Task Force 2:
**Our Common Digital Future: Affordable, Accessible
and Inclusive Digital Public Infrastructure**

OPEN DATA AS BACKBONE OF DIGITAL PUBLIC INFRASTRUCTURE: PRIORITISING HIGH-VALUE DATASETS FOR EFFECTIVE GOVERNANCE

July 2023

Gaurav Godhwani, Executive Director and Co-founder, CivicDataLab, India

Pencho Kuzev, Policy Advisor Data and Competition Policy, Konrad-Adenauer-Foundation, Germany

INDIA 2023

G20 2023 INDIA भारत

वसुधैव कुटुम्बकम्

ONE EARTH • ONE FAMILY • ONE FUTURE

Abstract

Open government data (OGD) plays a vital role in engendering trust, accountability, and inclusive socio-economic growth via services delivered through digital public infrastructure (DPI). Studies suggest that publishing granular and timely open data by governments increases transparency and evidence-based policies, improves efficiency and public service delivery, enhances economic growth, and empowers innovation. The COVID-19 pandemic further highlighted the urgent need for governments to publish near real-time open data to help combat the crisis. Though the ease of building and deploying DPI has improved, most OGD initiatives still struggle to sustain and grow. Based on empirical evidence and recent progressive legislations, this brief proposes a shared G20 OGD commitment for DPIs that will help countries prioritise economic growth, climate action, and accelerate progress towards the Sustainable Development Goals.

The Challenge

1

Open government data (OGD) refers to the data collected and published by governments that can be freely used, reused, and redistributed by anyone—subject only, at most, to the requirement to attribute and share-alike.[1] This describes government data that is digitally available in machine-readable and non-proprietary formats, either free or at no more than the cost of reproduction, and under explicit terms that permit re-use, removing all the technical and legal barriers to data access. OGD has numerous proven benefits, including improving efficiency of public service delivery, strengthening economic growth, and enhancing citizen trust and participation in governance.[2]

Countries are prioritising the creation of digital public infrastructure (DPI) to provide timely public service delivery and cater to the e-governance needs of their citizens. DPIs refer to core digital solutions that facilitate essential society-wide functions and services, such as digital identification, digital payments, and data exchange.[3] However, DPIs need to be developed on Open Data principles, define necessary governance standards, and develop policies to ensure that citizen rights and interests are well protected, while defining mechanisms that improve public service delivery. Despite its numerous benefits, OGD has remained mostly an afterthought in e-government and DPI discussions across nations. We see the following major challenges faced by the OGD landscape worldwide.

OGD agendas are alive, but not spreading

According to the 2022 Global Data Barometer (GDB) survey, an expert study conducted across 109 countries, there has been a growth of only 10.6 percent in the volume of government datasets that are truly 'Open' since the last survey conducted in 2016.[4] When a more flexible assessment model is implemented to include datasets with minor weaknesses (like missing data standards, appropriate formats, or licenses linkages) then the survey identifies an increase of 17 percent datasets that can be considered 'Open' (see Figure 1).

**Figure 1: Percentage of datasets surveyed and their availability by region based on Open Data criteria.**

Data availability by region (flexible criteria for open data)
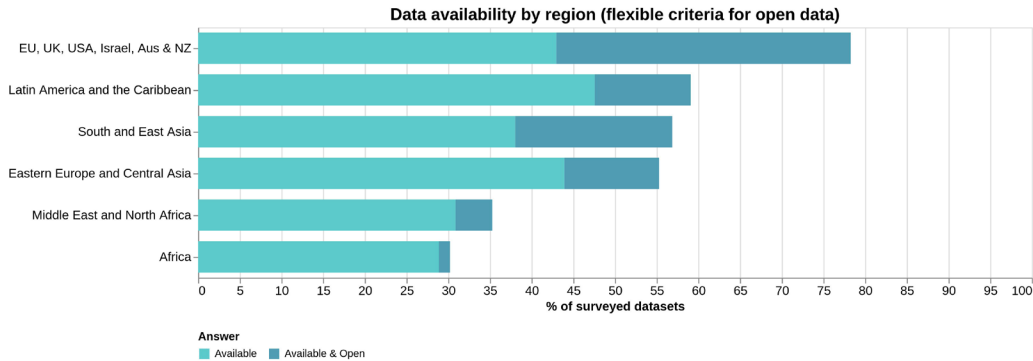
Source: Global Data barometer[5]

Lack of bulk data provision, even with the progress made with DPIs

As per the 2022 GDB Survey and Figure 2, the biggest factor limiting the spread of OGD appears to be the lack of bulk data provision. Bulk access to OGD fosters re-use of data, co-creation of data-driven innovations, and use-cases among stakeholders such as the academia, micro, small, and medium enterprises (MSMEs), businesses, and civil societies. The survey further shows that only 46.4 percent of online datasets were accompanied by some form of exploratory data analysis (EDA) or data visualisation tools, enabling easy access to insights from data.

Figure 2: Barriers to open data availability
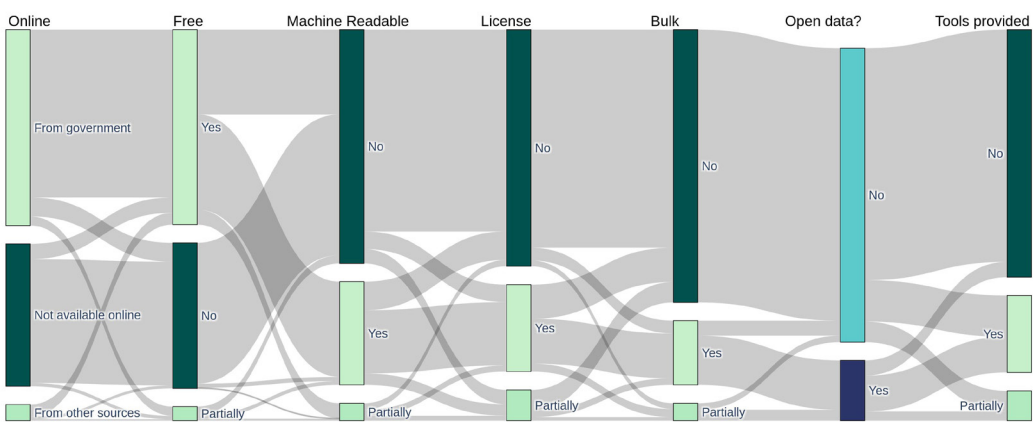

Source: Global Data barometer[6]

The sankey diagram shows 2022 GDB aggregate assessment of all datasets against the open definition criteria, as well as the presence of accessible tools to explore available data.

Limited progress on digital inclusion due to lack of disaggregated data:

Given the commitment made by United Nations (UN) member states of 'leaving no one behind' (LNOB),[7] countries need to operationalise digital services to eradicate poverty in all its forms, end discrimination and exclusion, and reduce the inequalities and vulnerabilities that leave people behind and undermine the potential of individuals and of humanity as a whole. Due to the lack of reliable disaggregated OGD published, it is difficult to measure progress towards LNOB commitments and understand population segments that still do not benefit from e-governance solutions like DPIs. Populations that experience discrimination and exclusion due to poverty, gender, disabilities, access, status within a marginalised group, or/and indigenous backgrounds are often missing in the assessment of e-governance related services, administrative data, and other kinds of OGD. The UN E-government Survey 2022[8] shows that relatively fewer countries are currently collecting and publishing gender-disaggregated user data of e-governance services (shown in Figure 3).

Figure 3: Proportion of countries collecting and publishing gender-disaggregated user data by region.
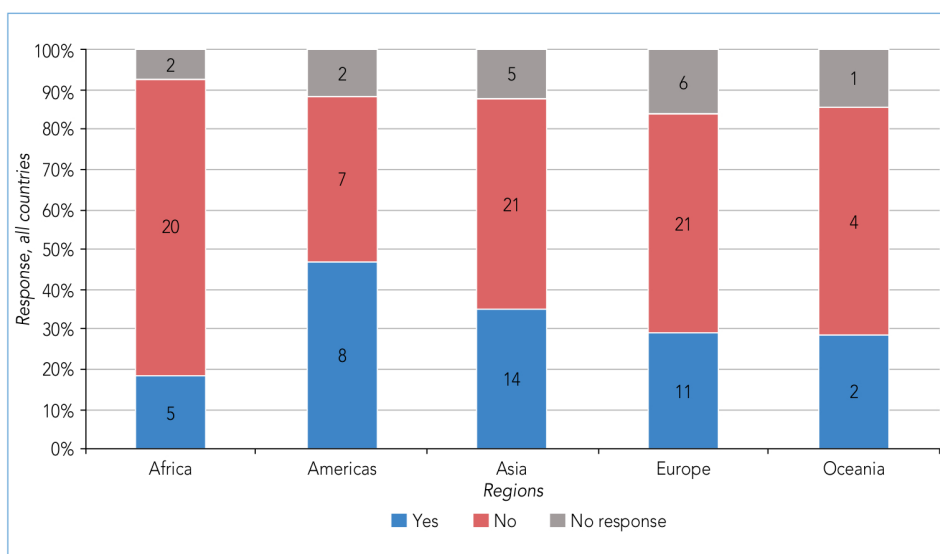
Source: UN E-government Survey 2022[9]

While countries must address these significant gaps in the collection of disaggregated data related to inclusion, there is also a growing need for comprehensive data protection laws similar to the EU General Data Protection Regulation.

Lack of data capabilities and staff capacity

Most government agencies appoint chief data officers (CDOs) who are responsible for managing the data lifecycle within the department and its jurisdiction. They are required to identify datasets that can be made public, cleaned, standardised, and published on time on OGD platforms, and engage with citizens to ensure uptake. Often, CDOs are just a one-person army trying to maintain these crucial datasets, with limited budget and capabilities.[10] In some countries, CDOs are still voluntary positions undertaken by public servants handling multiple responsibilities and portfolios,

leaving them overworked. This worsens when one moves from national to local government departments. Therefore, CDOs require regular capacity building programmes and skills development exercises to stay up-to-date with new technological advancements and better manage OGD lifecycles.

Critical government datasets lack good quality

Various stakeholders frequently struggle to meaningfully consume OGD because of lack of comprehensive metadata, data standards, timeliness, and completeness of the data. According to the 2022 GDB Survey, there are significant gaps in the quality of datasets published by governments. Datasets related to public procurement ranked better, while public health and climate change related datasets remained categories where the most gaps in quality have been observed, as shown in Figure 4.

Figure 4: Mean dataset quality score by category for datasets surveys in GDB 2022

Dataset quality score by category
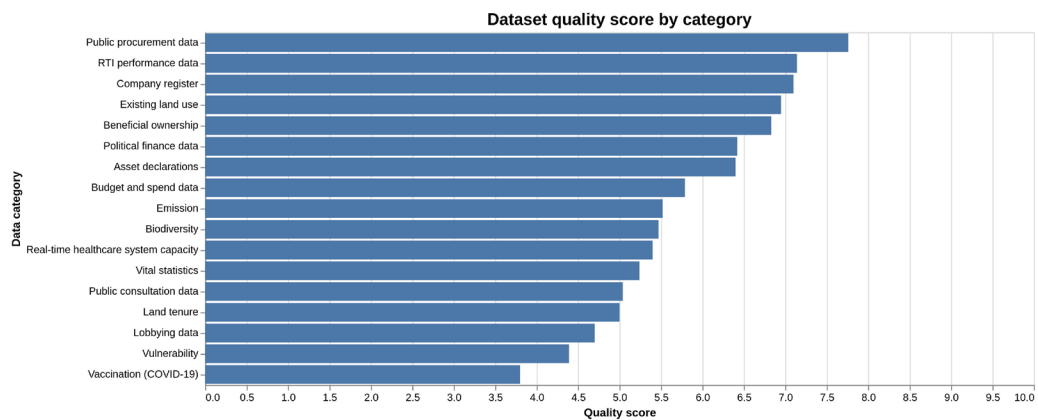
Source: Global Data barometer[11]

The G20's Role

2

The G20 countries are the world's largest economies shaping our shared socioeconomic development with almost 80 percent of the gross world product (GWP), 75 percent of international trade, two-third of the world's population, and 60 percent of world land area.[12] In 2014, the G20's Anti-Corruption Working Group established Open Data as one of the main focuses for the promotion of public sector transparency and integrity.[13] Under the Turkish G20 Presidency, the member states agreed to follow a set of principles based on the international Open Data Charter[14]:

i.   open by default;

ii.   timely and comprehensive;

iii. accessible and usable;

iv.   comparable and interoperable;

v.   for improved governance and citizen engagement;

vi.   for inclusive development and innovation.

The role of G20 countries is critical, because a common approach towards high-value datasets (HVD) within G20

would be a significant enabler of cross-border data applications and services. It would lower entry barriers to data-driven markets, and the availability of various datasets can make it easier to forecast the impact of policy measures. The availability of HVDs should rebalance the position of MSMEs in relation to big tech.

The G20 countries have the potential to help accelerate our shared progress towards Sustainable Development Goals (SDGs) and targets by prioritising a shared OGD commitment. This commitment must focus on emancipating critical datasets from existing silos, standardise, and harness them to proactively respond to the most pressing needs of the citizens and vulnerable population groups across the globe.

Many countries have under-deliveredon their pledge to help disadvantaged countries strengthen their data infrastructure. As part of SDG target 17.18, various countries pledged to help disadvantaged countries strengthen their data infrastructure till 2020. They were supposed to significantly increase the availability of high-quality, timely, and reliable data disaggregated by

income, gender, age, race, ethnicity, migratory status, disability, geographic location, and other characteristics relevant to the national contexts of disadvantaged countries.[15] Most countries have however failed in extending capacity-building support to the developing countries.

Recommendations to the G20

3

Based on empirical evidence and recent progressive legislations, we propose G20 countries to have a shared OGD commitment and action plan prioritising the following:

Adoption of a comprehensive regulatory framework: The G20 can co-create best practices and recommendations on better data-lifecycle management, sharing, standardisation, protection, and thematic guidelines. These best practices and recommendations should be both generic and sectoral in nature, including:

a. *Data governance and sharing:* There is a need for a comparative analysis of existing data governance policies to understand the relationship between Open and Restricted government data governance in various countries. G20 countries should conceptualise comprehensive data sharing frameworks for government-to-government (G2G), government-to-others (G2X), and others-to-government (G2X) entities. At present, a majority of the countries are only prioritising G2G data sharing, but this needs to evolve into

building global data partnerships. For instance, the European Data Act provides protection from unfair contractual terms, mechanisms for public sector bodies to access and use data held by the private sector, as well as enables users of connected devices to access the data generated by these devices.[16] More G20 countries should focus on creating appropriate governance structures for private sector data sharing.

b. *Data protection and ethics framework:* Various countries are drafting and adopting data protection laws and frameworks. Some of these frameworks however are limited to the jurisdiction of the national government and are restricted to specific sectors, providing very little or no safeguards against potential data abuse. A comprehensive data protection and ethics framework is essential to enable data re-use. For example, the UK government has published its Data Ethics Framework,[17] outlining how to use data responsibly when planning, implementing, and evaluating a new policy or service.

c. *Data standardisation and discovery:* There is a growing need for adopting metadata standards like Data Catalog Vocabulary (DCAT)[18] and sectoral data standards to make OGD and datasets contributed by other actors much more discoverable, accessible, and standardised. For example, the European Open Data portal has harnessed the DCAT metadata standard to make datasets discoverable and accessible from more than 170 data publishing portals across 36 countries in one place.

Prioritising 10 Sectors for proactively publishing HVDs: To fast-track progress towards SDG 2030 targets and expand the data-led economy, G20 countries must prioritise opening up of at least 10 HVD categories of OGD as part of their shared commitment and action plan. For example, the EU Open Data Directive defines the HVD as datasets associated with important benefits for the society, environment, and economy.[19] For maximum impact, HVDs should be made available for bulk download via application programming interfaces for re-use with minimal legal restrictions and free of charge.

HVDs must follow a detailed technical specification including a set of required data quality attributes like structure and semantics, completeness, granularity, and formats.[20] Globally, the following 10 HVD categories have maximum citizen demand and high impact potential:

a. *Public Finance:* Timely access to disaggregated government budget, treasury expenditure, receipts, and other fiscal data leads to efficient use of public funds, reduces corruption in public service delivery, and enhances sustained citizen participation in the budgetary process.[21]

b. *Public procurement:* Publishing open, accessible, and timely data on public procurement using standards like Open Contracting Data Standard[22] and sharing information about all significant stages of procurement such as tendering, awards, contract, and implementation helps governments engage all stakeholders, improve efficiency, effectiveness, and integrity.

c. *Land governance:* This includes data related to core land

administration functions of tenure, use, development, and value. The popular use-cases of land governance data include tracking of the ownership landscape, identification of land concentration, and better understanding of access to land and land tenure security.[23]

d. *Political integrity:* This broadly includes data on political party finance, information on the interests of political decision makers, lobbyists' interventions, and public consultation processes. It also includes data on the right to information or freedom of information systems, which enables accountability of elected representatives.

e. *Public health:* The Covid-19 pandemic has highlighted the need for reliable, accessible, and trusted data related to public health infrastructure and service delivery to enable timely and coordinated action. Some key public health datasets include urban and rural public health care infrastructure data, vaccination, vector borne disease spread, and mental health.

f. *Climate action:* Governments have been collecting and reporting a wide range of climate-relevant information, as required by the United Nations Framework Convention on Climate Change (UNFCCC). Thus, some datasets that need to be prioritised include greenhouse gas (GHG) inventories, energy production and consumption, climate change related vulnerabilities, mitigation targets, impacts, and adaptation.[24]

g. *Law and justice:* To better understand implementation of laws at various levels, it is imperative to publish timely and disaggregated data related to laws and amendments, courts, prisons or correctional institutions, policing and legal aid.[25]

h. *Company registries:* Private businesses fuel innovation and economic development but at the same time can pose risks in the form of illicit trades, and social and environmental harms. Thus, data related to company registration, ownership, and key activities can help map private businesses and their contributions to sustainable development.

i. Drinking water and sanitation: Access to safe water, sanitation, and hygiene is the most basic human need for health and well-being. Thus, it is crucial to share data on drinking water services, sanitation, and waste management.

j. Urban development: It is essential to collect and publish data related to urban development issues like infrastructure, transport, civic amenities, public service delivery, and citizen complaints to better plan and make cities more efficient.

Setting-up open data collaboratives to reach vulnerable populations: The UN E-government Survey 2022 highlights the need to create a robust legal, regulatory, and fiscal framework that supports cooperative efforts like Open Data Collaborative, where government agencies can build long-term partnerships with local stakeholders including the academia, NGOs, businesses, and volunteer groups and actively engage them in OGD related activities, policy implementation, and data-driven decision-making. These collaboratives can also act as an invaluable conduit for communication and provide accurate

information on the circumstances and needs of vulnerable populations.[26] For example, information systems around government welfare programs are being built in the Indian states of Rajasthan and Karnataka through regular citizen-led dialogues.[27] Along with this, it is critical to set up accountability mechanisms to protect the rights of vulnerable groups.

Enhancing national and sub-national data capabilities: Co-create capacity building plans and modules for CDOs and other staff members to improve their data management skills, capabilities, and toolkits. Open Data Collaboratives can be leveraged to co-create curriculum, Massive Open Online Courses, and regular training sessions on government data-lifecycle, EDA, data science techniques, visualisations, and community engagement. For example, Data.org and J-PAL have created resources and guidebooks to help local governments and other stakeholders use OGD for research and evidence-based policy.[28]

While unilateral best practices are appreciated, only harmonisation at the G20 level can enhance the value of existing data and open new

opportunities. DPIs have the potential to advance multifold in the journey to achieve 2030 SDG targets, if G20 countries adopt a shared OGD commitment to co-create DPIs. There is a need to develop a comprehensive action plan that enables G20 countries to unlock their legacy datasets and create DPIs that help other countries do the same. By co-creating a robust Open Data Collaborative at international, national, and sub-national levels, we can make government data much more actionable and prioritise the needs of our vulnerable population. Prioritising DPIs to open 10 HVD categories will also improve climate change action and resource management. With regular sharing of learnings from these activities, G20 countries should also help other countries plan their data-driven reforms and accelerate wider and sustained Open Data collaborations.

Endnotes

1   "What Is Open Data?," Open Data Handbook, 2015, https://opendatahandbook.org/guide/en/what-is-open-data/.

2   "What Is Open Data," European Commission, November 6, 2015, https://data.europa.eu/en/dataeuropa-academy/what-open-data.

3   Vyjayanti T Desai, Jonathan Marskell, Georgina Marin, and Minita Varghese, "How Digital Public Infrastructure Supports Empowerment, Inclusion, and Resilience," *World Bank Blogs*, March 15, 2023, https://blogs.worldbank.org/digital-development/how-digital-public-infrastructure-supports-empowerment-inclusion-and-resilience.

4   Tim Davies and Silvana Fumega, "Global Data barometer – First Edition," *Zenodo*, May 11, 2022, https://doi.org/10.5281/zenodo.6488349.

5   Davies and Fumega, "Global Data barometer"

6   Davies and Fumega, "Global Data barometer"

7   United Nations System Shared Framework for Action, *Leaving No One Behind: Equality and Non-Discrimination at the Heart of Sustainable Development*, New York, United Nations, 2017, 84, https://unsceb.org/sites/default/files/imported_files/CEB%20equality%20framework-A4-web-rev3.pdf.

8   United Nations Department of Economic and Social Affairs, *United Nations E-Government Survey 2022 - The Future of Digital Government*, New York, United Nations, 2022, https://desapublications.un.org/sites/default/files/publications/2022-09/Web%20version%20E-Government%202022.pdf.

9   UN, "United Nations E-Government Survey 2022"

10   Stefaan Verhulst et al., "The Emergence of a Third Wave of Open Data: How To Accelerate the Re-Use of Data for Public Interest Purposes While Ensuring Data Rights and Community Flourishing," *SSRN Electronic Journal*, 2020, https://doi.org/10.2139/ssrn.3937638.

11   Davies and Fumega, "Global Data barometer"

12   G20 India, "G20 Background Brief," https://www.g20.org/content/dam/gtwenty/about_g20/overview/G20_Background_Brief_06-03-2023.pdf.

13   G20 Turkey, "Introductory Note to the G20 Anti-Corruption Open Data Principles," 2015, http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf.

14   "G8 Open Data Charter and Technical Annex," GOV.UK, June 18, 2013, https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex.

15   UN Statistical Commission, "Cape Town Global Action Plan for Sustainable Development Data," UN, March 2017, https://unstats.un.org/sdgs/hlg/Cape_Town_Global_Action_Plan_for_Sustainable_Development_Data.pdf.

16   "Data Act: Commission Welcomes Political Agreement on Rules for a Fair and Innovative Data Economy'," *European Commission*, June 28, 2023, https://ec.europa.eu/commission/presscorner/detail/en/ip_23_3491.

17   "Data Ethics Framework," GOV.UK, September 16, 2020, https://www.gov.uk/government/publications/data-ethics-framework.

18   "Data Catalog Vocabulary (DCAT) - Version 3," W3C, March 7, 2023, https://www.w3.org/TR/vocab-dcat-3/.

19   European Union, "Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information (Recast)," June 20, 2019, https://eur-lex.europa.eu/eli/dir/2019/1024/oj.

20   National Open Data Coordinator of the Czech Republic, Department of the Chief Architect of the eGovernment, Ministry of the Interior of the Czech Republic, "Importance of Data Quality of High Value Datasets (HVDs)," February 2020, https://www.data.gv.at/wp-content/uploads/2020/02/CZ-Non-position-paper-Importance-of-data-quality-of-High-Value-Datasets-HVDs-EN-version_fin.pdf.

21   Gaurav Godhwani, "Making India's Budgets Machinable," in *Lives of Data: Essays on Computational Cultures from India*, Theory on Demand 39 (Amsterdam: Institute of Network Cultures, 2020), https://networkcultures.org/blog/publication/lives-of-data-essays-on-computational-cultures-from-india/.

22   "Open Contracting Data Standard v1.1.5 Documentation," Open Contracting Partnership, August 20, 2020, https://standard.open-contracting.org/latest/en/.

23   Charl-Thom Bayer and Keitha Booth, "Open Up Guide for Land Governance, Version 2.0," *Land Portal Foundation*, October 25, 2021, https://landportal.org/node/100820.

24   "Open Up Climate Data: Using Open Data to Advance Climate Action," Open Data Charter, 2019, https://open-data-charter.gitbook.io/open-up-guide-using-open-data-to-advance-climate-a/.

25   "Justice Hub," CivicDataLab, 2021, https://justicehub.in/.

26   Felix Dodds, "Multi-Stakeholder Partnerships: Making Them Work for the Post-2015 Development Agenda," *United Nations Department of Economic and Social Affairs*, 2015.

27   Social Accountability Forum for Action and Research and the Soochna Evum Rozgar Adhikar Abhiyan, "Jan Soochna Portal Breakthrough," SAFAR and SR Abhiyan, March 2021, https://safar-india.org/documents/3.%20Jan%20Soochna%20Primer_ebook.pdf.

28   "J-PAL Builds Partnerships, Tools to Make Government Data More Accessible and Actionable," *Data.Org*, December 10, 2021, https://data.org/stories/jpal/.

वसुधैव कुटुम्बकम्

ONE EARTH • ONE FAMILY • ONE FUTURE