status and healthcare situation. Although 46.3% of available vaccination datasets (67.7% of those published as open data) contain age disaggregation, just 33.7% (51.6% of open datasets) were disaggregated by sex or gender, and only 13.7% (29% of open datasets) provided a breakdown by disability status.

As the health module of the Barometer is entirely based on availability indicators, we can use the module scores to build a picture of comparative regional performance on health data availability. The table below shows the top five questions in each Barometer region, along with their module scores.

Africa	Eastern Europe and Central Asia	European Union, United Kingdom, North America, Israel, Australia and New Zealand	Latin America and the Caribbean	Middle East and North Africa	South and East Asia
Uganda (44.97), Burkina Faso (36.03), Kenya (34.03), South Africa (32.22), Rwanda (30.08)	Ukraine (55.15), Armenia (49.14), Moldova (44.77), Mongolia (41.75), Kyrgyz Republic (41.27)	USA (85.4), Germany (83.02), Slovakia (82.4), New Zealand (79.19), Italy (78.56)	Peru (77.08), Brazil (73.22), Chile (73.21), Argentina (63.63), Uruguay (58.11)	UAE (41.58), Qatar (41.09), Saudi Arabia (36.59), Oman (32.62), Bahrain (31.47)	Republic of Korea (75.62), Hong Kong (67.17), Taiwan (65.41), Bangladesh (55.69), India (55.06)

Climate Action

As the UN's 2030 Agenda makes clear, addressing the climate crisis is a globally agreed public good. Climate change and the actions that governments and publics can take to mitigate and adapt to climate change are matters of vital importance around the world. Data can be a powerful tool for prioritizing and assessing climate action, as well as a means to expand the number of groups who can track and propose climate action, take meaningful action in their own communities and organizations, and hold governments accountable for their actions or inaction.

The Open Data Charter's Open Up Guide for Climate Action[23] identifies a range of climate-relevant datasets that states could publish, from emission statistics to biodiversity indicators, power generation statistics, and data on climate risk exposure. The guide notes that although:

> "countries may already be required to report some of this information to a regional or international body … it is important to consider how the information that is already available or reported could be made more accessible and useful to third parties by its publication in open data format."

The Barometer indicators provide an insight into how far data is being made available at the national level, highlighting some bright spots and some significant gaps.

Dataset	Available online	Open data	Mean quality score	Mean openness score
Emission	73	28	5.52	5.81
Biodiversity	68	16	5.47	4.75
Vulnerability	58	14	4.39	5.49

Although 73 countries provide emissions data, and 28 do so as open data, there is significant variation in how this is provided, how timely the data is, and how detailed. In many cases, countries are reporting emissions data under the United Nations Framework Convention on Climate Change (UNFCCC), but not providing access to this data for national citizens. This suggests that when international reporting standards are developed, consideration should also be given to standards or recommendations for also making data available at the national level, maximizing the value realized from data collection exercises.

The data gaps are even larger when it comes to climate vulnerability data. In spite of the importance of localized data on climate vulnerability to support adaptation, just 53.2% of countries surveyed provided relevant data (14 as open datasets), and the datasets that we were able to locate were notably weak on the provision of information about how vulnerabilities may differentially affect marginalized populations.

The headline figure of 62.4% of countries providing biodiversity information online also hides notable variation in terms of the life-forms represented in available data. For example, 60.6% of countries provide data on endangered birds and 58.7% cover mammals, but just 37.6% provide data on fungi and lichen. This comparative lack of attention to fungi is particularly notable given the vital role they have been found to play in ecosystems around the world, as well as serving as a significant global carbon sink.

Early work by Barometer partners with preliminary data from the climate action module has drawn attention to the importance of using the qualitative insights from the survey to understand diverse practices of biodiversity monitoring, recognizing that there are many histories and practices of gathering information about biodiversity and endangered species, from scientific taxonomies to colonial inventories and local knowledge practices. In a point that can be adapted to many of the datasets covered by our survey, researcher Dr Jonathan Gray argues that: *"Making biodiversity data more widely available may not only serve to enable its re-use to communicate the state of biodiversity to broader publics, but also to facilitate public involvement, learning and surfacing different kinds of local expertise and action to understand and care for the life which data portrays."* (Source: Workshop on preliminary findings from the Global Data Barometer Climate Action module at Mozilla Festival on 7th March titled "GDB as a collective learning device to explore climate action".)

Political Integrity

The public good can best be served when there is an open, accountable, and equitable public sphere in which money doesn't distort fair decision-making or access to political office. Data can be a powerful tool to identify whose interests shape how governance decisions are made and implemented. Within democratic political systems, this involves transparency of political party finance, information on the interests of political decision makers, information on lobbyists' interventions, and information on public consultation processes in rule-making, as well as a robust right to information system that helps members of the public evaluate and hold to account those in power.

Our focus on political integrity data aligns with SDG 16: Peace, Justice, and Strong Institutions: particularly its targets around rule of law (16.3); transparent, accountable institutions (16.6); responsive, inclusive, participatory, and representative decision-making (16.7); and public access to information (16.10).

Dataset	Available online	Open data	Mean quality score	Mean openness score
Public consultation data	62	8	5.04	4.15
Political finance data	57	10	6.42	5.33
RTI performance data	43	8	7.14	5.62
Asset declarations	50	4	6.4	4.73
Lobbying data	19	4	4.7	5.19

Across this module, we find a very mixed picture of information and data availability.

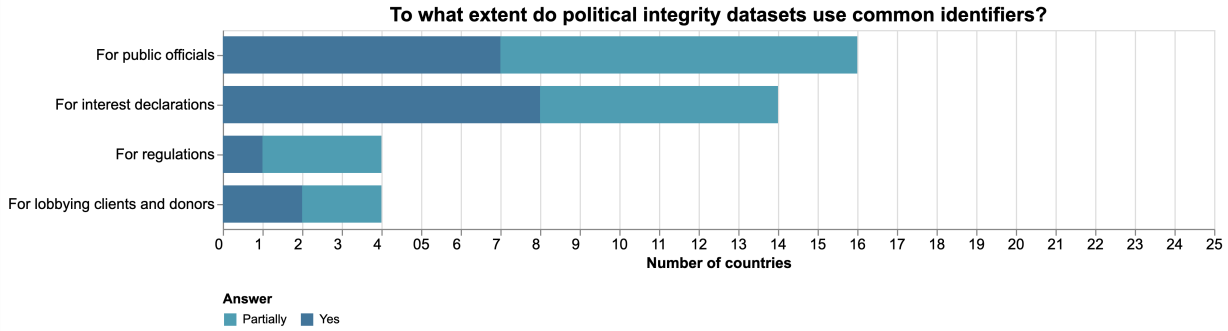
When it comes to public consultation, we find that although in 56.9% of countries surveyed, there is at least some information online that can be used to support engagement in, or scrutiny of, public consultation processes around executive rule-making (sometimes called 'secondary legislation'), very few countries (17.4%) provide machine-readable or bulk data (11%) in ways that might facilitate use of innovative tools to support citizen engagement or to monitor special interest influence through consultation processes. A number of the strongest performing countries in this indicator have national electronic platforms to facilitate rule-making, such as the USA's Regulation.gov and Estonia's Eelnõude Infosüsteem, although others take a more distributed approach with the responsibility on each agency to host details of consultations and consultation responses. Beyond the leading countries, there is a sharp drop-off in the quality of data provided with many countries that counted among those providing *some* data online either only covering a small set of possible consultations, lacking information about comments made, or missing other key data features: focusing primarily on advertising the opportunity to participate.

While the vast majority of the countries we have surveyed have some form of right to information (RTI) legal framework, the RTI availability indicator provides new insights into the number of countries that are publishing performance information for monitoring implementation of their RTI rules. It shows that 39.4% of countries have some performance information online, with 88.4% of these providing reasonably broad coverage of regions and agencies. Where information is available online, 97.7% of countries provide data on the number of requests made, although just 58.1% provide data on when and why information was withheld and 58.1% provide data on the timeliness of responses to requests, suggesting that much of the information provided will be of limited use to monitor compliance with statutory timelines. Data on appeals against RTI refusals is available in 58.1% of countries with information online, and in some cases, **only** appeals data is available as statistics are only collected for appeal processes, rather than for initial requests.

It is important to note that a high rating on the RTI availability indicator does not prove that a countries RTI regime *is* performing effectively in practice. However, the presence of performance information demonstrates a commitment to RTI implementation and provides a resource for scrutiny of how RTI rules are working. The supporting evidence and links gathered for this indicator (available through the Barometer website) have the potential to support further work on developing cross-country comparative metrics for RTI performance, as well as providing pointers to sources that could be used for more in-depth qualitative research.

Both Political Finance datasets and data on Interest and Asset Disclosures are particularly valuable when provided in bulk and machine-readable forms that allow analysis to find patterns or look for connections that might indicate political influence by particular companies or individuals. Although 52.3% of countries have some form of political finance data online, and 45.9% have asset declaration information, less than 10% of countries provide open datasets in either case with disclosure systems often paper-based or insufficiently digital to support provision of reliable machine-readable data. This, coupled with the low availability of common identifiers for donors (Just 29.8% of countries with information online), or the individuals making asset disclosures (40%), means that re-use of this data has to rely on intermediaries who can digitize resources and use various fuzzy matching techniques to support data exploration and investigations.

The lack of an interoperability infrastructure for political integrity data is illustrated by an indicator from our capability pillar that finds that very few countries have established widespread use of common identifiers that could tie together datasets. Although there are legitimate privacy concerns about enabling some forms of data linkage, there is significant scope for countries to develop more digital-first and joined-up approaches to integrity-related disclosure processes.



Presence of common identifiers to support political integrity data interoperability: Very few countries have interoperability infrastructures in place to support joined-up political integrity data: creating avoidable barriers to the use of political integrity data to promote greater oversight and accountability.

The missed opportunities and lessons from poor interoperability of political finance and asset disclosure data should be instructive for the design of future lobbying frameworks and data publication. The Barometer reveals, for the first time, the low availability of robust data on how private interests are seeking to influence public policy. With just 17.4 % of countries providing data online, and only 4 offering open data, this first edition of the Barometer offers a baseline to track progress if lobbying transparency gains, as we hope, greater profile in fora such as the Open Government Partnership in the coming years.

Overall, although comparing the governance indicators paired with each political integrity availability indicator reveals a clear positive correlation between the quality of rules requiring data publication and the availability of data (see Governance chapter), it also reveals significant implementation gaps. For example, while 103 countries have rules requiring interest and asset declarations, and 53 include requirements around structured data collection and publication, just 50 have any information available online with just 4 providing open datasets. Data on the Barometer website can help countries to identify best practice examples from among their peers to support action to close this gap in future.

Company Information

Private firms can be engines of development, innovation and the delivery of vital services and consumer goods. At the same time, company activities can cause environmental and social harms, corporate structures can be abused for money laundering, corruption and to hide wrongdoing, and opaque corporate structures can harm international trade. Information on company registration, ownership and activities can provide critical evidence for public understanding and regulation of the activities of companies and can support the functioning of a productive private sector that supports sustainable development outcomes.

Over recent years, international standards have developed that require minimum levels of data collection on the registration and beneficial ownership of companies, trusts and other legal arrangements, with civil society advocating for this information to be made publicly accessible, and regional and sectoral norms of public disclosure developing as part of moves towards greater transparency[24]. The open data movement has placed particular emphasis on company identifiers as a point of linkage between different public datasets, including public finance, public contracting, and political integrity datasets[25].

Dataset	Available online	Open data	Mean quality score	Mean openness score
Beneficial ownership	33	10	6.83	4.96
Company register	78	24	7.1	5.55

At present, early progress on establishing open beneficial ownership registers risks being held back by a lack of established open data practice amongst company registers.

Although 77 countries have at least basic online access to company registers, and 82.1% of these provide access to some of this information free of charge, just 46 countries provide machine readable company data, and only 24 countries provide data accessible in bulk under open licenses to meet the open data definition. In many other cases, the structured data is only available through commercial data sharing agreements.

Public Finance and Public Contracting

Data on government spending and procurement can be a powerful resource for improving efficiency and effectiveness of government, stimulating private sector innovation, facilitating increased control of corruption, and supporting greater public participation in decisions about public spending.

Data on budgets, spending and public procurement are among the highest scoring on availability indicators in the Barometer with examples of structured and machine-readable data available on every continent. The relatively high availability of well-structured open datasets appears to reflect the impact of long-standing global campaigns and the provision of technical assistance to countries to implement open data sharing by organizations, including our thematic partners the Open Contracting Partnership (OCP) and Global Initiative for Fiscal Transparency (GIFT). Public procurement and public finance systems have also been more widely digitized, making the release of data reliant on political will in many cases, rather than technical capacity. Notably, not all the countries with open public procurement data have open public finance data, and vice-versa.

Dataset	Available online	Open data	Mean quality score	Mean openness score
Budget and spend data	106	33	5.79	6.47
Public procurement data	100	37	7.76	6.57

The Barometer's indicators provide a breakdown of the features of available datasets, complementing studies, such as the Open Budget Survey[26], which have long tracked the extent to which key budget documents are made available, as well as the provision of opportunities for public participation in budget processes. Using this detail, we can see significant areas for improvement around the provision of budget disaggregation by cross-cutting programs or issues, such as the SDGs, climate action or gender budgeting (currently available for just 21.2% of open datasets), and inclusion of identifiers to link budget and spend to major projects (present in 45.5% cases of open data).

There are similar gaps in available procurement data, where although researchers located many cases where at least some tender and award information is published (81.7% and 79.8% of countries respectively), data on the planning phase (56%) and implementation or delivery of contracted goods or works (30.3%) data were much less likely to available. In general, many of the key features for understanding public contracts, such as a description of the goods, works or services being procured, the name of the company awarded the contract, and contract values, were identified as present in the majority of datasets checked. However, this does not guarantee that these data fields are *always* present. In 33.9% of cases, researchers reported either that there was evidence that a portion of mandated data was missing or there was evidence of widespread omissions in the data (procurement was among the datasets with the greatest number of countries with missing data reported). Many of the procurement data sources identified also only cover a limited number of agencies, or in some cases represent examples from leading cities that are yet to be replicated across the whole country. Qualitative responses also reported on a range of specific challenges around procurement data quality, such as missing award notices or notices with key variables unfilled. This highlights that, when looking at the availability of data for the public good, it is not only the fields listed in column headers that matter - but also the extent to which those fields are populated with meaningful data. A global instrument like the Barometer can surface indicators of likely data quality, but the full quality and value of any dataset can only be understood through data use.

Land

Land tenure data identifies who holds rights over land. This data can be used to understand the land ownership landscape in a country, to identify land concentration, and to understand access to land and land tenure security, as well as for anti-corruption purposes[27]. The Barometer looks at the availability and openness of land data and at how far available data covers different kinds of tenure, including urban and rural tenure, state owned land, common lands, company owned land and land held by private individuals.

Land use data describes the purposes to which land is put and has significant value for economic planning, environmental protection and work on climate change mitigation and resilience. The Barometer looks for structured datasets that detail the kinds of activities occurring in particular locations with associated geo-spatial references. We look for land use data that includes metadata describing how land has been classified, as well as information specifically on protected areas and forested areas.

Dataset	Available online	Open data	Mean quality score	Mean openness score
Land tenure	53	11	5	4.08
Existing land use	77	31	6.95	5.57

While a number of countries are publishing structured open data on land use, in many cases, online land-use maps are presented as PDFs or images without the underlying data. There is potential for significantly more land use data to be made available by building the capacity of key data owners to publish in machine-readable formats.

For land tenure data, there is a mixed picture across the globe. In some cases, data is not yet digitized. In others, data exists, but is only available following formal applications, or payment of fees. While some countries have developed robust models to balance the open provision of granular tenure data with protection of individual land owners' personal details, there is a pressing need for more consistent approaches to manage accessibility and protection of data. The Barometer provides a new benchmark for tracking future progress in making particular kinds of tenure information available online, including state owned lands (currently 41.3% of countries have some information available), corporate-owned land (36.7%), individual tenure (41.3%), and the land tenure of indigenous or marginalized people (15.6%).

How can countries improve data availability?

For each country, the Global Data Barometer website contains a profile that shows the breakdown of individual scores against availability indicators. This can be used to identify particular areas for improvement based on the data categories covered in this edition of the Barometer. Future editions of the Barometer will address a number of other sectors, as well as measure progress against baselines established in this edition.

Drawing on the quantitative and qualitative evidence captured for the availability pillar, we highlight the following common areas for action.

Develop data infrastructure

Low capability countries are often reliant on outdated, or externally provided, infrastructures for data publishing. This can constrain their ability to manage data and to publish it in accessible formats. Governments need support to develop integrated national data systems[28] and to make sure these are able to support open data flows.

For mid and higher capability countries, there should be a focus on the interoperability of data through adoption of common standards and identifiers that can allow links to be made between datasets, while paying attention to the potential risks of certain data linkage.

Refresh data publishing practices

Mid capability countries need to focus on filling data gaps and improving data quality. This requires a move from approaching open data as a tick-box exercise based around data portals to a focus on 'data as a service', recognizing that published datasets need to be designed around user need and need to be actively maintained. In some cases, this will involve wider processes of service redesign, looking at how data from land use systems and land registers, company registers, environmental reporting systems or other elements of national data infrastructure can be provided in sustainable open forms.

Federal countries, or countries where cities or states are leading on the open data agenda, may need to focus in particular on interoperability to move towards a situation in which data is reliably available across the whole country, rather than being restricted to particular urban areas.

Support provision of both data *and* interfaces

Even for higher capability countries, gaps in the provision of bulk data, restrictive licensing arrangements, and charging models for data, are holding back progress towards data being widely available for re-use. At the same time, a lack of accessible interfaces for exploring data means that a large pool of potential data users are not able to find the particular data points or facts that could be relevant to them.

Countries should focus on understanding the wider national and local use-cases for each major dataset they are providing and should look to provide different means of access for different audiences either directly or by actively supporting sustainable intermediaries.

Use and impact

The extent to which data is used to deliver socially valued impacts is notoriously difficult to measure. Over the years many estimates have been offered of the economic or social value that data can bring, presenting large figures on the potential value. However, disentangling public good benefits of data from GDP growth or other aggregate metrics is challenging: not least because aggregate economic statistics may describe revenue from both uses *and* abuses of data. Qualitative approaches also have their difficulties, although understanding stories of how data has been used can be valuable to inspire replication and adaptation of ideas to new contexts. However, when case studies are used as a measure of impact, this often captures how well countries write up their open data uses, rather than how widespread data use actually is.

In the first edition of the Barometer, we have explored an approach centred on representative use-cases in order to gain a rough comparative understanding of data use and impact. We set out four specific examples of data use that researchers were invited to look for. These uses, each linked to a particular thematic module in the Barometer, were designed to capture a range of forms of data use: from data enabling civil society scrutiny of government to private sector re-use of data to provide services that bring public benefits, and from data-enabled research and advocacy to promote greater equity to government use of data to deliver more effective services. Along with these primary metrics, we include one secondary indicator on international organization use of country data.

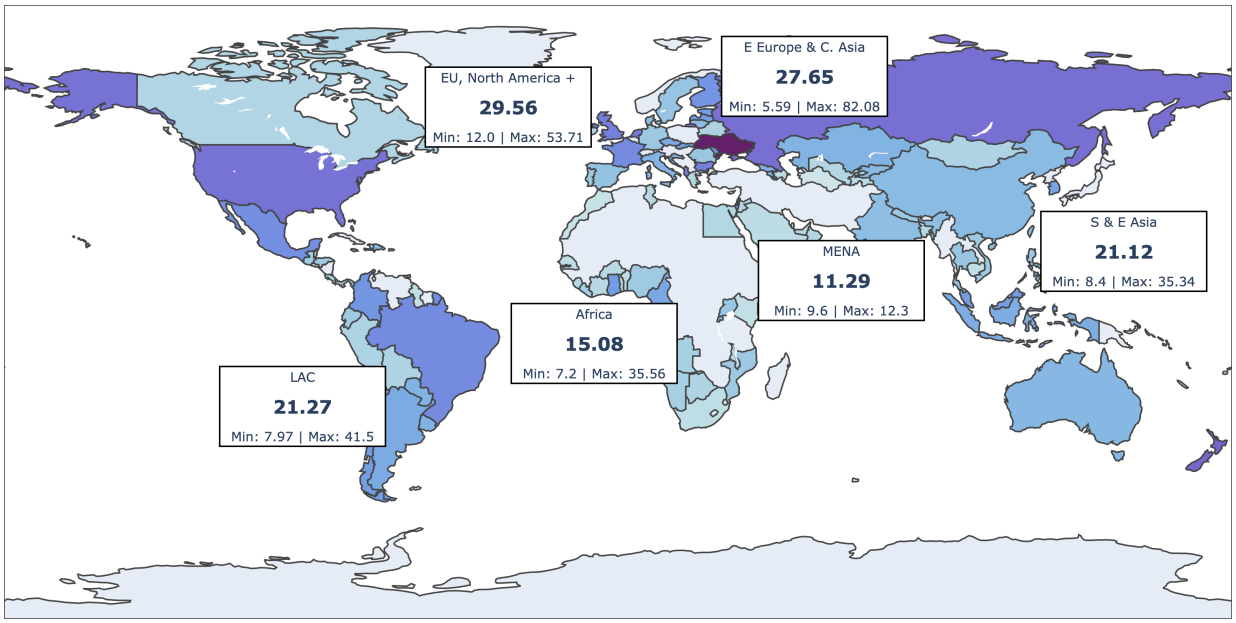
Although limited in scope and covering just a very small number of the possible uses to which data can be put, the data gathered for this pillar offers insights into the drivers of data use and the challenges faced by current and potential data users.

Summary

- Across the examples surveyed in the Barometer, there are relatively low levels of data use, with 20 countries reporting no notable data use against the example use-cases provided, and only 22 countries having uses to report across all four. However, more encouraging, in 56.3% of the 247 instances where data use was identified, at least some evidence of impact (e.g. the data use leading to policy change or social or political value) was identified.
- The majority of data uses identified were seen as isolated cases of use (for example, a single civil society organization using a dataset for a particular advocacy campaign), rather than representing established and widespread uses of data (for example, data being used by multiple Civil Society Organisations (CSOs) and media outlets in a number of instances). While higher and lower capability countries have roughly equal numbers of isolated data use-cases, widespread use appears to be correlated with higher capability. In particular, countries with higher sub-national data capability are more likely to see extensive data use.
- Many data users work around the low availability of structured datasets by bringing together data from multiple sources, converting and cleaning it. In some cases, civil society, academic and occasionally private sector users become intermediaries providing structured data for others to use. However, this is often not sustainable and can create barriers to scaling up data use.
- Media (36.4% of cases checked) and civil society organizations (37.6%) were most likely to be identified as data users in our survey with government (30.3%) and academia (25.7%) following. We found just 62 cases of private sector use, primarily in countries with strong technology sectors.

- We found limited evidence of data uses drawing upon Artificial Intelligence with less than 20 cases cited across the survey. In a number of cases, AI was used to clean, rather than analyse, available data.

Where is data being put to use?



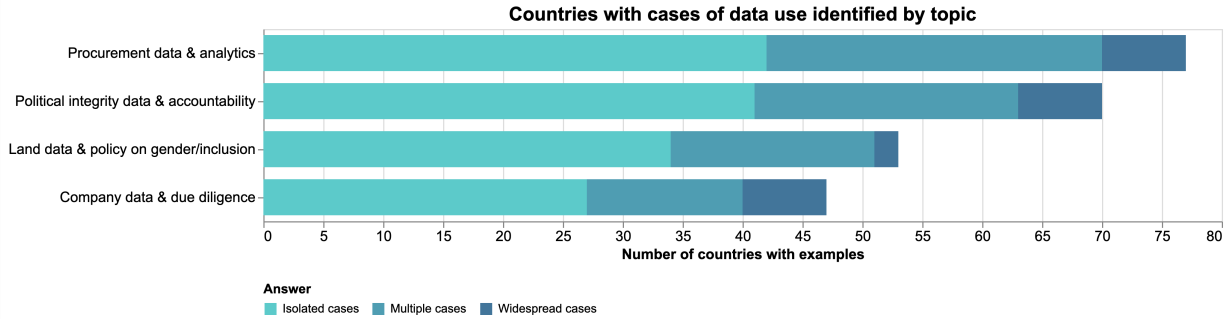
Use and impact pillar regional scores: low scores on the use and impact pillar reflect that relatively few of the specific cases of data use sought by the Barometer survey were identified in each region.

Africa	Eastern Europe and Central Asia	European Union, United Kingdom, North America, Israel, Australia and New Zealand	Latin America and the Caribbean	Middle East and North Africa	South and East Asia
Ghana (35.56), Cameroon (31.49), Rwanda (24.1), Uganda (19.32), Mozambique (17.82)	Ukraine (82.08), Russia (50.81), Albania (45.46), Moldova (36.77), Georgia (28.21)	New Zealand (53.71), USA (49.42), Netherlands (45.75), United Kingdom (45.12), Bulgaria (45.0)	Brazil (41.5), Mexico (40.33), Colombia (37.82), Chile (35.36), Argentina (30.29)	Jordan (12.29), Oman (12.0), UAE (12.0), Qatar (12.0), Egypt (12.0)	Malaysia (35.34), Republic of Korea (32.57), Hong Kong (30.08), Indonesia (29.77), China (24.28)

The use pillar is primarily based on four survey indicators, asking to what extent there is evidence of data being used to:

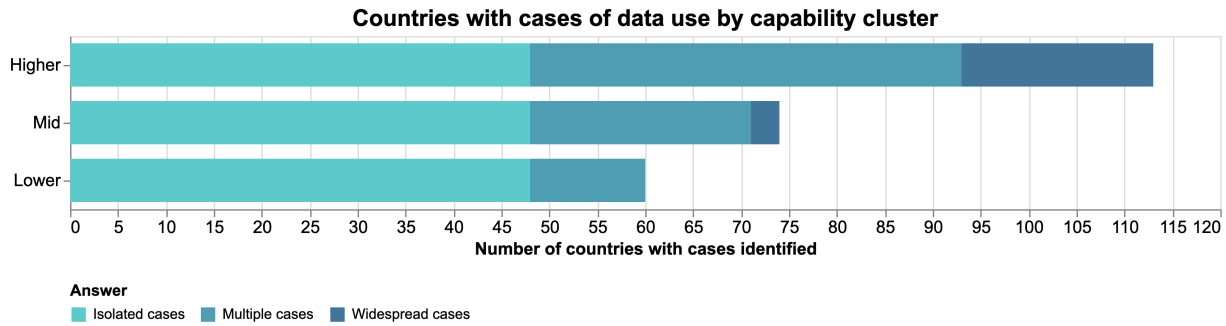
- **Improve procurement practice**: including through analytic dashboards- carry out red flag analysis, improve the diversity of procurement, or assess and improve the environmental impacts of procurement.
- **Identify, expose, or highlight failures of government**: with an emphasis on the use of political integrity data by a wide range of stakeholders, including civil society, journalists, academia and the private sector.
- **Influence policy in the interests of equitable and inclusive land tenure and use**: particularly through use by journalists, academics or civil society organizations.
- **Support corporate due diligence**: including through the provision of services for private sector organizations to check the credential of potential business partners, for governments to carry out checks before entering into contracts, or for media to report on company ownership.

For each indicator, researchers were asked to identify cases of use and to ascertain whether these were isolated examples or representative of widespread use. As the chart below shows, procurement data saw the greatest use with due diligence uses of company data identified in the smallest number of countries.



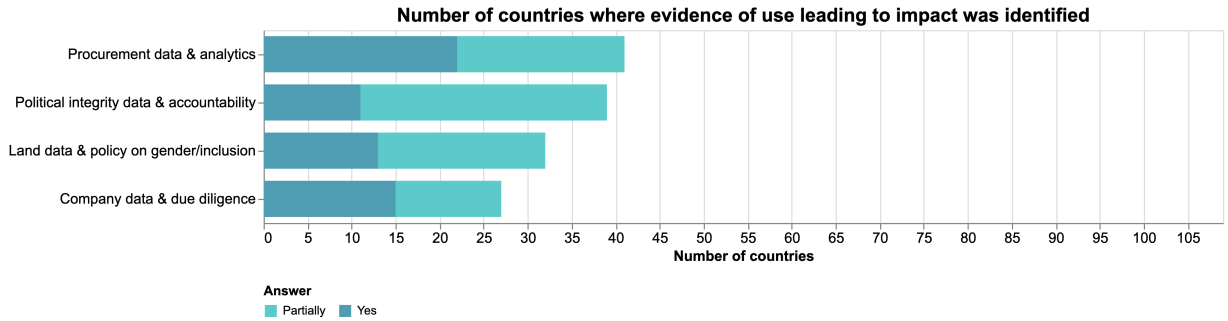
Cases of data use by topic: Procurement analytics use-cases were the most frequently identified, followed by use of data for political accountability and integrity purposes, and then data use related to land governance. Although fewer countries could demonstrate domestic private sector companies using company data to support due diligence activities, where the data and market to support this was available, it was more likely to result in widespread use of the data.

Examining cases of data use by capability clusters (see Capability chapter), we find surprisingly high levels of data use in lower capability countries, although most uses identified were isolated cases. A look at the qualitative justifications for these scores reveals that, while civil society or journalists in countries with limited data availability may be able to run pilot projects or one-off analysis, or while donors may support demonstrator initiatives in these contexts, sustaining and scaling data use requires both more sustainable supply of data from government and greater capability in the population as a whole to work with the data made available. Notably, over 50% of the cases of private sector data use identified came from high-capability countries.



Cases of data use by capability cluster: Higher, mid and lower capability countries all have similar numbers of isolated data use-cases, whereas widespread use of data is correlated with capability.

Researchers were also asked to identify whether there was evidence that the particular uses of data covered by the survey were having meaningful positive impacts and to assess the strength of the evidence. Notably, while there is little difference between indicators when we consider the presence of *any* evidence (some evidence exists in 56.3% of cases), when we look for *strong* evidence of impact, and high levels of confidence that data is driving impact, the scores for procurement analytics uses of data were double those for political integrity use. In short, it appears that while data is a powerful tool in identifying, exposing or revealing failures of government, the barriers to translating that information into impact are substantial, and persistent problems of impunity, and weak accountability mechanisms must not be underestimated when thinking about how data can drive change.



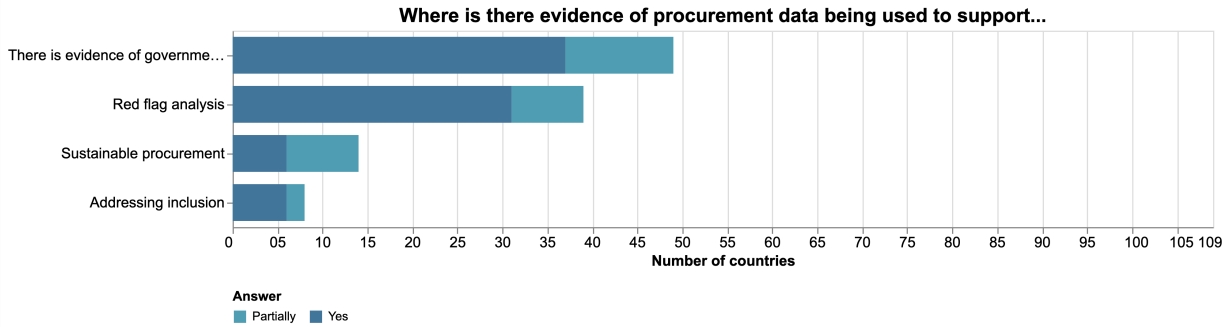
Evidence of impact: Strong evidence of impact was identified in twice as many countries for procurement analytic data use, as for impact of political integrity data use.

How is data use developing?

While this first edition of the Barometer cannot provide longitudinal data on use, the rich qualitative justifications, and collection of links and short summaries provided by researchers, can provide critical insights into how data use practices are developing around the world.

Dashboards & detection

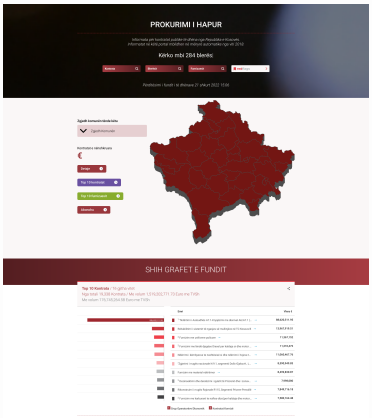
When it comes to procurement data use, we see significant focus on creating dashboards and red-flag analysis tools that can use data to indicate potential corruption risks. The Barometer survey identified over 50 examples of publicly accessible procurement dashboards produced by a range of stakeholders, from official public procurement agencies to civil society organizations, international research programmes and investigative journalists. Below you can find selected screenshots of portals, showing some of the comment elements and unique approaches taken in each. Particularly notably are the number of platforms that were either adapted for, or launched with a particular focus on COVID-19 emergency procurement, demonstrating the ability of open data platforms to respond rapidly to current events.



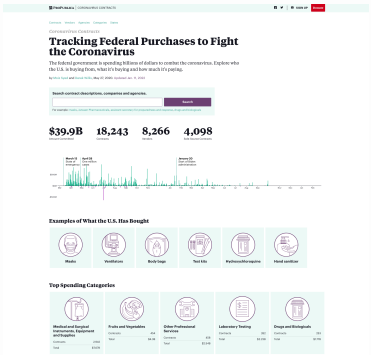
Uses of procurement data: Procurement data is widely used to provide dashboards, and for red-flag analysis, but there is less evidence of data being used to support inclusion or environmental procurement

Less evident were uses of procurement data in support of sustainable procurement or to promote more inclusive pattern of procurement (such as analysis of gender impacts of procurement decisions). Although a number of countries have developed ‘green public procurement’ policies, there was limited evidence of data being used to monitor these, with only the Republic of Korea and Latvia providing some evidence in our survey of reporting statistics on compliance with such policies. This highlights an area for future work around better integrating strong platforms of procurement data transparency with critical challenges of sustainable procurement.

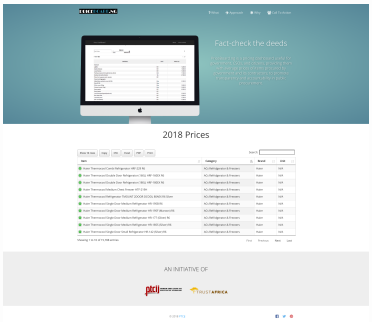
Procurement dashboards compared



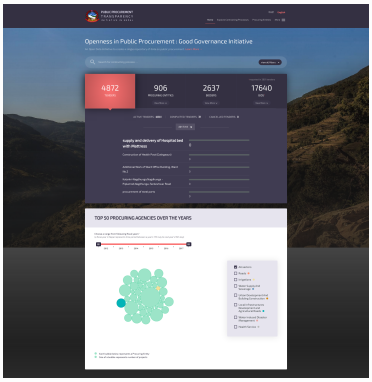
The Kosovo Open and Transparent Procurement Portal - (PPHT) is managed by a civil society organisation, with funding from USAID and cooperation from the government procurement agency. Recent developments to the portal have added the ability to search for contracts by the name of the owner of the company receiving a contract.



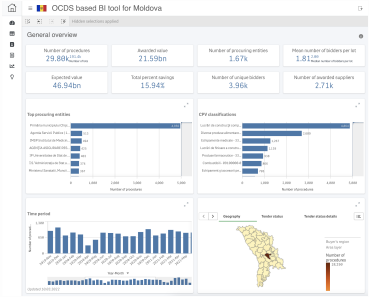
In the USA, investigative journalism newsroom ProPublica used procurement data to build a dashboard focussed on federal COVID-19 related contracts.



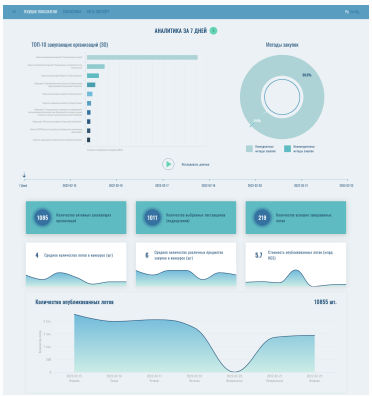
In Nigeria, the Premium Times Centre for Investigative Journalism have taken data from the Nigeria Bureau of Public Procurement to provide a dashboard of commodity prices, responding in particular to the corruption and cost inflation issues.



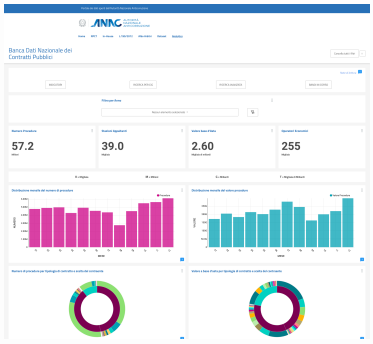
In Nepal, a platform initially prototyped by civil society, has been adopted by the official Public Procurement Monitoring Office, and is kept updated with data from e-procurement systems.



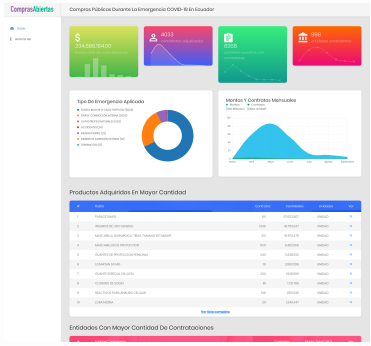
In Moldova, a range of dashboards exist driven by data in the Open Contracting Data Standard, including one based on the Qlik business intelligence dashboard software - replicating a model first developed by the ProZorro Business Intelligence platform in Ukraine



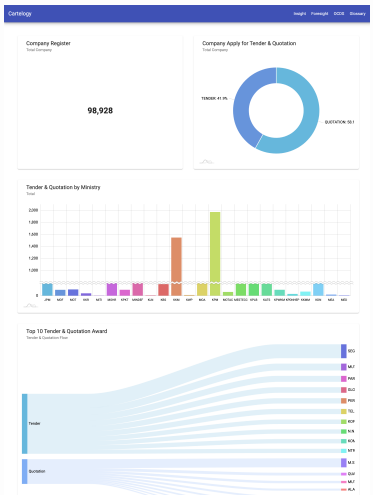
In the Kyrgyz Republic, a dashboard funded by the European Bank for Reconstruction and Development draws on data from the national procurement portal. Researchers report that this dashboard is occasionally used to support media reports, but is more commonly of use to businesses seeking procurement opportunities.



In Italy, the Authorita National AntiCorruzione (ANAC) hosts an analytic dashboard onto the National Database of Public Contracts

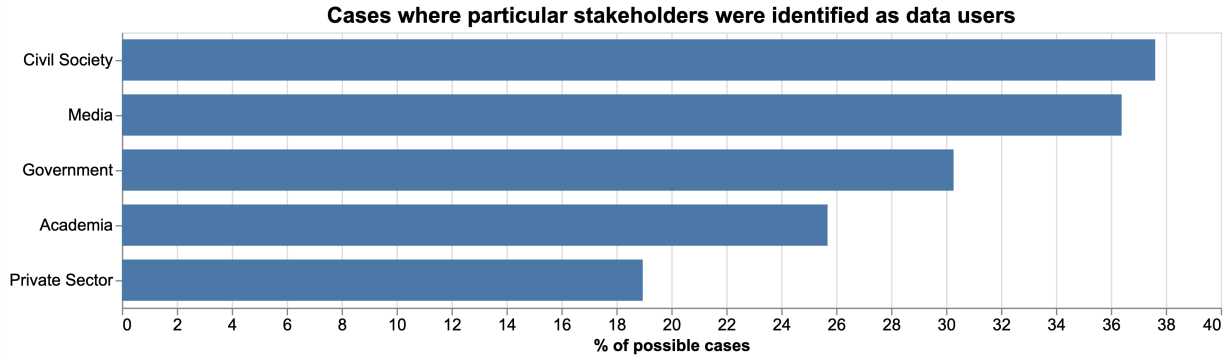


In Ecuador, a dashboard build by non-profit Fundapi, took data in Open Contracting Data Standard format from the National Public Procurement Service to visualize the use of emergency procedures in COVID-19 response



In Malaysia, the 'Cartelogy' platform was developed to connect procurement data, company profiles and political exposed persons data to create a "red flag" tool designed to help government officials to proactively detect potentially suspicious procurement activity.

Promising partnerships



Stakeholders using data: Civil society organizations were identified as data users in 37.6% of the cases where this was checked for, whereas private sector users were identified in just 19% of cases.

In the past, some of the key barriers to civil society and media data use have been related to skill gaps. While there remain significant data literacy gaps in traditional media and civil society organizations, we observe a number of promising trends, including the ongoing professionalization of data journalism, evidence of deeper partnerships between traditional civil society organizations and civic technology organizations, and use by media of commercial provided platforms to support both research and data-visualisation in complex media stories.

In the United States, political integrity data finds extensive use both by specialist data-journalism and investigatory journalism platforms, and within mainstream newsrooms, as well as being used by academics addressing wider trends in political lobbying or party finance. Extensive journalistic uses of political integrity data were also identified in countries including Ghana, Chile, Argentina and Nepal, albeit with journalists turning to right to information laws, rather than open data sources in a number of cases.

The ability of media reporting to lead to substantive impacts on political integrity is shaped by wider political environments. However, where official institutions are strong, the media can play an important role in drawing attention to issues that then spark enforcement activity. Writing prior to the illegal invasion of Ukraine in 2022, our researcher reported that media reports based on the register of asset declarations have been observed to trigger action by official anti-corruption bodies, who might check on the data used and sometimes initiate criminal cases.

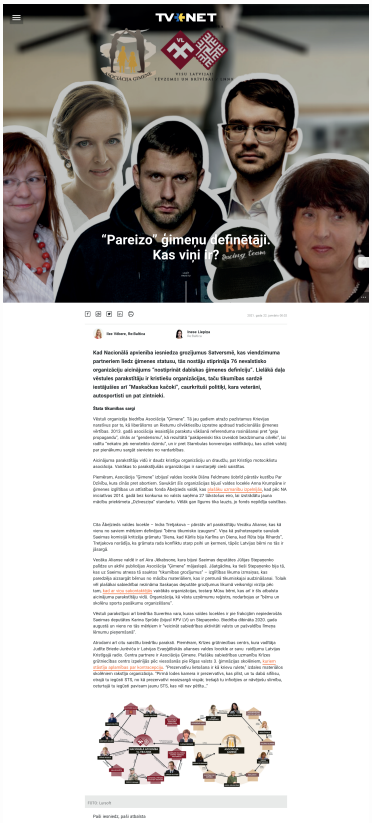
In reports on the use of corporate due diligence data, we observed a number of cases where media outlets were drawing upon commercially provided platforms in order to carry out research into company ownership, using visualisations generated by those platforms in their media reports. A number of platforms appear to have recognized the value of this by offering national civil society organizations and journalists free access to their platforms, while still operating a paid-for business model for commercial users.

A number of the cases of use identified in the Barometer story appear to involve collaborations between traditional, cause-focused civil society organizations and more technically-oriented civic technology organizations. This represents a maturing of the landscape, away from civic technology organizations running their own independent campaigns, and suggests a more sustainable model of practice in which organizations with a long-term commitment to a particular cause bring their domain knowledge together with the skills and knowledge of data specialists. For example, in France, a consortium of NGOs have worked with cooperative business Datactivist to develop a platform that uses machine-learning to analyse whether companies are publishing legally required plans on due diligence activities.

There is also evidence of public data sources being used by actors outside of countries, particularly in cases where domestic political freedoms are restricted. Although many countries report no use of political integrity data to hold governments to account, either because of a lack of data or a lack of

independent civic space, in a number of cases, researchers report that international non-governmental organizations, expatriate citizens or investigative journalists abroad, are making use of data to highlight government failures or issues of concern.

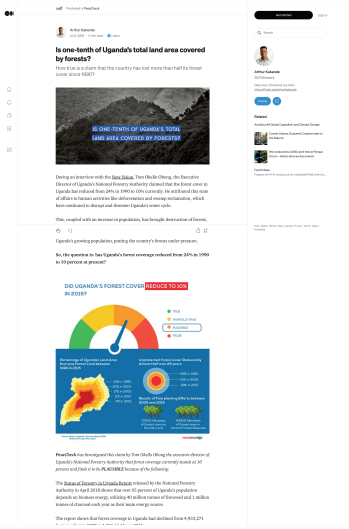
Examples of media data use



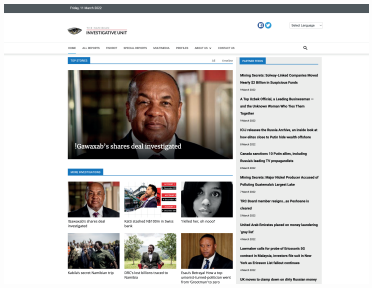
In Latvia, investigative reporters have made use of a commercially provided tool for exploring company networks structures for research and presentation of a news story about the signatories to a letter that opposed equality for same-sex partners. (Screenshot source: tvnet.lv)

In China, reporters have made use of the company data provided through the Tianyancha App to report on legal cases involving companies, and to add details on company ownership to other news stories. (Screenshot source: new.qq.com)

Data from the Moldovan Company register was used to investigate ownership of the ship that transported the hazardous chemicals involved in a catastrophic 2020 blast in Beirut, Lebanon. Screenshot source: balkaninsight.com



In Uganda, a data scientist has used public data to fact-check claims about deforestation. Screenshot source: pesacheck.org



In Namibia, the Investigative Journalism Unit had success in 2021 with two major stories about political integrity, drawing on data to track financial flows across countries including Iceland, Angola and Dubai. Screenshot source: investigations.namibian.com.na

In Chile, investigative journalists regularly draw on company data and political integrity registers to report on the interests of politically exposed persons. Screenshot source: ciperchile.cl

Working around data gaps

Using Barometer data to compare the availability of land data and evidence of land data use reveals a surprisingly weak correlation. Given we have more countries where data use is evident than countries that are providing fully open data, this suggests actors are making use of data *in spite of* low availability, rather than *because of* the available data.

Looking into detailed survey responses reveals many cases where media and civil society actors pieced together documentary evidence from land registers or official records and combined this with other sources to produce data-enabled analysis. For example, in **Paraguay**, the environmental NGO, IDEA, used right to information requests to obtain land ownership data and combined this with satellite images and information on licenses to detect illegal deforestation, and to then file an official complaint. As the country researcher noted, this case demonstrates *"more the achievement of the CSO that led the case, rather than the availability of the data for public use"*, particularly in terms of turning an observation of illegal deforestation into a formal complaint.

Our survey data also suggests that civil society currently play the biggest role in the use of land data to drive policy change (40 countries), followed by academia and media (in 35 countries), and lastly, by the private sector (in 19 countries). In many cases, where private sector use of land data was reported, the focus was more often on developing services for house buyers and sellers, rather than explicitly supporting greater equity and inclusion in land markets. In both France and Taiwan, researchers reported the main users of land data were private firms operating in this way, with Taiwan providing one of just 5 examples identified of AI being used to work with land data.

Questions of land ownership concentration, and of agricultural policy, feature highly among those addressed through data use. For example, in Brazil, the 'Map of Inequality' uses records on more than 5.3 million rural properties from the Atlas of Brazilian Agriculture to input into ongoing, and intensely political, discussions on agrarian reform. This project, using analysis of geographic data, offers the potential to provide alternative perspectives on land inequality from those available using only official statistical data. In another example, covering both agricultural policy and issues of land grabs, Mexican NGO, the Civil Council for Sustainable Forestry (CCMSS), draws upon data from the National Agrarian Registry to develop advocacy-oriented visualisations showing ecosystems, economic activities and land tenure. This feeds into both advisory support for producers through field work and into public policy advocacy informed by a combination of grassroots experience and data analysis.

The challenges of piecing together a picture of land ownership, even in highly digitized economies, is evident in use-cases from both the United Kingdom and the Netherlands. In both countries, where bulk access to land tenure information isn't available to the media or civil society, journalists and activists have pieced together records to present data on the largest landowners, challenging conventional thinking about the distribution of land ownership, including providing evidence on the extent of state-owned land. In one analysis, drawing on data from the Food and Agriculture Organisation (FAO) of the UN, journalists report that just 6% of Dutch land is owned by women.

In Rwanda, one of the few countries in our survey to collect and publish data on the gender of land owners, the Land Administration Information System, and the Rwanda Land Dashboard it drives, has seen use in policy analysis amongst academics, civil society, media and government. Our survey researchers noted that its creation since 2005 has been a significant achievement, but that challenges remain in refining the top-level statistical findings the dashboard presents into actionable information for policy-makers. This issue, that users are often reliant upon statistics released government agencies but lack the ability to access the underlying data to develop more detailed metrics of analysis, is a common theme in a number of countries and arguably limits the space for actors outside government to develop robust inputs into policy debates. In particular, researchers noted that when civil society has to rely on limited and ad-hoc data sources, this can lead to a fragmentation of policy debate and advocacy when, as our survey responses on land tenure data use from Sierra Leone reported *"advocacy often goes in all directions with different objectives simply because organizations each use non-standardized data"*.

How can countries improve data use in future?

For each country, the Global Data Barometer website contains a profile that shows the breakdown of individual scores against use and impact indicators. This can be used to identify possible areas for improvement, although it should be noted that future editions of the Barometer are likely to include a wider range of use-cases and impact assessment than are included this edition and may see some further development of the use and impact measurement methodology.

As a result, our recommendations here focus less on the specific use-cases covered by the Barometer survey and more on particular approaches to support diverse re-use of data oriented towards the public good. This builds on a recognition that demand for data, and data use, is a key driver in improving data quality, and that transforming national data infrastructures requires the active involvement of a wide range of data users.

Governments should make greater, and more transparent, use of their own data

With improved data management and sharing frameworks, by building internal capacity for data use, and by developing partnerships with a range of sectors, governments can increase the public value generated from the data they already collect and hold.

We found it was often challenging to identify how government is making use of data. However, to promote uses of data that build towards the public good, and to protect against those that may harm minority or marginalized populations, or that otherwise have negative impacts, it is important that governments are accountable to citizens for how data is being used. When publishing datasets, commissioning new analysis or tools based on data, or making use of algorithmic systems, governments should identify ways to make the public aware of how data is being used.

Support cross-sector partnerships for data use

Some of the most interesting and sustainable examples of data we identified involved partnership working between multiple sectors, including traditional civil society working with civic technology non-profits and companies or data platforms initially developed by non-governmental organizations being adopted and hosted on an ongoing basis by government agencies. Governments, funders and other stakeholders should all consider how to support, and remove the barriers to, these forms of collaboration.

Deepen emphasis on equity and inclusion

We found relatively limited evidence of data being used to address issues of equity and inclusion. Data use for the public good should address the needs of everyone. This may involve paying particular attention to promoting data use aligned with the sustainable development goals or setting frameworks that ensure data use is taking account of the particular needs of marginalized populations.

Regional Analysis

- África
- Eastern Europa and Central Asia
- The European Union, United Kingdom, North America, Isareal, Australia and New Zealand
- Latin America and the Caribbean
- Middle East and North Africa
- South and East Asia

In this section we explore Barometer findings across 6 global regions, selected to reflect the structure of our research hubs network: Africa; Eastern Europe and Central Asia; the European Union, United Kingdom, North America, Israel, Australia, and New Zealand; Latin America and the Caribbean; Middle East and North Africa; and South and East Asia. Below you will find summary data for each region, along with selected insights from the regional research hub or hubs responsible for coordinating research in that region.

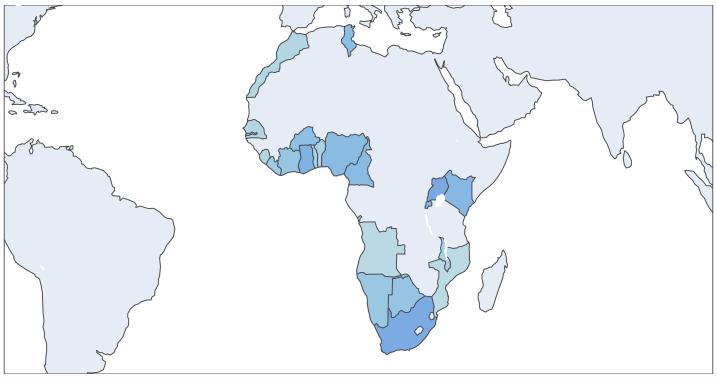
Reading regional tables

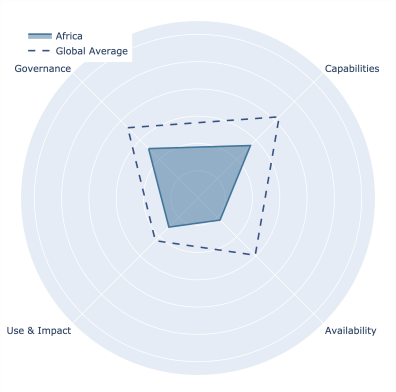
- **Strengths and weaknesses** are selected based on the primary indicator for which a country has the greatest deviation from the regional mean score. In other words, of all available indicators, the strength column shows the indicators where the country is most likely to have a good practice example to share with others in the region. The weakness column indicates areas where a country may face particular challenges, and/or where there is the greatest opportunity for rapid progress by learning from regional peers. These are determined by indicator scores only and do not reflect an editorial decision.
- **Regional (RL) and Global Leaders (GL)**. We use 'RL' against an indicator to show when a country has among the highest scores in the region on a particular indicator, and 'GL' when it has among the highest scores globally. This can be useful to locate countries that may demonstrate best practice.
- **Component scores**. The table includes the weighted score for each country on the Governance, Capability, Availability and Use and Impact components of the Barometer. Each of these components are assigned a score out of 10.
- **Open Data Policy and Initiative**. The tables include specific indicators scores for the presence of Open Data Policy and Open Data Initiatives (see Governance and Capability chapters). These indicators are broadly comparable with indicators from past editions of the Open Data Barometer offering insights into whether national open data policies and practices have improved or been neglected in recent years. Less than a 1 in 10 change in the indicator score is shown as ▣. An increase of more than 1 in 10 points on past ODB scores is shown with ▲, suggesting a likely improvement in policy or initiative quality. A fall of more than 1 in 10 is shown with ▼. ODB measurements may be from 2020 (Latin America), 2017 (ODB Leaders), or 2016 (Other countries). Where there is no past ODB data to compare with, ☐ is shown.

These tables should be read with the limitations of this first Barometer edition in mind. It is possible that in some cases the strengths and weaknesses displayed will represent measurement errors, or outliers present due to some particular aspect of country context that can only be fully understood by reading the supporting qualitative evidence. Similarly, although certain GDB and ODB indicators were designed for comparability, there are differences in the measurement method that may affect particular countries. The tables below are intended as the start, rather than the end of discussion and exploration. Given limitations of space, we only display one strength or weakness for each country.

Africa

The Barometer covers 22 countries in Africa. The region scores below the global average on all pillars, with significant need for investment in data governance institutions, robust and comprehensive data infrastructures, and in fostering broad capabilities to manage and use data for the public good. However, significant opportunities exist for peer-learning to take progress on topics such as public finance data management and sharing, and to transfer learnings to other sectors such as data for climate action. The region also has a number of solid, and improving, open data policy frameworks to build on, although sustaining resources and support for open data initiatives remains a key challenge for the future.





Overall Score	Governance	Capabilities	Availability	Impact and Use
19.61	25.6	27.19	11.43	15.08

	Country	Comparative Strength (Regional)	Comparative Weakness (Regional)	Overall Score	Governance	Capability	Availability	Use & Impact	Open Data Policy	Open Data Initiative
	Angola	(G) Public finance data	(A) Public procurement data	10.6	16.7	15.9	2.7	15.2	0.0 □	0.0 □
	Benin	(G) Data protection	(G) Data management	14.4	14.2	23.1	10.1	8.8	0.0 ▣	40.0 ▲
	Burkina Faso	(G) Public consultation data RL GL	(A) Public procurement data	22.5	36.2	32.6	8.2	10.8	54.0 ▲	72.0 ▣
	Botswana	(G) Data management	(G) Asset declarations	20.2	29.5	28.0	9.6	13.0	0.0 ▣	32.0 ▲
	Côte d'Ivoire	(C) Government support for re-use RL GL	(G) Asset declarations	19.8	25.2	41.3	4.6	10.8	0.0 ▣	51.0 ▲
	Cameroon	(A) Existing land use RL	(G) Data management	24.1	23.9	25.6	22.8	31.5	0.0 ▣	36.0 ▲
	Ghana	(G) Beneficial ownership RL GL	(A) Public procurement data	27.7	32.3	43.2	14.7	35.6	0.0 ▣	70.0 ▲
	Gambia	(G) Public consultation data RL GL	(A) Public procurement data	20.5	29.0	20.7	14.6	17.2	0.0 □	0.0 □
	Kenya	(G) Public consultation data RL GL	(C) Open data initiative	25.7	44.0	21.9	16.5	8.4	0.0 ▼	0.0 ▼
	Liberia	(G) RTI performance RL	(G) Data protection	17.2	24.6	14.7	13.6	14.9	0.0 □	20.0 □
	Morocco	(A) Company register	(G) Public finance data	12.4	10.8	23.0	7.9	7.2	0.0 ▣	0.0 ▼
	Mozambique	(G) Data sharing frameworks	(G) Data protection	10.3	14.0	13.5	5.0	17.8	0.0 ▣	0.0 ▣
	Malawi	(G) Data sharing frameworks	(G) Data management	14.6	27.3	17.3	4.3	11.6	28.0 ▲	16.8 ▣
	Namibia	(A) Vaccination (COVID-19) RL	(G) Public finance data	18.9	21.1	23.2	15.2	15.4	0.0 ▣	0.0 ▼
	Nigeria	(C) Sub-national RL	(G) Public finance data	24.3	25.5	40.0	15.1	17.5	6.3 ▼	80.0 ▲
	Rwanda	(G) Open data policy	(G) Data protection	24.8	32.6	39.1	11.1	24.1	63.0 ▲	72.0 ▲
	Senegal	(A) Public procurement data	(A) Budget and spend data	12.1	14.8	18.9	6.4	9.6	0.0 ▣	0.0 ▣
	Sierra Leone	(G) Political finance	(G) Data protection	12.1	13.1	22.8	5.3	11.6	0.0 ▼	28.0 ▼
	Togo	(G) Asset declarations	(G) Public finance data	14.6	15.5	20.8	10.6	11.6	0.0 ▣	14.0 ▼
	Tunisia	(G) Open data policy RL	(G) Public finance data	23.1	35.0	37.5	7.5	10.8	80.0 ▲	40.0 ▣
	Uganda	(G) Data management	(G) Beneficial ownership	31.4	39.6	37.8	23.0	19.3	0.0 ▣	50.0 ▲
	South Africa	(A) Budget and spend data RL GL	(C) Civil service	30.4	38.5	37.3	22.6	9.2	0.0 ▣	28.0 ▼

G = Governance pillar; C = Capabilities pillar; A = Availability pillar; U = Use and Impact pillar; RL = Regional Leader (among the highest scoring regionally) on this indicator; GL = Global Leader (among the highest scoring globally) on this indicator. ▲ = GDB score higher than ODB score on related indicator; ▼ = GDB score lower than ODB score on related indicator; ▣ = less than 1pt change in score; □ = No data for comparison.

Hub perspectives: Local Development Research Institute (Sub-Saharan Africa)

By Keziah Munyao

Regional profile

Sub-Saharan Africa is a diversified continent with a population of over one billion. It has human and natural resources with potential to generate equitable growth and alleviate poverty. While the continent continues to achieve significant advances, sadly the region still has the highest death rate for under-5s in the world; nearly a quarter of the population is malnourished; 42% of people lack a basic water supply; 72% lack basic sanitation; and 58% of 15 to 17 -year- olds do not attend school[29]. The area fails to achieve gender equality and ranks last among regions on the 2019 SDGs Gender Index[30],owing to high maternal mortality rates and uneven access to water and power. Sub-Saharan Africa also faces the fastest rates of population growth and a rapidly rising urban population: expected to reach 1.3 billion by 2050. In spite of significant advances in global health, particularly the fight against HIV/AIDS, many broader trends have potential to strain already overburdened food, health, education, and public infrastructure systems, as well as to overwhelm job markets that are unable to meet employment demand from growing youth populations.

The continent is also a region of dramatic political and socioeconomic disparities. In a few nations, autocratic governments, civil unrest, corruption, poor institutions, and sluggish political systems continue to stymie reform initiatives. Empowered women and youth continue to pursue larger involvement in politics within and via the African Union. International organizations have been able to assist by expanding regional cooperation, supporting democratic and security sector governance across Sub-Saharan countries, promoting constructive cross-party discourse and strengthening political party capability. Even though Sub-Saharan African countries began to open up to democracy three decades ago, the area is still characterized by a high variety of political systems. Fragile democracies frequently face multiple problems and flaws, and they share borders with some of the world's least democratic countries. Multi-party elections, almost non-existent in 1990, are now the norm, yet they still seldom result in power transfers. Recent tendencies of democratic recession have not gone unnoticed in the region but have affected distinct nations differently. Some unanticipated democratic shifts have occurred in parallel to broader democratic decline. Furthermore, while the coronavirus pandemic struck the region less severely than other regions of the world, its influence on democratic and human rights was considerable.

Findings from the field

The first edition of the Barometer was a learning experience for the Sub-Saharan Africa regional hub with some of the significant themes observed during the field study including that:

- There is little publicly available evidence of governments investing in the data literacy skills in the civil service outside of statistical units or national statistics offices;
- There is significant fragmentation of content across agencies, ministry and departments websites, making it difficult for the public and other stakeholders to find relevant data;
- Government platforms in some

countries lack easy ways of disseminating information to a larger audience.

In addition to these issues around data creation and dissemination, we found that low data literacy for interpretation of public data was a common issue in the region. For example, datasets on population civil registration and vital statistics systems (CRVS) were difficult for individuals to comprehend. This points to challenges on both the 'supply side' and 'demand side' of data for the public good. It also raises questions about how far data is being managed in ways that support analysis and re-use, both inside and outside government, rather than just allowing statistical reports to be published. Much of the data we found through the field study was not tailored to ongoing analysis, nor did it appear to be linked to processes inside government that would keep it updated.

The field work also identified gaps with respect to data protection or privacy standards in a number of countries, even where efforts are underway to promote wider data usage and openness. The absence of strong legal frameworks alongside new technological advancements seems to be a developing concern, particularly in countries where no frameworks exist to oversee the use of emerging technologies such as artificial intelligence (AI).

Finally, there appears to be a struggle with local governments when it comes to making data available and open to the public as the survey shows there has been minimal movement on the open data and open government front at regional or city levels beyond a few exceptional examples.

Future directions

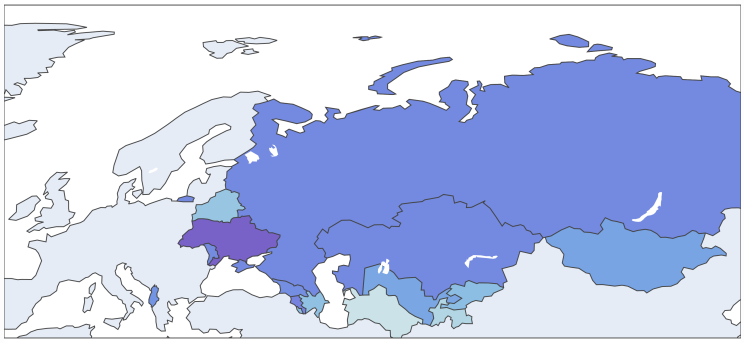
The Barometer study has been a critical activity for the region with the potential to help countries become more confident, to take decisions on data strategies, and be more proactive in developing data-driven policies and initiatives that can have a greater impact on people in the future. The Barometer can help countries to better understand the dynamics and complexities of generating, publishing, and analyzing data, and to bring together expert opinions and government officials' perspectives to understand practical steps forward. By focussing on constructive peer-learning in the region, and across the globe, we hope the comprehensive and balanced tools provided by the Barometer can help defuse the political considerations that so often delay or prevent data openness in Africa.

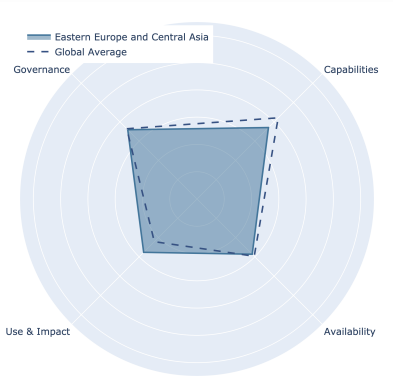
When we consider next steps, and priorities, African countries must secure and provide greater resources to deliver ongoing assistance, capacity building, and strengthening to national statistical offices, which are a driving force behind national open data projects. Supporting re-use calls for data that is accessible, comprehensible, and useable. To deliver this we must overcome various obstacles, such as technological, political, financial, and socio-cultural barriers. Recognizing this, greater focus should be placed on capacity development, peer learning, and training among the community of academics, journalists, and other practitioners who can drive demand for, and use of, data.

Finally, as the Barometer hopefully expands to cover further thematic areas in future, and given the current status of the global food system, which is under strain from climate uncertainties and a growing population, governments in Sub-Saharan Africa should consider placing a particular focus on sharing data about the agriculture sector for public consumption, supporting both government-led data sharing efforts, and working to make sure agricultural data from all stakeholders is available as a resource for the public good.

Eastern Europe and Central Asia

The Barometer covers 15 countries in Eastern Europe and Central Asia. The region scores marginally below the global average on the capabilities pillar with significant variability in how far countries have the skills, institutions and freedoms to use data for the public good. The region scores marginally above the global average on use and impact, reflecting civil society-led uses of data in particular. Strengthening data governance, and adopting more multi-stakeholder approaches to promoting management, availability and use of data for the public good, are key areas for development in the region in the years ahead.





Overall Score	Governance	Capabilities	Availability	Impact and Use
32.79	35.9	37.05	28.62	27.65

	Country	Comparative Strength (Regional)	Comparative Weakness (Regional)	Overall Score	Governance	Capability	Availability	Use & Impact	Open Data Policy	Open Data Initiative
	Albania	(A) Existing land use RL GL	(A) Emission	38.3	41.4	47.7	30.1	45.5	40.0 ▣	54.0 ▲
	Armenia	(G) Public finance data RL GL	(C) Open data initiative	44.6	54.7	27.7	48.5	28.0	0.0 □	0.0 □
	Azerbaijan	(A) Emission	(G) Asset declarations	21.8	26.7	29.5	15.1	9.6	10.0 □	0.0 □
	Belarus	(G) Public consultation data RL GL	(G) Data protection	19.5	12.7	28.4	19.7	13.7	0.0 ▣	0.0 ▼
	Georgia	(A) Emission RL	(G) Public consultation data	40.3	34.5	40.1	45.5	28.2	0.0 ▣	36.0 ▣
	Kazakhstan	(A) RTI performance data RL	(G) Asset declarations	41.7	42.8	55.5	34.3	27.6	50.0 ▲	80.0 ▲
	Kyrgyz Republic	(G) Beneficial ownership	(G) Public consultation data	23.5	24.5	28.3	19.9	25.0	0.0 ▣	35.0 ▲
	Moldova	(G) Asset declarations	(C) Civil service	41.4	45.4	40.4	39.6	36.8	36.0 ▼	36.0 ▼
	Mongolia	(A) Vulnerability RL	(A) Company register	32.8	34.8	32.7	33.4	13.5	54.0 □	60.0 □
	Russia	(C) Civil service RL	(A) Vital statistics	41.7	41.5	52.3	34.8	50.8	54.0 ▼	36.0 ▼
	Tajikistan	(G) Open data policy	(G) Public consultation data	12.2	15.5	18.5	6.5	10.8	37.8 ▲	7.0 ▣
	Turkmenistan	(C) Civil service	(G) Public consultation data	6.4	8.2	15.5	0.0	5.6	0.0 □	0.0 □
	Ukraine	(U) Corporate due diligence RL GL	(A) Public consultation data	55.5	61.1	56.5	48.3	82.1	80.0 ▣	80.0 ▣
	Uzbekistan	(C) Open data initiative RL	(G) Asset declarations	31.7	43.7	41.6	19.4	12.1	68.8 □	80.0 □
	Kosovo	(G) RTI performance RL	(A) Existing land use	40.5	50.8	41.0	34.2	25.5	54.0 ▲	40.0 ▼

G = Governance pillar; C = Capabilities pillar; A = Availability pillar; U = Use and Impact pillar; RL = Regional Leader (among the highest scoring regionally) on this indicator; GL = Global Leader (among the highest scoring globally) on this indicator. ▲ = GDB score higher than ODB score on related indicator; ▼ = GDB score lower than ODB score on related indicator; ▣ = less than 1pt change in score; □ = No data for comparison.

Hub Perspectives: IDFI (Central Asia and Eastern Partnership)

By Teona Turashvili and David Eristavi

Regional profile

Central Asian and Eastern Partnership countries have one major common denominator in the form of the Soviet past, the consequences of which are still fairly apparent in terms of governance and administration. These countries share many common characteristics of post-soviet governments to varying degrees, such as complex bureaucracies, authoritarian-leaning leadership, under-developed digital infrastructures, low literacy and challenges to freedom of the press. However, despite the similarities, there are vastly different data ecosystems, with fairly developed examples such as Ukraine at one end of the spectrum and more suppressed structures such as Turkmenistan and Belarus at the other.

When it comes to marginalized groups, most researchers defaulted to identifying ethnic minorities residing in their respective countries, such as the Uzbek population in Tajikistan, ethnic Armenians and Azeris in Georgia, various groups in Azerbaijan, etc. For conflict-affected areas such as Georgia and Ukraine, a common theme was the populations isolated in rural areas and internally displaced people. More generally, the following groups were mentioned as potentially marginalized groups in the case of several countries: women; the LGBTQ+ community; Roma people; people with disabilities; the elderly; and rural populations. An interesting trend that was uncovered had to do with official recognition of marginalized/vulnerable groups by the state. While some countries refer to such groups in their legislation/policy documents and are actively engaged in tackling the challenges faced by them, others do not seem to have taken any formal steps in this direction.

Findings from the field

The Barometer's use and impact category appeared to be the most challenging for the region, with the exception of strong performance in Ukraine. Researchers found only sporadic examples of data use-cases, mostly led by media and civil society organizations. This points to a lack of engagement with data from academia and the private sector actors in many countries. Even where media were using data, this often relied on civil society as intermediaries. With an absence of open datasets and analytical tools, carrying out in-depth data processing can be too time-consuming for fast-paced media outlets, and for this reason, they mostly cover the studies and reports published by CSOs, particularly those based on political finance data and public officials' asset declarations.

Furthermore, in several countries of the region, there are virtually no tools for public oversight of officials and governance processes. Their political systems remain closed for any meaningful socio-political engagement of citizens and civil society actors. There is a lack of genuinely independent institutions that can identify, expose and highlight the government's failures in all authoritarian-leaning countries of the region. The national researchers from these countries have reported that the very few activists that speak out against the government in their respective countries, regularly face intimidation, harassment and imprisonment.

Out of all data categories, procurement data seems to be the most available in the region. Most countries surveyed had innovative procurement portals in place and were actively using the Open Contracting Data Standard (OCDS) to varying degrees of adherence. In contrast, lobbying data seems to be the most challenging in the region, with most of the countries either not having a functional framework to collect such data or not

making it publicly available. Another challenge that stands out, especially in the post-pandemic recovery period, is the lack of disaggregated vaccination data. Although overall statistics are available in most countries of the region, there is a significant lack of machine-readable, open datasets in this regard.

Although this may be true for many other regions, political integrity data in Eastern Partnership and Central Asian countries was found to lack interoperability. Some countries are progressing in terms of improving access to open data legislation and practices, while some countries are stalled and no fundamental reforms have been initiated over the past years and others are even backsliding.

Future directions

First and foremost, countries in the region need solid regulatory frameworks on access to open data, since a majority of the countries lack common open data standards. To this end, there is a need for a strong political will, which will ensure that decision-makers are aware of the importance, benefits, and key enablers for solid open data ecosystems in the country.

Another important component is the capacity building of all relevant stakeholders in open data management: which includes data collection, processing, publication, and use. In particular, civil servant qualifications with respect to data management and open data also need to be prioritized, since they are responsible for generating or collecting most public data. In parallel, capacity building with other stakeholders, such as civil society, media and business, is needed in order to increase the impact generated through the use of open data.

Finally, it is of crucial importance for different stakeholders to establish partnerships with each other on the issues of open data collection, publication and use. Such multi-sector and multi-stakeholder cooperation will significantly increase the impact of data initiatives for the public good. Different stakeholders with different backgrounds and experience can better combine their efforts and potentially design innovative services and new products positively affecting citizens' well-being. Benefits to civil society and media are already apparent at this stage, however the economic potential of data does not seem to be fully generated or encouraged at present. Shifting focus towards the economic aspects of open data may incentivize and play a crucial role in enabling and advancing all open data practices around the region. Moreover, a focus on economic aspects of open data may help convince decision-makers to put open-data related reforms on the political agenda and may engage the private sector in dialogue with civil society and the public sector regarding the need for improved data management systems and increased access to public data for public good.

Hub Perspective: Open Data Kosovo (Kosovo and Albania)

By Dafina Olluri

Regional profile

Kosovo and Albania are both new democracies with a difficult historical past. Albania has suffered a long transitional state that began in the early nineties and followed half century of a communist dictatorship, during which time the government had total control over the economy and revoked civil liberties. The Republic of Kosovo, which has been in existence as a country for roughly thirteen years, also has a communist dictatorship past, as well as an economy recovering from the destructions of war. Both countries are currently democratic societies and have in force modern constitutions. Through these both countries

express: the determination to build free, democratic, and peace-loving countries for all citizens; commitment to the development of a state that ensures civil freedom, equality, and the rights of every citizen before the law; and commitment to promoting economic wellbeing and social prosperity.

Findings from the field

The process of opening data in Albania and Kosovo has intensified during the last four to five years, mainly due to drivers coming from the civil society, where non-governmental organizations such as Open Data Kosovo have played a crucial role. Open government portals are run by the respective agencies of the government for information society but are not populated sufficiently with data from other public entities, ultimately containing only a limited selection of datasets. However, open data is published by a variety of public entities and state government agencies in different formats, and these act as a valuable source of information and decision-making for citizens and officials. The statistical data on public procurement, public finance, political integrity (data published by the Anti-Corruption Agency) is quite complete and open, leading towards a more transparent country. Despite this, there is more that needs to be done in order to engage the private sector, as examples of open or shared data from the private sector as still very rare.

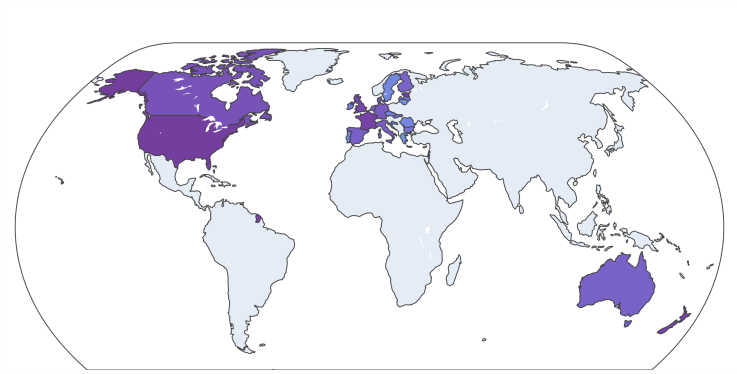
In addition to open data, we note that the legal framework regarding data protection and access to public information is quite advanced and is in accordance with European Union requirements. Kosovo is one of the few countries that openly publishes the properties, assets, incomes, and debts of all notable public officials. This data, available and easily accessible online for free, contains more than ten years worth of declarations, providing a control base for tracking the wealth of Kosovo's politicians. On the other hand, data regarding land tenure is not available as open data and thus there is no evidence that the data is being used for influencing policy in the interest of equitable and inclusive land tenure and use. In addition, the country does not have a health information system available to centrally host citizens' data.

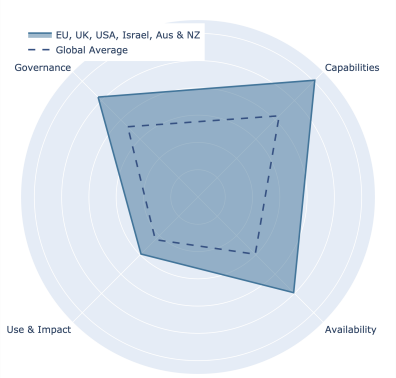
Future directions

Both Kosovo and Albania, despite their efforts to open all public data, are still in the initial stages of this process and a stronger legal requirement will be vital to help positively boost the open data process and support the use of data for the public good. In particular, attention needs to be given to how compliance with data governance, open data and data sharing rules will be ensured. Looking at the use of data for the public good more widely, specific efforts should be made to raise awareness about the data produced in the private sector and the importance of using and opening it to both public and private sector stakeholders.

European Union, United Kingdom, North America, Israel, Australia and New Zealand

The Barometer covers 20 European Union countries, which have been analysed alongside the United Kingdom, United States, Canada, Australia, New Zealand and Israel in a group defined by generally high levels of national income and data-related capabilities. Comparable governance scores hide some significant differences between countries with some countries strong on sectoral rules for data disclosure, while others are stronger on the universal applicability of data protection frameworks. Continued work on data interoperability, improving sub-national capabilities, addressing issues of equity and inclusion in data collection and use, and strengthening open data and data sharing related to climate action should be particular areas of priority for countries in this group.





Overall Score	Governance	Capabilities	Availability	Impact and Use
52.08	51.8	60.55	49.59	29.56

	Country	Comparative Strength (Regional)	Comparative Weakness (Regional)	Overall Score	Governance	Capability	Availability	Use & Impact	Open Data Policy	Open Data Initiative
	Australia	(A) Biodiversity	(G) Public consultation data	55.5	52.7	67.7	53.3	25.7	81.0 ▣	80.0 ▣
	Bulgaria	(U) Influencing policy for gender and inclusion	(A) Company register	49.7	59.0	54.4	40.7	45.0	70.0 ▲	70.0 ▣
	Canada	(A) Lobbying data RL GL	(G) Public consultation data	60.8	59.5	68.0	62.2	12.8	90.0 ▣	80.0 ▼
	Czechia	(G) RTI performance	(G) Public consultation data	45.0	45.1	59.7	37.9	31.4	80.0 ▲	72.0 ▲
	Germany	(C) Sub-national	(A) Company register	58.1	61.1	68.8	53.6	18.2	50.0 ▼	81.0 ▣
	Denmark	(A) Beneficial ownership RL GL	(G) Open data policy	58.2	45.2	67.3	65.5	23.6	0.0 ▼	51.0 ▣
	Spain	(C) Sub-national RL GL	(G) Public finance data	55.8	58.5	74.4	46.6	20.6	70.0 ▣	80.0 ▣
	Estonia	(C) Political integrity interoperability RL GL	(A) Real-time healthcare system capacity	67.4	61.9	91.2	60.7	35.1	50.0 ▲	90.0 ▲
	Finland	(A) Biodiversity RL GL	(A) Real-time healthcare system capacity	54.5	48.3	68.9	52.3	37.9	30.0 ▼	80.0 ▣
	France	(C) Government support for re-use RL	(C) Civil service	66.2	69.7	70.5	63.6	41.3	80.0 ▣	100.0 ▣
	United Kingdom	(A) Beneficial ownership	(A) Real-time healthcare system capacity	64.5	64.1	68.9	64.2	45.1	50.0 ▼	63.0 ▼
	Greece	(G) Beneficial ownership RL	(G) Political finance	36.6	37.3	47.4	32.3	12.0	60.0 ▣	63.0 ▣
	Croatia	(A) RTI performance data	(G) Open data policy	47.9	55.4	42.8	47.3	28.8	0.0 ▼	45.0 ▣
	Ireland	(G) Data management	(G) Public finance data	46.0	47.2	66.5	36.7	12.8	72.0 ▲	90.0 ▣
	Israel	(A) Company register	(A) Existing land use	42.1	38.0	49.6	41.3	35.5	53.6 ▣	76.5 ▲
	Italy	(A) Beneficial ownership	(C) Civil service	56.5	62.3	54.5	56.4	26.9	50.0 ▼	60.0 ▣
	Lithuania	(A) Real-time healthcare system capacity	(G) Public consultation data	37.3	36.8	49.7	31.9	23.0	36.0 □	56.0 □
	Latvia	(A) Beneficial ownership	(A) Existing land use	49.2	43.5	50.4	53.9	35.1	81.0 ▲	50.0 ▲
	Malta	(G) Public finance data	(A) Vital statistics	36.5	45.8	44.8	27.0	18.0	0.0 □	16.0 □
	Netherlands	(U) Corporate due diligence RL	(G) Asset declarations	54.0	49.3	70.4	48.8	45.8	42.5 ▼	90.0 ▲
	New Zealand	(A) Land tenure RL GL	(G) Beneficial ownership	65.6	63.7	62.4	69.8	53.7	81.0 ▣	70.0 ▼
	Portugal	(C) Civil service	(G) Public consultation data	41.9	43.8	50.2	38.2	17.5	70.0 ▲	50.0 ▣
	Romania	(G) Public consultation data RL GL	(A) Existing land use	43.0	55.1	44.8	35.6	19.1	30.0 □	51.0 □
	Slovakia	(A) Real-time healthcare system capacity	(A) Biodiversity	50.9	54.6	57.9	45.8	34.0	63.0 ▣	80.0 ▣
	Sweden	(A) Biodiversity	(G) Public consultation data	42.8	32.0	58.6	43.6	20.2	0.0 ▼	80.0 ▣
	USA	(A) Lobbying data	(G) Data protection	68.0	56.7	64.4	80.0	49.4	42.0 ▼	54.0 ▼

G = Governance pillar; C = Capabilities pillar; A = Availability pillar; U = Use and Impact pillar; RL = Regional Leader (among the highest scoring regionally) on this indicator; GL = Global Leader (among the highest scoring globally) on this indicator. ▲ = GDB score higher than ODB score on related indicator; ▼ = GDB score lower than ODB score on related indicator; ▣ = less than 1pt change in score; □ = No data for comparison.

Hub Perspectives: Access Info Europe (Europe and Beyond)

By Helen Darbishire and Marta Morcuende

Regional profile

Access Info Europe acted as regional hub for a group of 27 countries, most of them in the European region. From these 27 countries, 20 are member states of the European Union (EU) and share common policies in many areas, such as data protection. The United Kingdom, also located in the European region and included in the study, formally left the European Union in 2020, but it also shares many norms and policies with the EU member states that were approved before its departure. Outside of the European region, Access Info coordinated the research in several countries from different parts of the world, such as North America (Canada and the United States), Oceania (Australia and New Zealand), Western Asia (Israel) and Eastern Asia (Republic of Korea). Korea is included in the 'South and East Asia' tables and statistics, but all these other countries are treated in the Barometer as part of the 'EU, North America +' cluster.

These countries are, in general, some of the most developed countries in the world in terms of both economic and democratic development. They are countries which led the open data movement, and the European Union countries led the way with common standards on opening up public sector information as part of ensuring a level playing field for reuse of data, with a large and increasing number of companies making use of public data, as well as use by civil society and investigative journalists.

These are countries with strong reputations for good governance and successfully controlling corruption, although there are significant variations around the region, as can be perceived, for example, in the Corruption Perceptions Index, having better scores in the north and west of the European continent than in southern and eastern European countries. This has not stopped there being multiple corruption scandals linked to money laundering, offshore financing, stolen assets from other regions of the world and irregular public procurement during the Covid pandemic.

These countries are all members of the Open Government Partnership, which means every two or four years, they have to submit an action plan co-created with civil society that outlines concrete commitments to enhance transparency, accountability and public participation in government.

Findings from the field

Countries in this regional cluster perform, as a whole, higher than the global average. This includes on the four key pillars of the Barometer, as well as on all thematic modules. The region scores highest with respect to the capabilities pillar, with the lowest scores in the use and impact pillar. Political integrity is the lowest-scoring thematic module, reflecting that comparably fewer countries meet the high-bar of data governance and publication across the breadth of topics this module covers.

This region scores particularly well on the existence of data protection frameworks, thanks in part to mandatory EU data protection rules since 1995 and the strengthened General Data Protection Regulation of 2016, which came into force in all EU countries in May 2018. Legal frameworks requiring collection and publication of political finance and asset declarations data exist in most countries in the region, although it is important to note that most asset declarations data is not available in machine-readable formats.

Other data sets which were generally available in this region were those on Covid-19 vaccinations, asset declarations, public procurement, emissions, political finance, budget and spending data, and vital statistics. The survey also found that there are few regulations requiring collection and publication of right to information performance and lobbying data, and the lack of regulation directly translates into a lack of availability of this information in practice.

A number of further observations can be drawn from data for this group of countries:

- Having regulations requiring the collection and/or publication is important to guarantee that there is more data available. In most cases, the absence of a legal framework requiring data collection/publication translates into a lack of data availability in practice.
- Beneficial Ownership data is still not available in many countries, and where it is available, it is usually not available for free, under open licenses or in machine-readable formats. Company Information is available in more countries, but it is not fully free of charge in many of them.
- When company information and public procurement data are available, there is strong evidence that they are well-used by a range of stakeholders.
- Despite climate change being one of the most pressing issues, and the region having high capability to produce and use data, there is still suprisingly little data available on climate vulnerability, such as information on future natural hazards, extreme weather events, or climate variability.

Future directions

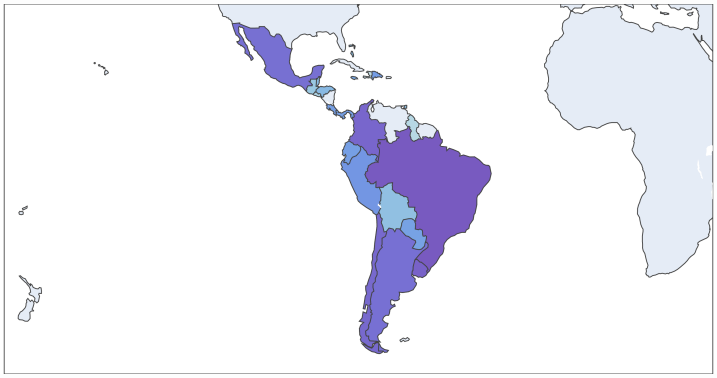
The countries in this group score better than the global average, but there is still significant room for improvement.

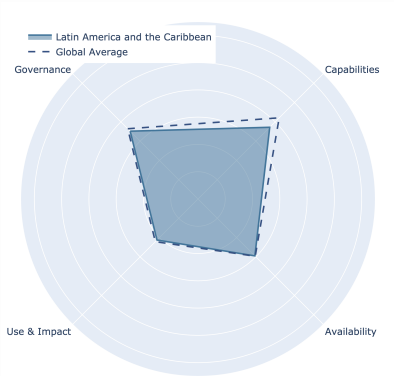
A gap between Northern/Western and Southern/Eastern European countries can still be perceived. Open data policies applying across all EU countries, such as the Open Data Directive, if properly implemented, would level the playing field and improve the data landscape across the region. A number of specific legal frameworks are also still largely missing, including lobby regulations: adopting rules requiring collection of data on lobby activities is a priority to enable monitoring of the influence of lobbying on decision-making. Where frameworks are already broadly in place, countries need to focus on improving availability, openness and usability of key datasets, including company and beneficial ownership registers in particular. This data has been established to have high value for users, and evidence on the use of this data has been found in almost 20 countries with its use key to effectively fight corruption and money laundering.

Governments also need to take more action to promote the reuse of public data to enhance the benefits that data brings to societies and economies. To this end, data should always be made available for free, under open licenses which allow for any type of reuse, in machine-readable formats and having machine-readable datasets available as a whole, released in a timely manner, updated, and with historical data available to allow users to track change over time, available in all the languages of the country, and having accessible and open tools to help users explore the data. Governments should also promote greater interoperability among datasets to truly unleash the full potential that open data can bring to society. This said, releasing open data alone is not enough. Relevant data strategies need to be implemented along the way, ensuring public officials are properly trained and that there is active promotion on the benefits of releasing and reusing data.

Latin America and the Caribbean

The Barometer covers 23 countries in Latin America and the Caribbean. Despite mean scores lower than the global average on governance and capability, the region matches global average for data availability and use and impact, reflecting, in particular, the role of community-led open data initiatives in supporting data publication and use. Further developing and embedding participatory and inclusive approaches to data policy, and developing stronger frameworks to govern data sharing and algorithmic use of data, should be on the agenda for the region in coming years, alongside work to strengthen sectoral open data initiatives through ongoing peer-learning.





Overall Score	Governance	Capabilities	Availability	Impact and Use
32.71	35.16	37.18	29.49	21.27

Country	Comparative Strength (Regional)	Comparative Weakness (Regional)	Overall Score	Governance	Capability	Availability	Use & Impact	Open Data Policy	Open Data Initiative
Argentina	(G) Political finance RL	(A) Emission	50.4	56.4	56.0	44.8	30.3	72.0 ▣	72.0 ▣
Bahamas	(G) Public consultation data RL GL	(C) Open data initiative	23.9	31.4	28.5	17.1	12.4	63.0 ▲	0.0 ▣
Belize	(G) Public consultation data RL GL	(A) Budget and spend data	24.4	30.1	31.3	17.3	14.6	36.0 ▲	35.0 ▲
Bolivia	(A) Existing land use	(G) Public consultation data	22.0	19.2	31.1	19.6	13.4	63.0 ▲	33.6 ▣
Brazil	(A) Emission RL	(G) Political finance	58.0	61.6	49.5	62.0	41.5	72.0 ▲	36.0 ▼
Chile	(A) Lobbying data RL	(G) RTI performance	52.9	48.7	50.0	59.2	35.4	50.4 ▲	70.0 ▣
Colombia	(C) Government support for re-use RL	(G) RTI performance	53.8	57.1	66.3	45.8	37.8	80.0 ▣	80.0 ▣
Costa Rica	(A) Emission	(G) Public consultation data	34.5	35.8	37.5	33.7	15.1	37.8 ▼	45.0 ▣
Dominican Republic	(G) Public consultation data RL GL	(A) Existing land use	35.2	31.7	40.3	35.7	24.0	54.0 ▼	80.0 ▣
Ecuador	(A) Public consultation data	(G) Data protection	34.6	37.3	34.2	34.4	17.7	76.5 ▣	56.0 ▼
Guatemala	(A) Budget and spend data	(G) Public consultation data	18.8	19.4	21.3	17.1	16.9	44.1 ▼	13.6 ▼
Guyana	(C) Civil service	(G) Public consultation data	11.2	5.9	27.3	6.2	8.0	0.0 ▣	24.0 ▲
Honduras	(A) RTI performance data	(G) Public consultation data	24.9	28.0	28.8	21.3	17.2	0.0 ▼	28.0 ▼
Haiti	(A) Existing land use	(G) Public consultation data	8.0	2.8	11.9	9.2	10.4	0.0 ▼	0.0 ▼
Jamaica	(G) Asset declarations RL	(G) Open data policy	31.0	39.5	30.7	25.2	29.6	0.0 ▼	25.5 ▼
Saint Lucia	(G) Open data policy RL	(G) Public consultation data	21.3	30.1	32.5	9.2	16.0	81.0 ▲	34.0 ▣
Mexico	(A) Vulnerability RL	(A) Emission	50.6	55.1	46.9	50.6	40.3	80.0 ▲	28.8 ▼
Panama	(G) Public finance data RL	(G) Public consultation data	34.6	36.0	46.4	28.9	12.7	63.0 ▣	72.0 ▣
Peru	(A) Real-time healthcare system capacity RL	(G) Public consultation data	37.6	38.3	37.6	39.5	13.4	54.0 ▼	28.0 ▼
Paraguay	(G) Public consultation data RL GL	(A) Vital statistics	33.5	43.7	30.4	28.6	27.5	34.0 ▼	16.8 ▼
El Salvador	(U) Accountability uses of PI	(G) Public consultation data	13.4	12.6	22.3	9.0	13.6	0.0 ▼	0.0 ▼
Trinidad and Tobago	(G) Public consultation data RL GL	(G) Public finance data	22.4	25.3	28.2	17.9	13.5	0.0 ▼	13.6 ▼
Uruguay	(C) Government support for re-use	(A) Company register	55.2	62.8	66.2	46.1	27.9	80.0 ▣	72.0 ▣

G = Governance pillar; C = Capabilities pillar; A = Availability pillar; U = Use and Impact pillar; RL = Regional Leader (among the highest scoring regionally) on this indicator; GL = Global Leader (among the highest scoring globally) on this indicator. ▲ = GDB score higher than ODB score on related indicator; ▼ = GDB score lower than ODB score on related indicator; ▣ = less than 1pt change in score; ☐ = No data for comparison.

Hub Perspectives: Iniciativa Latinoamericana de Datos Abiertos - ILDA (Latin America)

By Fabrizio Scrollini and David Zamora

Regional profile

Latin America is the most unequal region in the world. The problems the region faces are complex, the result of a legacy of weak political institutions, poor economic management, and inefficient and insufficient social and innovation policies. Nevertheless, Latin America is also a place where social and business innovation flourish, and where a set of new "digital unicorns" emerged in the last ten years. Most Latin American governments joined the Open Government Partnership (OGP), an alliance of government and civil society organizations, where many of them have played prominent roles leading the way in openness and digital transformations. The Web Foundation's Open Data Barometer - among other indicators - helped demonstrate these achievements. But openness is, at the moment, at a standstill in the region.

The last regional Open Data Barometer coordinated by ILDA shows limited advances from governments, and this reflects a larger and worrying picture currently emerging in Latin America. Moreover, democracy is on the backfoot. While the region always faced challenges, a new wave of governments and social movements has emerged with a tendency to disregard data as a valid source for political debate and policy decision-making. This in turn, fuels disinformation campaigns that affect political debate and political institutions.

Governance as a whole is now heavily debated and new forms of political representation and participation are needed. The environmental crisis is also hitting Latin Americans. Some governments have a poor record in releasing datasets about the environment and natural resources, but there are also capacity issues preventing them doing so. Without proper data on the environment, action against climate change remains wishful thinking. The emergence of AI has the potential to augment several social and economic risks. Latin American countries still do not have a shared approach that could strike a balance between privacy, productivity, innovation and justice.

Findings from the field

The mean score for Latin America on the Barometer's data governance pillar was similar to the global average, showing limited general progress in the region compared to the world. A closer look at individual governance indicators shows some performing better than others with higher scores present for data protection frameworks and open data policies. On the other hand, indicators on data sharing frameworks had among the lowest score in the region. Data sharing is one of the most complex areas to address but is critical to advance the use of data for development.

The mean score for Latin America for data capabilities was slightly lower than the world average. Specifically, the government support for data reuse presents a significant weakness, followed by concerns about the quality and resourcing of open data initiatives. A total of 14 of the 15 participating countries lack funding schemes to support data reuse, and 10 stated that the government does not conduct any form of information sessions about data reuse. And a total of 7 countries lack senior political leaders to back the open data initiative, and another 7 do not have an allocated budget for open data activities. Despite these regional weaknesses, more than half of the countries of the region (as per count) are above the world average. Capacity building

activities remain a priority for Latin America and an area where international cooperation, development banks, the private sector and governments can collaborate.

While globally the political integrity module had the lowest average scores, in Latin America this module was the second lowest, some points above the global average. The frameworks on assets declarations and political finance were the two indicators with the highest scores in the region, followed by the data availability on political finance. Opposed to these indicators, the interoperability of political integrity data, and the availability of lobbying data were the two indicators with the lowest average in the region. Also, the region has an important gap between the top performing countries and the ones with poor performance. This second group has 4 countries with significant low scores. In short, Latin America has great room to improve in this area.

The average regional score on climate action data was above the global average in the Barometer. The three assessed indicators (data availability on emissions, biodiversity and vulnerability) had a higher average in the region than in the world, with climate vulnerability data as the indicator where the region has the largest positive gap when compared to the global average. This is the result of some positive efforts of specific countries in publishing climate data. Despite this, the general results on climate action data cannot be considered satisfactory in the region, because overall scores still obtained only half of the possible points available, showing that there are still many pending tasks both on the openness of data and, for a couple of countries, making basic climate data available online in any form.

On the specific topic of privacy, the Barometer identified that 14 of the 15 participating countries of Latin America have a data protection framework, and 13 of these frameworks have the force of law, showing a general progress on the topic. Most of the frameworks provide data subjects with rights of choice or consent and the rights to access and correct data about themselves. However, none of them explicitly cover the protection of location-related data, and only one addresses in some context the algorithmic decision making, which, in general, gives the impression that the data protection frameworks in the region require updates to address modern privacy issues.

Future directions

Our recent regional event - Abrelatam Condatos - was focused on the future of the open data field. Amidst the pandemic, thinkers and doers from the region gathered to brainstorm about the possibilities the future could bring. This is part of what the world can learn from Latin America: regional forums to set up dialogues and set standards are amazingly effective. Synthesizing these dialogues, and drawing on Barometer research, we think there are three key areas where the field needs improvement.

First, openness needs to be defended. The openness record in the region has stalled and there seems not to be enough support from governments in order to improve the supply of open data in several fields. There is an urgent need to bring together a coalition of stakeholders able to focus on defending openness in sectors where the region needs it most: environment, public finances and integrity. To sustain openness, it is important to support an emergent group of civil society, media outlets and academics using these data in these sectors.

Second, data goes beyond openness. The way data is currently treated by businesses and governments is, to a certain degree, uncertain. Latin American privacy frameworks are evolving, but the region does not have a coherent and shared view of how openness, privacy and innovation are linked together. In the age of AI, this becomes crucial to reap the benefits of innovation. New governance models need to emerge based on evidence and respect of human rights.

Third, capacity and inclusion matter. Many design and implementation decisions are still in the hands of a small number of people who are not able or willing to create inclusive policy design processes and implement the emerging designs. The inclusion of criteria that are able to adequately and safely include different genders, as well as represent excluded communities such as indigenous populations, matter. Investment in robust and reliable public infrastructure to sustain the data field is also needed to continue to build capacity in the region.

All in all, Latin America has work to do if the data field is to contribute towards the genuine development of the region. The evolution will be inextricably linked with the evolution of political

institutions in the region, and as result, the critical question to ask is: to what degree will the data field serve democracy?

Hub Perspectives: Caribbean Open Institute (The Caribbean)

By Suzana Russell, Lila Rao-Graham, Maurice McNaughton

Regional profile

The Caribbean Open Institute (COI), based at the Mona School of Business & Management, UWI Mona Campus, is one of a global community of regional hubs that supported the first Global Data Barometer project through engaging in the data collection, data analysis and dissemination of the results as part of the 2021 global study. Eight Caribbean countries participated in this initial study: Bahamas, Belize, Guyana, Jamaica, Saint Lucia, Trinidad and Tobago, Haiti and Dominican Republic.

The Caribbean is one of the most culturally and politically diverse regions in the world. With a long history of cultural and commercial "openness". Situated as it is, astride the major East–West shipping lanes, colonial histories have given the Caribbean region an eclectic legacy of influences in political, social, cultural, and administrative institutions. All the countries covered in this region are considered to be Small Island Developing States (SIDS), raising particular challenges with respect to public resources for data-related work. When it comes to open data, the region has been characterized as 'high demand, but slow supply'.

Findings from the field

In terms of overall assessment, the Caribbean as a region ranked just below the global average across all four pillars. The margin from the average was greater for availability and least for use and impact. This is consistent with findings from prior studies (Open Data Barometer) that suggest that while businesses and entrepreneurs in the region demonstrate the capacity and readiness to take advantage of the social and economic opportunities offered by data for development, this is not sufficiently matched by a commensurate commitment and action on the part of governments in terms of appropriate data initiatives and policies.

Exploring the Governance pillar further revealed the Caribbean to be strongest in data protection mechanisms but weakest in data sharing frameworks. For the Capabilities pillar, while there is some evidence of active government-led open government data initiatives in several countries, related activities to support data capacity-building for civil servants, and to encourage wider data reuse were less evident. In terms of the secondary indicators reflecting capabilities, "political freedoms and civil liberties", "human capital" and "business use of digital tools", were all significant enabling factors for the Caribbean, while the weakest were "digital government" and "data institutions". This finding supports the narrative that the governments of the region need to be more proactive in establishing an enabling environment and building institutional capacity to support the data ecosystem.

Drilling down further in availability and into use and impact pillars through the thematic modules, the region scored highest with respect to procurement and health-related data. Surprisingly, the lowest thematic module was in the availability and use of Climate Action data, given the Caribbean's particular vulnerability to climate change.

Future directions

The most recent World Development Report[28] outlines that countries can only realize the full value of data for development by establishing an integrated national data system that is built on:

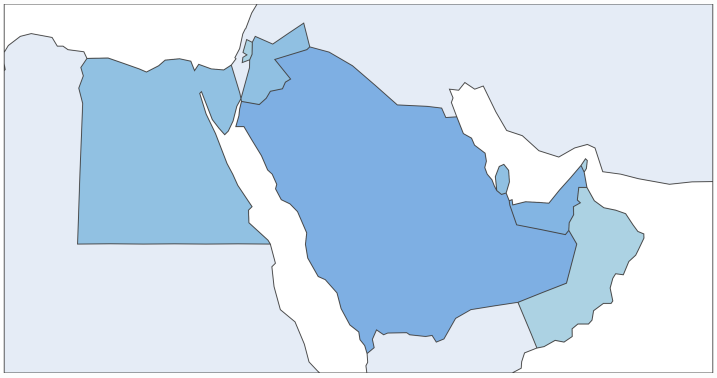
- intentional, whole-of-government and multi-stakeholder data governance and institutions that safeguard and protect data from misuse;
- highly skilled human resources in government, civil society, academia, and the private sector;
- available data that are produced, open, quality-controlled, used and reused.

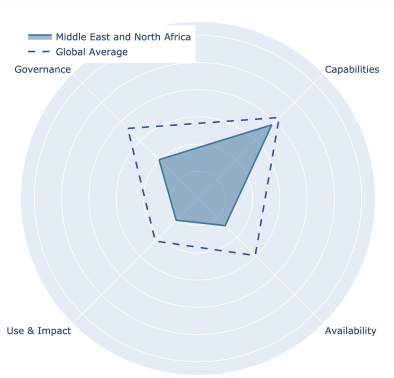
The Global Data Barometer is an important multi-dimensional Index for the Caribbean, as it helps regional policymakers and decision-makers in both public and private sectors to measure and benchmark progress across all of these dimensions. Healthy national and regional data systems are seen as a critical component of equitable digital ecosystems and an explicit pre-requisite for the realization of the sustainable development goals (SDGs) of the 2030 Agenda.

The early findings above suggest a particular need for adequate funding and targeted investments into data institutions and the technical skills of public sector agencies to coordinate data sharing efforts, as well as work to strengthen critical processes and service delivery channels related to data.

Middle East and North Africa

The Barometer covers 8 countries in the Middle East and North Africa region. While the region is significantly below the global average on governance, data availability and use and impact, it is comparatively strong in terms of both government and private sector capability to manage data, with limits on capability scores shaped by restrictions on political freedom. Further developing partnership models that can harness capacity in the private sector to promote use of data for the public good, particularly in relation to sustainable development challenges, offers one route to deepen support for data re-use. Through work on data sharing frameworks and capacity building activities, there is also a need to focus on ensuring equality of access to the benefits of data and to protection from potential harms.





Overall Score	Governance	Capabilities	Availability	Impact and Use
21.6	20.19	38.26	14.08	11.29

	Country	Comparative Strength (Regional)	Comparative Weakness (Regional)	Overall Score	Governance	Capability	Availability	Use & Impact	Open Data Policy	Open Data Initiative
	UAE	(C) Sub-national RL	(G) Data protection	26.7	16.1	58.2	17.6	12.0	20.0 ▣	72.0 ▲
	Bahrain	(G) Data protection RL	(C) Open data initiative	22.0	23.6	28.1	18.4	10.8	63.0 ▲	20.0 ▼
	Egypt	(A) Emission RL	(G) Open data policy	21.8	25.0	34.7	13.2	12.0	0.0 ▣	16.8 ▣
	Jordan	(G) Public consultation data RL GL	(G) Data protection	22.2	34.5	33.7	7.8	12.3	63.0 ▲	80.0 ▲
	Oman	(A) Budget and spend data RL	(C) Open data initiative	14.1	8.6	25.1	12.1	12.0	54.0 □	0.0 □
	State of Palestine	(C) Open data initiative RL	(G) Data protection	14.7	9.1	32.3	9.2	9.6	37.8 ▲	81.0 ▲
	Qatar	(G) Data protection	(A) Public procurement data	22.2	16.6	44.9	14.2	12.0	40.0 ▣	80.0 ▲
	Saudi Arabia	(G) Data management RL	(G) Asset declarations	29.0	28.0	49.1	20.1	9.6	63.0 ▲	80.0 ▣

G = Governance pillar; C = Capabilities pillar; A = Availability pillar; U = Use and Impact pillar; RL = Regional Leader (among the highest scoring regionally) on this indicator; GL = Global Leader (among the highest scoring globally) on this indicator. ▲ = GDB score higher than ODB score on related indicator; ▼ = GDB score lower than ODB score on related indicator; ▣ = less than 1pt change in score; □ = No data for comparison.

Hub Perspectives: CEE Birzeit University (Middle East and North Africa)

By Raed Sharif with Abed Khooli

Regional profile

Over the past decade, global changes and several rounds of uprising in the MENA region have fundamentally changed the government-citizen relationship. These shifts have reshaped what citizens now perceive as their rights and duties, especially in areas related to accountability, transparency and openness. More recently, the rapid digitalization of economies and societies in response to the COVID-19 pandemic has created pressures for faster and more effective data and open data strategies, tools and applications.

At the same time, it is important to acknowledge some of the fundamental infrastructure and policy challenges that the region is facing. According to the International Telecommunication Union (ITU), 55% of residents in the Arab countries had access to the Internet in 2019, with more than two thirds of them (67%) being youth between 15-24. Broken down by gender, 47% of women and 61% of men had access to the Internet in 2019. This means a gender disparity of 0.77. In terms of geography, 74% in urban areas (mostly 4G) and 34% in rural areas had access to the Internet, with rural areas mostly reliant on slower 3G mobile connections.

Further breaking down the data by sub-regions and countries reveals further disparity in both access to, and use of, the Internet. For example, the percentage of Internet users in the Gulf area is over 90%, between 60-65% in Northern African Arab countries (Morocco, Algeria and Tunisia) and less than 40% in other parts of the MENA region such as Sudan (30.9%), Yemen (26.7%), Syria (34.3) and Mauritania (20.8%). Similar trends also exist in terms of individual ICT skills (e.g., ability to use devices, software and apps).

At the legal and policy level, only six Arab countries have laws related to right to information (Jordan (2007), Morocco (2011), Yemen (2012), Sudan (2015), Tunisia (2016) and Lebanon (2017)), and three are members of the Open Government Partnership (OGP).

Findings from the field

Despite country-level differences, the survey revealed modest progress towards more data sharing, openness and usability in the MENA region. There is also a growing political will and official support for data and open data initiatives in countries such as UAE, Qatar, Oman and Saudi Arabia. Additionally, there are encouraging signs of civil society and other non-governmental actors playing a more active role in promoting effective data management, sharing and usability, and in helping governments in translating their policies and visions into tangible data products and services.

However, persisting capability gaps can have a negative impact on the potential impact of national initiatives. Limited or lack of infrastructure and weak digital literacy can hamper citizens' ability to fully utilize the available data and meaningfully participate in public consultations and decision-making processes. For example, one of the practices that has proved popular and effective in the region to incentivize citizens and key stakeholders to get involved in the open government and data domain (and the wider government digital transformation process) has been data ideathons and hackathons, where citizens, innovators, activists, researchers and data scientists collaboratively explore and test the

potential of a wide range of datasets to create solutions for the social, economic and environmental challenges their societies face.

Future directions

The Barometer survey revealed specific areas in need of action, including:

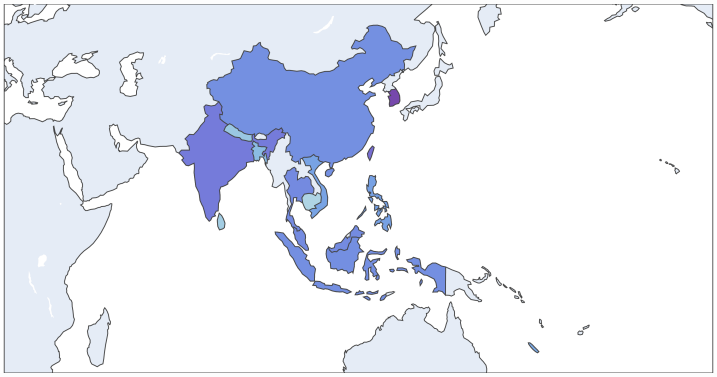
•

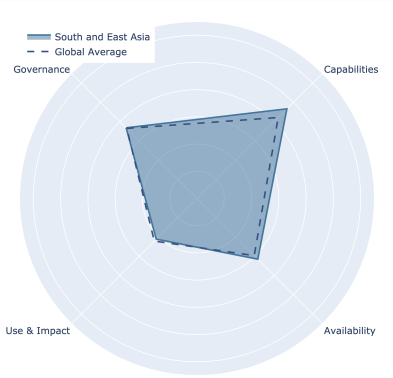
Additionally, we encourage data agencies and initiatives in the MENA region to form strategic engagements with civil society organizations, the private sector and other key stakeholders to build the capacities necessary for effective and strategic use of data that is made available. There is also a need for more data-related professional training and academic programs that can increase and improve the region's human capital around data for the public good.

Finally, we encourage data agencies and initiatives in the MENA region to explore the potential for regional collaborations, and the exchange of knowledge, skills and lessons learned, to build regional capacity in responsible AI and data analytics. This is especially important in areas related to Arabic language machine learning tools and techniques, advanced data analytics and data visualization, with more focus on infrastructure and regional regulatory frameworks.

South and East Asia

The Barometer covers 15 countries in South and East Asia. The diversity of countries within the region is also reflected in significant variation in regional scores across all four pillars. Despite widespread presence of open data policies and initiatives, data availability and open data publication of key datasets could be stronger: particularly in relation to political integrity, company information and climate action.





Overall Score	Governance	Capabilities	Availability	Impact and Use
32.79	35.9	37.05	28.62	27.65

Country	Comparative Strength (Regional)	Comparative Weakness (Regional)	Overall Score	Governance	Capability	Availability	Use & Impact	Open Data Policy	Open Data Initiative
Bangladesh	(A) Political finance data	(G) Public consultation data	23.8	21.1	32.1	22.1	10.8	63.0 ▲	49.0 ▣
China	(G) RTI performance	(A) Vaccination (COVID-19)	39.8	35.8	55.5	35.3	24.3	18.9 ▣	68.0 ▲
Hong Kong	(A) Emission RL	(G) Data management	48.6	37.5	58.2	52.9	30.1	54.0 □	80.0 □
Indonesia	(G) RTI performance RL	(G) Data protection	40.2	48.0	46.2	32.2	29.8	0.0 ▼	76.5 ▣
India	(G) Data management	(G) Data sharing frameworks	46.7	53.2	51.4	41.6	22.9	90.0 ▲	80.0 ▣
Cambodia	(C) Civil service	(G) Public consultation data	13.2	5.6	25.0	12.2	8.4	0.0 □	0.0 □
Republic of Korea	(A) Beneficial ownership RL	(A) Vulnerability	64.5	66.4	80.4	57.0	32.6	90.0 ▣	100.0 ▣
Sri Lanka	(G) RTI performance	(G) Public consultation data	16.3	13.4	35.0	8.2	12.0	0.0 □	40.0 □
Malaysia	(C) Government support for re-use	(G) Data management	41.6	44.8	68.8	24.2	35.3	54.0 ▲	63.0 ▣
New Caledonia	(G) Data protection RL	(G) Public consultation data	30.7	32.2	37.9	27.2	13.0	80.0 □	54.0 □
Nepal	(U) Procurement data analytics	(G) Public consultation data	18.9	16.7	23.0	18.3	18.6	0.0 ▣	0.0 ▼
Philippines	(A) Beneficial ownership	(A) Company register	34.0	44.7	33.4	28.3	18.2	50.4 ▼	40.8 ▼
Thailand	(G) Political finance	(G) Public finance data	41.7	42.2	55.0	35.9	19.7	70.0 ▲	80.0 ▲
Taiwan	(G) Lobbying register RL	(C) Civil service	51.1	48.9	55.4	52.7	23.4	72.0 □	80.0 □
Viet Nam	(G) Public finance data	(A) Real-time healthcare system capacity	33.3	40.7	41.8	24.7	17.7	36.0 ▲	40.8 ▲

G = Governance pillar; C = Capabilities pillar; A = Availability pillar; U = Use and Impact pillar; RL = Regional Leader (among the highest scoring regionally) on this indicator; GL = Global Leader (among the highest scoring globally) on this indicator. ▲ = GDB score higher than ODB score on related indicator; ▼ = GDB score lower than ODB score on related indicator; ▣ = less than 1pt change in score; □ = No data for comparison.

Hub Perspectives: D4DAsia (Southeast Asia)

By Khairil Yusof and Pyrou Chung

Regional profile

Asia is a populous and diverse region, with different languages, cultures, forms of government, levels of economic development and varying political context even within sub-regions. As such, when reviewing general trends and overall scores, one will need to have a deeper dive into each country and each pillar/module in detail to better understand the challenges and opportunities for data for development.

For Southeast Asia, as for many regions, in addition to health issues, Covid-19 also has resulted in rapidly declining civic space, particularly in Hong Kong, and to a lesser extent in Malaysia, Thailand and Philippines, with political instability and the use of emergency measures having impacts on data availability in a number of cases.

East and Southeast Asia is also a geopolitically important region for international trade and investments. Many disruptions in the global supply chain due to Covid-19 lockdowns can be traced to key manufacturing exports and shipping of goods from the region. The capacity of governments to provide quality data and digital services related to trade and investment will be key for economic recovery in the region.

At the end of 2021, Southeast Asia also faced disastrous floods and tropical storms, and, with the climate crisis, extreme weather conditions are likely to continue to affect the region more often in coming years. Tropical forests hold over 68% of world global carbon stocks, and Southeast Asia is home to 15% of the world's global tropical forests.

Findings from the field

Southeast Asia is an outlier among the other sub-regions of Asia, having higher capability than availability of data. In general the overall Barometer scores in the region correlate with the level of economic development (per capita/ income) of a country. The notable outlier within the region is Taiwan, which has high data availability, but poor scores for governance.

With the exception of South Korea, the top scoring Asian country in the study, the data capabilities of countries do not appear to translate well into impact. This raises questions about the kinds of interventions that might be needed to support wider use of data.

With respect to open data, countries in the region appear to have good legislative, legal and policies foundations for open data, with the notable exception of Cambodia, where data availability and capabilities are provided by both local and international civil society organizations, rather than by the government.

The diversity of the region is evident from the Barometer data. For each country in the region, there are a few modules or indicators where they score weakly in comparison to overall scores. However, country data strengths do not appear to always follow the most pressing issues they face, suggesting there may be a disconnect between the data infrastructures countries have and those they need to address future challenges. For example, a number of the countries recently affected severely by climate change have comparatively low scores in the climate action module when compared to their overall capabilities.

Future directions

Rather than focusing on overall scores, countries can use the Global Data Barometer website to take a deep-dive into indicators and modules where their scores are comparatively weak, looking at how individual measures might be targeted for improvement. Looking at the comparative strengths of other countries is equally important. Each country has something unique that it does well and that others could learn from. For example, South Asian countries such as Sri Lanka, India and Nepal have very strong right to information laws and capacity that other countries can learn from. There is great potential to further develop the opportunities for sharing good practices across the region.

Overall, Southeast Asia appears to have good capacity in theory to secure public good from data, but in comparison to other regions, there is a large gap between data collection and availability for re-use. The evidence collected by the Barometer survey provides a wealth of insight into where this locked up data may exist and can help in shaping strategies to unlock this data through government and civil society initiatives. In particular, for a region that has been affected by severe weather due to climate change, much more needs to be done in terms of data for climate action. Many countries have large populations living on flood plains and are affected annually by floods and tropical storms along with deforestation, but the scores registered in the Barometers' Climate Action module for the region were relatively poor.

As we look to the future uses of the Barometer, we also note that in a connected world, data of relevance to each country may also be found at the global level or across borders. For example, we should pay attention to global data publication systems, such as UN COMTRADE for trade statistics, and the US/NASA Global Ecosystem Dynamics Investigation mission (GEDI) that can provide forest coverage data for countries that do not have the capacity to generate or publish such data. Similarly, international non-government organizations and development banks have and should make available the data they gather in the countries they work in, providing alternative sources of data availability for countries that currently lack the capacity to provide key datasets.

Using the current Barometer, we can already look at potential cross-border data that could help solve domestic challenges, such as beneficial ownership and asset disclosures data available from countries with large financial sectors like the United Kingdom and United States, which can be used to gain insights into the foreign assets of politicians, or political financing flows from other countries. To support data-informed decision-making around East and Southeast Asia, we need to continue to develop both regional data governance and availability, but also to continue to develop awareness of the region-related data resources on climate change, vaccine exports, trade and many other topics being made available elsewhere in the world.

Hub Perspectives: Internet Society Hong Kong (Hong Kong)

Internet Society
Hong Kong Chapter

By Benjamin Zhou

Regional profile

Hong Kong, officially the Hong Kong Special Administrative Region of the People's Republic of China (HKSAR), is governed under the "one country, two systems" principle, which allows the city to maintain a legal system different from the mainland of China. In practice, the city government enjoys a high level of autonomy for local affairs including formulation of data policies. Hong Kong is also an international financial hub recognizing Chinese and English as official languages, and both are widely used.

Regarding data policies and measures, the Hong Kong government has adopted a sector-specific rather than a holistic approach. There is no law or policy document laying out a strategy of collecting, sharing and protecting data, nor evidence indicating that they plan to make one. Except for the data privacy law - Personal Data (Privacy) Ordinance - enacted in 1996, the government's data-related initiatives are mostly administrative measures without legislation. For example, while an open data portal was launched in 2011 in a pilot scheme, it wasn't until December 2017 that the government officially addressed "open data" in a high level policy paper and in the chief executive's annual policy address. The portal data.gov.hk is maintained by the Office of Government Chief Information Officer (OGCIO) as a platform for all government departments, public organizations and some private companies (voluntarily) to publish data. Meanwhile, some other departments built their own data platform, such as an on-going project on Common Spatial Data Infrastructure developed by the Lands Department.

Findings from the field

The assessment results show that, among all four pillars of the Barometer, Hong Kong performs better on data availability than the other three pillars by means of global ranking, while governance is the weakest one in this regard.

The datasets examined by the Barometer's thematic modules are mostly available, except for three: beneficial ownership, lobbying and political finance. The beneficial ownership indicator in the company information module scores zero, as Hong Kong's Companies Ordinance (amended in 2018) which requires companies to maintain the information of their significant controllers does not mandate a centralized beneficial ownership register. Hong Kong also fails to receive credit on lobbying and political finance indicators in the political integrity module as there is no legislation regulating political parties including their financing, though the issue has been debated publicly on and off for many years. In fact, such data is not available from the government or any authorized third party, but from some political parties' as a result of voluntary disclosure. Compared to other jurisdictions in the Barometer assessments, Hong Kong's data availability is good in general.

The major issue of Hong Kong lies in governance: it has lower scores on governance indicators than in other pillars including capability as the city has yet to set up a governance structure or provide any data strategy. Among all indicators in this module, only the data protection indicator scores higher than 80% due to the privacy legislation in place. Open data policy exists, but simply asks each government department to make an annual open data plan by themselves and make sure datasets published on the open data portal fulfill a few criteria (i.e. machine-readable, timely, metadata and open license). Indicators for data management and sharing frameworks score the lowest as they are not clearly documented, if not unavailable. Some sector-wide data sharing initiatives exist, for example, a platform developed by the Hospital Authority to share healthcare data from public hospitals for research purposes, but no data sharing framework for the society at large was found during research. A data management framework, regardless of whether it is open data or not, may exist in the form of national/local data strategy, guidance or standards according to the Barometer guidelines, however, none of these policy tools are publicly available in Hong Kong.

Future directions

Although the evidence shows a good foundation of data availability, merely pushing government departments to publish more datasets is inadequate to address challenges in the big data era and to unlock data values. A lack of data governance structure and strategies in the administration risks impairing current achievements and government capability to deliver public service, facing an increasing amount of data all around. The Hong Kong government should invest more in formulating a long-term roadmap on coordinating collection, management, sharing, publishing, and utilization of data across all sectors in the society at large. Vision and leadership are essential for the governance structure to ensure the strategies and policies will be constantly reviewed and adapted to address new challenges.

Conclusions and future work

This report reviews data from the first edition of the Global Data Barometer, organized around four pillars: data governance, data capabilities, data availability, and data use and impacts. Through a global survey and secondary data, the Barometer has explored different dimensions of data governance, including the prevention of data mis-use, the management of data to make sure it is fit for use, and the promotion of data availability for re-use. It has addressed the extent to which data related resources and skills, and the freedom and opportunity to deploy them, exist and are distributed across each country. It has looked at the extent to which data needed to address pressing issues is shared or open in each country, including data related to the climate crisis, COVID-19 pandemic, political integrity, company ownership, land use and management, and the oversight of public procurement and finance. It has gathered examples of data use, along with evidence of impacts, across four use-cases with relevance to public, private, civil-society and media sector stakeholders. The evidence and indicators gathered through our survey fill critical knowledge gaps, and the Barometer's networked model of regional research hubs and data collection by country researchers has supported capacity building for more holistic action on data for the public good.

Through a weighted index the Barometer summarizes hundreds of data-points for each of the 109 countries covered within detailed ratings. In this report, we've used those quantitative scores, and the individual metrics that generate them, as an entry point for comparative analysis and exploration. We have been able to cover only a small subset of the evidence gathered by the Barometer. This report sits alongside the Barometer dataset, country and thematic profiles, and in-depth data stories, all available at https://www.globaldatabarometer.org, as well as the analysis and outputs being produced by regional hubs. We invite and encourage readers to engage with these companion resources in order to dig deeper into Barometer findings and to interrogate how they can inform your own priorities and practice.

Data for the public good: from goal to reality?

If we take the highest score achieved on each Barometer indicator and imagine those scores were all achieved by a single country, it would score 95.92 out of 100 on the overall Barometer weighted index. Given that the Barometer sets a high benchmark, based, in most cases, on globally agreed norms and standards rather than the de-facto practice of any particular country or region, this is a striking finding. Across each theme explored, it is *possible* for data to be governed, made available, and used, in ways that meet broadly agreed aspirations and that maximize the opportunity for public good outcomes. However, the average overall score across all countries was just 34.38 out of 100, showing a significant gulf between the ideal benchmark and current realities.

In short, conditions for data to help deliver the public good *can* be created, but they are by no means guaranteed. Policy choices and practical actions over the coming years must focus explicitly on building the capabilities, institutions and infrastructures around data that will ensure individuals and communities have effective protection from data-enabled harms, while critical datasets are reliably collected, managed, shared and published with the features that are needed for widespread, collaborative and high impact re-use.

In some cases, there are quick wins to be had. By looking to the practice of nearby peers, or focusing on missing fields and features from published open data, governments may be able to make rapid progress against selected Barometer indicators. In other cases, making progress requires larger and longer-term interventions, such as identifying opportunities to reform legal frameworks to better reflect current risks and opportunities related to data and to put disclosure of structured data on a statutory footing or investing in sustained and inclusive capacity-building programmes for data production and

use. Global initiatives that support peer-learning and that provide practical assistance to governments in translating policy commitments into practical data provision appear to have been particularly valuable in accelerating progress in a number of sectors such as public finance and public contracting, and there is a growing case for donors to invest in data infrastructure building. The baselines provided by this first edition of the GDB will be invaluable to track how far other emerging agendas, such as transparency of company ownership, use of data for land governance, or greater disclosure of political lobbying, are able to embed and spread good practices around structured, open and interoperable data publication in the coming decade, and to create diverse data ecosystems that help us reach the Sustainable Development Goals.

The challenge of ensuring that data works for the public good should not be underestimated. The relatively low scores seen across Barometer pillars and indicators provide a reality-check. There are relatively few countries that are anywhere near having interoperable public datasets that cover a breadth of topics. Many low- and middle-income countries have significant data-divides to overcome, and for higher income countries, legacy systems can frustrate attempts to create modern joined-up data infrastructures. Looking also at low scores against inclusion-related questions in the Barometer reveals data-divides within countries can be as significant as those between them. In many cases, the public datasets or skills that might be needed to drive new algorithmic systems, to support analysis to design more inclusive policies, or to enable scrutiny of public decision making through data, are simply not present in any comprehensive forms, let alone widely available. A drive towards data-driven decision making may have very different dynamics in a country that already has reasonably inclusive public data and active community of data users from one where the data needed is only available from private actors, and where data was originally collected with solely private intent. Similarly, the impact of new data collection efforts may vary substantially between countries that have robust data protection frameworks and institutions from those that do not. The Barometer offers a means to critically interrogate claims about how data might transform development in different settings and provides a starting point for the design of more tailored and context-appropriate interventions.

Comparing evidence from this first edition of the *Global* Data Barometer with comparable indicators from five past editions of the Open Data Barometer (2012 - 2020) reveals an open data agenda that is still alive, but more-or-less stalled. With a number of past open data leaders losing their focus on openness, and a churn among the countries rated as having strong policies or initiatives, we see open data struggling to retain attention. Although some countries are building open data principles into sectoral legislation, there is a significant risk that the idea that public data should be ‘open by default’ will continue to lose traction, and gains in making public data open could be lost. If approached carefully, work to strengthen frameworks for data sharing could provide an opportunity to also restate the importance of openness, framing data sharing as the fallback alternative when ‘open by default’ cannot be adopted.

The challenge of making sure data is available for meaningful re-use is well illustrated by data gaps around climate action and COVID-19. Disaggregated data, available for problem-solving practical use by national, and particularly sub-national, stakeholders, can and should be available. Yet, in hundreds of cases, data that’s reported in aggregate to global stakeholders is not accessible to communities, journalists or entrepreneurs in-country or working at the grassroots. While our survey of data use and impact revealed many examples of the work-arounds and accommodations that users may employ for missing or unstructured data, the myriad of barriers to effective data access and use continue to limit how far data-enabled innovation and problem-solving can scale. In this report, we’ve identified a range of potential ways forward to close some of these gaps, from governments on focusing more on the use of their own (open) datasets, and developing more robust data infrastructures, through to increased partnership and collaborative working both around the supply and use of data. Just as the detail of the public good is ultimately defined in each country and community, data infrastructures also need to be defined and shaped through engagement and dialogue.

In summary, this first edition of the Global Data Barometer is ultimately an invitation to dialogue. It is our hope that you may be able to:

- Use the findings and evidence presented in this report to spark discussion and debate about the way data is governed, made available and used in your region, country or sector;
- Explore the visualizations or data stories on the Global Data Barometer website to access new insights, questions or ideas, and to find potential peers to talk with and learn from;
- Download the detailed survey data and use it as part of new research, analysis or exploration - whether that involves statistical analysis of indicator scores, or taking the URLs provided be researchers as the starting point for more globally representative qualitative study;
- Share your feedback with us to help us improve the data and methodology of the study. We know there is a lot to be done to improve both the method and data of the study. Look for details on the website of how to input to shape future iterations of the Barometer.
- Keep in touch with developments by signing up for updates on the Barometer website.

Methodology reflections and future directions

There are, of course, significant limitations in this first edition. As discussed in the methodology section, constraints caused by COVID and the scale of the research challenge in a first edition (where all the research involves *new* evidence collection, rather than *updating* past evidence), mean that we have covered fewer thematic areas than we had originally planned, and we were not able to develop negative indicators looking directly at evidence of harms from data mis-use or poor governance frameworks, or that more robustly measure issues of governance *in practice* rather than *in law*. We faced particular challenges in sourcing secondary data and developing robust primary survey questions to explicitly address issues of gender and equity, and we are committed to continued work to address these gaps. We also identified that we would not be able to do justice to issues of artificial intelligence within the Barometer and scaled back our planned focus on this topic. We look forward to the upcoming Global Index on Responsible AI which we hope will be able to inform future editions of the GDB. It is our intention that in future editions of this study, we will be able to include new indicators (both primary and secondary) and refine some of the existing measures to further develop an approach to measuring not only *conditions* to govern and use data for the public good, but to provide better evidence on whether or not all residents of countries across the world are living in contexts where the net effect of data in society is to drive sustainably and equitable development that is aligned with globally and community-defined visions of the public good.

We have identified areas where a number of our indicators, and the training for researchers to operationalize them, could be strengthened: and at times this has had impacts on the quality of the data generated. With 107,389 data points in the survey, we cannot guarantee that every value is error-free. Indeed, we are certain that there are responses that remain open to question: whether due to different interpretation of guidance across researchers and reviewers, false negatives when sources went undiscovered, or false positives when a source has been interpreted over-generously. While the review process has resolved many of these issues, it has by no means caught them all. Rather than hide these limitations of the data by keeping our source material closed, we prefer the path of sharing our full research data, errors and all, so that re-users can make their own assessments on the accuracy of particular indicators and calibrate how they use the data for their own particular use-cases. Ultimately, when considering the use of data for the public good, all data should be approached critically, ours included.

Appendix: Methodology

This appendix contains details of the weightings applied for calculation of Global Data Barometer scores. More details on the survey methodology used for data collection can be found in the research handbook which is not reproduced here.

Structure

The Barometer is structured around four pillars and is composed of 39 primary indicators from an expert survey and 14 secondary indicators.

Governance	Capabilities	Availability	Use and Impact
			
Covering data management; data protection; data sharing and open data, as well as looking at how data is addressed within sectoral regulations.	Covering foundations (e.g. connectivity; education), government data capability and institutions, and capability within private sector & civil society.	Addressing data existence, features (quality), openness, and coverage (extent) across a number of thematic areas.	Preliminary analysis based on example use-cases.
Weight: 30%	Weight: 24%	Weight: 42%	Weight: 4%

Each indicator belongs to one pillar and to one thematic or cross-cutting module (with one exception: the secondary indicator on 'Data use by international organizations' which is only in the use pillar and not in a module). This supports the calculation of an overall score, pillar scores and module scores.

Each primary indicator is calculated based on the weighted responses given to closed sub-questions organized in a set of subsections. Each primary indicator has a written justification and source links. Indicator sub-questions may have supporting open questions requesting evidence.

With a few exceptions, indicators are structured in the following subsections:

- **Existence** - assessing whether there is evidence that a governance framework, capability, type of data, or data use exists in the country, and the nature of that existence.
- **Elements** - generally split into two parts to assess:
    - **Quality related features** of a law, dataset or capability. Wherever possible, the selection of features was based on both widely agreed international norms and on clear use-cases.
    - **(Open) data related features** of a law or dataset, using a common set of sub-questions to assess issues such as references to/presence of structured data and licenses.
- **Extent** - assessing whether the governance, capability or data evaluated is applicable, relevant or useful to across the whole country and for all citizens, whether it has limitations,

or whether data use has related evidence of impacts.

Barometer data is therefore organized into the following structure, with different information captured or calculated at each level.

- **Overall score** - Weighted score
  - **Pillars** or Module - Weighted score
    - **Indicator** - Weighted score, justifications and evidence
      - **Sub-section** - Weighted score or multiplier
        - **Sub-question** - Raw responses, scores, weighted scores, supporting data

The full research handbook, with details of each primarily indicator, sub-questions and supporting data points requested, can be found at https://handbook.globaldatabarometer.org/2021/.

Question level multipliers and weights

The Barometer design aims to provide both a range of discrete data points, that can support exploratory or hypothesis testing research about **specific** datasets, data uses, capabilities or governance rules, *and* to provide indicators and metrics that support higher-level comparison across countries and contexts. For this reason, we both present the raw data collected from our expert survey and, to calculate comparable indicators, we apply a number of weighting and multiplier rules to data points. In this, we seek to strike a balance between respecting the uniqueness of each issue or theme surveyed and adopting a common approach.

- **Fixed weights** are assigned to each **elements** sub-question.
- **Multipliers** are used to modify the overall indicator score based on **existence** and **extent** sub-questions.

Element scoring and weights

The element section of each indicator is initially scored on a range from 0 to 100 before multipliers are applied.

The element section is divided into groups (e1, e2 & e3). Within the group, sub-questions are generally equally weighted with minor adjustments on theoretical grounds. The weight of each group is set based on the question justification and seeks to balance the relative importance of the sub-questions to an overall evaluation of the indicator drawing on established frameworks and precedent, as well as supporting comparability between indicators. For example, in Availability indicators, upwards of 60% of the element score is made up from a common checklist of properties of the surveyed dataset, while around 40% of the score comes from dataset specific elements.

For some indicators, a fourth set of element questions is present (labelled ‘eb’) which count *against* the score on this indicator. These generally can remove up to 20 points from the elements score.

The weights assigned to each individual element group and sub-question can be found in the `weight` column in the Barometer dataset.

Most individual sub-questions can be answered ‘Yes’ (1), ‘Partially’ (0.5) or ‘No’ (0). The research handbook provides general guidance on when to use the partially response, and specific guidance is provided for certain questions (detailed in the handbook under those sub-questions).

Worked example

The following example shows the elements for the indicator 'To what extent do relevant laws, regulations, policies, and guidance provide a comprehensive framework for protection of personal data?'.

Variable Name	Sub-question	Weight	Note
G.GOVERNANCE.DPL.e	Elements	**100**	The highest score on all elements would sum to 100
G.GOVERNANCE.DPL.e.e1	(Element group) Rights and responsibilities	**80**	For this indicator the e1 element group is worth 80 of the 100 points available. There are five elements in this group, so each is worth 16 points.
G.GOVERNANCE.DPL.e.e1.CONSENT	The framework provides data subjects with rights of choice or consent.	16	-
G.GOVERNANCE.DPL.e.e1.ACCESS_CORRECTION	The framework provides data subjects with rights to access and correct data about themselves.	16	-
G.GOVERNANCE.DPL.e.e1.DATA_HOLDER_RESPONSIBILITY	The framework sets out clear responsibilities for data holders.	16	-
G.GOVERNANCE.DPL.e.e1.REDRESS	The framework provides rights of redress	16	-
G.GOVERNANCE.DPL.e.e1.BREACH	The framework requires data controllers to notify an appropriate authority of data breaches.	16	-
G.GOVERNANCE.DPL.e.e2	(Element group) Specific considerations	**20**	For this indicator, the e2 element group is worth 20 out of the 100 points available. There are two elements in this group, so each is worth 10 points.
G.GOVERNANCE.DPL.e.e2.LOCATION	Frameworks explicitly cover the protection of location-related data.	10	-
G.GOVERNANCE.DPL.e.e2.AI	The framework addresses algorithmic decision making.	10	-
G.GOVERNANCE.DPL.e.eb	(Element Group) Negative scoring	**-20**	For this indicator, the eb element group can remove up to 20 points from the score gained
G.GOVERNANCE.DPL.e.eb.COVIDEXCEPTIONS	Exceptions to the usual data protection framework have been made as part of the country's COVID-19 response.	-20	-

If we imagine a country where the researcher answers 'Yes' to CONSENT (1 (score) x (weight) 16), ACCESS_CORRECTION (1 x 16) and DATA_HOLDER_RESPONSIBILITY (1 x 16), 'Partially' to REDRESS (0.5 x 16), BREACH (0.5 x 16) and LOCATION (0.5 x 10) and COVIDEXCEPTIONS (0.5 x -20), and 'No' to AI (0 x 10), then we would calculate the element score as 59.

Existence multipliers

A multiplier is applied to the elements score based on existence questions. Different multiplier approaches are taken in the different pillars of the Barometer.

Governance

Core governance questions ask to what extent a particular governance framework exists and what form it takes. Where researchers found no framework exists, they were asked “In the absence of a strong legal framework, are there alternative norms or customs that play this role in the country?” and in some cases, have provided structured data about these. To remove scoring values from responses where no framework is present, the total indicator score is multiplied by 0 in such cases.

Where a framework exists but lacks the force of law, the overall indicator score is marginally reduced. This means that if there are two countries with equal features in their frameworks, but one has the force of law, and the other does not, the former country would score higher.

Response	Multiplier
No framework exists.	0
A framework exists but lacks full force of law.	0.9
A framework exists and has the force of law.	1

Thematic governance existence questions come in two parts, asking whether frameworks exist or are in draft, and asking about the strength of data-related rules.

The following multipliers are applied:

Response	Multiplier
Frameworks do not exist.	0
Frameworks are being drafted, or are not yet implemented.	0.6
Frameworks exist and are operational.	1

Response	Multiplier
There is no mention of data or the publication of data in relevant laws, policies, or guidance	0.6
Requirements to publish data are set out in non-binding policy or guidance	0.85
Requirements to publish data are set out in binding policy, regulations, or law	0.95
Requirements to publish this information as open data are set out in binding policy, regulations, or	1

The ‘Accessibility coverage & data’ governance questions do not use an existence multiplier, and instead, the two existence questions are treated the same as elements, allocating up to 40 points to the base indicator score.

Capabilities

Capabilities existence questions use either a three or four-point scale, with multipliers assigned accordingly.

Response	Multiplier in 3 point scale	Multiplier in 4 point scale
0	0	0
1	0.8	0.6
2	1	0.8
3	-	1

Availability

Availability indicators ask whether data is available online, and whether that is as a result of government action or not. The highest scores are reserved for cases where data is provided by government, but researchers are able to use element questions to describe the features of alternative methods of data access, offering key qualitative insights into how data is provided where governments are not reliably offering access to it. The following multipliers are applied:

Response	Multiplier
Data is not available online	0
Data is available, but not as a result of government action	0.5
Data is available from government, or because of government actions	1

Use

Use indicators ask whether there are cases of use identified, and whether these are isolated or widespread cases. The highest indicator scores are reserved for widespread data use.

Response	Multiplier
No evidence	0
Isolated cases	0.5
A number of cases	0.9
Widespread cases	1

Extent multipliers

Indicators have different extent questions, based on whether a full extent might be considered to be national coverage, coverage of all ministries and agencies, and so-on. Some indicators have a single three-element extent question. Others have two extent questions, leading to 6 or more possible values in the extent score. The calculation of the multiplier is based on a threshold such that:

- The most limited level of extent as assessed by the sub-questions receives a weight of 0.7
- Responses reaching the mid-level on extent receives a weight of 0.85
- Where there are more than three extent options, responses between the mid and highest level receive 0.9
- The highest values of extent assessed by the sub-questions receives a weight of 1

Applying multipliers: worked example

Using the worked example from earlier, the governance indicator 'To what extent do relevant laws, regulations, policies, and guidance provide a comprehensive framework for protection of personal data?' has:

- An existence question asking whether the framework has the full force of law
- Two extent questions, asking about how broadly the framework applies (e.g. to different sectors) and about geographical coverage (i.e. does it apply across the whole country?).

Let us imagine that the best available data protection framework in the country lacks full force of law (existence multiplier of 0.9), and that it applies across the whole country, but that it does not apply to every sector (giving a response between the mid, and highest available level, so an extent multiplier of 0.85).

To get the final indicator score we multiply the elements score as follows:

> Element Score x Existence Multiplier x Extent Multiplier

In our example this gives:

> 59 * 0.9 * 0.85 = 45.135

So the final indicator score is 45.135.

Finding sub-question weights

All the question weights used can be found in the published Barometer dataset (Available at https://www.globaldatabarometer.org) which contains a number of key fields that combined show the weights that have been applied.

To find the weight of each sub-question:

- Filter on the `hlevel` column to `hlevel=4` (hierarchy level = 4 = sub-questions)
- Filter on the `data_type` column to `data_type='response'` (to see responses rather than justifications, supporting data, examples or other content)

Weights can then read from the following fields:

field	description
response	The answer selected or entered by the reseacher. For `data_type='response'` this will generally be a fixed option provided by the survey.
response_value	The numeric value assigned to the score. For questions with 'No' (0), 'Partially' (1), 'Yes' (2) answers this will be in the range 0 - 2. For some Existence and Extent questions this can range 0 - 3 or 0 - 4.
normalized_response_value	Each score is re-scaled on a 0 - 1 scale
weight	A multiplier applied to calculated score to get the weighted contribution of this question to the indicator. With minimal exceptions, sub-question weights are held constant within the same

field	description
	subsection of the indicator.
score	normalizad_response_value*weight

Indicator level weights

The contribution of each indicator to module and pillar scores is also weighted. Taking into account issues of data quality, the sensitivity of each indicator, and the number of indicators in each pillar or module, and following discussions with the advisory board, seven weighting principles have been used:

- Primary data from one of the two GDB core modules (governance and capability) receive the first weight priority.
- Primary data from Governance core module weights slightly more than Capability core module.
- Primary data from one of the GDB thematic modules (from any of the four pillars) receive the next weight priority.
- Primary data from Availability weights slightly more than the rest of data in the GDB thematic modules.
- Primary data from Use and Impact weights less than the rest of the data in the GDB thematic modules.
- Secondary data presented as an index (processing several variables) receive the sixth weight priority.
- Secondary data presented as a metric (processing few variables that typically belong to an index) receive the seventh weight priority.
- Secondary data presented as single dichotomous variables receive the eighth weight priority.

Based on these rules, the number of indicators inside each pillar, and the exploration of various weighting options and alternatives, the individual pillars have been weighted as follows: governance 0.30; capability 0.24; availability 0.42; and use and impact 0.04

The tables below show all the indicators in each pillar, along with their pillar weight.

Governance

The governance pillar of the Barometer is made up of 14 indicators (13 primary, one secondary). One indicator (Language coverage & data) is not included in the pillar scoring, because post-survey checks on data quality suggested responses were not robust enough to rely upon, although the qualitative data for this indicator is still included in the Barometer dataset to support future work.

Indicator Name	Module	Type	Question / Source	Weight in Pillar	Weight in Module
Data protection	Governance	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a comprehensive framework for protection of personal data?	0.1	0.2000
Open data policy	Governance	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a comprehensive framework for generating	0.1	0.2000

Indicator Name	Module	Type	Question / Source	Weight in Pillar	Weight in Module
			and publishing open data?		
Data sharing frameworks	Governance	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a comprehensive framework for data sharing?	0.1	0.2000
Data management	Governance	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a comprehensive framework for consistent data management and publication?	0.1	0.2000
Language coverage & data	Governance	Primary	To what extent do relevant laws, regulations, policies, and guidance require that data collection and publication processes be available in the country's official or national languages? If the country has no official or national languages, are these processes available in the languages used in the country?	0	0
Accessibility coverage & data	Governance	Primary	To what extent do relevant laws, regulations, policies, and guidance require that data collection and publication be accessible to people with disabilities?	0.1	0.2000
Beneficial ownership	Company Information	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a basis for collecting and publishing beneficial ownership data on companies?	0.065	0.275
Political finance	Political Integrity	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a basis for collecting and publishing data on campaign and party finance?	0.065	0.075
Asset declarations	Political Integrity	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a basis for collecting and publishing data on the interests and assets of public officials?	0.065	0.075
Lobbying register	Political Integrity	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a basis for collecting and publishing data on lobbying activities?	0.065	0.075
Public consultation data	Political Integrity	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a basis for collecting and publishing data generated through and about public consultation on rulemaking?	0.065	0.075
RTI performance	Political Integrity	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a basis for collecting and publishing data on the performance of right to information (RTI) / freedom of information (FOI) processes?	0.065	0.075
Public finance data	Public Finance	Primary	To what extent do relevant laws, regulations, policies, and guidance provide a basis for collecting and publishing data on public finances? (E.g., government budgets, government spending, debt, and borrowing.)	0.065	0.45
RTI framework	Political Integrity	Secondary	RTI Rating: RTI Rating	0.045	0.052

Capability

The capabilities pillar of the Barometer is made up of 5 primary indicators, and 10 secondary indicators.

Indicator Name	Module	Type	Question / Source	Weight in Pillar	Weight in Module
Civil service	Capabilities	Primary	To what extent is the government providing training to develop civil servants' data literacy and data skills?	0.12	0.09500
Open data initiative	Capabilities	Primary	To what extent is there a well-resourced open government data initiative in the country?	0.12	0.09500
Government support	Capabilities	Primary	To what extent is there evidence that government is	0.12	0.09500

Indicator Name	Module	Type	Question / Source	Weight in Pillar	Weight in Module
for re-use			providing support for data reuse?		
Sub-national	Capabilities	Primary	To what extent do city, regional, and local governments have the capability to effectively manage data?	0.12	0.09500
Political integrity interoperability	Political Integrity	Primary	To what extent is political integrity data interoperable across different political integrity datasets, as well as other datasets associated with relevant information flows?	0.08	0.075
Digital Government	Capabilities	Secondary	World Bank - DGSS dataset: Is there a DG/GovTech Strategy?; Is there a dedicated GovTech institution; Is there a national strategy on disruptive technologies?; Is there a government cloud (shared platform)?; Is there a government service bus / interoperability platform in place?	0.0575	0.07250
Government online services	Capabilities	Secondary	UN E-Government Survey: UN eGov Online Service Index (2020)	0.0575	0.07250
Human capital	Capabilities	Secondary	UN E-Government Survey: UN eGov Human Capital Index (2020)	0.0575	0.07250
Political freedoms and civil liberties	Capabilities	Secondary	FreedomHouse: Political Rights score & Civil Liberties score	0.0575	0.07250
Business use of digital tools	Capabilities	Secondary	WE Forum: Business use of digital tools	0.035	0.05500
Data institutions	Capabilities	Secondary	World Bank - DGSS dataset: Is there a government entity in charge of data governance or data management?, Is there a data protection authority?	0.035	0.05500
Internet access	Capabilities	Secondary	International Telecomunication Union (ITU): Fixed broadband basket as a % of GNI p.c; Individuals using the Internet, total (%)	0.035	0.05500
Knowledge-intensive employment	Capabilities	Secondary	Global Innovation Index / ILO: Employment in knowledge-intensive occupations (% of workforce)	0.035	0.05500
Use of standards and methods in statistic offices	Capabilities	Secondary	Statistical Performance Indicators: Dimension 5.2: Standards and Methods	0.035	0.05500
Digital skills	Capabilities	Secondary	WE Forum: Digital skills among active population (score)	0.035	0.05500

Availability

The availability pillar is made up of 17 primary indicators and 2 secondary indicators. One more secondary indicator is included in the dataset, but it was assigned zero weight during index review because it was found not to adequately track differences between countries.

Indicator Name	Module	Type	Question / Source	Weight in Pillar	Weight in Module
Beneficial ownership	Company Information	Primary	To what extent is company beneficial ownership information available as structured open data?	0.0570	0.325
Company register	Company Information	Primary	To what extent is company information available as structured open data?	0.0570	0.325
Land tenure	Land	Primary	To what extent is detailed land tenure information available as open data?	0.0570	0.45
Existing land use	Land	Primary	To what extent is existing land use information available as open data?	0.0570	0.45
Political finance	Political	Primary	To what extent is political finance information available	0.0570	0.095

Indicator Name	Module	Type	Question / Source	Weight in Pillar	Weight in Module
data	Integrity		as open data?		
Asset declarations	Political Integrity	Primary	To what extent is interest and asset declaration information available as open data?	0.0570	0.095
Lobbying data	Political Integrity	Primary	To what extent is lobby register information available as open data?	0.0570	0.095
Public consultation data	Political Integrity	Primary	To what extent is public consultation information available as open data?	0.0570	0.095
RTI performance data	Political Integrity	Primary	To what extent is detailed RTI performance information available as open data?	0.0570	0.095
Budget and spend data	Public Finance	Primary	To what extent is government budget and spending information (budget execution) available as structured open data?	0.0570	0.55
Public procurement data	Procurement	Primary	To what extent is detailed structured data on public procurement processes available as open data?	0.0570	0.825
Emission	Climate Action	Primary	To what extent is emissions information available as open data?	0.0570	0.3334
Biodiversity	Climate Action	Primary	To what extent is information on endangered species and ecosystems available as open data?	0.0570	0.3333
Vulnerability	Climate Action	Primary	To what extent is climate vulnerability information available as open data?	0.0570	0.3333
Vital statistics	Health & Covid-19	Primary	To what extent is civil registration and vital statistics (CRVS) information available as open data?	0.0570	0.26
Real-time healthcare system capacity	Health & Covid-19	Primary	To what extent is information about the real-time capacity of the healthcare system available as open data?	0.0570	0.26
Vaccination (COVID-19)	Health & Covid-19	Primary	To what extent is COVID-19 vaccination information available as open data?	0.0570	0.26
Healthcare system capacity	Health & Covid-19	Secondary	ODW Open Data Index: Health facilities	0.02	0.14
Testing data (COVID-19)	Health & Covid-19	Secondary	World Health Organisation (WHO): Cases - cumulative total	0.011	0.08

Use and impact

The use and impact pillar is made up of four primary indicators and one secondary indicator.

Indicator Name	Module	Type	Question / Source	Weight in Pillar	Weight in Module
Corporate due diligence	Company Information	Primary	To what extent do products or services exist that use open company data to support due diligence?	0.22	0.075
Influencing policy for gender and inclusion	Land	Primary	To what extent is there evidence that land data is being used to influence policy in the interests of equitable and inclusive land tenure and use?	0.22	0.1
Accountability uses of PI	Political Integrity	Primary	To what extent is there evidence of political integrity data being used to identify, expose, or highlight failures of government?	0.22	0.023
Procurement data analytics	Procurement	Primary	To what extent is there evidence of government procurement data being analyzed to improve procurement practice?	0.22	0.175

Indicator Name	Module	Type	Question / Source	Weight in Pillar	Weight in Module
Data use by international organizations	Use and Impact	Secondary	Statistical Performance Indicators: Dimension 1.5: Data use by international organizations	0.12	-

References

1. UN IEAG. A world that counts–mobilising the data revolution for sustainable development. UN Secretary-General's Independent Expert Advisory Group on the Data Revolution for Sustainable Development; 2014. Available from: https://www.undatarevolution.org/report/
2. Kitchin R. The data revolution: Big data, open data, data infrastructures and their consequences. Sage; 2014.
3. Multiple Authors. Public good (economics). In: Wikipedia. Wikipedia; 2022. Available from: https://en.wikipedia.org/w/index.php?title=Public_good_(economics)&oldid=1074988450
4. Multiple Authors. Network effect. In: Wikipedia. 2022. Available from: https://en.wikipedia.org/w/index.php?title=Network_effect&oldid=1081034493
5. Open Data Institute. The Data Spectrum. 2016. Available from: https://theodi.org/about-the-odi/the-data-spectrum/
6. Baack S, Maxwell M. Alternative Data Governance Approaches: Global Landscape Scan and Analysis. Mozilla Foundation; 2020. Available from: https://assets.mofoprod.net/network/documents/DataGovernanceApproaches.pdf
7. De La Chapelle B, Porciuncula L. We Need to Talk About Data: Framing the Debate Around the Free Flow of Data and Data Sovereignty. Internet & Jurisdiction Policy Network; 2021. Available from: https://www.internetjurisdiction.net/news/aboutdata-report
8. World Wide Web Foundation. Open Data Barometer. 2013–2017. Available from: https://opendatabarometer.org/
9. Open Knowledge Foundation. Open Data Index. 2011–2017. Available from: https://index.okfn.org/
10. ILDA. Regional Open Data Barometer Latin America and the Caribbean 2020. 2020. Available from: https://barometerlac.org/wp-content/themes/odbpress/reporte-ILDA-EN.pdf
11. Davies T, Walker SB, Rubinstein M, Perini F. The state of open data: Histories and horizons. African Minds; 2019. Available from: https://stateofopendata.od4d.net/
12. Greenleaf G. Sheherezade and the 101 Data Privacy Laws: Origins, Significance and Global Trajectories. Rochester, NY: Social Science Research Network; 2013. Available from: https://papers.ssrn.com/abstract=2280877
13. Greenleaf G. Global Data Privacy Laws 2017: 120 National Data Privacy Laws, Including Indonesia and Turkey. Rochester, NY: Social Science Research Network; 2017. Available from: https://papers.ssrn.com/abstract=2993035
14. Union E, Europarat, editors. Handbook on European data protection law. 2018 edition. Luxembourg: Publications Office of the European Union; 2018.
15. Council of Europe. Artificial Intelligence and Data Protection. Council of Europe; 2019. Available from: https://rm.coe.int/2018-lignes-directrices-sur-l-intelligence-artificielle-et-la-protecti/168098e1b7
16. EC. European Data Strategy. 2022. Available from: https://digital-strategy.ec.europa.eu/en/policies/strategy-data
17. Dang H-AH, Pullinger J, Serajuddin U, Stacy B. Statistical Performance Indicators and Index: A New Tool to Measure Country Statistical Capacity. The World Bank; 2021. Available from: http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-9570
18. Micheli M. Public bodies' access to private sector data: The perspectives of twelve European local administrations. First Monday 2022. Available from: https://firstmonday.org/ojs/index.php/fm/article/view/11720

19. OKF. The Open Definition - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge. 2016. Available from: https://opendefinition.org/
20. Brandusescu A, Iglesias C, Robinson K, Alonso JM, Fagan C, Jellema A, et al. Open Data Barometer : Global report : Fourth edition. 2017. Available from: https://idl-bnc-idrc.dspacedirect.org/handle/10625/57676
21. Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, et al. Coronavirus Pandemic (COVID-19). 2022. Available from: https://ourworldindata.org/coronavirus-testing#source-information-country-by-country
22. Mathieu E. Commit to transparent COVID data until the WHO declares the pandemic is over. Nature 2022. 602(7898):549–9. Available from: https://www.nature.com/articles/d41586-022-00424-9
23. Open Data Charter. Open Up Guide to Climate Action. Open Data Charter; 2020. Available from: https://open-data-charter.gitbook.io/
24. Davila J, Barron M, Law T. Towards a Global Norm of Beneficial Ownership Transparency - A scoping study on a strategic approach to achieving a global norm. 2019. Available from: https://adamsmithinternational.com/app/uploads/2019/07/Towards-a-Global-Norm-of-Beneficial-Ownership-Transparency-Phase-2-Paper-March-2019.pdf
25. Garcia Aceves R. Open Up Guide: Using Open Data to Combat Corruption (Version 2.0). Open Data Charter; 2018. Available from: https://open-data-charter.gitbook.io/open-up-guide-using-open-data-to-combat-corruption/
26. International Budget Partnership. Open Budget Survey. 2020. Available from: https://www.internationalbudget.org/open-budget-survey
27. Bayer C-T, Booth K. Open data for improved land governance. Land Portal and Open Data Charter; 2021. Available from: https://landportal.org/library/resources/open-guide-land-governance-version-20-following-public-comments
28. World Bank. World development report 2021: Data for better lives. The World Bank; 2021. Available from: https://www.worldbank.org/en/publication/wdr2021
29. Interaction. Regional Trends: Sub-Saharan Africa. 2020. Available from: https://www.interaction.org/globalsituationreport2020/global-situation-report/sub-saharan-africa/
30. Equal Measures 2030. 2019 SDG Gender Index Report. Equal Measures; 2019. Available from: https://www.equalmeasures2030.org/2019-sdg-gender-index-report/