Special Collection "Contributions of Citizen Science to the SDGs and other International Development Frameworks of Citizen Science: Theory and Practice"

Supplemental File 1. Appendix A to Citizen science: What is it in it for the official statistics community?

[1]Elena Proden, [2]Dilek Fraisl and [2]Linda See

1 United Nations Institute for Training and Research (UNITAR), Geneva, Switzerland

2 International Institute for Applied Systems Analysis (IIASA), Laxenburg, A-2361, Austria

Copy of the questionnaire for Official Statistics Track A

Annex 1. Questionnaires for official statistics community, policy-makers/data users and citizen scientists

Dear respondent,

This survey is designed to help improve the understanding of the current awareness around the use of citizen science data for monitoring the SDGs, identify good practices, successful examples and limitations, examine key data quality considerations in relation to citizen science data and explore how some of the obstacles preventing a broader and more effective use of citizen science data could be addressed upfront.

The survey will take around 10-15 min. to complete. It includes three separate tracks: one for the representatives of the official statistics community, one for data users such as policy-makers, planners, civil society representatives, and one for those who can be broadly considered member of the citizen science community. The survey is voluntary and anonymous unless you decide to provide some optional information such as your country and contact details. You don't have to indicate them if you don't wish to. If you wish to share a good practice or a case study but want to preserve anonymity for the rest of the survey responses, you will be given an option to contact us separately by email at a2030@unitar.org instead of providing your contact details/email address.

This survey is conducted by UNITAR as part of the CROWD4SDG project funded by the European Commission and benefits from advisory support of experts from UN Statistics Division, UN Convention on Biodiversity, International Institute for Applied Systems Analysis, Global Partnership for Sustainable Development Data, UN Women and Citizen Science Center Zurich.

The results of the survey will be used to prepare recommendations for National Statistical Systems to leverage citizen science data more effectively for monitoring the SDGs. They will be published in the dedicated Crowd4SDG report and policy brief, used for guiding the work on citizen science pilots and will inform the feedback that will be provided for the citizen science data related innovations developed as part of the CROWD4SDG project.

We thank you again for your valuable answers.

Section 0. Basic information

What is your sex?

What is your year of birth?

What group do you belong to? Please note that you will be taken to the relevant track based on your answer below. If you feel you belong to more than one, kindly chose the one that describes the bulk of your daily activities.

- Representative of the official statistics community
- Data user / policy-maker / non-official data producer/provider
- Representative of the citizen science community

For how long have you been a member of this community?

FOR OFFICIAL STATISTICS (TRACK A)

Section 1. Your role

Please indicate which part of the official statistics community you are from:

- National Statistical Office (NSO)
- National Statistical System outside of NSO – Sectoral statistics
- National Statistical System – Sub-national level
- International Organization
- Regional Organization

Please describe below in 1/2 sentences one or two main functions you have (e.g., management roles and/ or technical specialization – what - and area of statistical activity, e.g., demographer, as applicable).

_____

Section 2. Familiarity with non-traditional data sources and citizen science data

This section explores your familiarity and experience with non-traditional data sources. It is assumed that some of the respondents may have only a vague idea about some of the non-traditional data sources, others - only conceptual understanding while still others may have direct experience of running experimental data projects.

For the purposes of this questionnaire, under non-traditional data sources we will include mobile phone tracking data, call detail records, textual data (e.g., web-scrapping and social media), sensor data, etc. Some of the other non-traditional data sources with well-established practices and already broadly used by NSOs such as administrative data or earth observation data will be excluded.

(Earth observation will be considered here only when used in combination with citizen science data, i.e., where volunteers participate in one way or another in the production of data from earth observation (satellite imagery)).

How would you qualify your **level of familiarity** with **non-traditional data** sources (administrative and earth observation data excluded)?

- Rudimentary.
- Basic awareness but no direct or indirect experience.
- Some indirect experience (have seen or witnessed successful examples).
- Direct experience (contributed to projects involving non-traditional data sources).
- This is one of my main areas of responsibility / I have led one or more projects involving non-traditional data sources.

What kind of non-traditional data sources were used in the **successful examples you may be aware of**?

What kind of **non-traditional data** sources have you worked with directly or indirectly?

How would you rank your level of knowledge about **potential use of the following types of data** for SDG indicators?

How would you **describe citizen science data**? Please also provide examples of such data that could be useful for SDGs.

The project under which this study is undertaken aims to connect the official statistics and citizen science communities by raising awareness on both sides. As part of this survey, we are collecting baseline information on the awareness around the potential of citizen science data within the official statistics community. To help us in this task, we would be grateful if you could answer the following question: Based on your current experience, knowledge or information you may have heard/read, please select **all statements** that you think are true from the ones below:

- None of the SDG global indicators are using citizen science data yet.
- Several global SDG indicators are being tracked already with the help of citizen science data.
- Citizen science data are data from community monitoring activities.
- One example of citizen science data is citizens' taking samples of water or monitoring air to measure water and air pollution respectively.
- On example of citizen science data is truth-checking on the ground to confirm results of satellite imagery analysis in cases of damage assessment.
- One example of citizen science data is classifying images from social media to help build machine learning algorithm to support real time social media analysis.
- Citizen science data could be potentially used to support data production for more than 70 SDG indicators.
- Citizen science data can be helpful for environmental indicators but less so in other areas.
- Citizen science data cannot be used by NSOs because there are concerns related to data quality.
- Citizen science data are validated by the scientific community.
- Citizen science data can only be used for some of the global SDG indicators.

Section 3. State of play on the use of citizen science data

Citizen science projects are defined in this and several other studies as "intentional collaborations in which members of the public engage in the process of research to generate new science-based knowledge" (Shirk et al., 2021) including any initiative that produces scientific knowledge through the participation of volunteers, such as, among others (Fraisl et al., 2020):

- community-based monitoring (environmental monitoring and adaptive management activities by citizens in local communities as studied Conrad and Hilchey 2011),
- community-based participatory research (knowledge co-creation with researchers, practitioners and community members working together to address issues relevant to communities, often those with a history of marginalization, Asaba and Suarez-Balcazar 2018),
- participatory action research (qualitative research methodology where communities are involved in all stages of defining the research process and with social change as main goal that played already an important role in the area of education, MacDonald 2012),
- citizen-generated data sets ((*data that people or their organisations produce to directly monitor, demand or drive change on issues that affect them, actively given by citizens, providing direct representations of their perspectives,* Datashift 2017),
- crowdsourcing (outsourcing the function to a lrage, undefined group of individuals through an open call, Howe 2006),
- volunteered geographic information (digital geographical information that is generated and shared by individuals, Sieber and Haklay 2015), and
- participatory sensing (an open practice of data capture, analysis and sharing through digital devices and platforms, Ganti et al. 2011; Bria et al. 2015, Coulson et al. 2018),, .

In the context of this study, the above includes activities related to data collection and data analysis (the latter understood as the human recognition and classification of imagery, for example, in conjunction with machine learning algorithms for social media or other related activities or confirmation on the ground).

Are you aware of any **projects run by your Organization** involving the use of **citizen science data**? If yes, please describe each project in 1-2 sentences mentioning relevant SDG or national indicators and possible partner Organization/s facilitating citizen scientists' engagement.

Are you aware of any **projects run by your Organization** involving the use of **other non-traditional data sources**? If yes, please provide the data sources and describe each project in 1-2 sentences mentioning relevant SDG or national indicators.

If any of citizen science data projects are already considered **a success story**, would your Organization be interested in having it featured in the good practices and policy brief prepared by Crowd4SDG / UNITAR in collaboration with UNSD, GPSDD, IIASA, UN Women and UNCBD?.

If yes, please provide us with the contact email address or send an email to a2030@unitar.org .

Are there any other **successful citizen science projects run by official statistics producers** in other countries or by other Organizations you would like to recommend for this "good practice" publication? If yes, please mention the country name, the Organization, contact persons' details and the specific area of the project/data it helps to produce.

If yes to first question in section 3,

What has **sparked this collaboration** between NSO/other official statistics producer and citizen science community?

What factors, in your view, explain the **interest** and **success of the citizen science community** in addressing this measurement problem/indicator?

What factors, in your view, explain the **high level of engagement of the citizen science community in your country on data issues overall**?

Section 4. Impediments, limitations and quality considerations for the use of citizen science data

In this section, you will be asked questions pertaining to quality considerations for the use of citizen science data based on your experience with citizen science data, if any, or other relevant experience you may have pointing to possible impediments and limitations. Questions will be adapted based on your previous responses.

<u>If yes to first question in section 3 (existing citizen science data projects),</u>

Did you encounter any of the below challenges with the **quality of data** generated by these **citizen science projects**? For your selected challenges, please briefly describe the project/s and relevant indicator/s.

- Limited access to data
- Legal issues with access or use of data
- Sustainability of access to the data source
- Incoherent use or lack of use of statistical standard concepts, definitions and classifications undermining accuracy, reliability, coherence and comparability of the resulting statistics
- Non-compliance with UN Fundamental Principles of Official Statistics such as professional independence and commitment to quality;
- Utilization of data may put confidentiality and privacy of individuals, households and businesses at risk;
- Selection bias due to data not being representative;
- Interference and bias in the statistical production process;
- A lack of information about how the data are being produced;
- The data may not provide the information required by users;
- The data may not correctly describe the phenomena that are to be measured;
- Lack of IT capacities
- Lack of human capacities or relevant skills
- Inability to quantify uncertainty
- Insufficient information to apply weights
- Incomplete/partial information for some entries
- Other, please specify

How did you/your colleagues **address** the above challenges?

Did you encounter **other challenges** as part of these **citizen science projects** in general? Please specify which challenge corresponds to what project and what indicator if more than 1.

How did you/your colleagues **address** the above challenges?

Did data providers contribute to **methodological decisions** regarding the use of the above data?

Do you have **specific rules of access and confidentiality measures** to treat these datasets?

No

Yes, please specify

<u>If yes to second question in section 3 (other types of non-traditional data sources),</u>

Did you encounter any of the below challenges with the **quality of data** generated by these other **non-traditional data source-based** projects? For your selected challenges, please briefly describe the project/s and relevant indicator/s.

- Limited access to data
- Legal issues with access or use of data
- Sustainability of access to the data source
- Incoherent use or lack of use of statistical standard concepts, definitions and classifications undermining accuracy, reliability, coherence and comparability of the resulting statistics
- Non-compliance with UN Fundamental Principles of Official Statistics such as professional independence and commitment to quality;
- Utilization of data may put confidentiality and privacy of individuals, households and businesses at risk;
- Selection bias due to data not being representative;
- Interference and bias in the statistical production process;
- A lack of information about how the data are being produced;
- The data may not provide the information required by users;
- The data may not correctly describe the phenomena that are to be measured;
- Lack of IT capacities
- Lack of human capacities or relevant skills
- Inability to quantify uncertainty
- Insufficient information to apply weights
- Incomplete/partial information for some entries
- Other, please specify

How did you/your colleagues **address** the above challenges?

Did you encounter **other challenges** as part of these projects in general? (Please specify what kind of data this was and related indicators)

How did you/your colleagues **address** the above challenges?

Did data providers contribute to **methodological decisions** regarding the use of the above data?

Do you have **specific rules of access and confidentiality measures** to treat these datasets?

<u>For all,</u>

In your view, what are the **main impediments** to the **broader use of citizen science** data by NSOs or NSSs related to the fitness for purpose or quality of data?

What are the **main impediments to the broader use of citizen science data** by NSOs or NSSs related to other factors?

- lack of awareness
- fear of misinterpretation and misuse for citizen science data and other non-traditional data sources
- absence of vibrant citizen science community
- irrelevance of citizen science data to main data requirements
- fear of biasedness due to design of data collection
- inability to ensure the use of statistical standard concepts, definitions and classifications
- technological limitations
- lack of methodological guidance

- lack of human capacities to run experimental statistics projects
- lack of technical and financial support to use citizen science data for the production of official statistics
- statistical legislation does not enable/ prevents engagement
- normative loophole: no mechanism or protocols on key issues related to confidentiality, mandate, impartiality, etc.
- sustainability of access to the data source
- other, please specify

Please explain in more detail some of the obstacles you have personally encountered.

Are representatives of citizen science, civil society or organizations that generate other types of innovative data sources invited to participate in the deliberations of your Statistical Council or in the user-producer dialogues to inform the production of official statistics?

The UN National Quality Assurance Frameworks for Official Statistics identifies requirements along 3 areas: Outputs, Processes, Enabling Environment. It also recommends that the national quality assurance framework be applied to all data and statistics produced outside of the national statistical system that are disseminated with the help and support of a member of the national statistical system or that are used for government decision-making

What kinds of quality issues do you expect to arise from citizen science projects in relation to the following criteria**:**

Outputs:

- **Relevance**
- **Accuracy and reliability**
- **Timeliness and punctuality**
- **Coherence, comparability and integrability**
- **Accessibility and clarity**

Processes:

- **Sound methodology**
- **Appropriate statistical procedures**
- **Non-excessive burdens on respondents**
- **Cost effectiveness**

Institutional environment:

- **Professional independence**
- **Coordination**
- **Mandate for data collection and access to data**
- **Adequacy of resources**
- **Commitment to quality**
- **Statistical confidentiality**
- **Impartiality and objectivity**

Please explain your answer above.

Section 5. Needs and opportunities

This is the last section of the survey dedicated to the opportunities that citizen science represent in meeting specific SDG data needs relevant in your country/area of work.

What **global SDG indicators** are you aware of that **already benefit** from citizen science data?

- dropdown list of all indicators
- None

What **other kinds of opportunities** does citizen science data offer for monitoring the SDGs?

- Providing more granular data (any specific indicators in mind?)
- Providing timelier data (any specific indicators in mind?)
- Providing data on indicators where there are significant gaps (any specific indicators in mind?)
- Providing data on indicators measuring the perception of citizens on various aspects (any specific indicators in mind?)

Please explain your response above.

In your view, what other SDG goals could benefit from citizen science data? Do you see any opportunities when it comes to tracking progress on indicators under:

GOAL 1: No Poverty

GOAL 2: Zero Hunger

GOAL 3: Good Health and Well-being

GOAL 4: Quality Education

GOAL 5: Gender Equality

GOAL 6: Clean Water and Sanitation

GOAL 7: Affordable and Clean Energy

GOAL 8: Decent Work and Economic Growth

GOAL 9: Industry, Innovation and Infrastructure

GOAL 10: Reduced Inequality

GOAL 11: Sustainable Cities and Communities

GOAL 12: Responsible Consumption and Production

GOAL 13: Climate Action

GOAL 14: Life Below Water

GOAL 15: Life on Land

GOAL 16: Peace and Justice Strong Institutions

GOAL 17: Partnerships to achieve the Goal

Please explain your response above specifying indicators as relevant.

This is an optional question and you can skip it, if you wish to keep your anonymity. What country are you from?