Open Data Capacity Development
Module 5: Citizen-generated data (CGD) and open data
Training Syllabus

This training module was developed under the guidance of the United Nations Statistical Division as part of the Data4Now Initiative.

Introduction

This module on Citizen Data is part of a course on Open Data practices in official statistics. The course is intended for National Statistical Offices and other parts of the national data system. It focuses on data dissemination and communication tools to ensure data are fully and effectively utilized, including by applying Open Data principles, making data and statistics available to users in more accessible and readable formats. The course is based on a data value chain approach, where users' needs are considered from the conceptualization and design of the data collection phase to the data uptake and utilization. In addition, as NSOs find themselves at the center of increasingly complex systems, with multiple data producers within and outside the official statistical system, the course also addresses issues related to metadata, data exchange and interoperability so that NSOs can enable effective data exchange within the system and play a stronger data stewardship role.

Learning outcomes

The expected learning outcomes for this module include:

- Becoming familiar with the recent developments in the area of citizen data.
- Understanding how citizen data can contribute to official statistics or be used outside the official statistical system.
- Becoming familiar with the Copenhagen Framework on Citizen Data, its roadmap for implementation and the role of the Collaborative on Citizen Data.
- Understanding how citizen data can be used to meet users' needs on specific topics that might not necessarily be covered by foundational data collection programmes.

Note to trainer: Depending on the pace of the trainer and trainees, it is expected that training for this module can be delivered in 2-3 hours, including exercises and group discussions.

Contents

Why citizen data?

Citizen engagement throughout the data value chain[1] is critical to overcoming many of the data challenges we face today. Citizen generated data can enhance data availability, timeliness, quality, openness, and inclusiveness; they empower citizens to be agents of change; and they provide information on population groups that remain invisible in other data sources.

Citizen data can address the serious shortcomings in covering difficult-to-reach populations and give a voice to those who are underrepresented because of their vulnerabilities or biases and misconceptions in the design and implementation of data collection programs. They are also an opportunity for citizens and communities to be involved in data production, gain some ownership of the data, and influence how data are used to address their needs, resulting in a more open and inclusive data governance system and greater data impact.

Citizen contributions to data are particularly relevant to the '*leave no one behind*' principle of the 2030 Agenda for Sustainable Development and more broadly can support policymaking and actions aimed at reducing inequalities.

Engagement of citizens can be initiated by community leaders, civil society organizations (CSOs), academia or citizens and communities themselves, national statistical offices (NSOs), other state agencies within national statistical systems (NSSs), national human rights institutions (NHRIs), and can take place at different stages of the data value chain.

For instance, through the SDG process, CSOs take part in national and international consultations on statistical methods, choice of indicators and other indicator requirements, engaging as experts in their respective domains. Some NSOs use data generated by citizens for SDG reporting or to complement SDG reporting with additional information. In other cases, CSOs and citizens are leading data collection efforts in the communities they serve focusing on policy issues that matter to them or the population groups they serve, such as the needs of marginalized populations and other broader concerns related for instance to the environment, poverty, health, or education.

Citizen data are produced through a variety of tools, including surveys, social media, community mapping, and qualitative methods such as community dialogues, scorecards, focus group and key informant interviews. These data can help to improve the documentation of lived experiences of marginalized populations, identifying drivers of vulnerability, and monitoring the level of recognition and implementation of their rights.

---

[1] The data value chain describes the evolution of data from collection to analysis, dissemination, and the final impact of data on decision making.

It is also important to note that advances in citizen data are happening within a broader context of transformation of national statistical systems. NSOs are increasingly focusing on building inclusive data systems by implementing new practices in traditional data collection programmes, such as participatory data collection and inclusion. New data sources are being used and integrated with traditional ones, such as from social media, cell phone records, earth observations, administrative data, geospatial data, and other sources.

In the context of data stewardship, NSOs are expected to operationalize data equality and inclusion within data governance. Engaging actively with civil society while maintaining the adherence to quality standards, the protection of privacy and confidentiality helps NSOs to perform the new role as the "data steward" within the national data ecosystem.

What are citizen generated data and how can they be used?

Definition of citizen generated data

The term citizen generated data or CGD has been used to describe different approaches to data production and initiatives by civil society organizations, communities, and citizens. Often called citizen data, different terms have been used over the last few years to identify diverse types of data produced with the involvement of citizens, either through organized platforms or community-based and participatory approaches. These include citizen science, community science, citizen-generated data, crowdsourcing, community-based monitoring, participatory mapping, community-driven or community-generated data.

There is still no globally agreed definition of citizen generated data and various organizations still use different approaches and language to describe citizen data. The CIVICUS Datashift programme refers to citizen generated data as data that "people or their organizations produce to directly monitor, demand, or drive change on issues that affect them. It is actively given by citizens, providing direct representations of their perspectives and an alternative to datasets collected by governments or international institutions." [2] The World Bank also uses the CIVICUS definitions in its 2021 World Development Report where CGD are described as data that people or their organizations produce to directly monitor, demand, or drive change on issues that

[2] https://civicus.org/images/ER%20cgd_brief.pdf

affect them, often to fill gaps in public and private sector data or when the accuracy of existing data are in question.[3]

Other definitions are broader and include "data that individuals consciously generate and that are openly available for use in the public domain." In that context, citizen-generated data are a subset of user-generated data, where user-generated data are described as "data from individual users that are either generated consciously – for example, Facebook posts, Google searches, Airbnb ratings – or unconsciously – such as geodata from smart phones and cookies". (Meijer and Potjer, 2018)

CGD as defined in the Copenhagen Framework for Citizen Data

Recently, a definition was proposed in the context of the development of the new Copenhagen Framework for Citizen Data[4]. The operational definition is shown in box 1 and identifies three key elements defining citizen data: "initiate"," sufficiently engaged," and "in the design and/or collection stages of the data value chain" corresponding to the three key characteristics of citizen data discussed below.

Box 1. Operational definition of citizen data

Citizen Data are defined as data originating from initiatives guided by a set of key principles, where citizens either initiate or are at the minimum sufficiently engaged in the design or collection stages of the data value chain, whether or not these data are integrated into official statistics.

The operational definition recognizes that citizens' role in design and collection will differ for different types of data. For example, community-driven data typically demand a higher level of citizen engagement in both the design and collection stages. The design of citizen science data is often spearheaded by researchers or the scientific community, while citizens may participate in formulating research questions, collecting, or analyzing data.

[3] World Development Report. 2021. *Data for Better Lives.* Washington, D.C.: World Bank. https://www.worldbank.org/en/publication/wdr2021

[4] Collaborative on Citizen Data, Citizen data: a conceptual framework, UN Statistical Commission 2024 at https://unstats.un.org/UNSDWebsite/statcom/session_55/documents/BG-4c-CGD_Framework-E.pdf

Using this definition, citizen data do not include the following:

- Citizens' contributions to regular data collection processes conducted by national statistical offices or other state agencies, such as censuses or surveys, which engage citizens with the main purpose of increasing the response rate or making data more inclusive, are not considered in this framework. Similarly, citizens' contributions only to the analysis, dissemination, communication, and use of data produced by NSOs and other non-citizen actors where citizens did not play a role in the design and production are also not included in this framework.
- Any data consciously or unconsciously generated by citizens in social platforms, apps or websites collecting ratings or other information, geocoded information derived from cell phone use, and so forth/

Although the definition included in the Copenhagen framework is more restrictive than the one proposed in the past by other partners, there is still ongoing discussion on how this could be expanded as the work on the implementation of the framework advances.

It is important to note that in the context of citizen data, "citizen" refers to all individuals and is not restricted to those holding citizenship in the country where they reside. Refugees, asylum seekers, migrants, stateless persons, and other marginalized groups might not have citizenship in a country they live in, but they are nonetheless rights-holders under international and national laws and face specific needs that are equally important to be captured by adequate data and to actively contribute to the data production process. Therefore, they are all included within the scope of this work under the umbrella term "citizen."

Uses of CGD

CGD can be fully integrated into official statistics or used to complement them as additional data outside the official statistical boundaries or to challenge official statistics and offer an alternative description of experiences, especially by groups that feel misrepresented by the existing approaches to data collection.

A review[5] of how citizen data has been used in recent years reveals that citizen data initiatives have helped respond to a wide range of demands

1. **Responding to data needs of groups or communities** that do not feel represented by the data available in their national or local data ecosystem and understanding complex

[5] Collaborative on Citizen Data, Citizen data: a conceptual framework, UN Statistical Commission 2024.

issues, dynamics, and powers prevailing in communities, which may cause increasing marginalization. These initiatives address data gaps on population groups that are not adequately covered through existing data collection programmes due to resource and practical constraint in data collection methods, lack of awareness, difficult access, or issues of discrimination or biases.

2. **Evaluating or monitoring service access and supporting legal and policy action.** These are initiatives aimed at ensuring equitable and effective service delivery in various sectors, such as healthcare, education, social services, etc. This involves assessing the availability, affordability, and quality of services, and using the gathered data to inform legal and policy interventions.
3. **Assessing and monitoring the level of recognition and implementation of rights.** These initiatives are aimed at safeguarding and promoting human rights to ensure that individuals and communities can enjoy their basic freedoms and entitlements. Citizen data are collected to document the extent to which legal rights and international human rights standards are acknowledged, effectively put into practice, and experienced by rightsholders.
4. **Challenge official data when government reporting may misreport or deliberately mislead on specific topics.** Sometimes called shadow reporting, these are efforts by civil society to offer alternative reports from the official government ones, including for instance to existing international treaties and committees or to challenge government reports and data in the national context.
5. **Increasing public understanding and awareness**, mobilizing actions and empowering local communities. These are intertwined processes that can drive positive change and foster greater resilience and well-being.
6. **Providing evidence for scientific studies and research**. These initiatives are aimed at collecting data to help establish credibility, validate hypotheses, and contribute to the advancement of knowledge.

A few examples

The 100 Hotspots project[6] in India covers 35 vulnerable communities across India providing data on hard-to-reach groups such as migrant workers, fisherfolk of Loktak Lake, urban poor, persons with disabilities, families of victims of extrajudicial killings. The project focused on

[6] https://sdgs.un.org/partnerships/100-hotspots

generating evidence through community-driven data. Volunteers received a two-day training on data collection and practiced on the app-based platform.

In the Philippines, the Concerned Citizens of Abra for Good Governance (CCAGG)[7] includes citizens working alongside volunteer engineers who provide training on how to monitor roads and infrastructure projects. They monitor the completion and quality of public infrastructure projects in the province of Abra. The resulting data and report are sent to the national planning and development agency, National Economic and Development Authority (NEDA) and the Ministry of Budget and Management (MBM), as well as local agencies to facilitate progress monitoring. The group also sometimes provides recommendations to state agencies based on their data.

In Peru[8], the National Water Authority has supported participatory water monitoring programmes for watershed planning with the community involved in data collection. Results were used for national data reporting on SDG 6 ('Clean water and sanitation'). In the Andean region of Peru, local stakeholders, academic institutions, and NGOs part of the Regional Initiative for Hydrological Monitoring of Andean Ecosystems (iMHEA) co-developed a robust and standardized water monitoring protocol, and in partnerships with local universities provided resources for training, equipment calibration, and data analysis and management.

Strategies for leveraging citizen data for policymaking[9]

There are still challenges in fully leveraging citizen data. Below are a few strategies to make the best use of citizen data.

- **Establish a common framework for citizens' contribution to data**. There are different ways citizens have been contributing to various stages of the data value chain, and with different objectives, and overall, there is a lack of a common understanding of what citizen data is. In this context, a framework helps bring coherence and offers a guide on how to best use citizen data. The Copenhagen Framework on Citizen Data introduced

[7] https://thedatashift.civicus.org/wp-content/uploads/2015/07/case-study-ccagg.pdf

[8] https://www.nature.com/articles/s41893-019-0390-3

[9] Extracted from "Background paper: Towards a framework to harness data by citizens, for citizens, and about citizens", https://unstats.un.org/sdgs/files/meetings/harnessing-data-by-citizens-for-public-policy-and-SDG-monitoring/Citizens-contribution-to-data-background-paper-202211.pdf

below is a good way to offer a common and shared approach to citizen data in all its various forms.

- **Foster collaboration between state and non-state partners.** It is important to promote an understanding of the value of citizen data in the broader data community. This may entail specific actions such as adopting a broader mandate for NSOs on data stewardship, embracing new data sources; establishing a legal framework to clarify roles and responsibilities of all stakeholders when working with non-traditional data; nurturing trust across different parts of the statistical system.
- **Improving capacity of NSOs and CSOs**. Building capacities of CSOs and NSOs for better utilization of citizen data is key. For CSOs, this would focus on increasing statistical literacy overall, and improving their technical skills for data collection and analysis. For NSOs, capacity needs to be improved on how to approach and engage with CSOs; and on how to assess quality and integrate citizen data into official statistics.
- **Improving the quality of citizens' engagement**. The ability for state actors to work with and collaborate with citizens is key for the future development of citizen data. For instance, it is important to make official data more open and the data dissemination process more transparent to build trust. Similarly, efforts could focus on building trust and fostering public confidence in the institutions, including NSOs and CSOs.

The Copenhagen Framework for citizen data and its implementation

Background

In response to the recommendations from a UN Expert Meeting[10] and a decision of the 54th Session of the UN Statistical Commission[11], a Collaborative on Citizen Data was established to foster systematic collaboration among civil society, national statistical systems, academia, and other relevant stakeholders. The collaborative is tasked to advance the sustainable coordination, production, and use of citizen data for impact.

[10] Report of the United Nations Expert Group Meeting on Harnessing data by citizens for public policy and SDG monitoring: a conceptual framework, 10-11 November 2022, Bangkok, Thailand. Available at https://unstats.un.org/sdgs/files/meetings/harnessing-data-by-citizens-for-public-policy-and-SDGmonitoring/CDG_EGM_report_final_public.pdf

[11] https://unstats.un.org/UNSDWebsite/statcom/session_54/documents/2023-37-FinalReport-E.pdf, Decision 54/102, page 11.

Since its establishment, the collaborative has focused on the development of a new conceptual framework for citizen data and a roadmap for its implementation. The collaborative will be responsible for the implementation of the framework and will serve as venue for a community of sharing and learning exchange.

The framework is intended to help conceptualize and agree on the different ways that citizens can play a role in data; and to provide coherence and coordination of actions for the community to further foster the development and use of citizen data. It is called the "Copenhagen Framework on Citizen Data" as it is the result of a review and discussion at the second United Nations Expert Group Meeting on citizen data, held in Copenhagen in September 2023.

The term "citizen data" is used in the Copenhagen Framework for consistency and inclusivity and is meant to cover various types of citizen data that are often referred to with different terms including citizen science, community science, citizen-generated data, crowdsourcing, volunteered geographic information, citizen observatories, citizen engagement in social innovation, community-based monitoring, participatory mapping, participatory action research, community-driven or community-generated data.

The Framework

In the framework, citizen data are defined based on the following three characteristics[12]:

- Levels of citizen participation
- Stages of the data value chain at which citizens engage
- Types of initiative of data collection

Data can be considered citizen data when they are produced through initiatives driven by citizens, communities or CSOs, whether independently or in collaboration with other state or non-state actors. Citizens are either fully engaged or in partnership in all stages of the data value chain. When initiatives are not driven by citizens, citizens must be sufficiently engaged in the design or data collection stage. Data produced by actors without collaboration with citizens, or with only limited or insufficient engagement of citizens are not considered citizen data.

[12] For details on the taxonomy for citizen data and the levels of citizens' participation in the different stages of the value chain, see the Collaborative on Citizen Data, Citizen data: a conceptual framework, UN Statistical Commission 2024 at https://unstats.un.org/UNSDWebsite/statcom/session_55/documents/BG-4c-CGD_Framework-E.pdf

While acknowledging the diversity and plurality of citizen contributions to data, the framework focuses on data that are produced with a sufficient engagement of CSOs, communities or citizens at the stages of design or data collection stressing the importance of the role citizens play in effectively influencing the process.

Engagement of citizens takes place in many other ways, including in traditional data collection programmes carried out by NSOs. In these instances, citizens do not have influence over the design of data collection but might be able to provide inputs at other stages of the value chain. These forms of collaboration and engagement are not now included in the framework.

Citizen science

Citizen science is a particular case of citizen data and can be defined as "the involvement of citizens in scientific research and/or knowledge production".[13] Citizen science projects take place in different scientific fields, such as life sciences, physical sciences, medical research, engineering, social sciences and humanities, and the terminologies used may vary depending on the specific domain.

The Australian Citizen Science Association for instance defines citizen science as involving "public participation and collaboration in scientific research with the aim to increase scientific knowledge" and a "way to harness community skills and passion to fuel the capacity of science to answer our questions about the world and how it works". UNEP defines citizen science as "entailing the engagement of volunteers in science and research". [14]

Some citizen science projects have been successful in providing data to monitor the implementation of the SDGs especially for the goals related to the environment and can be used to supplement data produced on some of the indicators through traditional data collection tools.[15] In general, citizen science projects are initiated by academic or research institutions, environmental agencies, international agencies, and other organizations not necessarily linked to CSOs or community organizations.

A few examples of citizen science projects

---

[13] Fritz, S, et al., Citizen science and the United Nations Sustainable Development Goals at https://www.nature.com/articles/s41893-019-0390-3

[14] Shanley, L. A., Hulbert, J. & Auerbach, J. Definitions of Citizen Science (GitHub, 2019); https://github.com/lshanley/CitSciDefinitions

[15] Fritz, S. et al.

Drinkable Rivers (DR) by Drinkable Rivers Citizen Science Program (CSO)[16]: The Drinkable Rivers program runs a global 'citizen science program,' with 60 hubs in 20 countries, including recent expansions into America and Asia, where volunteers monitor river health, ultimately aiming at improving rivers health and drinking water quality.

The Citizen Observation of Local Litter in coastal ECosysTem (COLLECT) project carried out between 2021 and 2022, supported by the Partnership for Observation of the Global Ocean (POGO) is a citizen science project aimed at collecting data on plastic litter in seven countries of Africa and Asia. Data were collected by trained local students aged 15 to 18 on macro-, meso- and micro- plastics in beach sediments. The results are expected to contribute to information on coastal plastic pollution, the identification of hotspots of coastal plastic litter, and the increased awareness of local communities to the potential consequences of plastic pollution.[17]

Principles of citizen data

The principles for citizen data are established to ensure that data production and use are responsible, professional, and ethical. These principles are important as they are an integral part of how citizen data are defined and how citizen data initiatives should be implemented. They also serve as a framework for data governance, management, and protection, in the context of organizations that collect and use citizen data.

1. Independence: Data collection should be initiated and conducted free of any unwanted political pressure.

2. Relevance: Data collected should directly respond to the issues identified or valued by the citizens or the organizations (community level and/or CSO) representing them.

3. Participation and informed consent: All groups of interest should be involved, including those that are vulnerable and marginalized, and participation should be free, open, equitable, accessible, and transparent. Individuals should provide informed and voluntary consent before their data are collected. They should have control over their data and understand the purpose and use of the data collected.

---

[16] Organizational Engagement with citizen-generated data, DV 443, LSE Development Management Consultancy Project, 2024.

[17]Catarino, A. I., et al. Addressing data gaps in marine litter distribution: Citizen science observation of plastics in coastal ecosystems by high-school students, METHODS article Front. Mar. Sci., 06 February 2023

https://www.frontiersin.org/articles/10.3389/fmars.2023.1126895/full

4. Professional standards: Data collection, processing and dissemination should be conducted according to professional and methodological scientific considerations to ensure data quality and fitness for use and wider societal or policy impact.

5. Data security: Copyright, intellectual property, data sharing agreements, and data security should be clearly defined and implemented.

6. Self-definition and self-identification: population of interest should be self-defined and personal identity and characteristics should be assigned to individuals through self-identification (at individual's discretion).

7. Transparency: Data collectors should provide clear, openly accessible metadata and paradata about their operations, including research design, data collection methodology and description on how data will be used and shared.

8. Ethical and safe production and use: Protection of human rights, safety and wellbeing should be a primary concern throughout the data value chain. Proper use of data should be ensured in order not to harm, discriminate or stigmatize any individual or groups (do no harm). Data producers are accountable for upholding human rights in their operations and for the impact of their data collection operations.

9. Confidentiality, privacy, and data attribution: Data referring to individuals, whether they refer to natural or legal persons, are strictly confidential and should be used exclusively for statistical purposes. Information that identifies individuals or discloses an individual's personal characteristics should never be made public as a result of data dissemination.

10. Openness, and accessibility: Data should be made accessible and publicly available to all including persons with disabilities and those with limited access to technologies, and where possible, results should be published in an open access format for use and re-use.

Citizen data as open data

In the previous sections, we offered a review of what citizen data are and their critical importance as a data source to supplement data from traditional programmes or to fill gaps where other data collection tools are inadequate to represent all parts of the population and their needs and circumstances.

Citizen data initiatives can also foster trust in data and connect CSOs and various communities to the official statistical system. They can empower citizens and communities to have a direct

impact on issues that matter to their lives and have control on how data are communicated and used. Engaged citizens are more likely to understand and trust data, including official statistics.

When published and disseminated within the national statistical system, citizen data should be open and fulfill all the principles discussed above. In addition, since the value of citizen data is in the level of engagement of citizens at the various stages of the value chain, it is essential that the communities or CSOs who conducted the initiatives are fully involved in the interpretation of results and all other stages of data communication and dissemination.

Also, as the use of citizen data by the NSO may offer the opportunity to increase trust in official data, it is important to involve CSOs in data dissemination efforts more broadly. CSOs can help through campaigns and educational opportunities and can encourage further citizens' engagement by demonstrating the value of their data and popularizing the concept in the relevant communities.

When communicating and disseminating citizen data, particular attention should be given to offering a clear description of the initiative, including the data collection design and methodology. Also, citizen data should be made especially accessible to and usable by the concerned groups, including persons with disabilities and those with limited access to technologies.

Finally, given the value of citizens' engagement in the production of citizen data, it may also be beneficial to offer tools to citizens to explore and better understand the data and how they were produced. These could be limited to open-source software or include more specialized training initiatives for data literacy.

Summary and key takeaways

1. Citizen generated data contribute to better data availability, timeliness, quality, openness, and inclusiveness
    a. Engagement of citizens throughout the data value chain empowers citizens to be agents of change
    b. Citizen data help fill data gaps on difficult-to-reach populations and those who are underrepresented because of their vulnerabilities or biases and misconceptions in the design and implementation of data collection programs
    c. Citizen data offer the opportunity to citizens to have some ownership of data, influence how data are used and help achieve better impact
2. Multiple terms and definitions have been used to describe the different forms of citizens' engagement in data production and use
3. The Copenhagen Framework, developed by the Collaborative on Citizen Data offers a structured approach to citizen data and an operational definition, based on

   a. Levels of citizen participation
   b. Stages of the data value chain at which citizens engage
   c. Types of initiative of data collection
4. Citizen data have been used effectively for different purposes, including for addressing very specific users' needs, safeguarding and promoting human rights, challenging official data, increasing public awareness on issues important to different communities, etc.
5. The production and use of citizen data should be guided by the principles for citizen data to ensure data production and use are responsible, professional, and ethical.