

# Background paper: Towards a framework to harness data by citizens, for citizens, and about citizens

## Table of Content

1. Introduction.....	2
2. A summary of current practices on citizens’ contribution to data.....	4
<b>2.1 Terminologies and definitions</b> .....	4
<b>2.2 Objectives</b> .....	10
<b>2.3 Engaging citizens in the data value chain</b> .....	15
3. About data quality.....	17
4. An overview of key challenges and strategies to using citizen data for policy.....	20
3.1 <i>Establishing a common framework for citizens’ contribution to data</i> .....	20
3.2 <i>Fostering collaboration between state and non-state partners</i> .....	21
3.3 <i>Improving capacity of NSOs and CSOs</i> .....	22
3.4 <i>Improving the quality of engagement: how to gain trust from and engage with citizens?</i> .....	23
5. Key action points: fostering more engagement of citizens and collaboration of the state and non-state partners .....	23
References.....	25
Annex 1.....	28

## 1. Introduction

Citizens' contribution to data, broadly defined as the engagement of citizens in multiple processes in the data value chain [1–3], from specifying needs to use of the data to inform policy. Citizen's contribution to data is increasingly recognized for its unique ability to help overcome many data challenges of our times. A recent report suggests that National Statistical Offices (NSOs) and the broader official statistical community have started to change their perceptions of citizen data over the past five years, leading to new efforts to integrate data that was previously considered “non-traditional” into official statistics [2]. In the context of the 2030 Agenda's principle to address the needs of people who have been left behind, citizen data helps fill critical data gaps for groups suffering from data marginalization and ensure that their experiences are reflected in national statistics [4].

Beyond filling data gaps, some citizen data initiatives further advance important values such as fairness, inclusiveness, openness and transparency in statistics. Engaging citizens in policy discussions that are grounded in data can be seen as a measure to enhance public participation in government decision-making and ensure responsiveness and inclusive decision-making at different levels of government.

The background paper provides an overview of current practices of citizens' contribution to data and shows the diversity and plurality in how citizens are engaged in different stages of the data value chain and at different levels (global, regional, national and local) (Chapter 2). The Chapter organizes the existing examples into broad themes including terminologies and definition, objectives and the contribution of citizens to various stages of the data value chain, to highlight the potentials of citizen data as well as the intrinsic complexities and pluralities around citizens' contribution to data. The paper further outlines challenges faced by both citizens (mostly represented by civil society organizations) and national statistical offices in fostering collaboration between the two parties, in order to serve various purposes from improving advocacy, to filling data gap and informing public policy (Chapter 3). Strategies to overcome those challenges are also covered, based on experiences and discussions around success stories and lessons learnt from projects and practices engaging with citizens. In Chapter 4 we cover the discussion on quality standard. We then conclude the paper in Chapter 5 by offering some key action points to be further discussed and considered at the upcoming Expert Group Meeting.

To prepare for the paper we draw on two main sources of information to review the ongoing practices in citizen data. First, we send an open-ended questionnaire (Annex 1) to around 20 key experts from organizations including academic researchers, national statistical offices (NSOs), civil-society organizations (CSOs), and international organizations (IOs) who are actively pursuing various forms of citizen data projects. Second, we conduct a literature review using keywords related to citizen data (such as citizen science, citizen-generated data, etc.) to capture published studies stemming from the topic. Our approach ensures that we can capture, for as much as possible, the diverse perspectives on citizen data and understand how challenges and successes look like from the vantage point of multiple stakeholders.

The paper, however, does not aim to be comprehensive in its review. This is because citizens' contribution to data could happen at different levels (global, regional, national and local), is organized by different stakeholders (scientists, community leaders, civil society organisations, or individual citizens), serve different objectives and contribute to data and policy in many different ways. The diversity and plurality of citizens' contribution to data make it challenging for any entity to have a full grasp of everything that is happening in the ground. This certainly is a challenge for us to have a comprehensive view of what has happened in the ground but it also demonstrates the huge potential of citizens' contribution.

## 2. A summary of current practices on citizens' contribution to data

### 2.1 Terminologies and definitions

The contribution of citizens to the data process takes several forms and serves different purposes. Generally, it is helpful to start with two overlapping lineages: “citizen science” as the effort to integrate the citizen perspective into science [5], and “citizen-generated data” as the relatively newer effort to ensure citizens are directly involved in the design, collection, analysis, and use of data that describes them [4].

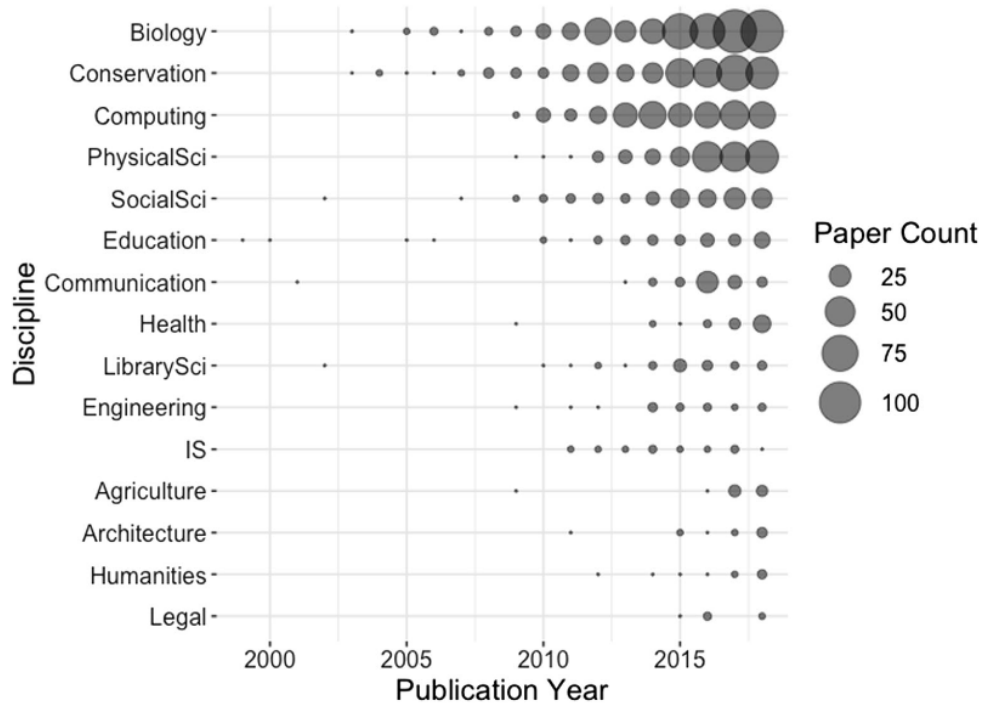
Also known as “public participation in scientific research,” [6,7] the overarching motivation for citizen science is to bring together public participation and knowledge production. In the simplest sense, the motivation can be described as pragmatism, such as in the use of citizen volunteers in field-based observations of the natural world [8]. For example, the [Cornell Lab of Ornithology](#) has been operating numerous citizen science projects that engaged citizens in collecting and submitting data on bird observations, reading about project findings, visualizing and analyzing the collected data. Those projects have been successful in advancing scientific knowledge on the bird population changes over time and space and how environmental changes impacted on bird population and their behavior.

At the other end of the spectrum, citizen science can be seen as a strategy to democratize science by ensuring that citizens are involved in the design and management stages of scientific research projects [9,10]. It was argued by some that citizen science is a way to tapping into the potential of those who have not been engaged in the “elitist science” world. The European Commission<sup>1</sup> further suggests that citizen science is an important mechanism to reinforce societal trust in science through increasing transparency, science literacy, and confidence of the public in research.

Published research under the term “citizen science” has grown quickly in the past 30 years[11]. As shown in Figure 1 below, large numbers of papers are published annually between 2010 and 2015. The broad fields of Biology and Conservation have been leading in terms of published research, but we are also seeing growing activities in social sciences, education, health and humanities.

---

<sup>1</sup> Citizen Science Factsheet, 09 September 2020, accessed on November 1, 2022 at <https://op.europa.eu/en/publication-detail/-/publication/d1768147-f17a-11ea-991b-01aa75ed71a1/language-en/format-PDF/source-152465380#>



**Figure 1. Publications on citizen science over time from Lukyanenko et al. 2020 [11]**

*Note: IS stands for “information system”*

Because citizen science takes place and can be rooted in many different scientific fields, such as life sciences, physical sciences, medical research, engineering, social sciences and humanities, many terminologies used expressed the focus of the specific fields, e.g., volunteered geographic information [12], citizen observatories [13], citizen engagement in social innovation [14].

While some citizen science projects can be geared towards public policy monitoring and advocacy goals, a big proportion focuses on scientific research. In contrast, citizen-generated data was typically defined with the objective of advocacy and public policy monitoring in the center. The definition by CIVICUS which has been widely used by many CSOs specifies that citizen-generated data are data “that people or their organizations produce to directly monitor, demand or drive change on issues that affect them” [15]. A related concept—community-driven or community-generated data—further emphasizes that citizens engaged belong to specific social groups typically marginalized groups which could have been kept invisible from national or official statistics. These data are crucial to make citizens’ voices heard and their wellbeing reflected in the data that can be used to drive for change in the public policy arena.

The term “crowdsourcing” has also been used by many, sometimes interchangeably with citizen science or citizen-generated data. More recently, national statistical offices have adopted the

term to denote the efforts to collect data from a large community of citizens. Within this exercise, citizens are considered the “experts within their local environments”<sup>2</sup>.

In some examples the definition of citizen data can be expanded to cover all types of non-state actors, meaning that data collected and shared by private companies and corporations<sup>3</sup> may also be utilized under the broad concept of citizen data.

So far the discussion on citizen science, citizen-generated data and other engagement of citizens has been focusing heavily on data collection. The contribution of citizens, however, does not only limit to data collection. As will be further discussed later in the paper, citizens have been engaged in all stages of the data value chain, ranging from specifying data needs to using data to evaluate and make policy decisions. Similarly for citizen science projects, citizens’ contribution may expand from conceptualization to drawing research conclusions. As citizen science expands to the field of social science and humanities, the line between citizen science and citizen-generated data has also becoming more blurred.

Given the variations in how citizens are engaged in advocacy, research and policymaking and the variations in terms used as well as their definitions, it would be a difficult task to propose one term for or a strict definition on citizens’ contribution to data. Nor is it possible to delineate between various terminologies such as citizen science, citizen-generated data, crowdsourcing, or other terms that have used in the science and data field. Even within the same community, there are also wide-ranging interpretations in terms of what citizen science is, or what citizen-generated data is. A recent study finds a wide array of interpretations when it asks 333 experts to provide ratings to 50 pre-written vignettes, representing the full ranges of citizen science. Further, the experts not only provide diverging classifications of the vignettes as citizen science or not, but they also express varying levels of confidence regarding their classification for close to half of the vignettes [5].

It is however useful to outline the dimensions that have stand out in literatures and country case studies when citizens’ contribution to data is discussed. For example, reviews of citizen science practices highlight the importance of **data collection** as a core component [5,7] such that projects without data collection tend to not be recognized by experts as citizen science [5]. This certainly would not be the case or agreed upon by others. In the past decade, NSOs and CSOs have started to collaborate to integrate aspects of citizens’ interests and perspectives into national official statistics. The term “participatory data collection” defines this process as one in which traditional and non-traditional data producers, particularly citizens, work side by side to enable a better data collection process and content [16]. This process could go in either direction: NSOs consulting with citizens when defining survey categories for population groups, or citizens receiving advice and guidance from the NSOs on their citizen-generated data collection processes. As the scopes vary, it means that some projects only involve citizen engagement at the

---

<sup>2</sup> StatCan definition of crowdsourcing, retrieved from <https://www.statcan.gc.ca/en/our-data/where/crowdsourcing>

<sup>3</sup> Examples: (1) Malaysia, (2) UK ONS use webscraped price data from supermarkets to track changes in price of key grocery items, report published in May.

initial planning stages, while others involve citizen participation at multiple stages in the data value chain (we offer a discussion of the data value chain later in Section 2.3).

**Level of engagement** has also been a key aspect that many have debated on. Projects may range from passive engagement where the data are simply drawn from citizens. For instance, the UK ONS used web-scraping technique to track the price of common grocery items to monitor changes in costs of living. Citizen data may also be contributed to in semi-active manner, such as a part of another activity. In the Pulse Lab initiative, the data are collected from Indonesian citizens who report a complaint through SMS, website, mobile app, or Twitter [17]. Other examples of semi-active participation include data collected through compulsory education or training programs provided for citizens volunteering in the projects. Many studies recruit students as volunteers through their schools and science teachers. Some studies present participation as an optional homework [18], and in some studies the students acquire science education credits by volunteering in citizen science projects [19].

Even for active engagement where citizens are aware of the data collection and give active and continuous consent, we may still identify varying levels of activeness on the citizen part. Some projects simply involve wearing a tracking device or carrying a smart phone, whereas others require significant training so that citizens can accurately report the data back to the research team [20,21].

Therefore the question is: at what level of citizen's engagement would be required for the data to be considered "citizen data"? Would data passively collected from citizens through webscraping, or by mobile phone or social media companies qualify? In the case of a community-led data collection effort, some members may be more active in designing the data collection while the others are only survey respondents. The former engaged more actively while the latter can be considered "semi-active".

**Informed and active consent** is considered integral to citizen data collection by some, distinguishing the field from "big data" because participants in citizen data project must be aware that they are participating in data collections, while data "indirectly" provided by citizens in "big data" are generated without informed and active consent from citizens themselves.

Interestingly putting aside the debate on types of entities that have ownership of the data, that is, whether data are owned by citizens and the community representing them, or by a private company, some also view social media as **a safe space (or a community)** for individuals to express their opinion. Therefore social media data can also be perceived as citizens' contribution for a better understanding of a phenomenon that might be too sensitive to gather through other platforms [22].

**Voluntarily** participating in data collection, analysis or dissemination is required by some. A recent review further points out that the view of citizens as volunteers hinges on whether and how much citizens are **compensated financially** for their contribution [5]. Projects vary by the amount of financial compensation and the types of compensation provided, e.g., some projects compensate for participants' time while others only cover direct costs such as meals and transportation. Beyond the financial aspects, volunteers staying with a project for a long time

may alter the relationships between citizens and researchers. In the Geological Survey of Brazil,<sup>4</sup> the technicians who are present at the hydrological station every 3 months are considered family members of the citizen-volunteer. The technicians are welcomed like old friends, invited to lunch, they cannot fail to have a snack and a cup of coffee made by the owner of the house.

**Level of expertise** of the citizens is another interesting dimension that has stirred debate. For some citizen science project, “non-professional” scientists contribute to the project [5]. This is particularly true when democratization of science is the main objective of the project when those typically excluded from the science field have a chance to be part of the research. When the citizens are seen as the helping hands to follow well-established data collection protocols developed by the researchers, then the citizens could be seen as non-experts with no specialized knowledge. Example of this non-expert assumption includes projects where the volunteers wear tracking devices in daily-life activities to help researchers collect data [23].

A relatively more common view of citizens’ contribution acknowledges their contextual knowledge while recognizing citizens are not experts in research or in official statistics. Under this assumption, it is common for researchers to develop a research plan and provide training for citizens to collect and report data. Various environmental and wildlife monitoring projects fit this description [24]. In the RinkWatch project, for instance, the citizens comprise of highly experienced individuals who built ice rinks in their backyards every year for personal use. In order to collect data on the ice conditions and assess winter warming trends, the research team provides guidelines for participants to pin the location of their rinks on an interactive map and report the skating conditions daily in the winter based on a drop-down menu of choices [25]. In the Statistics Canada’s crowdsourcing exercise, citizens are considered the “experts within their local environments”.

Similarly for data projects undertaken by the Housing Data Coalition, a New York City based CSO (or formed by a group of CSOs), volunteers engaged with the project are often technological savvy and highly skilled in their own professional field including data science, statistics, and public policy; though not necessary in official statistics. In this case, the only aspect that distinguishes them from official statistics is that they are “non-official” (or “non-government”). This points to another important dimension which is “**official**” versus “**non-official**”.

Additionally, experts from the citizen science field also tend to agree that the projects must involve some **educational aspects** to demonstrate that citizens learn something from the scientific process. This can take many forms, such as attending webinars, self-education, peer-to-peer learning, structured training by researchers, or on-the-field practical experiences [26]. This aspect however is not always agreed by all experts [5]. But this certainly does not completely align with other practices while the expectation is to gain support from citizens that can offer additional insights, either technical skills or knowledge that the official statistics community lacks. The educational aspects may still be there when the project is seen as providing mutual benefits but not the main focus.

---

<sup>4</sup> Source: questionnaire response from CPRM – Geological Survey of Brazil



Citizen data projects also vary in terms of **how citizens are recruited**, whether as individuals or as representatives of communities. Many projects are based on a general call for volunteers or through convenient outreach such as by leveraging personal relationships, social media, site visits, or through partnered programs at schools or colleges [27]. A known problem with this approach is that citizens participation can be very uneven, leading to inconsistency in available data as well as their quality [28]. In contrast, projects undertaken under the definition of community-driven data tend to be built upon collaborations with CSOs and community leaders, who help to ensure that the community perspective is well represented.

For published works based on citizen science, the **scientific peer-review process** is a key mechanism for quality assurance, that usually includes “implicit” standards for design and validations to be covered by the research teams. For example, the research team in the marine invasive species example collected and re-counted the samples collected by citizens to check for accuracy and reliability of the reported data [29]. Expert assessment may sometimes be used to benchmark the data collected by citizens. In a 2018 study of image interpretation and classification using images from Hurricane Matthew, the research team compares image classifications made by experts and citizens and reports that they agree around 92% of the time [21].

For citizens’ contribution outside the purview of scientific research, the mechanisms for **quality assurance** are more diverse and sometimes completely absent. This has been identified as a challenge, especially for projects led by community leaders with limited knowledge on the data work. Collaborative partnerships play an important role in this case. Academic researchers have provided guidance and leadership in many community projects, for instance, the Ruff Institute of Global Homelessness<sup>5</sup> represents one such partnership between researchers from DePaul University in Chicago, USA and DePaul international, an international CSO that provides direct services and advocacy for homeless people in the UK, Ireland, Ukraine, Slovakia, USA, and France. Having scientists on board ensures that data collection and analysis efforts done by the organization can be validated with other data. At minimum, the documentation process could cover key aspects to facilitate data validation and benchmarking.

Some partnerships may involve both academic researchers and NSOs. The effort to collect data on marine plastic litter in Ghana<sup>6</sup> have been supported by research scientists from the International Institute for Applied Systems Analysis (IIASA) as well as the Ghana Environmental Protection Agency and the country NSO (Ghana Statistical Service). Through the partnership, various stakeholders discuss and align the citizens’ data collection methodology with global standards (such as SDG reporting requirements, protocols) as well as various NSOs practice on marine litter data. The involvement of academic researchers may link the effort back to peer-review as an additional quality assessment channel.

When NSOs are involved in the partnerships, more specific quality assurance standards are considered. For example, the SDG Team in Canada worked with Statistics Canada quality

---

<sup>5</sup> <https://ighhub.org/about-us>

<sup>6</sup> <https://dataforchange.net/strengthening-measurement-of-marine-litter-in-Ghana>

secretariat and developed a new tool to assess the quality of civil society data to see if the data could be used for SDG reporting on the [Canada SDG data hub](#). Kenya National Bureau of Statistics included a set of quality criteria for validating citizen generated data within its newly issued [Kenya Statistical Quality Assurance Framework](#). UK Office for National Statistics SDG Team developed a [protocol for non-official sources](#) that is used to access both the raw data and the final data output from a citizen science project that report SDG indicator 14.1.1(b) on plastic marine debris density. Many reviews have also noted, however, that involvement of NSOs may be “light” (as opposed to “intensive”) where the NSOs only provide some initial guidance and not strictly applying a full quality assurance framework in all cases [2]. Given the large number of ongoing citizen data project, assessing quality does require an increased capacity within national statistical offices, as noted by some respondents to the questionnaire.

To conclude this section, we observe that the field of citizens’ contribution to data is inherently pluralistic with wide-ranging definitions and interpretations by many stakeholders. Additionally, each effort is also characterized by multiple guiding questions as we outlined in this section to which there are also a wide range of decisions to be made both in terms of project design and quality assurance. The project’s leadership, which can sometimes be a combination of stakeholders from NSOs, CSOs, academic researchers, and private entities, tend to have the strongest influence on shaping its goals and the subsequent decisions [5].

## **2.2 Objectives**

As citizen data is characterized by inherent plurality and diversity of interpretation, the objectives in engaging with citizens also vary greatly from one project to another. The review showed that the objectives can be broadly categorized into 3 levels<sup>7</sup>:

- (1) to increase public understanding and awareness, mobilize actions around specific issues and empower local communities especially those the marginalized population groups.
- (2) to provide evidence for scientific studies and research, and
- (3) to evaluate and/or monitor impact of interventions and to support legal and policy actions.
- (4) Other qualitative and cross-cutting goals.

### **(1) Increase public understanding and awareness; and mobilize actions**

Many examples from the National Statistical Offices indicate that their engagement of citizens is to increase public awareness and to consult with them. Citizens are actively engaged in the initial consultation stage before data collection. For example, a survey questionnaire on LGBTIQ+ by Colombian National Department of Statistics (DANE) has gone through wide consultation with civil society organizations before being finalized. Similarly, Japan engages with various stakeholders including CSOs in its SDGs promotion round table meetings for exchange of ideas and inputs. The types of engagement outlined are in line with the movements of NSOs in making data more inclusive. As part of the Response by the National Statistician of UK to the Inclusive Data Taskforce [Report and Recommendations](#), systematically partner with others

---

<sup>7</sup> Adapted from a similar structure adopted by U.S.’s Environmental Protection Agency (EPA) [30]

including the civil society has been covered as one important aspect.<sup>8</sup> For NSOs, engaging citizens actively helps increase public understanding of the data collection activities, and potentially leads to improving response rate and inclusiveness of official statistics. It creates a sense of ownership of data for those engaged and is welcomed by communities that are concerned about ownership and uses of data as reflected in mottos such as “Nothing about us, without us”<sup>9</sup>.

Empowering and mobilizing actions are also embedded in the work of CSOs and community-driven projects. As highlighted by the Management and Sustainable Development Initiative (MSD) in Viet Nam, grassroots level engagement of citizens is crucial to ensure that the communities especially the marginalized, understand the concept of SDGs and how data can be used to address issues relevant to the communities.

New insights from citizens are also central to many community-based citizen data projects, as the experiences of marginalized communities are unknown to both researchers and policy makers. The 100 Hotspots project in India,<sup>10</sup> for instance, covers 35 vulnerable communities in India to overcome the data void on communities and in difficult to reach terrains, such as migrant workers, fisherfolk of Loktak Lake, urban poor, persons with disabilities, families of victims of extrajudicial killings. Official statistics that typically relies on household surveys and sometimes administrative data, would not be able to provide information at this level of granularity.

## **(2) Provide evidence for scientific study and research**

In many scientific fields such as biology, data collection involves significant human resource to observe the environment in fine timescales and covering many small areas. From a pragmatic perspective, citizens as volunteers make the task possible. For example, a 2010 project engaged citizens to gather data on wildlife in the northeastern United States. The research team asked up to 6,000 homeowners to report whether they have seen or heard a coyote on their property and also to record the home locations. Citizen data enables researchers to generate a map predicting human-coyote interaction, successfully covering a large geographic area in a short period [18].

Beyond human resources, citizen data could also be seen as lower-cost alternatives to high-tech solutions. A project seeking to monitor air quality in Germany and Niger suggests that the use of citizen volunteers carrying low-cost sensor devices helps to create a dense network of measuring points, making it possible to apply machine-learning methods to generate accurate and real-time information on air quality. In this way, citizen participation can be seen as a low-cost alternative to fixed monitoring stations [31].

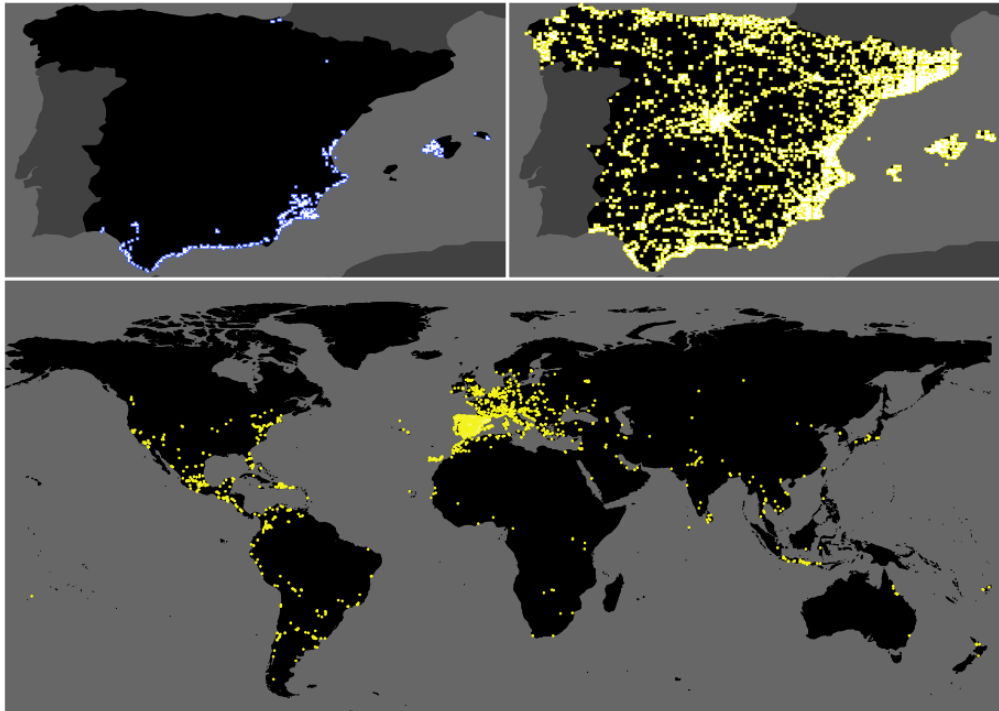
---

<sup>8</sup> [Response by the National Statistician to the Inclusive Data Taskforce \(IDTF\) Report and recommendations – UK Statistics Authority](#)

<sup>9</sup> <https://www.un.org/development/desa/disabilities/international-day-of-persons-with-disabilities-3-december/international-day-of-disabled-persons-2004-nothing-about-us-without-us.html>

<sup>10</sup> <https://sdgs.un.org/partnerships/100-hotspots>

Scalability is a related, but slightly different dimension. In some contexts, it would be much more feasible to expand the network of volunteers than to build technology-enabled monitoring stations. To track disease-carrying mosquitoes, for instance, a project compares the scalability of Ovitrap, a traditional tiger mosquito surveillance device, and citizen scientists using their own Android devices. Figure 2 below demonstrates how citizen data can provide much more data points with much lower set-up efforts.



**Fig. 8** Ovitrap and citizen scientist locations. Top left: blue markers show locations of sampling cells containing the 1558 ovitrap used in the analysis, accounting for nearly all ovitrap placed in Spain during 2014–2015. Top right and bottom: yellow markers show locations of sampling cells in which Mosquito Alert participants were randomly sampled in Spain (top right, excluding the Canary Islands, Ceuta and Melilla) and around the globe (bottom) during 2014–2015. Sampling occurred on Android devices five times per day at random times between 7:00 am and 10:00 pm (and only if the participant did not opt out of the background tracking feature). In all cases, white square at centre of marker shows actual sampling cell size, while coloured border is used to ease visualization. Boundary data from Natural Earth ([naturalearthdata.com](http://naturalearthdata.com))

**Figure 2. An example of citizen data scalability, from Palmer et al. 2017 [32]**

Some projects engage citizens because they contribute new insights to the data collection methodologies, or so called “foster grassroots innovations”. In a tick surveillance study, the research team engages citizens to test the effectiveness of various tick identification methodologies. While the researchers offer a standard methodology, they also suggest that citizens could use any other methodologies that they are familiar with. The study finds that the varying methodologies generate similar results, but citizens’ methodologies tend to create more sustained data collection over time, compared to the standard methodology [33].

### **(3) Evaluate and/or monitor impact of interventions and to support legal and policy actions**

This type of efforts can be seen commonly in the work of CSOs seeking to support their advocacy work with data and evidence. The Housing Data Coalition,<sup>11</sup> for instance, is a group of individuals and organizations who collaborate on their use of public data to further housing justice in New York City (USA). Their work focuses on maintaining and expanding a central database for public data related to housing in New York City to help inform advocacy efforts by other non-profits, tenant organizers, and members of the public. The group also advocates for city and state agencies to improve the quality of public data.

In the Philippines, Concerned Citizens of Abra for Good Governance (CCAGG)<sup>12</sup> represents a continuous citizen monitoring effort which has been in place for the past 20 years. The group includes citizens working alongside volunteered engineers who provide training on how to monitor roads and infrastructure projects. 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. Sometimes the group also provides recommendations to state agencies based on their data.

In recent years, some collaborations between CSOs and NSOs have been forged around SDG reporting. While there are strong variations in the projects quality, a few efforts have been formally supported by NSOs, such as the marine litter monitoring project in Ghana and water quality monitoring in Peru [24]. In Peru, community involvement in data collection for watershed planning have been supported by the National Water Authority of Peru, which aggregates and reports national data related to SDG 6 ('Clean water and sanitation'). In the Andean region of Peru, local stakeholders, academic institutions, and NGOs have formed the Regional Initiative for Hydrological Monitoring of Andean Ecosystems (iMHEA) to improve management of local water resources. The iMHEA network has co-developed a robust and standardized water monitoring protocol, and leverages partnerships with local universities to provide resources for training, equipment calibration, and data analysis and management.

As this type of citizen data projects connect directly to policy monitoring and policy advocacy, some NSOs are tasked with data validation and quality assurance to ensure that the data and analysis line up with expected quality standards for official data. A few NSOs have developed quality assurance protocols to work with non-official data, e.g., the UK<sup>13</sup> and Canada (work in progress). Another approach is to consider citizen data as experimental statistics, e.g., Columbia.<sup>14</sup>

#### **(4) Cross-cutting objectives**

Beyond the three broad levels that correlate with different requirement for quality assurance in each subsequent set of purposes, we also observe several citizen data initiatives that do not fit

---

<sup>11</sup> <https://www.housingdatanyc.org/>

<sup>12</sup> <http://ccagg.org/>

<sup>13</sup> <https://www.ons.gov.uk/economy/environmentalaccounts/methodologies/uksustainabledevelopmentgoalsuseofnonofficialsources>

<sup>14</sup> <https://www.dane.gov.co/index.php/estadisticas-por-tema/estadisticas-experimentales>

into one specific level. Some projects may have **cross-cutting purposes**, especially as the project grows overtime. In the case of the 100 Hotspot projects in India, for instance, raising awareness and advocating for excluded communities are the central starting point for the citizen data initiative. As the project cumulates a larger volume of data and streamlines their training as well as data collection efforts, the data becomes a resource to support both policy monitoring of several important SDGs including 1, 2, 3, 4, 5, 6, 7, 8, 11 and 16.<sup>15</sup> Additionally, the real time and disaggregated data can complement official statistics and help local governments develop policy measures and monitor policy impacts.

A similar example, the COVID-19 Disability Right Monitor<sup>16</sup> was developed to call for urgent actions by States and the international community to halt the catastrophic failure to protect the lives, health, and rights of persons with disabilities during the COVID-19 pandemic. With raising awareness and advocacy at its heart, the initiative has collected data from over 2,000 respondents from 134 countries around the world. The data collection methodology, in particular, are considered rigorous with intensive trainings provided to citizens and participating organizations, and as such, the data could also contribute towards monitoring SDGs related to disability.

The 2015 Pacific Disability Forum represents an example of relatively success, in which CSOs advocated for changes in the data collection and analysis of the Kiribati national census. Not only were new census questions added to cover population on disability, but the group also pushed for disability disaggregated data analysis and reporting across a range of outcomes including health, education, and the lived experiences of people with disability. In the same year, the NSO of Kiribati carried out the analysis and developed the “Kiribati Disability Monograph.”<sup>17</sup> Similar analysis has been conducted in Samoa, Palau, Tonga, Fiji, and Tuvalu.<sup>18</sup>

Some initiatives further emphasize the importance of **qualitative data** as complementary to official statistics which tend to be quantitative. The Danish Institute for Human Rights, for instance, suggest that monitoring mechanisms should combine both quantitative and qualitative data, with qualitative data including context-specific analysis, advices, as well as information about vulnerable groups and sensitive issues that are hard to capture through common statistical data [16].

The UK ONS conducted an Inclusive Data Consultation to request feedback from the public (including citizens and organizations) in 2021 and reported large gaps in understanding the experiences of various vulnerable populations, such as people with learning disabilities, LGBT+, Muslim women, unpaid caregivers.<sup>19</sup> Accordingly, future analysis from the UK ONS would combine quantitative and qualitative data to shed lights on the lived experiences of such groups.

The broad categorization of the objectives that citizens’ contributions serve cannot cover all the activities covered under the citizen data umbrella given the diversity of the activities. The

---

<sup>15</sup> <https://gcap.global/wp-content/uploads/2019/06/WNTA-The-100-hotSPOTS-r5.pdf>

<sup>16</sup> <https://covid-drm.org/assets/documents/Disability-Rights-During-the-Pandemic-report-web.pdf>

<sup>17</sup> [https://sdd.spc.int/digital\\_library/kiribati-disability-monograph-2015-population-and-housing-census](https://sdd.spc.int/digital_library/kiribati-disability-monograph-2015-population-and-housing-census)

<sup>18</sup> Jose Viera’s Presentation at the 13<sup>th</sup> Meeting of the IAEG-SDGs in Bangkok, Thailand on November 9, 2022.

<sup>19</sup> <https://uksa.statisticsauthority.gov.uk/publication/findings-from-the-online-inclusive-data-consultation/>

categorization, however, is useful when considering the quality requirement attached to different objectives of the citizen engagement projects.

Official statistics that are produced to inform policy have its quality standard, typically framed under the official quality assurance framework<sup>20</sup>. When working with data generated by citizens that are considered “other data sources”<sup>21</sup> for policy and monitoring purposes, a number of national statistical offices also established a set of quality criteria to assess the quality of those data, with the understanding that CSOs and community-based organisations function very differently from national statistical offices and typically serve different purposes. In this context the “fit-for-purpose” quality requirement has been consistently noted by contributing experts.

Citizens’ contribution for scientific and research purposes is usually subject to different set of quality requirement. As pointed out in Section 2.1, published scientific studies do go through scientific peer-review process that usually consists of certain quality requirement.

Data produced for the purpose of advocacy and raising awareness probably would not require the same level of quality rigorousness but the line between advocacy and informing policy is not always clear-cut. Another important aspect to note is that fit-for-purpose can be a slippery slope towards a neglect of quality assurance.

### ***2.3 Engaging citizens in the data value chain***

Several scientists and IOs have incorporated ideas from the data value chain to illuminate the many possible venues for engaging citizens in the data and statistical process. Similar to the scientific research process, the data value chain describes the evolution of data from the initial consultation and design, collection to analysis, review, dissemination, and the final impact of data on decision making [1,3]. The data value chain may be illustrated as a stage-wise process, starting with design, and ending with decision making. It may sometimes be represented as a circular process<sup>22</sup>, where decision-making becomes the initiator for new data planning and collection; or using the evaluation of the data and the processes (stage 6) to inform any future data collection (stage 1). A stage-wise illustration<sup>23</sup> of the stages in the data value chain is presented in Figure 3.

---

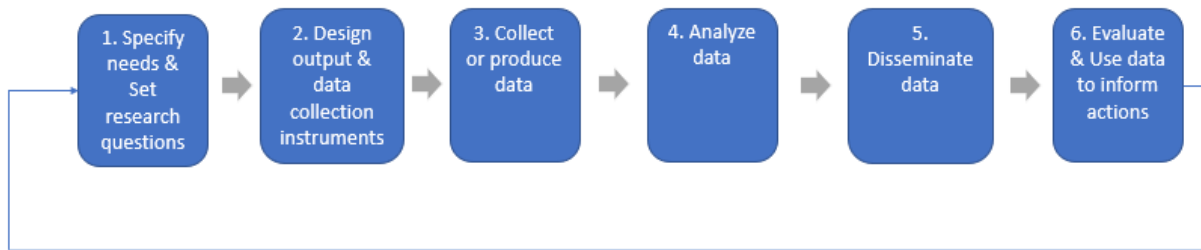
<sup>20</sup> At the international level, the United Nations National Quality Assurance Frameworks Manual for Official Statistics (UN-NAQF) provides guidance on framework for countries to implement national quality assurance framework for official statistics.

<sup>21</sup> UN-NAQF, para. 7.6

<sup>22</sup> For instance, the UK Civil Service office visualizes the GSBPM data value chain as a circle:

<https://analysisfunction.civilservice.gov.uk/wp-content/uploads/2018/02/Generic-Statistical-Business-Process-Model.pdf>

<sup>23</sup> Adapted from GSBPM, with the additional last stage.



**Figure 3. Contribution of citizens at different stages in the data value chain**

Thinking in terms of the data value chain clarifies that citizens’ contribution can happen at any stage, contributing to a wide range of purposes among citizen data projects. The most common stage of contribution is **data collection or production (Stage 3)**, and this holds true for both citizen science and citizen-generated data projects. Contribution of citizens in data collection stage is sometimes also considered a “contributory” process [4,7], relative to two other processes that also involve data collection alongside other stages.

Citizen engagement activities range from simple tasks like wearing a tracking device or carrying a smart phone to more complex ones requiring significant training so that citizens can accurately report the data back to the research team [20,21]. In a U.S. based study of marine invasive species, for instance, citizen scientists must identify crab species and determine the gender of the observed crabs. Because determining the gender of the crab is particularly challenging, the citizens are given an hour-long training session with relevant tools including magnifying glasses, buckets, rulers, and field guides. After each training session, the citizens enter a sampling session where they collect and identify the crabs they collected. All of the crabs collected by citizens were checked and re-counted by the research team to verify the degree of accuracy of data collection. The training helps improve the accuracy of citizens’ contribution.

When a great extent of interpretation and judgement by citizens are involved, data collection may overlap strongly with other stages in the data value chain (**Stage 4: Data analysis**). In the Forest Watchers initiative,<sup>24</sup> for instance, citizens help producing data on deforested areas (**Stage 3: Data collection and production**) by turning satellite images into data on deforested areas around the world. They either help by cleaning satellite images or by marking the deforested area. Studies have reported that variations in the quality of data stemming from different coding schemes or identification strategies impact the analytical results [34,35]. Variations in coding and interpretation may also require the researchers to change the project and training design [19], hence impacting the preceding stage of the data value chain (**Stage 2: Design**).

While relatively less common, citizens’ contribution does happen at other stages of the data value chain. At **Stage 1: Initial consultation (specify needs)**, national statistical offices have

<sup>24</sup> Daniel Lombraña, “ForestWatchers.net A Citizen Project for Forest Monitoring”, Open Knowledge Foundation Blog, October 1, 2012, accessed on November 1, 2022, <https://blog.okfn.org/2012/10/01/forestwatchers-net-a-citizen-project-for-forest-monitoring/>



stepped up in reaching out to CSOs and communities as part of the consultative process to identify data needs. Guidance on tools such social audit, scorecards, and action plan for various government offices to consult with citizens and capture the range of data needs are available [36]. Coalitions such as the Inclusive Data Charter<sup>25</sup>, established in 2018 and supported by 12 NSOs, has its first principle that “All populations must be included in the data.” Following this principle, the NSOs can develop action plans to engage marginalized communities to ensure that such communities are included in data, and their specifics reflected in data and the instrument that is used to collect data.

At **Stage 2: Design output and Data collection instruments**, some NSOs have started to incorporate citizens’ perspective in their designs. For instance, with support from UN Women in the conduct of its third Violence Against Women survey, Albania’s NSO (INSTAT) incorporated service-providers, local governments and CSOs in the survey design. Also supported by UN Women, GROOTS in Kenya has served as one of the members of the Inter-Agency Committee on Gender Statistics and has been influential in the decision-making and improvement of Census planning and design, as well as time-use survey.

The example of the New York City Housing Data Coalition works with a group of data scientist and statisticians on analyzing administrative data from the city administration (**Stage 4: Data analysis**), disseminates unified and easy-to-explore database (**Stage 5: Data dissemination**) and the analysis and data are used to evaluate the level for which the government has implemented its legislation bills (**Stage 6: Evaluate and use data to inform policy**).

Finally, in any stage of the data value chain, an additional dimension for variations among citizen data initiatives from the vantage points of non-state actors lies in *how central the data is*, relative to other goals of the non-state actors. The previous examples all have data as an explicitly stated objective. The opposite type of efforts, where there is not much discussion of the data work, may well exist. Consider the work of The Bronx Defenders, a public defender nonprofit whose work focuses on representing low-income and marginalized population in criminal justice system. Gleaning from their stated mission and descriptions of their work, it would be hard to consider if and how play a role in their work. However, the group also critically examines data from the New York city government to bolster their advocacy, e.g., for bail reform.<sup>26</sup> While this fits well in Stage 6 of the data value chain, their work may not be recognized as citizens’ contribution with data both from the views of NSOs and CSOs.

### 3. About data quality

Data quality always seems to be the biggest hurdle when discussing the use of citizen data for public policy. Therefore it would be useful to have a broad framework on quality for citizens’ contribution to data. However, this framework needs to be flexible and adaptable given the diversity of citizens’ contribution to data. An one-size-fits-all approach to quality would not be useful. As discussed in Session 2.2, the data quality requirement would be different for data

---

<sup>25</sup> <https://www.data4sdgs.org/initiatives/inclusive-data-charter>

<sup>26</sup> <https://www.bronxdefenders.org/the-bronx-defenders-statement-in-response-to-governor-hochuls-proposed-10-point-safety-plan/>

produced to serve different objectives. Data produced to raise awareness is likely to have a very different (potentially lower) quality requirement than those produced to inform policy. This point relates to the call for “fit-for-purpose” quality standard by many respondents to the questionnaire.

The quality framework also needs to be adapted depending on the stage of the data value chain citizens contribute to. For example, data produced by citizens with the aim to inform policy might be subject to similar standard as statistics produced by national statistical offices. But for citizen projects that only contribute to data analysis or dissemination, the official statistics quality standard is probably only partially applicable. For instance the validation would be more on the soundness of analytical method rather than on the accuracy and coverage of data.

In any case, when citizens’ contribution to data aims to inform public policy, a broad quality framework that consists of basic principles for a sound statistical product could be informative. According to the UN National Quality Assurance Framework Manual for Official Statistics (UN-NAQF) [37], members of the *extended data ecosystem* including all types of citizen-generated data and statistics (para. 2.4) could use the set of quality recommendations covered by UN-NAQF, under specific circumstances, as established by countries. UN-NAQF further noted that the quality principles and associated requirements apply to all data and statistics, regardless of data sources (citizen-generated data is listed under “other data sources”). The UN-NAQF includes 19 quality principles that are grouped in 4 levels: managing the statistical system, managing the institutional environment, managing statistical processes and managing statistical outputs.

Many countries have their own national quality assurance frameworks and quality standards but the most relevant ones are those that specifically apply to citizen-generated data or non-official data. The following table shows the 19 quality principles of UN-NAQF, together with three national examples: a tool developed by the SDG Team in Canada to assess the quality of civil society data to see if the data could be used for SDG reporting on the [Canada SDG data hub](#); the set of quality criteria for citizen-generated data covered by the [Kenya Statistical Quality Assurance Framework](#); and a [protocol for non-official sources](#) developed by the UK Office for National Statistics SDG Team.

<b>UN-NAQF Quality principles</b>	<b>Canada: Tool to assess civil society data</b>	<b>Kenya: Statistical Quality Assurance Framework</b>	<b>UK: Protocol for non-official sources</b>
	Benefit to Canadians	Need	Need
	Fairness and Do no harm		
1. Coordinating the national statistical system			
2. Managing relationship with data users, data providers and other stakeholders			
3. Managing statistical standards			
4. Assuring professional independence			

5. Assuring impartiality and objectivity			
6. Assuring transparency	Transparency and accountability		Transparency and accountability
7. Assuring statistical confidentiality and data security	Privacy Confidentiality		Ethics and privacy
8. Assuring commitment to quality			
9. Assuring adequacy of resources			
10. Assuring methodological soundness		Methodological soundness	Methods
11. Assuring cost-effectiveness			
12. Assuring appropriate statistical procedures			
		Credibility	
13. Managing the respondent burden			
14. Assuring relevance	Relevance	Relevance	Relevance
15. Assuring accuracy and reliability	Accuracy and reliability	Accuracy	Data quality
16. Assuring timeliness and punctuality	Timeliness and punctuality	Timeliness	Timeliness and availability
17. Assuring accessibility and clarity	Accessibility and clarity	Accessibility	
	Interpretability	Interpretability and clarity	
18. Assuring coherence and comparability	Coherence and comparability		
19. Managing metadata			

**Table 1. Quality criteria/principles, UN-NAQF, Canada, Kenya and UK**

There are overlaps across the 4 frameworks but also deviations from each other. Some national frameworks are also not linear: certain criteria are considered more important than others. For Canada, benefit to Canadians, fairness and do no harm, transparency and accountability, privacy and confidentiality are the quality gateway dimensions that all data need to pass to move on to the assessment of other quality criteria. The gateway dimensions are interpretability and clarify, credibility and trust and need for Kenya; and ethics and privacy, transparency and accountability and need for UK. Share elements across the three countries include “need”, “ethics” and “transparency”.

While the above frameworks can be adjusted and adapted to work broadly for citizen data, operationalize the framework could still be rather challenging. One challenge that has been repeatedly raised by experts from their response to the questionnaire was the lack of human resource to review a large number of data coming from civil society organisations. Under the current setting, many national statistical offices could barely cover their regular work programme. A significant boost of resources within the office is needed to take on the task in a sustainable way.

The other challenge relates to the nature of the citizens' contribution to data. As mentioned in Chapter 2, the diversity and plurality of citizens' effort in the field are immense. Having a full understanding of citizens' data requires NSOs to improve their knowledge of data coming from the non-official side and their technical capacity to understand the data and then to assess the quality of those data.

#### **4. An overview of key challenges and strategies to using citizen data for policy**

In this section, we draw primarily from the open-ended questionnaire responses by 20 stakeholders to take stock of the existing challenges identified in various citizen data initiatives and summarizes existing strategies to address the challenge. Such discussion helps identify collective efforts that are needed to support better channel and use citizen data to fill data gap and support data-driven policy.

##### *3.1 Establishing a common framework for citizens' contribution to data*

A summary of current practice on citizens' contribution to data in Chapter 1 shows that it is inherently pluralistic with wide-ranging definitions and interpretations. There are many different ways on how citizens have been contributing to different stages of the data value chain, with different objectives: ranging from advocacy and empowerment to scientific research and informing policy. Within this context, most respondents outlined that a major challenge when considering citizens' contribution to data is a lack of a common or shared understanding of what citizen data is, what it can do and how it should be used to fill data gap and inform policy.

The lack of shared understanding can sometimes be described as a lack of trust from all stakeholders involved. One respondent suggests that “is still a great barrier to convince NSOs to take the time to work with citizen-generate data producers.” The underlying issue is that citizen data quality is assumed to be “not good enough simply because they are not collected by experts”. Additionally, one respondent points out that the quality question stems from multiple levels: “Policy makers claim for a verification of data through NSOs, while the latter bemoan the lack of quality assurance mechanisms for CGD. Also, at times it is argued that civil society data could be biased because NGOs follow a specific political agenda.” On the other side, citizens and CSOs may feel frustrated because “they are not taken seriously or are not heard by international organizations or governments or that in some cases they felt that official statistics may be compromised (not produced in an impartial way)”. These reflections clearly show that the mistrust might stem from a lack of understanding of how citizens' contribution or engagement functions and the diversity of the practices.

Data quality does play some roles in exacerbating the distrust, including (a) the incoherent or lack of use of statistical concepts; (b) selection bias; (c) lack of information on how the data are being produced; and (d) lack of understanding of the data producers on the overall data quality. With that said, many respondents suggest that the more important issue lies in deciding which quality standards should citizen data be subjected to. One respondent suggests that the quality conversation be grounded in the key message that “citizen science data is best used to fill gaps – and data of known quality (even if the quality is not super-great) is infinitely better than no data

at all.” The same respondent also expresses that as many NSOs responded to the pandemic by relying on biased crowd-sourced samples, it might make them more open to incorporating citizen data. However, the longer-standing hyper-fixation on data quality may re-emerge as these pandemic experiences fade.

Therefore, it would be helpful to develop a broad framework for citizens’ contribution to data, outlining various terminologies and definitions used, objectives and the types of contributions, based on national practices. While developing such a framework, one needs to ensure such framework is broad and multi-dimensional to accommodate the goals of different stakeholders. Currently, different stakeholders have varying perspectives on this, e.g., scientists use peer-reviewed publications, NSOs focus on data quality, CSOs emphasize on inclusiveness, etc.

Towards creating a shared conceptual framework for citizen science, we recognize some overarching suggestions on its desirable characteristics. One suggestion is to have a conceptual framework alongside an operational one. While the conceptual component can be seen as time-invariant, the operational component can be adapted to an evolving data system incorporating citizen data to official statistics and other purposes.

Many respondents further emphasize the need to build trust through transparent data validation and assessment procedures. Perhaps this will include the creation of new “accreditation” entities to assess non-traditional data sources, which could be different from the official statistics quality assurance framework. Beyond data quality, it would also be crucial to develop standards for assessing other key dimensions such as quality of citizens’ contribution, or quality of data usage. At the global level, thematic groups could be formed to review and endorse non-traditional data sources and their methods on specific thematic areas. All of this will help foster a cultural shift in the statistical community to acknowledge and value citizen data for what it is and contribute towards stronger collaboration between state and non-state partners.

### *3.2 Fostering collaboration between state and non-state partners*

On the topic of collaboration, many respondents lament that there is a lack of political will, a legislative base, and a culture in countries for citizen data to be incorporated into the official data. Therefore, it is important to foster a cultural shift in the statistical community to acknowledge and value citizen data for what it is. This means accepting its limitations or differences in comparison to official data sources and understanding its respective added value according to the nature of the citizen data. This may entail:

- Adopting a broader mandate for NSOs on data stewardship, such as embracing new data sources (including the modernization of their legal and policy framework for statistics.)
- Establishing a legal framework to clarify roles and responsibilities of all stakeholders when working with non-traditional data sources (e.g., data ownership by CSOs and conditions regarding proper data usages by NSOs and others).
- Nurturing trust across different entities through a third/impartial party that both sides trust. Academics and scientists can play a role here. For instance, the coordinated work of state and non-state actors in Ghana’s marine plastic litter monitoring effort had achieved

a remarkable result. Citizen science data has been incorporated by Ghana Statistical Service in their official monitoring and reporting of SDG and also in relevant government policies in the country.

- At the international level, promoting and inspiring data collaborations through financial support and through fostering a structured spaces for exchange of good practices. This can also help steer actors of national level and offer technical support where needed.
- Improving accessibility of official data for citizens and/or CSOs to explore through opening official data and providing open-source software.
- Improving dissemination and communication strategies for citizen data. Two respondents suggest that lack of awareness is a barrier for participation, especially if “people do not know if and how their data has been used and whether it has made a difference.” The lack of awareness also applies to NSOs as they may not be able to locate citizens and CSOs to work with on specific topics of concerns. It may also be worthy here to note that NSOs is one among various state actors, including local and national governments and the various branches. As such, the culture shift should involve state actors in systemic manner and not focused solely on statistical offices at the national level.

### 3.3 *Improving capacity of NSOs and CSOs*

There have been various discussions on the need for capacities building of CSOs and NSOs to foster better utilization of citizen data and being incorporated into official statistics, where appropriate.

For CSOs, what came out clearly was the training on statistical literacy. To produce data, citizens and groups tend to not have a good understanding of official statistics or the indicators of quality required. Capacity-building on statistical literacy can be directed towards data collection as well as data usage, as both dimensions are connected. Citizens who are able to understand existing official statistics well would be in a better position to contribute to data planning and data collection.

For NSOs, capacity needs to be improved on (a) how to approach and engage with CSOs; and (b) how to assess quality and integrate citizen data into official statistics. Financial support to NSOs to allow for additional resources (including additional manpower) to engage with citizens was also mentioned.

Additionally, one respondent points out that different levels of decision-making must be addressed. WNTA India suggested that their information and policy briefs are used by the community to advocate with the government at the local level, but it was still a challenge to gain inroad into the government at the state or national levels. Related to this point, it is crucial to recognize that official data may also be derived from local units (e.g., city or province government), therefore NSOs should be placed within an interactive network of state actors and not as the sole representative for state actors.

Finally, the sustainability of citizen data projects also need to be considered. The conceptual framework may help to reinforce the fit-for-purpose as one key indication of success, helping CSOs and citizens themselves to establish a better sense for their projects' growth and development given the core purpose.

### *3.4 Improving the quality of engagement: how to gain trust from and engage with citizens?*

Considerations about quality of engagement tends to be viewed as separate from the data quality discussion, however, several respondents point out that the ability to work with and collaborate with citizens is key for the field's future developments. For many stakeholders, engaging with citizens to enrich official data or to inform policy is still a new concept that is yet to be embraced by citizens themselves. It requires a lot of effort to explain the concept, methods, and objectives to gain trust from citizens.

On this point, the respondents also suggested some strategies:

- Ensuring more inclusiveness by thoroughly understanding citizens, their attitudes, motivations and identifying the topics, communications channels and influencers that are most relevant to them.
- Engaging citizens in project designs, interpretation of results and all other stages of the data value chain.
- Nurturing the growing appreciation of data among citizens through campaign, education and training opportunities.
- Building trust with citizens and maintaining public confidence of the institutions, including NSOs and CSOs.
- Involving CSOs in the process, especially at the grassroots level, to make widespread the concept to the community, especially the marginalized population, to encourage their engagement.
- Continuing piloting in different contexts to demonstrate the value of citizen data and gradually popularize the concept to the community.
- Making official data more open & data dissemination process more transparent. Consider opening up synthetic data to make available data that are otherwise off limit for confidentiality reasons
- Offering/promoting the use of open-source software for citizens to explore. Build on and expand training for data literacy and data appreciation.

## **5. Key action points: fostering more engagement of citizens and collaboration of the state and non-state partners**

The key action points outlined below are extracted from the experts' contribution, on what can be done to foster more engagement of citizens in data and improve the collaboration of the state and non-state partners. This is to achieve the ultimate objective: to fill data gaps, to support data-drive policymaking and to make data and the data process more inclusive.

- 5.1. Establishing a common framework on types of citizens' contribution to the data value chain, related terminology and quality requirements
- 5.2. “Accreditation” of data quality
  - 5.2.1 Countries might consider forming “accreditation” entities to assess certain level of awareness of standards with non-traditional data sources (more relaxed), as opposed to using the official statistics quality assurance framework.
  - 5.2.2 At the global level, thematic groups could be formed to review and endorse non-traditional data sources and their methods on specific thematic areas.
- 5.3. Fostering a culture shift in how different communities work together
- 5.4. Providing guidance to NSOs, CSOs and other entities
  - 5.4.1. Ensuring that the guidance is scalable and aligned with the diverse knowledge and practice of civil society data-generation, and with the existing practices at the national and local level
- 5.5. Knowledge-sharing
  - 5.5.1. At the international level, providing a sustainable peer-learning mechanism such as a Community of Practice, for countries to exchange best practices. Examples of initiatives with cross-cutting objectives and a diverse set of stakeholders can help envision the community of practice for the field moving forward. This can also help steer actors of national level, offer technical support where needed and identify areas that need guidance.
- 5.6. Consider bringing the topic to UN Statistical Commission.



## References

- [1] OECD. Development Co-operation Report 2017: Data for Development. OECD; 2017. <https://doi.org/10.1787/dcr-2017-en>.
- [2] Proden E, Bett K, Chen H, Duerto Valero S, Fraisl D, Gamez G, et al. Citizen science data to track SDG progress: Low-hanging fruit for Governments and National Statistical Offices 2022. <https://iiasa.dev.local/> (accessed August 18, 2022).
- [3] UNECE Secretariat. Generic Statistical Business Process Model (GSBPM) Version 4.0. Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS): 2009.
- [4] Jungcurt S. Citizen-Generated Data: Data by people, for people. *Int Inst Sustain Dev* 2022.
- [5] Haklay M, Fraisl D, Tzovaras BG, Hecker S, Gold M, Hager G, et al. Contours of citizen science: a vignette study 2021:24.
- [6] Bonney R, Ballard H, Jordan R, McCallie E, Phillips T, Shirk J, et al. Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. 2009.
- [7] Shirk J, Ballard H, Wilderman C, Phillips T, Wiggins A, Jordan R, et al. Public Participation in Scientific Research: a Framework for Deliberate Design. *Ecol Soc* 2012;17. <https://doi.org/10.5751/ES-04705-170229>.
- [8] Bonney R. Citizen science: a lab tradition. *Living Bird* 1996;15.
- [9] Del Savio L, Prainsack B, Buyx A. Crowdsourcing the Human Gut. Is crowdsourcing also “citizen science”? *J Sci Commun* 2016;15:A03. <https://doi.org/10.22323/2.15030203>.
- [10] Chen S-L. How Empowering Is Citizen Science? Access, Credits, and Governance for the Crowd. *East Asian Sci Technol Soc Int J* 2019;13:215–34. <https://doi.org/10.1215/18752160-7497711>.
- [11] Lukyanenko R, Wiggins A, Rosser HK. Citizen Science: An Information Quality Research Frontier. *Inf Syst Front* 2020;22:961–83. <https://doi.org/10.1007/s10796-019-09915-z>.
- [12] Haklay M. Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In: Sui D, Elwood S, Goodchild M, editors. *Crowdsourcing Geogr. Knowl.*, Dordrecht: Springer Netherlands; 2013, p. 105–22. [https://doi.org/10.1007/978-94-007-4587-2\\_7](https://doi.org/10.1007/978-94-007-4587-2_7).
- [13] Freitag A. A typology for strategies to connect citizen science and management. *Environ Monit Assess* 2016;188:519. <https://doi.org/10.1007/s10661-016-5513-y>.
- [14] Gharesifard M, Wehn U, van der Zaag P. Towards benchmarking citizen observatories: Features and functioning of online amateur weather networks. *J Environ Manage* 2017;193:381–93. <https://doi.org/10.1016/j.jenvman.2017.02.003>.
- [15] CIVICUS. What is citizen-generated data and what is the DataShift doing to promote it? 2015.
- [16] The Danish Institute for Human Rights. Human rights and data. 2017.
- [17] Jakarta PL. Alternative Use of Traditional Data in Times of COVID-19. *Pulse Lab Jkt* 2021. <https://medium.com/pulse-lab-jakarta/alternative-use-of-traditional-data-in-times-of-covid-19-558914f23f7f> (accessed August 19, 2022).
- [18] Weckel ME, Mack D, Nagy C, Christie R, Wincorn A. Using Citizen Science to Map Human–Coyote Interaction in Suburban New York, USA. *J Wildl Manag* 2010;74:1163–71. <https://doi.org/10.2193/2008-512>.
- [19] Cox TE, Philippoff J, Baumgartner E, Smith CM. Expert variability provides perspective on the strengths and weaknesses of citizen-driven intertidal monitoring program. *Ecol Appl* 2012;22:1201–12. <https://doi.org/10.1890/11-1614.1>.

- [20] Fehri R, Bogaert P, Khelifi S, Vanclooster M. Data fusion of citizen-generated smartphone discharge measurements in Tunisia. *J Hydrol* 2020;590:125518. <https://doi.org/10.1016/j.jhydrol.2020.125518>.
- [21] Danylo O, Moorthy I, Sturn T, See L, Laso Bayas J-C, Domian D, et al. The Picture Pile Tool for Rapid Image Assessment: A Demonstration Using Hurricane Matthew. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2018;IV-4:27–32. <https://doi.org/10.5194/isprs-annals-IV-4-27-2018>.
- [22] Mancini T, Imperato C. Can Social Networks Make Us More Sensitive to Social Discrimination? E-Contact, Identity Processes and Perception of Online Sexual Discrimination in a Sample of Facebook Users. *Soc Sci* 2020;9:47. <https://doi.org/10.3390/socsci9040047>.
- [23] Tironi M, Valderrama M. Unpacking a citizen self-tracking device: Smartness and idiocy in the accumulation of cycling mobility data. *Environ Plan Soc Space* 2018;36:294–312. <https://doi.org/10.1177/0263775817744781>.
- [24] Fritz S, See L, Carlson T, Haklay M (Muki), Oliver JL, Fraisl D, et al. Citizen science and the United Nations Sustainable Development Goals. *Nat Sustain* 2019;2:922–30. <https://doi.org/10.1038/s41893-019-0390-3>.
- [25] Malik K, McLeman R, Robertson C, Lawrence H. Reconstruction of past backyard skating seasons in the Original Six NHL cities from citizen science data. *Can Geogr Géographe Can* 2020;64:564–75. <https://doi.org/10.1111/cag.12640>.
- [26] National Academies of Sciences E. Learning Through Citizen Science: Enhancing Opportunities by Design. 2018. <https://doi.org/10.17226/25183>.
- [27] Davids JC, Devkota N, Pandey A, Prajapati R, Ertis BA, Rutten MM, et al. Soda Bottle Science—Citizen Science Monsoon Precipitation Monitoring in Nepal. *Front Earth Sci* 2019;7:46. <https://doi.org/10.3389/feart.2019.00046>.
- [28] Hsu A, Weinfurter A, Tong J, Xie Y. Black and Smelly Waters: how citizen-generated transparency is addressing gaps in China’s environmental management. *J Environ Policy Plan* 2020;22:138–53. <https://doi.org/10.1080/1523908X.2019.1654365>.
- [29] Delaney DG, Sperling CD, Adams CS, Leung B. Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biol Invasions* 2008;10:117–28. <https://doi.org/10.1007/s10530-007-9114-0>.
- [30] United States Environmental Protection Agency (EPA). Handbook for Citizen Science Quality Assurance and Documentation 2019.
- [31] Lepenies R, Zakari IS. Citizen Science for Transformative Air Quality Policy in Germany and Niger. *Sustainability* 2021;13:3973. <https://doi.org/10.3390/su13073973>.
- [32] Palmer JRB, Oltra A, Collantes F, Delgado JA, Lucientes J, Delacour S, et al. Citizen science provides a reliable and scalable tool to track disease-carrying mosquitoes. *Nat Commun* 2017;8:916. <https://doi.org/10.1038/s41467-017-00914-9>.
- [33] Lewis J, Boudreau C, Patterson J, Bradet-Legrís J, Lloyd V. Citizen Science and Community Engagement in Tick Surveillance—A Canadian Case Study. *Healthcare* 2018;6:22. <https://doi.org/10.3390/healthcare6010022>.
- [34] Steinke J, van Etten J, Zelan PM. The accuracy of farmer-generated data in an agricultural citizen science methodology. *Agron Sustain Dev* 2017;37:32. <https://doi.org/10.1007/s13593-017-0441-y>.
- [35] van der Velde T, Milton DA, Lawson TJ, Wilcox C, Lansdell M, Davis G, et al. Comparison of marine debris data collected by researchers and citizen scientists: Is citizen

science data worth the effort? *Biol Conserv* 2017;208:127–38.  
<https://doi.org/10.1016/j.biocon.2016.05.025>.

- [36] Misra A, Schmidt J. Enhancing trust in data – participatory data ecosystems for the post-COVID society. OECD; 2020.
- [37] United Nations. National Quality Assurance Frameworks Manual for Official Statistics: Including Recommendations, The Framework and Implementation Guidance. New York: United Nations; 2019.

## Annex 1

### **Harnessing data by citizens for public policy and SDG monitoring**

### **Questionnaire on concept, definition and how citizens are engaged**

#### Background

The questionnaire is prepared in the context of the upcoming United Nations Expert Group Meeting Harnessing data by citizens for public policy and SDG monitoring: a conceptual framework, 10-11 November 2022, with the following objectives:

- Foster a better understanding on the type of citizen data that you have been working with, the term you use, how it is defined and the values and challenges in working with data generated by citizens
- Support the development of a background paper that will be used to facilitate the discussion during the Expert Group Meeting
- Identify areas for which support is needed from the international community

Your contribution is very important to our work in formulating a conceptual framework for citizen data and identifying additional work from the global statistical community to better leveraging the contribution of citizens for public policy and SDG monitoring.

#### Introduction about your organization

1. Name of the organization
2. Focal point: contact information
3. Briefly describe the work on Citizen Data (find a broader term) in your organisations

#### Think about the projects that you have been supporting/working on:

1. What are the objectives in engaging with citizens (check all that apply):
  - Advocacy
  - Research
  - Policy monitoring
  - Measuring accountability
  - Others, please specify:
2. Which SDG Goal(s) or cross-cutting issues citizen engagement is supporting?

3. Stages of the data value chain for which citizens are engaged (check all that apply):
  - Initial consultation/community engagement
  - Data collection
  - Data processing

- Data analysis
- Connect/Engage with data users/citizens or communities on data use
- Making an impact/informing policy
- Others, please specify:

4. Please specify term(s) (check all that apply) you use for the engagement of citizens; for all terms selected above, please provide your definition and if available, share the documents that those definitions have been covered/discussed.

Citizen-generated data, please provide definition and link to document(s) with discussion on the definitions

Citizen science data, please provide definition and link to document(s) with discussion on the definitions

Crowdsourcing, please provide definition and link to document(s) with discussion on the definitions

Community-driven data, please provide definition and link to document(s) with discussion on the definitions

Others, please specify. Also , please provide definition and link to document(s) with discussion on the definitions

5. What are the key lessons learned and success stories about engaging with citizens and in connecting citizens' contribution to public policy and/or official statistics?

6. What are the key challenges you have encountered in engaging with citizens and in connecting citizens' contribution to public policy and/or official statistics?

7. What can we do to overcome challenges? What additional work can be done at the global level?

Preparation for the Expert Group

8. What are the key areas you believe we should be covering during the meeting (draft concept note for the meeting is attached in Annex A).

9. Please list additional partners we should bring on board for the meeting

Partners and follow-ups

10. Provide a list of partners that you think we should be reaching out to respond to the questionnaire

11. Provide additional documents, reference materials and links to additional materials that would support the discussion on citizen engagement.

12. Are you available for a follow-up discussion in August/September 2022?

Yes

No