The 2030 Agenda's data challenge

APPROACHES TO ALTERNATIVE AND DIGITAL DATA COLLECTION AND USE

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Global Partnership for Sustainable Development Data

This is a joint publication by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH and the Global Partnership for Sustainable Development Data (the Global Partnership). The analysis was undertaken by Charu Vijayakumar on behalf of the Global Partnership in collaboration with Sabrina García (Sectoral Department) and Thomas Wollnik (Partners For Review) on behalf of GIZ. Gratitude is expressed towards Victoria Kohlrusch for her dedicated support in finalizing this publication.

CONTENTS

EXECUTIVE SUMMARY

The 2030 Agenda for Sustainable Development charts a path to a better world over the next decade. To effectively and efficiently address the breadth and depth of the Sustainable Development Goals (SDGs), governments are increasingly adopting alternative or non-traditional data sources and methods to fill data gaps and complement traditional data. While harnessing the data revolution for official statistics is a developing and evolving process, this report highlights the key non-traditional data sources and methods currently being used and the opportunities, challenges, and considerations associated with incorporating them into SDG monitoring and implementation.

This report is a practical resource that provides brief overviews of topics of interest and provides links to more detailed information. It highlights key areas of consideration for using alternative data sources and approaches. The four key alternative data sources and approaches highlighted are geospatial data, citizen-generated data, privately-held data, and administrative data.

In addition to highlighting key opportunities such as timeliness, coverage and scale, consistency and comparability, quality, and time and cost savings, the report also observes challenges associated with using alternative data such as limited technical capacity, technical infrastructure, data sharing, and privacy and public consent.

The core of the report focuses on collective learnings developed through Global Partnership for Sustainable Development Data (the Global Partnership)-facilitated and broader partner activities within the network. These learnings are structured to highlight the key normative, structural, and process-related considerations for successfully incorporating alternative data into official statistics, which are further illustrated through specific country-level examples. The normative shifts refer to using a multi-stakeholder approach and recognizing data as a public good; structural considerations include governance structures and data infrastructure and interoperability; and process-related considerations are coordination and collaboration mechanisms and standardization of definitions and methodologies.

The report provides seven recommendations to inform the programmatic approach to incorporating alternative data sources and methods for SDG monitoring.
These include:

1. Scope existing technical capacity, infrastructure, and needs;
2. Develop an enabling environment through institutionalization;
3. Facilitate multi-stakeholder engagement and collaboration;
4. Align with national mandates and identify policy implications;
5. Translate global-level legal and policy guidance to address practical national-level needs;
6. Coordinate with existing efforts to amplify results; and
7. Showcase success to build momentum and investment.

INTRODUCTION

The 2030 Agenda for Sustainable Development set 17 ambitious goals, with 169 targets and 232 corresponding indicators reflecting a global consensus and commitment to address the world's most acute and pressing challenges. The achievement of these goals relies heavily on the availability and use of relevant and timely data to understand the gaps, target solutions, and measure progress. Many traditional data sources and methods tend to have strong limitations with respect to data production and data use, with coverage gaps across space, time, and people. However, the digital revolution and new technologies, combined with existing data sources and methods offers the opportunity to fill these gaps and produce and use more robust and timely data for monitoring and decision-making that supports the achievement of the SDGs.[1] Planners and decision-makers need timely and granular data to monitor the SDGs as well as to implement them, which can include overlapping or distinct data sources and methods. For example, the data necessary for a ministry of agriculture to decide when a livestock insurance program should make specific payouts is not necessarily the same data that helps monitor SDG 2.3 – farmer productivity. Alternative data offers opportunities to support both data needs.[2]

This document aims to highlight the landscape of possibilities for key alternative or non-traditional data sources that can be used to implement and monitor the Sustainable Development Goals (SDGs) and the associated opportunities, challenges, and considerations when incorporating them into official statistics. There are many different ways to categorize the universe of non-traditional data types available and being used to implement and monitor the SDGs. For example, the GovLab uses the following four categories: Disclosed personal data, observed personal data, disclosed non-personal data, and observed non-personal data.[3] A UN Statistics Division representative uses two broader categories: transactions data – to capture digital interactions with people that result in large-scale data collection, and sensing data – to include remote sensing, satellites, sensors, job vacancy postings, and other passive observation and monitoring data.[4] A group of researchers present yet another grouping to include commercial data, official sensor networks, citizen-generated data, spatial data infrastructure, and Earth observations.[5] While there are many different ways to categorize the various alternative data sources, for clarity and accessibility to inform practical use, this paper approaches them in four categories: geospatial data, citizen-generated data, privately-held data, and administrative data. While these categories are presented as discrete categories, there are areas of overlap. For example, open mapping[6] can be categorized as both geospatial data, given the mapping aspect, and as citizen-generated data, given the crowd-sourced or citizen-engagement component.

The document provides a brief summary of some key data sources, tools, and methods within each category that serve as useful reference points in understanding the opportunities for using such alternative data. The document also highlights general advantages and disadvantages of using alternative data for SDG monitoring and the key considerations for incorporating such alternative data sources into official statistics for SDG monitoring. This information is a synthesis of publicly available research products and learnings from the Global Partnership's experience in the data for development space.

This is intended to be a user-friendly document that gives the practitioner a quick picture of the topic, in appropriate depth, with practical examples. It is not intended to be an academic or comprehensive synthesis and analysis of all potential alternative or non-traditional data sources.

ALTERNATIVE DATA SOURCES AND APPROACHES FOR MONITORING THE SDGs

Geospatial Data

Geospatial data broadly refers to any data that has a geographic or location component to it and "combines it with attribute information (characteristics of the object, event or phenomena concerned), and often also temporal information (the time or life span at which the location and attributes exist)."[7] There are many ways to collect, synthesize, and analyze such data and therefore, many ways to describe categories of geospatial data including by data source, type, and products. The United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM) has defined 14 Global Fundamental Geospatial Data Themes in support of SDG indicator monitoring[8,9]:

- Global Geodetic Reference Frame[10]
- Addresses
- Buildings and Settlements
- Elevation and Depth
- Functional Areas
- Geographical Names
- Geology and Soils
- Land Cover and Land Use
- Land Parcels
- Physical Infrastructure
- Population Distribution
- Orthoimagery[11]

Geospatial data can be a value add at various points in the data value chain[12] – from production to analysis. The UNSD Open SDG Data Hub[13] provides geospatially referenced data by SDG goal and indicator, including a variety of geospatial data types and sources.

EARTH OBSERVATION (EO) DATA

A subset of geospatial data, Earth observation data is defined as "the gathering of information about planet Earth's physical, chemical, and biological systems using remote sensing technologies such as satellites and aerial sensors supplemented with ground-based observations. It is used to monitor, detect and assess changes in natural and built environments."[14]

EO data can be used to monitor 16 of the 17 SDGs. This table (pg. 4)[15] highlights the targets and specific indicators for which EO data can be used.

Illustrative Practical Use Resources

Type	Resource
EO Data Sources/Infrastructure	Satellite Table below + MODIS Open Data Cube Global Earth Observation System of Systems (GEOSS) Digital Earth Africa Africa Regional Data Cube (ARDC) NASA's Earth Observing Systems Data and Information Systems (EOSDIS) ESA Earth Online Google Earth Engine
Population/Human Settlements Related	GRID3 POPGRID WorldPop Global Rural-Urban Mapping Project (GRUMP) GEO Human Planet Initiative
Agriculture and Water-Related	Synthesis Table of Land Cover Datasets, Page 7 FAO's AQUASTAT GEO Global Agricultural Monitoring (GEOGLAM) Trends.Earth Global Fishing Watch
EO Capacity Building Resources	GeoforAll Group on Earth Observations Inventory of Resources

VOLUNTEERED GEOGRAPHIC INFORMATION (VGI)

Crowd-sourced geospatial data or volunteered geographic information (VGI) is geographic information generated from individuals or about individuals. As mentioned, this is also considered citizen-generated data, further explored below.

Open mapping, primarily through the OpenStreetMap (OSM) project, is a global movement to create free and open geographic data. OSM emphasizes local knowledge and "contributors use aerial imagery, GPS devices, and low-tech field maps to verify that OSM is accurate and up to date."[16]

The open mapping community includes people from all parts of society and across the globe, from experts in geographic technologies to those in their own neighborhoods. It is powerful for people from around the globe to work in the same database. To date, more than 3 million people have registered, and approximately 30,000–40,000 make contributions during a given month. This approach has been very successful at creating maps in under-mapped places, serving critical needs, as in the response to the Nepal and Mexico[17] earthquakes and mapping local schools in informal settlements.[18]

"Open mapping can be a powerful tool to supplement and complement official sources of data for measuring progress against the SDGs, but it is not a cure-all. It is especially useful for assessing 'proportion of population' SDG indicators as they relate to observable characteristics of the built or natural environment, such as:

- **Indicator 6.1.1:** Proportion of population using safely managed drinking water services
- **Indicator 7.1.1:** Proportion of population with access to electricity
- **Indicator 9.1.1:** Proportion of the rural population who live within 2 km of an all-season road
- **Indicator 11.1.1:** Proportion of urban population living in slums, informal settlements, or inadequate housing
- **Indicator 11.2.1:** Proportion of the population that has convenient access to public Transport[19]"

Illustrative Practical Use Resources

- → Open Mapping for SDGs: A practical guide to launching and growing open mapping initiatives at the national and local levels
- → Crowdsourced Geographic Information Use in Government
- → Open Data for Resilience Initiative: Planning an Open Cities Mapping Project

Examples of free satellite data which can be useful for SDG monitoring

Satellite	Spatial resolution	Mission objectives	SDG
Sentinel 1	C-band Radar Strip Map Mode: 80 km swath, 5 x 5 m spatial resolution Interferometric Wide Swath: 250 km swath, 5 x 20 m spatial resolution Extra-Wide Swath Mode: 400 km swath, 20 x 40 m spatial resolution Wave-Mode: 20 x 20 km, 5 x 5 m spatial resolution.	Land monitoring of forests, water, soil and agriculture Emergency mapping support in the event of natural disasters Marine monitoring of the maritime environment Sea ice observations and iceberg monitoring Production of high-resolution ice charts Forecasting ice conditions at sea Mapping oil spills; Sea vessel detection Climate change monitoring.	SDG 2: Agriculture SDG 6: Water SDG 11: human settlements monitoring SDG 15: Forest, Biodiversity, Land degradation
Sentinel 2	13 spectral bands: four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution. The orbital swath width is 290 km	Monitoring agriculture, forests, land-use change, land-cover change; mapping biophysical variables such as leaf chlorophyll content, leaf water content, leaf area index; monitoring coastal and inland waters; risk mapping and disaster mapping	SDG 2: Agriculture SDG 6: Water SDG 11: human settlements monitoring SDG 15: Forest, Biodiversity, Land degradation
Sentinel 3	21 spectral bands 300 meters/1270 km swath	Supporting global land and ocean monitoring services, in particular: sea/land color data and surface temperature; sea surface and land ice topography; coastal zones, inland water and sea ice topography; vegetation products.	SDG 2: Agriculture SDG 14: Coastal eutrophication SDG 15: Forest, Biodiversity, Land degradation
Landsat 7/8	15m / 30m / 100m (panchromatic/multispectral/ thermal)	Provide data continuity with previous Landsat missions Offer 16-day repeat coverage of the Earth Build and periodically refresh a global archive of sunlit, substantially cloud free, land area and coastal images Make data widely and freely available. As of 2008, Landsat data with standard processing are available at no cost Support Government, international, and commercial communities Promote NASA's EOS interdisciplinary research via synergism with other EOS observations – By orbiting in tandem with NASA's Terra satellite to obtain near coincident observations.	SDG 6: Water SDG 11: human settlements monitoring (and multi-temporal comparisons) SDG 15: Forest, Biodiversity, Land degradation

Source: http://ggim.un.org/documents/Report_Global_and_Complementary_Geospatial_Data_for_SDGs.pdf; page 5

Citizen-Generated Data (CGD)

Citizen science or generated data is defined as "data that people or their organizations produce to directly monitor, demand or drive change on issues that affect them."[20]

"Citizens can generate data in many ways, not only by producing new data, but also by enriching and analyzing existing (official) data. This can include to compile formerly unstructured data on a database, to classify, format, annotate, mediate, translate or otherwise engage with data."[21]

CGD includes a wide range of methods from actively collected survey data to passively produced data through air pollution sensors, including overlaps with other data categories presented in this report such as geospatial data via mapping and big data via sensors and tracking devices. CGD also includes social accountability approaches such as citizen surveys and community scorecards. Given the nature of citizen engagement, citizen empowerment is an important byproduct of using CGD, which is particularly relevant for stakeholders working towards the Leave No One Behind (LNOB) agenda[22] through data.

Illustrative Practical Use Resources

- → Choosing and Engaging with Citizen Generated Data – A guide to help understand if CGD is suitable for a proposed project
- → Advancing Sustainability Together? Citizen-generated data and the Sustainable Development Goals – An overview of CGD methods and how each relates to implementation, monitoring, and achieving the SDGs
- → Citizen science and the United Nations Sustainable Development Goals – Guidance on global, national, and local implementation of citizen science projects
- → Acting Locally, Monitoring Globally – How to link citizen-generated data to SDG monitoring
- → WeObserve – an ecosystem of citizen observatories for environmental monitoring
- → SciStarter – a search engine for citizen science projects

Examples of CGD Methods and their Scalability

CDG Examples	Type of Method	Scaling Enabled
Classifying/Annotating: HOT web editor	Web-based method	Small groups may identify many data points by classifying image content. Contributors from around the world can contribute (large-scale) datasets, that are already produced, and can be further analyzed.
On-site observation: Social auditing, HOT field survey **Sample collection:** Water sample collection	Field-based method	Small or large groups of people are required, depending on the size of the territory.
Automated, stationary sensing: Weather observation stations	Tech-aided method	Real-time and longitudinal sensing in different intervals possible in fixed location.
Automated, mobile sensing: Sensor technologies implemented in cars, and other consumer devices	Tech-aided method	Real-time and longitudinal sensing in different intervals possible in location where people use consumer devices.

Source: Choosing and Engaging with Citizen Generated Data, page 29

Privately-held Data

Privately-held data refers to data that is commercially or privately owned and not publicly available. This includes a variety of data types and sources such as mobile data, financial transactions data, and commercial sample surveys or polling data. Often, privately-held data overlaps with big data such as transportation data and social media data. There are also multiple categories of big data that overlap with data categories already covered such as crowdsourcing data and remote sensing data, that can be privately-held such as monitoring or reporting applications and commercial drone data. Regardless of the data source or type, the use of privately-held data requires the development of public-private partnerships.

Accessibility of Data to the Public

MORE RESTRICTED

1. Data provider mines own data for insights to share publicly
2. Data provider brings external researchers into own trusted network to analyze data and report indings publicly
3. Data provider works with other trusted data providers to form data cooperatives, aggregating diferent sources
4. Data provider shares data with trusted external researchers
5. Data provider allows for direct public access to datasets (often through anonymized data, data samples, and/or data tools)

MORE ACCESS

Source: https://www.urban.org/research/publication/data-philanthropy-unlocking-power-private-data-public-good/view/full_report, page 9

BIG DATA

Data Source Topic	National	International	Project Topic	National	International
Web scraping	22	4	Prices	22	4
Scanner	20	1	Population/migration	10	4
Mobile phone/CDR	14	18	Transport/mobility	9	11
Social media	8	23	Geographical/spatial	8	7
Satellite imagery	6	7	Labour market	7	2
Smart meter	5	1	Agriculture/Land use	6	4
Credit card	3	1	Tourism	5	1
Road sensor	5		Health/disease	4	7
Health records	5	2	Energy/Environment	4	6
Ship identification	2		Crime/Corruption	2	4
Criminal record	1	2	Disaster risk reduction		8
Other	20	31	Other	31	24
Total	**111**	**90**	**Total**	**109**	**91**

Source: The Big (data) Bang: Opportunities and Challenges for Compiling SDG Indicators

Illustrative Practical Use Resources

- → Big Data for Development Research Network – A southern-led partnership with the objective of developing policy-relevant research on big data for development
- → Data Philanthropy: Unlocking the Power of Private Data for Public Good – a report outlining potential pathways and use cases for using privately-held data for good
- → UN Global Pulse – UN initiative to accelerate discovery, development, and scaled adoption of big data innovation for sustainable development and humanitarian action through its network of Pulse Labs
- → UN Global Pulse and BBVA Partnership – and new project measuring economic resilience to disasters with financial data

MOBILE DATA

A subset of privately-held data, mobile data can be used for a variety of SDG-related statistics including tourism and events, population and migration, commuting and traffic flows, employment, health and education, and managing disasters and emergency aid statistics and information.

Comparison of Mobile Data Sources

Data Sources	Accessibility	Major Features	Adequacy for Statistics
Cell activity	Easy, standard outlet from the operator's system	Phone use intensity at the cell level	Low
Call detail records (CDR)	Privacy problem, software development needed	Call activities	High
Probes	Privacy problem, major software development needed	Call activities and handover logs; personal features from the operator	High
Mobile positioning system	Easy for small samples, usually requires opt-in	Accurate positioning, custom frequencies; questionnaire with the respondent	High, can be used for benchmarking

Source: Handbook on the use of Mobile Phone Data for Official Statistics, page 26

Illustrative Practical Use Resources

- → Mobile Big Data Solution for a Better Future – Report on mobile big data's potential through case illustrations, highlighting principles that apply to using mobile data for social good, and offering a framework for showcasing benefits
- → GSMA Big Data for Social Good Initiative
- → UN Big Data Global Working Group Task Team Activities on Mobile Phone Data
- → Handbook on the use of Mobile Phone Data for Official Statistics – A draft guidance document developed by the UN Global Working Group on Big Data for Official Statistics
- → Flowkit – a suite of software tools designed to enable access and analysis of mobile data for humanitarian and development use cases by DIAL, GSMA, and Flowminder

Administrative Data

Administrative data refers to the routine data collected by governments and service providers in the course of their day-to-day business. In addition to new data sources, such as those described above, administrative data has emerged as a key alternate data source for implementing and monitoring the SDGs. Administrative data includes registers such as population registers, business registers and real-estate registers, service delivery information, tax records, crime reports, and much more.

The Asia and Pacific SDG Progress Report for 2019 notes that "nearly half of the SDG indicators from administrative sources have sufficient data in the Asia-Pacific region, while only 32 percent of the SDG indicators coming from surveys have sufficient data."[23]

Illustrative Practical Use Resources

- → Use of Administrative Data for Official Statistics: The Global Perspective – a UN Department of Economic and Social Affairs (DESA) workshop presentation
- → Advancing Administrative Sources of Data for Monitoring Gender-specific Sustainable Development Goals in Africa – A study presenting best practices in the use of gender statistics generated from administrative data for policy interventions
- → An Administrative Data Maturity Model – building national administrative data capacity to produce results for children
- → Compendium of Good Practices in Linking Civil Registration and Vital Statistics (CRVS) and Identify Management Systems – synthesis of case studies and good practices

OPPORTUNITIES AND CHALLENGES OF ACCESSING AND USING ALTERNATIVE DATA

Opportunities

Timeliness – Unlike traditional survey data, most alternative data types can be generated and accessed on a frequent basis or even in near real-time to provide insights when needed, based on information that is current or up-to-date. This is not only useful for monitoring and policymaking in short, medium and long-term planning, but also for implementation. For example, emergency and natural disaster management and response[24], education services targeting children with disabilities, and clinic diagnostics to avoid drug stockouts.

Coverage and Scale – Many alternative data sources offer better coverage and ability to scale over space and time. For example, Earth observation and mobile data provides coverage at local, national, regional, and global levels, including the ability to generate data on difficult to reach places. Similarly, these data sources can provide long time series and continuity, allowing robust assessments of change over time.

- Caveat: Alternative data sources can also pose coverage and scale limitations such as privileging people with access to digital technology creating potentially biased samples that leave out the poorest populations. Similarly, time series data can be dependent on the lifespan and scope of companies' business model and longevity.

Consistency and Comparability – given that many of the alternative data sources already have or are in the process of developing methodologies that are standardized per the data type to some degree, it allows for comparison across areas, countries, and time because formats and approaches are similar or standardized.

- Caveat: It is possible that privately-held data might not be consistent, particularly if a company goes out of business or a data sharing agreement is rescinded.

Quality - the digital nature of many of the alternative data sources and types means that a certain degree of quality assurance can be built into the data collection and often synthesis phase, minimizing human error.

- Caveat: This is largely dependent on the development of robust methodologies throughout the data value chain. For example, the quality of insights can be significantly affected by algorithmic biases.

Time and Cost Savings – Although the use of alternative data can require initial short-term investments in time and funding, they can provide both time and cost savings in the long run. For example, using satellite and mobile data to supplement the population census can reduce personnel costs by lowering the number of enumerators necessary and the amount of time it would have taken to administer door-to-door surveys.

Challenges

Limited Technical Capacity – with the exception of administrative data, many of the alternative data sources and types are new to the official statistics community or involve the use of unfamiliar technology and methods. Accessing and using these data types involves a certain degree of introducing and understanding the technology, its benefits, and capacity-building along the data value chain and among the different stakeholders involved in the process. It also requires capacity-building in the legal and financial aspects of engaging with these data. Effective use of alternative data also depends on high quality, strong capacity in traditional data systems. Use of many alternative data sources and types requires traditional data for ground truthing and training models, therefore the strength of the traditional data directly affects the quality of insights from alternative sources.

Technical Infrastructure – in order to access and analyze big data, such as satellite and mobile data, it is necessary to first develop and put in place the necessary infrastructure with the appropriate formats, computational capacity, and safeguards to enable data processing. For example, 10 percent of the world's population is not covered by mobile broadband, and for the 90 percent who live in "covered" areas, mobile Internet connectivity is not equitable;" just over 40 percent of the LMIC (low and middle income countries) population (around 2.6 billion people) are connected, compared to almost 75 percent of the population in high-income countries."[25] Similarly, the use of satellite data requires access, such as cloud credits and sufficient computer processing capacity; and the data interoperability for all data types including administrative data requires adequate digitization and appropriate formatting.

Data Sharing – Using alternative data largely relies on accessing it. While more data is becoming open, access to most data requires some form of data sharing agreement to ensure appropriate use and address additional concerns such as confidentiality. Sometimes, even when legal mandates are in place, data is not shared in practice, particularly in relation to administrative data, which requires more trust building among the different stakeholders involved. In other instances, there is a lack of data protection and privacy regulations to guide potential data sharing practices. Therefore, a fit-for-purpose legal framework is an important element of the data value chain.

Privacy and Public Consent – Ensuring data privacy is paramount to not only accessing data, as it relates to data-sharing, but also to securing public consent for the use of such data. This is particularly true for privately-held and administrative data. Agreeing to the parameters of what constitutes sufficient data protection can often be a lengthy process, especially in contexts where there are limited or no legal guidelines in place. In addition to developing legal and regulatory frameworks, this could require public sensitization efforts.

Table of illustrative challenges by stakeholder types:

Challenges Integrating Non-Official Data Sources into Official Statistics			
National Government	**Local Government**	**Private Sector**	**Civil Society Organizations**
Lack of national level coverage	Lack of standards and guidelines on data	Lack of data protection and privacy regulations guiding data sharing	Generally small sample sizes
Often inconsistent frequency of data production and limited documentation of methodologies	Uncoordinated data collection resulting in duplication	Lack of centralized data verification processes to ensure authenticity	Lack of data production guidelines
Lack of data non-official data and effective coordination with civil society	Inadequate documentation of data production and use methodology	Integration of data	Negative connotation of the term „unofficial" often interpreted as useless data
Need to additional capacity (technical, legal, and financial) to effectively collect, validate, analyze, and use different types of data			

Source: Ghana and Kenya Peer-to-Peer Learning Exchange on SDG Monitoring Report, page 6

INCORPORATING ALTERNATIVE DATA IN SDG MONITORING

Normative Shifts

Multi-stakeholder Approach – incorporating alternative data into official SDG monitoring requires statistical offices to engage in a multi-stakeholder approach where the national statistical offices are not necessarily the primary data producers or holders. This requires collaboration within the whole of government, as well as between state and non-state actors and the recognition that data producers, holders, and users span across the various types of stakeholders and will need to work together to achieve shared objectives. This also requires defining value proposition across the various stakeholders to ensure all stakeholders are adequately invested.

Data as a Public Good – regarding data as a public good, rather than private property will drive governments, and perhaps eventually companies, to operate on the default that data should be open, unless there is a reason not to. Open data is data that is freely available online for anyone to use and republish for any purpose and is recognized as a resource with high economic and social value. Using and promoting open data will help governments, civil society, and private sector stakeholders to build trust and transparency, and access and use data effectively towards the SDGs. This will require clear legal and licensing frameworks, protection of privacy rights, and clear metadata.[26]

Structural Considerations

Governance Structures – identifying and defining a governance structure for the incorporation of alternate data into SDG monitoring is important. This can take many different forms based on the nature of the collaboration and the stakeholders involved. The GovLab's Data Collaboratives[27] describe the various types of collaborative structures that can be formed based on the type of data and outcomes involved. Identifying the institutions and focal points, describing the governance mechanisms, and highlighting mandates and incentives for all stakeholders involved helps promote adoption and use more effectively. This is particularly important for building political buy-in and understanding the role of the national statistical office within these partnerships. The Africa Regional Data Cube Lessons Learned report highlights how incorporating alternative data involves the development of governance structures that include both political and technical engagement, and importance of embedding them within existing structures to ensure sustainability. It is also important for governance structures to articulate how data privacy will be protected and managed. These processes are important in increasing mutual understanding and subsequently building trust among a diverse set of stakeholders.

Data Infrastructure and Interoperability – incorporating new data or strengthening existing data systems such as administrative data to allow seamless data inputs, access, and use requires data interoperability, particularly across multiple stakeholders. The "Data Interoperability: A Practitioner's Guide to Joining up Data in the Development Sector" explains that interoperability is both a characteristic of good quality data and a concept that can be used to help frame data management policies and includes the technology layer, data and format layers, human layer, and institutional and organizational layers. Building a robust and sustainable data infrastructure includes strengthening technical infrastructure such as digitization, operating systems, and computing power as well harnessing or developing appropriate legal frameworks that guide the production, access, use, and sharing of data.

Process-related considerations

Collaboration and Coordination Mechanisms – in tandem with developing a governance structure, it is important to develop and define collaboration and coordination mechanisms and the various stages of the collaborative process to incorporate alternative data. Some specific examples of practices include delineating expectations in MOUs and agreeing on communication mechanisms. For example, WhatsApp groups are more effective in certain contexts than in-person meetings, and delineating distinctions between technical and political leadership can be helpful, if applicable. The Contracts for Data Collaboration[28] project shares examples of data-sharing agreements. Below is an example of a collaborative process map for incorporating CGD data.

Standardizing Definitions and Methodologies – Statistical offices have clear and established definitions and methods for their data collection and analysis processes. When incorporating new and supplemental data, it is important to ensure that definitions and methodologies have been considered and are comparable. For example, the definition of a "household" can be different in the census versus a CGD initiative. The UN National Quality Assurance Frameworks Manual for Official Statistics offers guidance on data sharing, legal frameworks, and policies.

What Process Should Producers of CGD follow to make their data official?

BEFORE DATA COLLECTION

CSOs should liaise with NSOs to inform them about the CGD project including the scope of the exercise

Develop the metadata, guidelines for data collecion, definiions and methodology, and recruit enumerators

NSOs can conduct enumerator trainings on data collection

This can be done through a Trainer of Trainers exercise to make it sustainable

NSOs can also help review data collecion tools, methodology, and sample design

DURING DATA COLLECTION

Utilize quality assurance guidelines shared by NSOs to ensure that the right data is collected and validated

Proper documentaion of the process/methodology

Proper management of the data collected including supervising enumerators

AFTER DATA COLLECTION

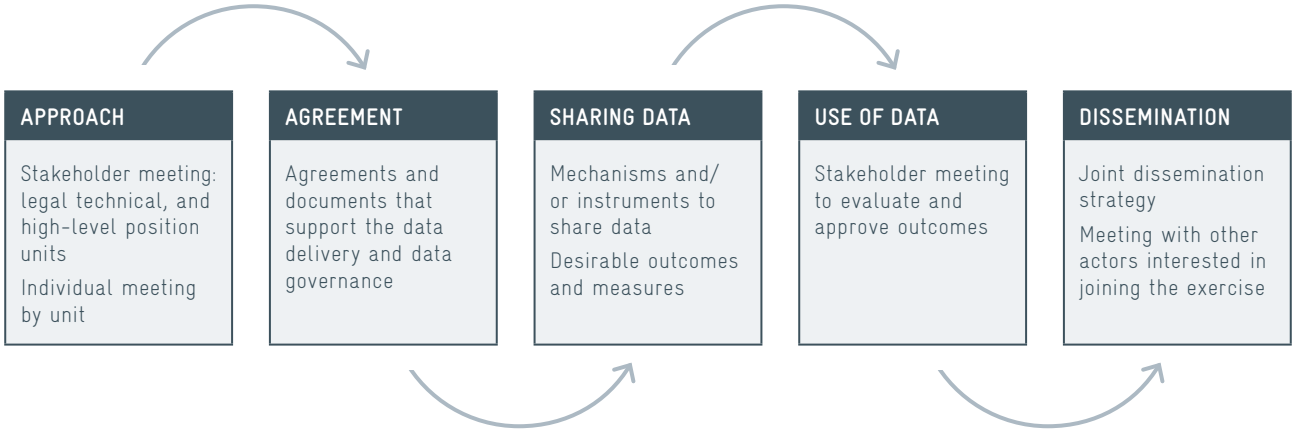
Submit methodology report to staisics office - this allows the data to undergo quality assurance tests to ensure it adheres to the required steps

Effecive dissemination of results and disclosure. Non-state actors can find more innovative ways to communicate the findings

Data may now be recognized as official and can be made available on government repository if it fulfills the requirements of the NSOs' statistical act or other internationally agreed standards (but data does not necessarily need to be published on government repository to be official)

Source: Ghana and Kenya Peer-to-Peer Learning Exchange on SDG Monitoring Report, page 9

Example of a General Data-sharing Flow Chart



Source: GPSDD-Cepei Facilitated LAC-Africa Regional Administrative Data Peer Exchange

Country-level Examples

Using Earth Observation Data to Save the Mangrove Ecosystem in Sierra Leone

The eroding shorelines in Sierra Leone have led to landslides destroying homes and lives. To address this, the Environmental Protection Agency (EPA) decided to use satellite data to better understand how mangroves, which protect and provide stability to the coastline, have been changing over time. As one of the five pilot countries for the Africa Regional Data Cube, Sierra Leone's EPA was able to access analysis-ready satellite data to map mangroves around Freetown and Yelibuya Island from 2010 to 2016. Access to this data, along with technical training, enabled the team to identify a high net loss in mangroves in Yelibuya Island, insights that aligned with those from the Global Mangrove Watch (GMW). This validation has built confidence in the ability to use the ARDC for additional analysis and has informed the EPA's drafting of an integrated coastal management policy and will contribute to SDG 14.5 on coastal mangroves. In particular, it has informed the EPA's plan to propose mangrove conservation of Aberdeen creek, where it plans to produce a buffer zone to protect the mangroves.[29]

Mapping SDG-Relevant Facilities and Their Accessibility in Uganda with OpenStreetMaps

Through the use of open source technical tools such as OpenStreetMaps and community-based methodology, the Humanitarian OpenStreetMaps Team produced comprehensive data on infrastructure and services where refugees and host communities reside in Uganda. Through this project, more than 1.5 million buildings and 36,000 meters of roads were digitized and over 4,000 facilities and services related to SDGs were mapped for the first time in Arua and Yumbe, home to the majority of the refugee population in the country and where previously the lack of coherent data on camp and non-camp locations and services hindered timely and informed decisions to support refugee response. This mapping effort has led to the development of base layer maps, which are now available through OpenStreetMap and can be used to guide government agencies and organizations in designing and implementing interventions.[30]

Strengthening CRVS Data and Systems to Combat Child Marriages in Bangladesh

Bangladesh has one of the highest rates of child marriage in the world despite having legal age of marriage. This has been traced to the root issue of the lack of adequate birth registration. For example, an estimated two-thirds of girls are married before reaching 18, largely because they are not registered and their ages can be easily falsified. In order to address this, the Bangladeshi government has passed the 2004 Birth and Death Registration Act, which went into effect in 2006 and mandated use of birth certificates to access services, including legal registration of marriages. An online birth registration system implemented in 2009 has simplified the process and has contributed to the increase in birth registration in the country from 12 percent in 2006 to 31 percent in 2011.[31]

Humanitarian Response to the 2017 Mexico Earthquake Using Satellite and Mapping Data

In September 2017, a series of three earthquakes hit Mexico, triggering the need for rapid disaster response and relief efforts. Hours after the September 19th earthquake, the General Director of Open Data in the Coordination of National Digital Strategy of the Office of the President of Mexico put out a call for "high quality satellite images" to help the Mexican government map and understand the degree of damage. In response, members of the Global Partnership for Sustainable Development Data network collaborated with each other and the Mexican government to activate the International Charter- Space and Major Disasters, an international collaboration between owners and operations of EO missions to provide rapid access to satellite data to assist rescue authorities in the event of disasters; access and use satellite data from ChinaGEOSS; and connect with Humanitarian OpenStreetMap to provide mapping services. As a result of this experience, the Mexican government prioritized coordination and collaboration around use of open data and technology in disaster response and established a data portal to publish and share strategic data in open formats and real time.[32]

Incorporating Mobile Phone Data into Official Statistics in Ghana

The Ghana Statistical Service (GSS), Vodafone Ghana, and Flowminder Foundation have partnered to produce official statistics using de-identified (anonymized) mobile phone data to strengthen humanitarian and development decision-making. A data-sharing agreement was established with GSS to incorporate insights from de-identified mobile data provided by Vodafone Ghana into the production of official national statistics. "In order for such an agreement to be viable, it was necessary for GSS and Flowminder to obtain approval from the Data Protection Commission, the national regulatory body in Ghana to ensure compliance with Ghanaian data protection legislation. Proof of compliance with U.K. legislation and European Union's GDPR was also required. Absolute clarity about the roles and responsibilities of each party with respect to data governance, as well as agreement about the technical aspects of data security, were also key elements that lead to a contract being successfully finalized."[33]

The data, in combination with more traditional sources such as household surveys, will provide useful information on the mobility and characteristics of the population, which can be used for a wide range of applications including health, disaster preparedness, and transport planning.[34]

KEY STAKEHOLDERS TO CONSIDER WHEN INCORPORATING ALTERNATIVE DATA

The table below provides an illustrative list of key stakeholders to consider including in projects, activities, or partnerships involving the incorporation of alternative data into official statistics. Given the objective of official statistics, the national statistics office should always be considered a key stakeholder as they play the role of lead coordinator, aggregator, standard setters, and quality assurance, among others.

Alternative Data	Stakeholders to Consider
Geospatial Data	– Space Agencies (national, international with open data) – Private companies (i.e. Digital Globe, Planet, etc.) – Academia (i.e. CEISIN) – Domestic geospatial institutes
Citizen-generated Data	– Civil society/ social accountability organizations – Grassroots community activities – Citizen science groups – Open mapping community
Privately-held Data	– Private Data-holders – Mobile Network Operators; Banks; Social Media Companies; Polling Companies, etc. – Regulatory bodies – International Telecommunication Union (ITU); domestic regulators – Industry associations (GSMA) – Civil society and academic groups that develop the operational models for unlocking privately held data
Administrative Data	– Relevant ministry, department or agency that collects and may analyze the data – Frontline service delivery partners – including civil society and private sector service providers (i.e. clinics, schools, utility companies etc.)

LESSONS LEARNED AND RECOMMENDATIONS

Scope existing technical capacity, infrastructure and needs.

In order to understand the most effective investments in technical infrastructure and capacity building efforts, it is important to first understand the existing landscape. For example, for the use of satellite data, it is important to understand the extent of Internet accessibility for users, the level of spatial resolution necessary, and level of user GIS skills, among others. This will help inform mitigation strategies and design activities to best fit needs.

Develop an enabling environment through institutionalization.

Institutionalizing the alternative data activity through governance structures is important to ensure buy-in and sustainability. It is helpful to approach institutionalization from both political and technical perspectives. Doing so effectively often involves identifying and supporting institutional and individual champions within institutions to drive the process through continuous engagement and advocacy to build ownership.

Facilitate multi-stakeholder engagement and collaboration.

Bringing together stakeholders from a variety of sectors (i.e. government, private-sector, civil society, academia, etc.) and institutions (i.e. agencies within government) to collaboratively discuss challenges and identify solutions helps build mutual understanding and trust. This also ensures that solutions are not only approached from a technical perspective but are designed to be holistic and sustainable.

Align with national mandates and identify policy implications.

It order to align interests, encourage time and funding investments, and encourage use, it is important to identify and leverage the national and institutional mandates and development priorities within country to facilitate operationalization. In addition, it is important to identify specific questions the alternative data can help answer or specify the policy issues it can help address. Identifying the policy implications in advance of data analysis and use can help users develop more effective ways of communicating results to a policy audience.

Translate global-level legal and policy guidance to address practical national-level needs.

While global-level guidance and frameworks exist, they are often not leveraged as accessible or practical resources at the country level. It is important to consider how global guidance can support country-level application. In addition, while legal guidance and frameworks help share data, it is also necessary to address changes in perceptions and attitudes around data sharing through multi-stakeholder coordination and building awareness of mutual benefits.

Coordinate with existing efforts to amplify results.

Given the need for, interest in, and availability of various alternative data, there are many activities already ongoing or being conceptualized by a range of stakeholders at the national, regional, and global levels. In order to reduce the burden on government agencies and statistical offices, avoid re-creation, and support interoperability, it is important to coordinate and collaborate with existing or developing initiatives. This can help consolidate available resources for users and reduce reliance on proprietary platforms and approaches.

Showcase successes to build momentum and investment.

In order to continue building high-level buy-in and support for users to invest time and resources into using the alternative data sources, it is important to showcase the benefits of accessing and using those data. This can be done by documenting and showcasing results at national and international venues and convenings and through technical publications as well as through more general, mainstream platforms and outlets. It is important to increase visibility of utility to encourage increased use.

REFERENCES

1 https://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf

2 http://www.data4sdgs.org/index.php/resources/agriculture-data-shaping-policy-and-changing-lives-kenya-tanzania

3 https://datacollaboratives.org/static/files/data-collaboratives-intro.pdf; page 11

4 https://unstats.un.org/sdgs/files/meetings/vnr-workshop-dec2019/5.2-UNSD.pdf; page 6

5 https://www.nature.com/articles/s41893-019-0390-3.pdf; page 3

6 "Open mapping is a global movement to create free and open geographic data. Beyond open data, it is a broad community partnership to collaboratively create a critical data resource for monitoring and meeting the Sustainable Development Goals (SDGs) by making available fundamental, detailed, and timely information on where things are in our world." - http://documents.worldbank.org/curated/en/967511563521548873/pdf/Open-Mapping-for-the-SDGs-A-Practical-Guide-to-Launching-and-Growing-Open-Mapping-Initiatives-at-the-National-and-Local-Levels.pdf; page 5

7 https://www.sciencedirect.com/topics/computer-science/geospatial-data

8 https://undesa.maps.arcgis.com/apps/Cascade/index.html?appid=4741ad51ff7a463d833d18cbcec29fff

9 http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/Fundamental_Data_Publication.pdf

10 "The Global Geodetic Reference Frame (GGRF) is a generic term describing the framework which allows users to precisely determine and express locations on the Earth, as well as to quantify changes of the Earth in space and time." – (http://ggim.un.org/knowledgebase/KnowledgebaseCategory37.aspx). For more information on GGRF, refer to: http://www.unggrf.org/

11 "An orthoimage is a uniform-scale image where corrections have been made for feature displacement such as building tilt and for scale variations caused by terrain relief, sensor geometry, and camera tilt." - https://www.usgs.gov/centers/eros/science/usgs-eros-archive-aerial-photography-high-resolution-orthoimagery-hro?qt-science_center_objects=0#qt-science_center_objects

12 https://opendatawatch.com/reference/the-data-value-chain-executive-summary/

13 http://unstats-undesa.opendata.arcgis.com/

14 https://www.rezatec.com/data-platform/earth-observation-remote-sensing-geospatial-data/;
For more information on what EO data comprises of, refer to: https://www.earthobservations.org/g_faq.html

15 https://www.earthobservations.org/documents/publications/201704_geo_unggim_4pager.pdf

16 https://www.openstreetmap.org/about

17 http://www.data4sdgs.org/resources/september-2017-mexico-earthquake-response

18 http://www.data4sdgs.org/sites/default/files/services_files/OpenMappingfortheSDGsGuide.pdf

19 http://documents.worldbank.org/curated/en/967511563521548873/pdf/Open-Mapping-for-the-SDGs-A-Practical-Guide-to-Launching-and-Growing-Open-Mapping-Initiatives-at-the-National-and-Local-Levels.pdf, page 7

20 http://civicus.org/images/ER%20cgd_brief.pdf

21 http://www.data4sdgs.org/sites/default/files/services_files/Choosing%20and%20Engaging%20with%20CGD_The%20Guide_0.pdf, page 10

22 https://sustainabledevelopment.un.org/content/documents/2754713_July_PM_2._Leaving_no_one_behind_Summary_from_UN_Committee_for_Development_Policy.pdf

23 https://www.unescap.org/sites/default/files/Part_III_Data_source_gaps_Asia-Pacific_SDG_Progress_Report2019.pdf

24 https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf

25 https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/07/GSMA-State-of-Mobile-Internet-Connectivity-Report-2019.pdf

26 http://pubdocs.worldbank.org/en/741081441230716917/Open-Data-for-Sustainable-development-PN-FINAL-ONLINE-September1.pdf

27 https://datacollaboratives.org/

28 https://contractsfordatacollaboration.org/

29 http://www.data4sdgs.org/news/using-earth-observation-data-save-mangrove-ecosystem-sierra-leone

30 https://www.hotosm.org/projects/bridging-data-gaps-mapping-refugee-contexts-in-east-africa/

31 http://www.data4sdgs.org/sites/default/files/2018-12/CRVS%20Case%20Study.pdf

32 http://www.data4sdgs.org/sites/default/files/services_files/CS3_Mexico%20Earthquake%20Response_FINAL.pdf

33 https://digitalimpactalliance.org/wp-content/uploads/2019/02/FlowKit_UnlockingthePowerofMobileData.pdf, page 22

34 https://www.graphic.com.gh/business/business-news/3-bodies-partner-to-promote-big-data.html

Published by:
Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

Dag-Hammarskjöld-Weg 1 - 5
65760 Eschborn, Germany
T +49 61 96 79-0
F +49 61 96 79-11 15

E agenda2030datachallenge@giz.de
I www.giz.de

Author:
Charu Vijayakumar (Global Partnership for Sustainable Development Data)

Responsible:
Sabrina García

Layout:
www.kromativ.de, Berlin

Photo credits/sources:
Cover: pixabay.com
page 6: unsplash.com@mauromora
page 10: pixabay.com
page 13: unsplash.com@ahmetsali
page 20: unsplash.com@chuttersnap

Eschborn 2020

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

Friedrich-Ebert-Allee 36 + 40
53113 Bonn, Germany
T +49 228 44 60-0
F +49 228 44 60-17 66

Dag-Hammarskjöld-Weg 1 – 5
65760 Eschborn, Germany
T +49 61 96 79-0
F +49 61 96 79-11 15

E info@giz.de
I www.giz.de