Citizen Science: What is in it for the Official Statistics Community?

ELENA PRODEN (iD)

DILEK FRAISL (iD)

LINDA SEE (iD)

*Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

Citizen science data are an example of a non-traditional data source that is starting to be used in the monitoring of the United Nations (UN) Sustainable Development Goals (SDGs) and for national monitoring by National Statistical Systems (NSSs). However, little is known about how the official statistics community views citizen science data, including the opportunities and the challenges, apart from some selected examples in the literature. To fill this gap, this paper presents the results from a survey of NSS representatives globally to understand the key factors in the readiness of national data ecosystems to leverage citizen science data for official monitoring and reporting, and assesses the current awareness and perceptions of NSSs regarding the potential use of these data. The results showed that less than 20% of respondents had direct experience with citizen science data, but almost 50% felt that citizen science data could provide data for SDG and national indicators where there are significant data gaps, listing SDGs 1, 5, and 6 as key areas where citizen science could contribute. The main perceived impediments to the use of citizen science data were lack of awareness, lack of human capacity, and lack of methodological guidance, and several different kinds of quality issues were raised by the respondents, including accuracy, reliability, and the need for appropriate statistical procedures, among many others. The survey was then used as a starting point to identify case studies of successful examples of the use of citizen science data, with follow-up interviews used to collect detailed information from different countries. Finally, the paper provides concrete recommendations targeted at NSSs on how they can use citizen science data for official monitoring.

CORRESPONDING AUTHOR:

Elena Proden

UN Institute for Training and Research, CH

elena.proden@unitar.org

INTRODUCTION

The United Nations (UN) Sustainable Development Goals (SDGs) are a set of 17 Goals with accompanying 169 targets that aim to guide global development efforts in addressing the world's greatest challenges from hunger to climate change (UN 2015). Progress towards the SDGs is monitored through 231 indicators developed by the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) (UN 2020). Custodian agencies, which are UN bodies or other international organizations, are responsible for developing global standards and methodologies for the monitoring of these indicators and for compiling and verifying country data and submitting these to the UN Statistics Division (UNSD) (IAEG-SDGs 2018). UNSD then publishes the regional and global aggregates in the SDG Global Database along with these country data and metadata (UN 2023). Countries drive these monitoring processes through their National Statistical Systems (NSSs), which comprise the National Statistical Offices (NSOs), line ministries, and other national agencies responsible for official monitoring. This means that the countries decide which data and metadata they will share with the custodian agencies and to what extent they should be published (IAEG-SDGs 2018). Furthermore, countries can choose to submit their Voluntary National Reviews (VNRs) on the SDGs (HLPF 2023) to the High-Level Political Forum (HLPF), which is the central platform for the review of the 2030 Agenda for Sustainable Development at a national level (UN 2012).

Countries and the custodian agencies often use traditional sources of data such as censuses and surveys in their SDG monitoring and reporting activities (ISWGHS 2019). However, it is widely acknowledged that these data have some limitations, such as outdatedness and insufficient coverage both spatially and temporally, among others (Fritz et al. 2019; Dang and Serajuddin 2020; Fraisl et al. 2020). Additionally, measuring 231 indicators using only traditional sources of data can be quite costly (Fraisl et al. 2020). As a result, many countries are turning to new sources of data, such as satellite imagery, social media, and mobile data as alternative sources or to complement traditional sources of indicator data (UNSD 2022). In addition, citizen science data, which are also often referred to as citizen-generated data by NSOs and civil society organizations (CSOs) (Higgins and Cornforth 2015; PARIS21, PSA, and PSRTI 2020), have emerged as a new data source in this process (MacFeely and Nastav 2019; Campbell et al. 2020).

Citizen science broadly refers to the practice of public engagement in scientific research and knowledge production (Eitzel et al. 2017; Fraisl et al. 2022a). It has neither an agreed definition nor one single term that is used to describe a wide range of activities considered as citizen science (Haklay et al. 2021). In addition to citizen-generated data, examples of other terms used to describe citizen science activities include crowdsourcing, community science, participatory action research, and volunteered geographic information, among others (Conrad and Hilchey 2011; Eitzel et al. 2017; Haklay et al. 2021; MacDonald 2012; Shirk et al. 2012; Sieber and Haklay 2015). The emergence of these diverse terms and approaches is related to the different contexts in which they originated and to the diverse disciplines in which they are applied. Collecting biodiversity observations, recording disease symptoms, reporting on sexual harassment, and classifying galaxies are just a few examples that show the broad spectrum of citizen science. The diversity of approaches is also reflected in the methodologies that are used in citizen science, from hypothesis-driven scientific approaches to practices that involve local knowledge for addressing a political or social issue or for improving transparency and accountability of public authorities (Shirk et al. 2012; Auerbach et al. 2019; Haklay et al. 2021).

Even though the diversity of terms, definitions, and methodologies has hindered the ability to leverage citizen science data for official monitoring and official statistics, it has been acknowledged that citizen science can provide valuable data to address many of the challenges facing our world and to inform policies and actions (Fraisl et al. 2022a; Proden et al. 2022a; UNEP 2019, 2021a). For example, in their systematic review, Fraisl et al. (2020) show that citizen science data can contribute to the monitoring of 33 percent of the SDG indicators, and the goals that can benefit from citizen science data the most, in order of relevance, are SDG 15, Life on Land; SDG 11, Sustainable Cities and Communities; SDG 3, Good Health and Wellbeing; and SDG 6, Clean Water and Sanitation. For example, Fritz et al. (2019) highlighted the potential of citizen science to address data gaps related to various environmental issues covered in the SDG indicator framework, such as Indicator 14.1.1b on plastic debris density. Subsequently, the first example of the integration of citizen science beach litter data into the official statistics of Ghana occurred, which demonstrated that citizen science data can help not only to address data gaps and needs, but also mobilize action and raise awareness of global issues related to marine plastic pollution, and inform policies in a more cost-efficient way in comparison to traditional sources of data (Fraisl et al. 2022b; Olen 2022).

Although the above-mentioned studies and others have shown the potential and challenges of using citizen science data for SDG monitoring and official statistics, they have not covered the perspectives of the official statistics communities, in particular the NSSs, in a comprehensive

way. Gaining a better understanding of this perspective is one of the main objectives of the Crowd4SDG project, funded under the Horizon program of the European Union, which aimed to identify and demonstrate the potential of citizen science for SDG monitoring and implementation. Conducted as a study within the Crowd4SDG project, the aim of this paper is to assess the current awareness and the perceptions of NSSs regarding the potential of citizen science data for official monitoring and reporting purposes, including the types of data gaps that citizen science data could fill, the perceived and actual impediments faced by NSSs in the use of citizen science data, and issues related to the quality of the data. In addition, it explores the types of citizen science data that are used, common institutional set-ups that govern the relationship between the official statistics and citizen science communities, and the related enabling factors.

METHODOLOGY

There are two components to the methodology: an online survey and case studies. The online survey had three tracks: (i) the official statistics community, (ii) policy makers, and (iii) citizens. Here we present the results from the first track with input primarily from NSSs. The questions in the survey were drafted by the lead author and organized into 6 sections. These include 0. Basic information, 1. Respondent's role, 2. Familiarity with non-traditional data sources and citizen science data, 3. State of play on the use of citizen science data, 4. Impediments, limitations, and quality considerations for the use of citizen science data, and 5. Needs and opportunities. Sections 0 to 1 were designed to collect information on the profile and role of the survey respondents. Among these questions were, for example, "Please indicate which part of the official statistics community you are from" with possible options being "NSO," "NSS outside of NSO - sectoral statistics," "NSS outside of NSO – sub-national level," "International organization," and "Regional organization."

In Section 2 there were questions about awareness of and experience with nontraditional data sources (including citizen science data) to help understand how open and willing NSSs were to leverage these new data sources; this excluded administrative data and data from Earth observation because they have well-established practices and are already broadly used by NSSs. Respondents were also requested to provide their definition of citizen science data with a view to comparing a self-assessment of their familiarity with citizen science data and the current state of knowledge and definitions in this area.

Questions in sections 3 to 5 were focused mainly on citizen science data. Section 3 asked respondents to outline any projects run by their organizations that use data generated by citizens along with the enabling factors that made this possible, including the motivation of the citizen science community to contribute to official monitoring. As the main mandate of NSOs is to produce high-quality data, in section 4, questions were asked about the quality-related challenges encountered as part of the implementation of these projects. Respondents were asked to review key areas of the UN National Quality Assurance Framework for official statistics, which includes criteria across three dimensions—the enabling environment, processes and outputs, and is used by many NSOs to define their own approaches. Respondents were then asked to identify those areas where they expected to encounter challenges regarding citizen science data. They were also asked to share their views on the common impediments to the broader use of citizens science data, covering the perspectives of those who ran such projects as well as those who have not. Finally, in section 5, respondents were asked about the perceived opportunities offered by citizen science data for NSOs, both in terms of data for specific SDGs, as well as other advantages such as more disaggregated data, timelier data, or data on citizen perceptions. Respondents could also provide additional explanations and add areas of potential use beyond the SDGs. A copy of the questionnaire is provided in Supplemental File 1: Appendix A.

To provide feedback on the questionnaire, the survey and the research study an Advisory Group was set up by the lead author, which was comprised of eight representatives of the official statistics community including several UN agencies such as the UN Statistics Division (UNSD), the UN Convention on Biodiversity (UNCBD), UN Women, and the United Nations Conference on Trade and Development (UNCTAD), as well as one representative from the Global Partnership for Sustainable Development Data (GPSDD), two representatives from the International Institute of Applied Systems Analysis (IIASA) (and co-authors of this paper), and one representative from the Citizen Science Center Zurich. The consultation process on the study design and the survey questionnaire ran for three months between November 2020 and February 2021. The questionnaire was not tested before release.

The survey was then sent to the Chief Statisticians of all 193 UN Member States, 1 Observer country, and 7 British Overseas Territories via official letters by email, asking them to invite members of NSOs and NSSs to complete the online survey. The survey was conducted from 12 February to 12 March 2021. The survey was also sent on 18 February 2021 to focal points or representatives of 44 custodian

agencies for SDG indicators, 5 regional commissions and 4 regional organizations, inviting them or their colleagues to participate. There was no limit to the number of answers to the survey per country or organization, as the aim was to gather varied perspectives and insights across NSSs at a global scale on the awareness, use, and perceived potential of citizen science data. Hence, in some cases, several respondents from the same NSS participated in the survey.

The second part of the methodology involved the collection and analysis of case studies, which were used to gather additional insights into the types of data gaps that citizen science data can help to address, as well as enrich understanding of the enabling factors and major challenges related to the use of citizen science data by NSSs. The responses from the survey were used as a starting point to gather these case studies; thirteen survey respondents expressed an interest in having their citizen science initiatives included as case studies. Survey respondents whose projects fit the definition of citizen science data used in this study were then contacted, and three were subsequently interviewed. Note that some survey respondents had more than one case study to share. Other case studies were suggested by the members of the Advisory Group and the Crowd4SDG project. Information about the case studies was gathered through a set of semi-structured interviews conducted mainly through videoconferencing between March and April 2021 with the NSOs or the members of NSSs; one case study is based on a detailed country report (PARIS21, PSA, and PSRTI 2020), and another on an interview that took place in April 2022 (Proden et al. 2022a). During the interviews, participants were first invited to talk about their work related to the use of citizen science data without any prompting questions. They were then asked to clarify certain points or

discuss aspects related to the enabling factors, common impediments, and the solutions used to ensure the success of a citizen science data project.

RESULTS
RESULTS FROM THE SURVEY

The results of the survey presented here are a selection of the most relevant outcomes to the aim of this paper. More details can be found in Proden (2021a) and Proden et al. (2022a).

In total, 121 representatives from the official statistics community answered most of the questions in the survey (2). These included questions on their background; awareness about citizen science data; experience with nontraditional data sources, including citizen science data; the existence of citizen science projects run by their organizations; and the related lessons learnt. From this total, 97 people also answered the questions on the expected impediments to the use of data from citizen science and on the issues related to quality, and there were 91 responses to the potential of citizen science data to fill SDG data gaps.

More than half of the survey respondents (52%) had 10 or more years of experience working in official statistics. The largest participation was from Latin America and the Caribbean (26%) followed by Europe and Asia-Pacific, both with 11% (Figure 1). 44% of the participants chose not to provide their location in the survey as this was not mandatory to ensure a higher participation rate.

Figure 2 shows that the majority of respondents (75%) were from NSOs followed by around 18% from broader NSSs in line Ministries or at the sub-national level. Finally, there was some small representation from international and regional organizations.
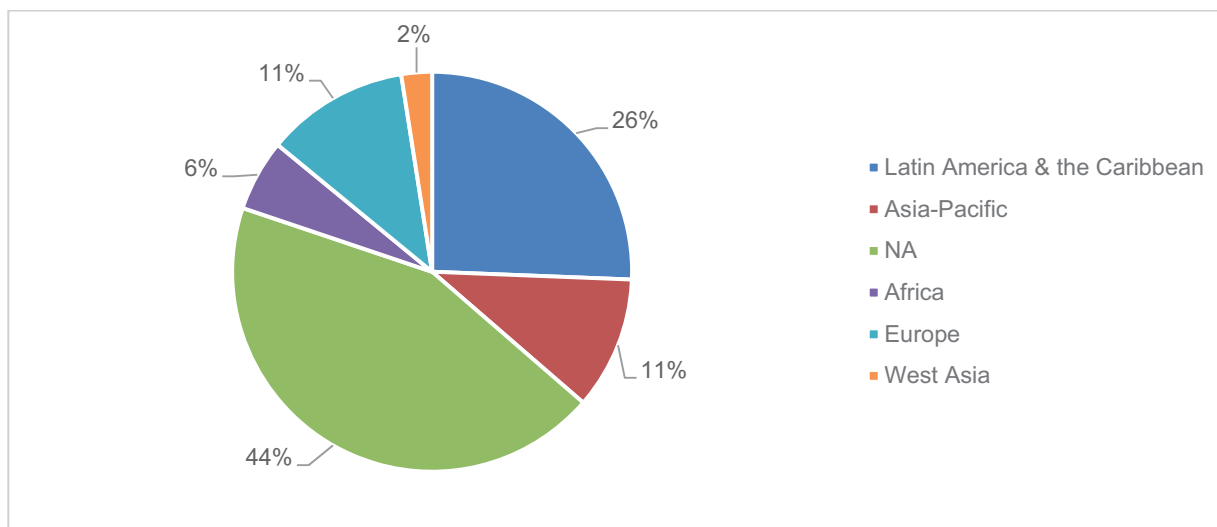

Figure 1 The regional breakdown of the survey respondents as a percentage of the total (n = 121).

Figure 3 illustrates the results related to the level of awareness of citizen science, with around 60% of respondents reporting a rudimentary or basic level of awareness and no experience with nontraditional data. Of the remaining 40%, around half of the participants had indirect experience, that is, they have seen or witnessed successful examples of the use of citizen science data, whereas the remaining half had direct experience, that is, they have contributed to projects involving nontraditional data.

To gain a better understanding of the types of nontraditional data sources that the survey respondents have worked with directly or indirectly, examples of different sources of data were provided. As summarized in Figure 4, mobile phone data was by far most used data source (26%), followed by text data (17%) and transport

data (16%). Only 13% of the respondents worked with citizen science data in the past.

Respondents were asked to clarify the actual or potential data gaps that citizen science data could help to address in terms of the SDG indicators and other areas of official monitoring. Summarized in Figure 5, the indicators from SDG 5, Gender Equality; SDG 6, Clean Water and Sanitation; SDG 13, Climate Action; SDG 1, No Poverty; SDG 15, Life on Land; SDG 11, Sustainable Cities and Communities; SDG 3, Good Health and Wellbeing; and SDG 16, Peace, Justice and Strong Institutions were identified as the SDGs with greatest potential for benefiting from citizen science data by the respondents. The lefthand column of Table S1 in Supplemental File 2: Appendix B lists individual indicators that were listed by respondents for which citizen science data could contribute or be useful. Respondents were
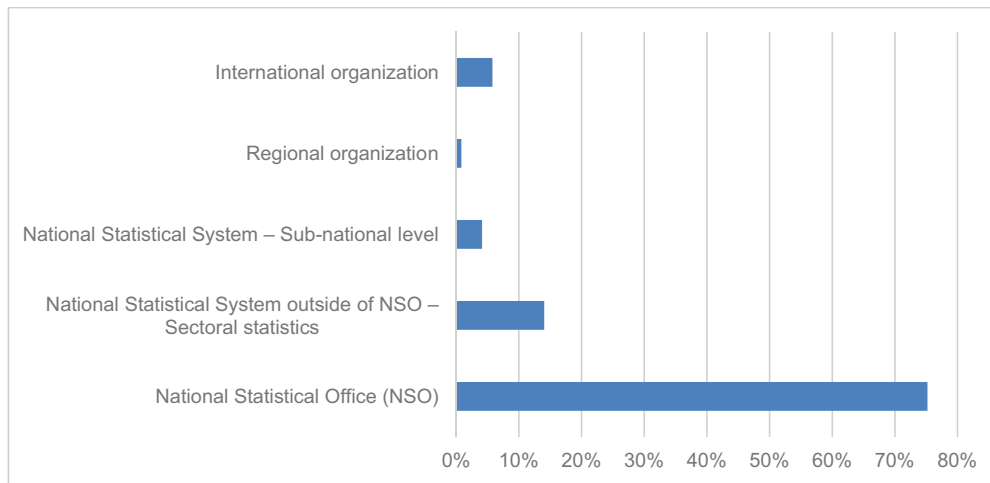

Figure 2 The distribution of survey respondents as a percentage of the total by the type of organization within the official statistics community that they are in (n = 121). Regional organizations refer to UN Regional Commissions or other regional organizations such as development banks while international organizations included UN agencies and other organizations that serve as custodians for SDG indicators.
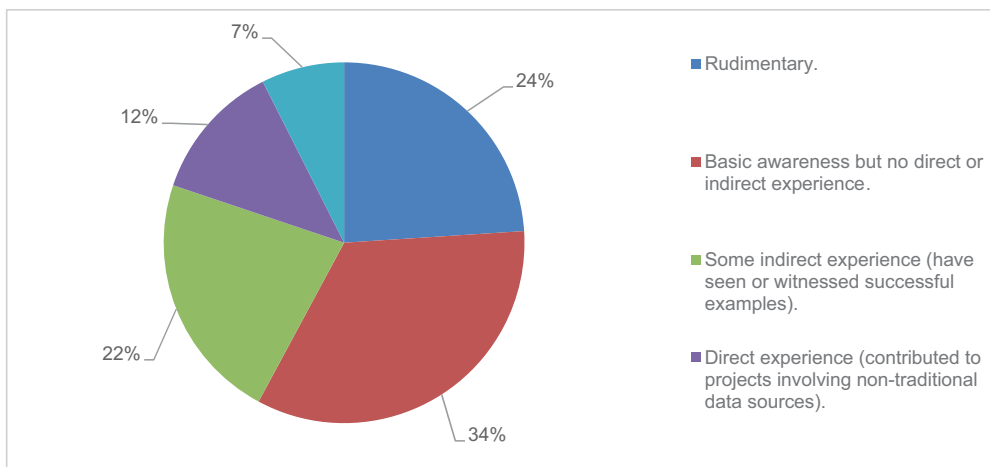

Figure 3 The degree of familiarity with nontraditional data sources (with administrative and Earth observation data excluded) among survey respondents as a percentage of total responses (n = 121).
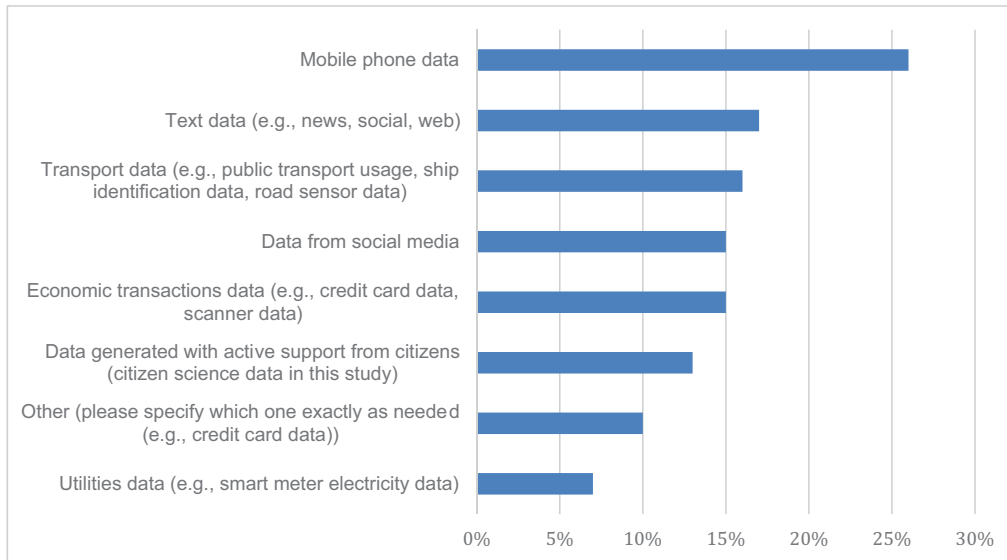
Figure 4 The types of nontraditional data sources that survey respondents have worked with, either directly or indirectly, as a percentage of the total number of respondents (n = 121).
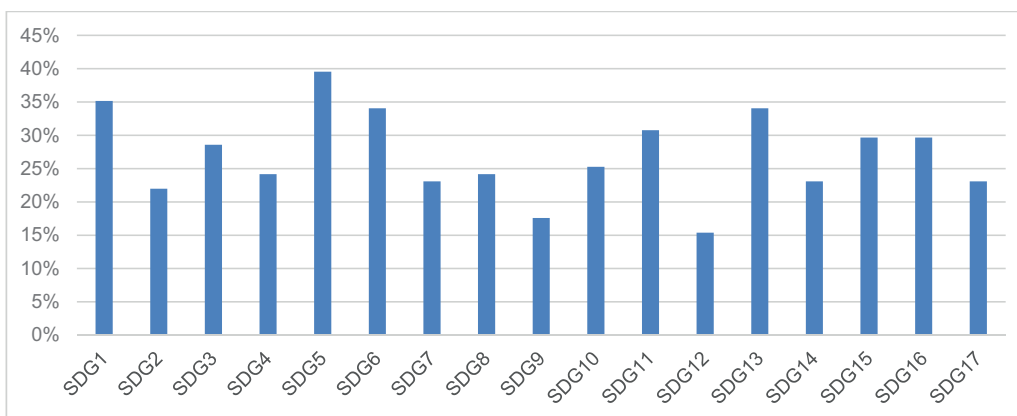

Figure 5 SDGs with indicators to which citizen science data could contribute identified by survey respondents as a percentage of the total number of responses (n = 91). SDG: Sustainable Development Goals.

also asked for general areas where citizen science could fill data gaps, which included beach litter, adequate housing, access to food, gender rights, land use, and land cover monitoring. The righthand column of Table S1 in the Supplemental File 2: Appendix B provides the responses obtained.

Figure 6 summarizes the responses regarding the perceived usefulness of citizen science data for official monitoring. Almost 50% felt that citizen science data could provide data for SDGs and national indicators where there are significant data gaps, while around 40% felt that citizen science data could be useful for indicators that involve the measurement of citizen perceptions. Increased spatial and temporal coverage were also perceived as areas that could benefit from citizen science data.

Respondents were also asked to rank the main impediments to the broader use of citizen science data as a

potential data source of official statistics. Figure 7 shows the results from 105 respondents who have not dealt with citizen science data before, so these are perceived impediments. The top impediments identified were lack of awareness of citizen science data, the lack of human capacity to use the data and the lack of methodological guidance. However, many other impediments were identified by more than 25% of respondents. The irrelevance of citizen science data was ranked near the bottom, indicating that most do not hold this view.

The key impediments identified by respondents with citizen science data experience compared with those without are somewhat different (Figure 8). Experienced respondents highlighted the key issues as potential biases in the data; uncertainty regarding the sustainability of data source access, which refers to regular data production and use so that one can construct time series and
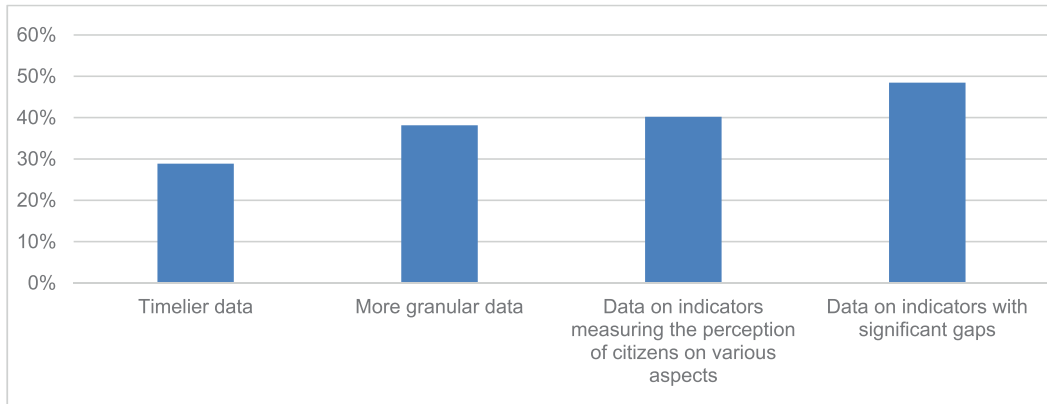
Figure 6 Ways in which citizen science data could be useful to NSSs identified by survey respondents as a percentage of the total number of responses (n = 91). NSS: National Statistical Systems.
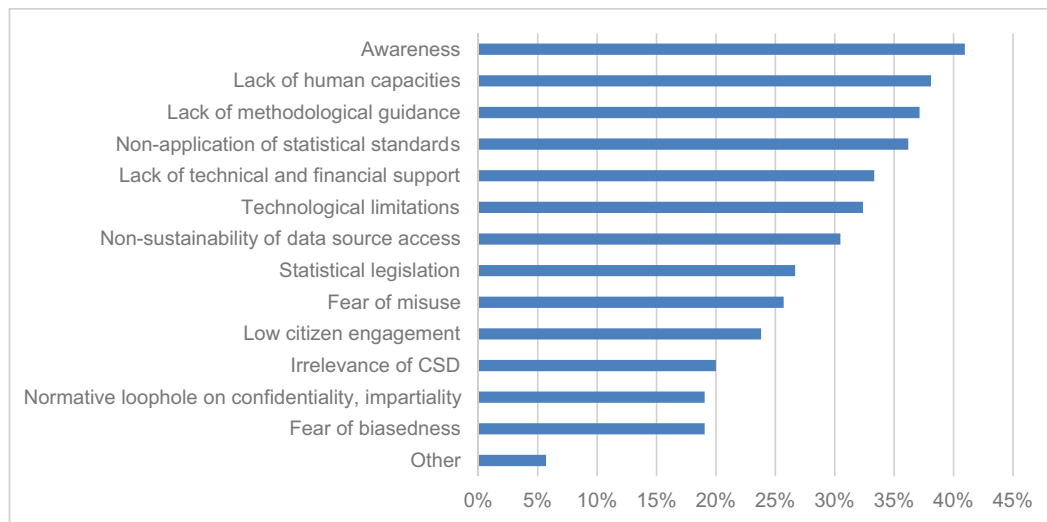

Figure 7 The perceived impediments to the use of data from citizen science by respondents with no experience as a percentage of the total number of responses (n = 105).
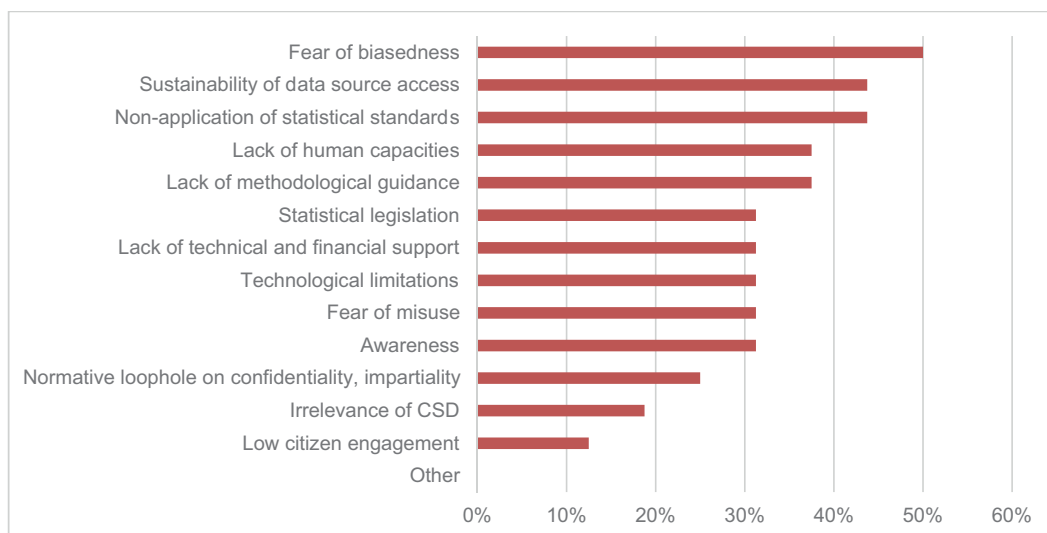

Figure 8 The identified impediments to the use of data from citizen science by respondents with experience in this area as a percentage of the total number of responses (n = 16).

compare the data over time; and the non-application of statistical standards set by NSSs, which is in contrast to the top responses from respondents with no experience with citizen science data (Figure 8), that is, the need for greater awareness and methodological guidance, and the lack of human capacity. However, the latter two are also mentioned by 37.5% of "experienced" respondents (Figure 8). Similar to those who had no experience in this area, they also believe that irrelevance of citizen science data and low citizen engagement are not as important. Note that for ease of comparison, Figures 7 and 8 have been combined into a single Figure S1 in the Supplemental File 3: Appendix C, which shows the differences in the perceptions of the two groups of respondents.

Under other possible impediments, it was noted that metadata are usually not available for the existing data sets or are not aligned with statistical standards, and there is a discrepancy between the concepts and methodology used by CSOs and the quality guidelines. Similar to those with no experience, it was noted that the lack of data source sustainability was a key consideration for NSSs to decide on the time investment required to work with citizen science data. Additional qualitative feedback from respondents with and without experience with citizen science data is included in the Supplemental File 4: Appendix D.

All respondents were asked to share their expectations in terms of the types of quality issues that could arise from citizen science projects by selecting all relevant areas from the list of the requirements within the three areas of The UN Quality Assurance Framework for Official Statistics: institutional environment, outputs, and processes (UNDESA 2019). Summarized in Figure 9,
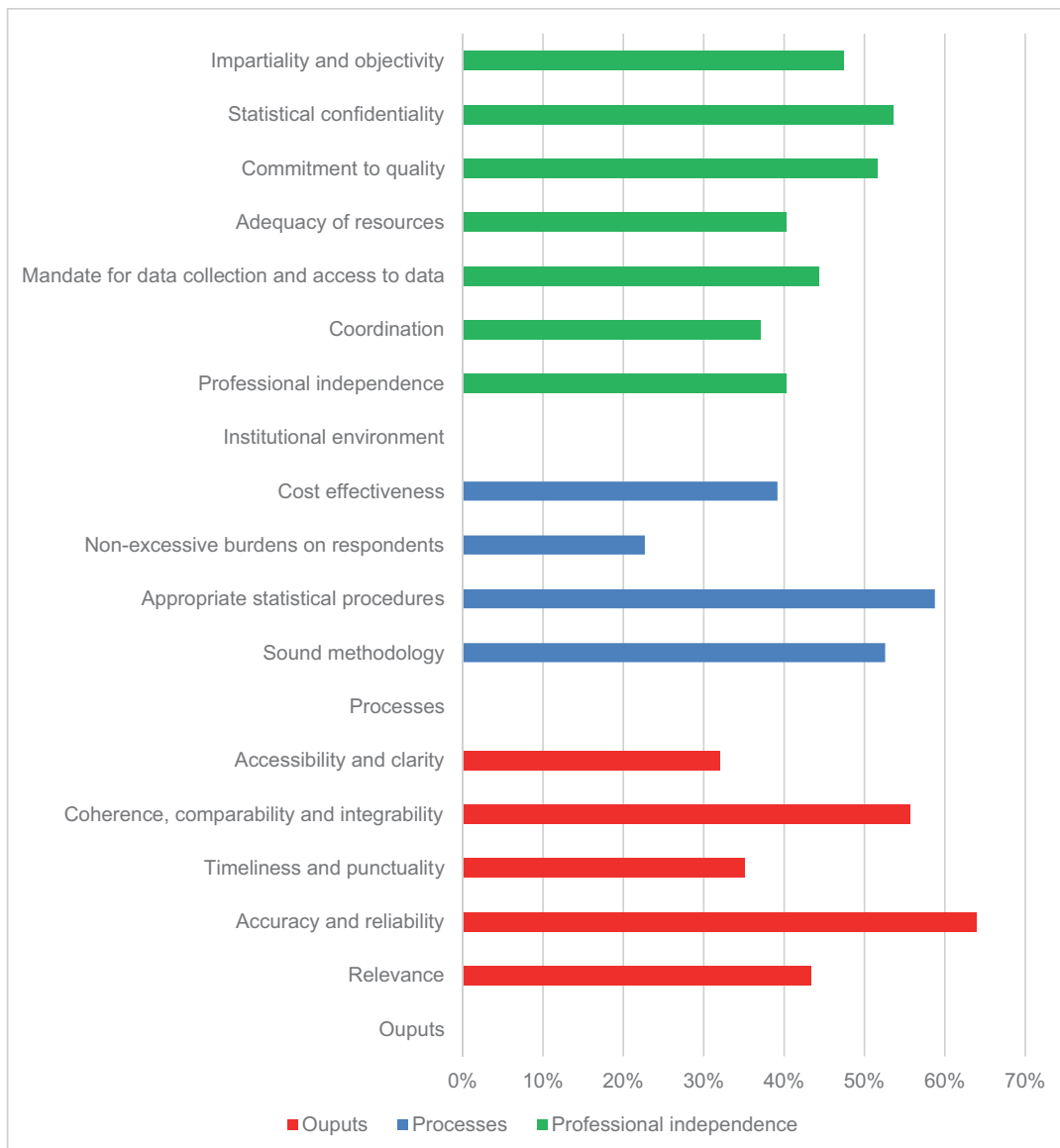

Figure 9 The types of quality issues that survey respondents expect may arise in citizen science projects, grouped into three categories: institutional environment, processes, and outputs, as a percentage of the total number of responses (n = 91).

the results show that the accuracy and reliability of the outputs is ranked as the top issue, followed by whether appropriate statistical procedures are followed, and by the coherence, comparability, and integrability of the data. The lowest-ranked issue is the non-excessive burden on participants, followed by accessibility and clarity of the outputs.

In the survey, we asked whether respondents were aware of any citizen science projects run by their organization, that is, NSOs or another government entity member of their NSS. Of the 121 responses to this question, only 17% were aware of such a project run by their organization. Hence, the question on the top challenges related to the quality of citizen science data were answered only by a subset of respondents (Figure 10). These challenges were a lack of information about the data production process, selection biases (e.g., only certain types of individuals may respond), and legal issues related to the access or use of data. The other two issues common to such projects were limited access to data, and incoherent or lack of use of statistical standard concepts, definitions, and classifications.
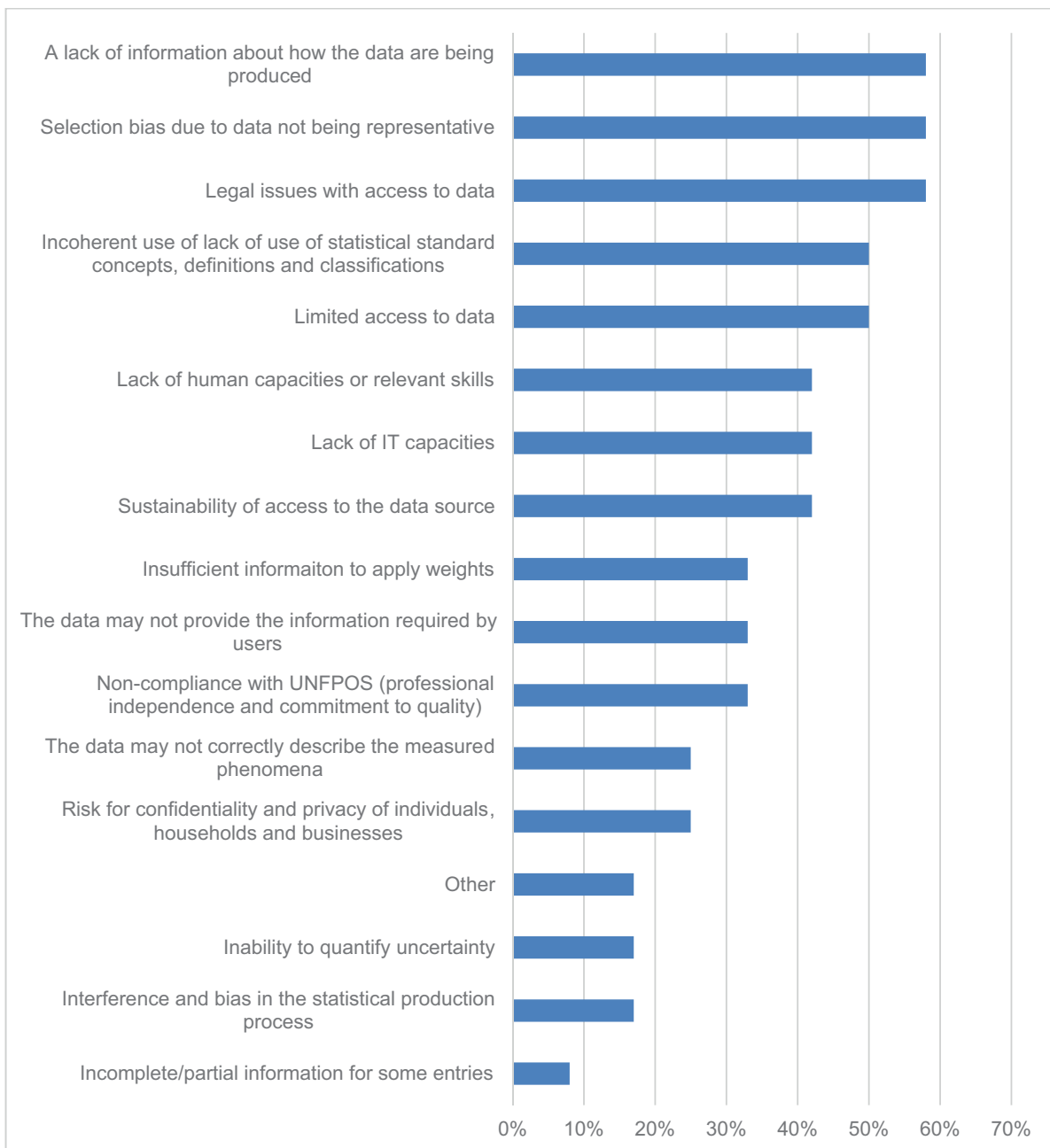

Figure 10 Challenges encountered with regards to the quality of data generated by the citizen science projects run by the organizations of the survey respondents (n = 12).

The survey also asked respondents to list enabling factors for the use of citizen science data for official monitoring. These included:

- the openness of leadership to innovation through experimental statistics, a type of statistical output that is in the testing phase and not yet entirely developed but has a potential to become official statistics, and engagement with new stakeholders;
- the presence of an enabling legal framework, including a specific mandate for NSOs to engage with stakeholders outside of government;
- the related need for fit-for-purpose institutional arrangements and the modernization of statistical legislation, including the review and alignment of all norms and guidance documents to provide a solid legal basis for leveraging new data sources;
- awareness around opportunities offered by citizen science data to fulfil global reporting requirements such as the SDGs or close-to-critical data gaps for the country;
- ensurance of data confidentiality; and
- partnerships with universities on this topic.

In addition, one respondent felt that it was important for the NSO to maintain its role as the leader on quality assurance of statistical information and to embrace this responsibility for data coming from non-official sources. Other respondents were also of the opinion that it could offer added advantages to NSOs by being timelier, more granular, and easier to obtain or free, and in general, that sharing data between NSOs and CSOs could be beneficial to both sides. More qualitative feedback is included in Supplemental File 4: Appendix D.

Finally, we asked respondents to provide solutions to the use of citizen science data for official monitoring. Some common solutions included brokering partnerships abroad to address their capability constraints, introducing validation procedures for the data coming from citizens, conducting technical meetings and stakeholder workshops with citizens, using citizen science data in combination with other sources of information, and NSOs actively providing guidance to other institutions.

RESULTS FROM THE CASE STUDIES

The case study results are reported here as brief summaries of the main findings and are summarized in Table S1 in Supplemental File 5: Appendix E. Organized into four different categories, they include case studies that: (i) showcase quality assurance frameworks or protocols developed by NSOs to leverage data from nontraditional sources, including citizen science; (ii) fill in a specific SDG

indicator or national development plan (NDP) indicators; (iii) include informal sectors or apply new data collection methods to produce key statistics; and (iv) generate new types of measurement of well-being of the society. Further details can be found in Proden (2021a) and Proden et al. (2022a).

Case studies on quality assurance frameworks for citizen science data

The first case study is from the UK, which publishes citizen science data on its SDG reporting platform, labelling it as non-official data after the data set has passed a dedicated quality assurance protocol. The protocol includes two stages. The first stage serves as a screening stage and reviews ethics, privacy, transparency, and accountability as mandatory requirements. The second stage uses a scoring approach that aims to ensure that the data set can be sufficiently useful, meeting, as much as possible, requirements related to relevance, methods, coverage, timeliness, and quality assurance. If some requirements are not fully met (0 or 1 on a 3-point scale), others should score high (2 or 3) to compensate and to ensure a sufficient average score of more than 1.5 points for the data set to be accepted. The non-official data protocol is aligned with the UK Statistics Authority Code of Practice (2023), and its voluntary application procedure is applied to non-official data sources.

In the case study from Colombia, a different approach is used. By law, Colombia's NSO already has the possibility to widen the scope of official data producers beyond the public sector. It has a conceptual and legal framework enabling the NSO to work on the production of official statistics with different stakeholders. At the same time, the NSO has developed quality assurance guidelines for experimental statistics building on its quality assurance framework for official statistics. The framework includes relevance, accessibility, interpretability, transparency, coherence, and timeliness as key criteria. It does not, however, include accuracy and reliability, thereby distinguishing experimental statistics from the official statistics produced from traditional data sources. Unlike in many other countries, in Colombia, experimental statistics are considered official statistics according to Decree 2404 from 2019. The third and final case study in this area is from Kenya, where the NSO has been working on quality assurance guidelines for other data sources. Quality criteria for citizen-generated data and a validation approach have been developed by the NSO with support from PARIS21, a partnership in statistics for development that promotes the better use and production of statistics across the developing countries. These criteria are included in the national Quality Assurance Framework for NSS (KNBS 2022).

Case studies to fill gaps in indicators (Sustainable Development Goals and/or national priorities)

In the first case study in this category, Colombia used citizen science data by teaming up with CSOs and with support from the Office of the United Nations High Commissioner for Human Rights (OHCHR) and other partners to produce the SDG Indicator 16.10.1 on the number of verified cases of killing, kidnapping, enforced disappearance, arbitrary detention, and torture of journalists, associated media personnel, trade unionists, and human rights advocates. It also ran a pilot using social media data to inform SDG Indicator 16.b.1. on persons having felt discriminated against or harassed in the previous 12 months on the basis of a ground of discrimination prohibited under international human rights law. While the pilot did not involve citizens directly, there is potential to improve the methodology involving citizens. In a case study in the UK, the NSO used citizen science data for SDG Indicator 14.1.1 (b) on plastic debris density. Ghana's NSO ran two pilot projects funded by GIZ using the mobile applications Let's Talk Ghana and CleanApp Ghana (both available from the Google Playstore) as well as crowdsourcing to produce data on gender-based violence and waste management. The National Institute of Environment and Public Health in the Netherlands uses air pollution data from citizen sensors to produce more geographically disaggregated data on PM2.5 and PM10 pollutants for the country, and makes it available on the Working Together platform (https://ict-research.nl/about/, Proden et al. 2022a). In the final case study in this area, the NSO in the Philippines, supported by PARIS21, has conducted exhaustive research into all the citizen science data available in the country and identified 81 indicators from their NDP that could be monitored using such data.

Case studies to apply new data collection methods or gather data on the informal economy

This category includes case studies from Italy and Kenya. The NSO in Italy has used citizen science data as part of new trusted smart surveys that leverage mobile phone data by drawing on the online data collection work as part of time use and household budget surveys. Such surveys may be useful for producing some of the SDG and other social indicators, such as the time spent on unpaid domestic and care work. This was done as part of the broader European Statistical System initiative to set up an EU smart surveys platform, and Italy was rolling out trusted smart survey pilots, combining traditional survey sampling techniques with sensor data from mobile devices. In the second case study, the Kenyan NSO partnered with a CSO to produce statistics on the economy of sex workers, which would not be typically covered by traditional data sources such as household surveys.

Case studies to generate new types of measurement of well-being of the society

Two case studies, one in Colombia and one in Mexico, have leveraged citizen science data to measure societal well-being in new ways. Colombia's NSO provided support to a pilot led by the University of Warwick in partnership with local institutions to develop a qualitative risk and vulnerability monitoring methodology using citizen science data in an informal settlement context, and applied it in one of their provinces. Mexico leveraged citizen science data for sentiment analysis of the daily mood of people based on social media trends. Students from one of the Mexican technology-oriented universities contributed to data classification as part of machine learning algorithm developments.

DISCUSSION

Our research is the first comprehensive analysis on the use of citizen science data for official and non-official statistics from an NSS perspective, with the following opportunities identified: (i) filling data gaps in the SDG indicator framework, (ii) measuring perceptions on various issues such as the quality of public services, (iii) collecting data on sensitive topics such as gender-based violence or corruption, and (iv) providing more granular and disaggregated data than is currently available to show trends in specific locations or among different population groups.

Our results showed similarity to the findings of Fraisl et al. (2020) regarding the areas to which citizen science data could contribute. Both studies identified biodiversity-, land use– and land cover–, and beach litter–related indicators as having high potential for benefiting from citizen science data. However, in terms of the SDGs that can benefit from citizen science data the most, the respondents identified SDG 5, Gender Equality; SDG 1, No Poverty; and SDG 6, Clean Water and Sanitation; as opposed to SDG 15, Life on Land; SDG 11, Sustainable Cities and Communities; SDG 3, Good Health and Wellbeing; and SDG 6, Clean Water and Sanitation highlighted by Fraisl et al. (2020). This difference may simply reflect areas where official data are lacking for SDG and national monitoring, or it may be related to the definition of citizen science data. Here, we used the term citizen science to describe the practice of citizen engagement in scientific research and knowledge production, whereas several respondents consistently used the term "citizen-generated data," which is more commonly used by CSOs and the official statistics community (PARIS21, PSA, and PSRTI 2020). Hence, some respondents may have selected SDGs with social dimensions in terms of the potential of citizen science for official statistics.

We identified data quality as one of the main barriers to the use of citizen science data by NSSs. However, we found that, in some cases, this is mostly based on assumptions rather than on actual experiences. In other cases, there are easy ways to deal with some of the quality issues. This may include improving data accessibility and providing detailed metadata, and—for the data yet to be generated—the alignment of concepts and improvements in coverage or sampling to ensure the data are representative. Yet in other cases, the issues of representativeness or the sustainability of these data source may be more difficult to address, but in the case studies, the advantages of using the data have outweighed possible limitations. Overall, our findings related to data quality are consistent with other studies that report on the need for ensuring and communicating data quality and promoting consistent data collection across citizen science initiatives (Fritz et al. 2019; Campbell et al. 2020; Bishop et al. 2020; UNEP 2021a, b; Downs et al. 2021; de Sherbinin et al. 2021; Price-Jones et al. 2022; Fraisl et al. 2022b).

Our findings also indicated that the potential non-compliance of citizen science data with the confidentiality and impartiality principles of the UN Fundamental Principles of Official Statistics (UNFPOS 2014) is a concern for the NSSs to leverage these data for official or non-official statistics. Confidentiality concerns are often related to the lack of knowledge or capacity to ensure that the data are anonymized and cannot be traced back to individuals. The rules on confidentiality are typically defined in the Statistics Acts of each country, where the data producers outside of an NSS are not bound by such legal requirements. Issues on impartiality are mainly concerned with the advocacy role of some CSOs as this may cast doubt on impartiality in data production. However, transparency about these processes, the application of robust statistical procedures, and improved methodologies, as well as their proper communication, can help to address such concerns.

Our results from the case studies showed that citizen science data can be used by the NSSs in three ways: (i) as input for the production of official statistics; (ii) as non-official statistics; and (iii) as experimental statistics, which can help close data gaps or answer important questions about societal well-being. In the first case, standard quality assurance procedures for official statistics apply, and the NSSs that are involved in such projects may have an influence over the data production process or be confident that the data will meet their quality requirements. The second case occurs when an NSS attempts to leverage already existing citizen science data when they were not involved in the production process. Several countries have developed quality assurance frameworks or guidelines for the use of such data, which aim to strengthen the capacities of citizen

science practitioners and CSOs to produce data in a way that comply with the minimum data quality requirements of the NSSs. Such frameworks and protocols can also help NSSs to assess the quality of existing citizen science data before they are disseminated on official platforms as non-official statistics. In the third case, citizen science data are usually part of an experimental statistics project, which are typically initiated and implemented by the NSOs and are not considered as official statistics. However, as an exceptional case, Colombia regards experimental statistics as official statistics, and uses the data produced by other stakeholders as part of the experimental statistics portfolio. It thus has a specific quality assurance framework for the use of experimental statistics similar to ones developed by other countries for non-official statistics.

The quality assurance frameworks used by NSSs for assessing the quality of a data set, including those from citizen science, may vary in scope and approach. For example, while Colombia identified relevance, accessibility, interpretability, coherence, timeliness, and transparency as the key criteria (Proden 2021a), PARIS21 recommended accuracy, credibility with "no attachment of interest pushed by the data producer," and frequency in addition to those adopted by Colombia (PARIS21, PSA, and PSRTI 2020). The UK also added ethics and privacy, as well as the methods, coverage, and quality assurance to their framework. Based on our analysis, we recommend the use of the standard criteria from the United Nations National Quality Assessment Framework (NQAF) for official statistics (UNDESA 2019), that is, accessibility, timeliness, frequency and sustainability, accuracy and reliability, coverage, relevance, metadata, coherence, comparability and integrability—along with the additional criteria on transparency, impartiality, and confidentiality—in order to assess the quality of a citizen science data set more comprehensively. However, it is important to keep in mind that the standard NQAF criteria may not always be fully met by all citizen science data sets, and thus a scoring system with a few mandatory screening criteria similar to the above-mentioned UK approach can be used to measure the overall quality rather than attributing equal weight for each criterion. These criteria could be complemented by "self-identification," which is a principle of the Human Rights-Based Approach to Data (HRBAD) and highlights that the respect and protection of personal identity is central to human rights, and that the individuals should decide whether to disclose or withhold the information about their personal characteristics (UN OHCHR 2018). The approach we propose here was tested on a number of citizen science data sets as part of the Crowd4SDG project, and the results are published in various reports (Proden 2021a, 2021b; Proden, Zodeougin-Quist, and Imaralieva 2022b).

Our study also showed that the lack of legal arrangements can also be an impediment to the use of citizen science data. However, modern statistical laws can grant legal authority to the NSOs to regularly access and use data from stakeholders across and outside of government agencies, such as from the CSOs and private companies as in the case, for example, of the Ghana Statistical Service Act (Republic of Ghana 2019). Nevertheless, our results also showed that having an up-to-date legal framework is not necessarily sufficient. Its effective implementation depends on the leadership and overall organizational culture of the NSO, including openness to innovation, and willingness to build new partnerships and interest in the use of new data sources. Overall, awareness about the potential of citizen science data and whether it can fill in a specific data gap is a key enabling factor pushing NSSs to engage with it.

One of the main limitations of our study is that even though the questionnaire was sent to all NSOs globally, our results may not be fully representative of different geographical regions and actors within the official statistics community. The response rate of our survey may have been influenced by lack of capacity to respond or a lack of interest in exploring new data sources within the NSS. As a result, responses to the questions on awareness may show a more positive picture than the true situation across all NSSs. However, our findings are still useful for understanding the perceived and actual potential of citizen science data by NSSs and for NSSs, and to our knowledge, is the first attempt to explore the perspectives of the NSSs on this topic.

RECOMMENDATIONS AND NEXT STEPS

Drawing on our research, we provide the following eight recommendations for how to leverage the potential offered by citizen science data for SDG monitoring and reporting as well as for official monitoring more broadly:

1. Citizen science data can be utilized as official or non-official statistics after a data validation process to ensure that the data are of good quality for their intended purpose. It can be used alone or integrated with other data sources for the production of official statistics.
2. There are three ways in which NSSs can work with citizen science data: (i) using already existing citizen science data, (ii) designing and implementing a new citizen science project, or by (iii) combining both approaches. For example, NSOs can use already

existing and openly available data sets from beach cleanups for monitoring marine plastic litter or they can launch a citizen science initiative to address a specific data gap such as gender-based violence. As a third approach, they can use an existing data set on a specific topic such as freshwater quality to create a baseline, while at the same time initiating a new citizen science project on freshwater quality with key stakeholders to ensure that the future data collection activities meet their data quality requirements, including the requirements of the global methodology of the relevant indicator, particularly if the purpose is SDG monitoring.
3. NSOs interested in using nontraditional data should work towards modernizing their statistical legislation and improving their governance factors to facilitate access to the data held by CSOs and other potential stakeholders.
4. The mapping of existing citizen science data and initiatives at a national level can help NSSs to identify relevant data sets, as well as the data producers with whom they can build new partnerships.
5. NSOs could organize periodic and thematic stakeholder workshops based on the needs and the context with a view to institutionalize a culture of collaboration with partners from both across and outside the government, including those producers and holders of nontraditional data.
6. Established processes, such as quality assurance frameworks, guidelines and protocols, can enable NSOs to decide on which data sets could be used as official or non-official statistics.
7. In addition to the quality assurance frameworks and guidelines for non-official data producers, NSOs could offer training and capacity building activities, particularly for citizen science practitioners to raise their statistical literacy, including their understanding of the issues related to compliance with privacy, confidentiality, and metadata, among others.
8. NSSs could offer open data reporting platforms for citizen science practitioners to contribute data in a transparent way.

Finally, our analysis shows the need for a typology of citizen science data and projects that can help NSSs identify their needs and data quality requirements for different citizen science projects and adopt a more active role in building citizen science data partnerships. In future research, this typology will build on the key characteristics presented in Table S1 (Supplemental File 5: Appendix E).

DATA ACCESSIBILITY STATEMENTS

Data are not publicly available, but aggregated results are provided in the publication.

NOTES

SUPPLEMENTAL FILES

The supplemental files for this article can be found as follows:

- **Supplemental File 1.** Appendix A. DOI: https://doi.org/10.5334/cstp.584.s1
- **Supplemental File 2.** Appendix B . DOI: https://doi.org/10.5334/cstp.584.s2
- **Supplemental File 3.** Appendix C . DOI: https://doi.org/10.5334/cstp.584.s3
- **Supplemental File 4.** Appendix D . DOI: https://doi.org/10.5334/cstp.584.s4
- **Supplemental File 5.** Appendix E . DOI: https://doi.org/10.5334/cstp.584.s5

ETHICS AND CONSENT

All the survey participants have been informed about the purpose of the survey, its voluntary nature, and the treatment and use of the resulting data before choosing to participate. All collected data on individuals are treated as confidential. Individuals who participated in the interviews have provided their consent to the publication of the resulting case studies in Proden 2021a, and Proden et al. 2022a.

ACKNOWLEDGEMENTS

FUNDING INFORMATION

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Elena Proden orcid.org/0000-0002-1239-7696
UN Institute for Training and Research, CH

Dilek Fraisl orcid.org/0000-0001-7523-7967
International Institute for Applied Systems Analysis, AT

Linda See orcid.org/0000-0002-2665-7065
International Institute for Applied Systems Analysis, AT

REFERENCES

Auerbach, J, Barthelmess, EL, Cavalier, D, Cooper, CB, Fenyk, H, Haklay, M, Hulbert, JM, Kyba, CCM, Larson, LR, Lewandowski, E and **Shanley, L.** 2019. The problem with delineating narrow criteria for citizen science. *Proceedings of the National Academy of Sciences*, 116(31): 15336–15337. DOI: https://doi.org/10.1073/pnas.1909278116

Bishop, IJ, Warner, S, van Noordwijk, TCGE, Nyoni, FC and **Loiselle, S.** 2020. Citizen Science Monitoring for Sustainable Development Goal Indicator 6.3.2 in England and Zambia. *Sustainability*, 12(24): 10271. DOI: https://doi.org/10.3390/su122410271

Campbell, J, Neuner, J, See, L, Fritz, S, Fraisl, D, Espey, J and **Kim, A.** 2020. The role of combining national official statistics with global monitoring to close the data gaps in the environmental SDGs. *Statistical Journal of the IAOS,* 36(2): 443–453. DOI: https://doi.org/10.3233/SJI-200648

Conrad, CC and **Hilchey, KG.** 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment*, 176(1–4): 273–291. DOI: https://doi.org/10.1007/s10661-010-1582-5

Dang, H-AH and **Serajuddin, U.** 2020. Tracking the sustainable development goals: Emerging measurement challenges and further reflections. *World Development*, 127: 104570. DOI: https://doi.org/10.1016/j.worlddev.2019.05.024

de Sherbinin, A, Bowser, A, Chuang, T-R, Cooper, C, Danielsen, F, Edmunds, R, Elias, P, Faustman, E, Hultquist, C, Mondardini, R, Popescu, I, Shonowo, A and **Sivakumar, K.** 2021. The Critical Importance of Citizen Science Data.

Frontiers in Climate, 3: 650760. DOI: https://doi.org/10.3389/fclim.2021.650760

Downs, RR, Ramapriyan, HK, Peng, G and **Wei, Y.** 2021. Perspectives on citizen science data quality. *Frontiers in Climate*, 3: 615032. DOI: https://doi.org/10.3389/fclim.2021.615032

Eitzel, MV, Cappadonna, JL, Santos-Lang, C, Duerr, RE, Virapongse, A, West, SE, Kyba, CCM, Bowser, A, Cooper, CB, Sforzi, A, Metcalfe, AN, Harris, ES, Thiel, M, Haklay, M, Ponciano, L, Roche, J, Ceccaroni, L, Shilling, FM, Dörler, D, Heigl, F, Kiessling, T, Davis, BY and **Jiang, Q.** 2017. Citizen science terminology matters: Exploring key terms. *Citizen Science: Theory and Practice*, 2(1): 1. DOI: https://doi.org/10.5334/cstp.96

Fraisl, D, Campbell, J, See, L, Wehn, U, Wardlaw, J, Gold, M, Moorthy, I, Arias, R, Piera, J, Oliver, JL, Masó, J, Penker, M and **Fritz, S.** 2020. Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science*, 15(6): 1735–1751. DOI: https://doi.org/10.1007/s11625-020-00833-7

Fraisl, D, Hager, G, Bedessem, B, Gold, M, Hsing, P-Y, Danielsen, F, Hitchcock, CB, Hulbert, JM, Piera, J, Spiers, H, Thiel, M and **Haklay, M.** 2022a. Citizen science in environmental and ecological sciences. *Nature Reviews Methods Primers*, 2(1): 64. DOI: https://doi.org/10.1038/s43586-022-00144-4

Fraisl, D, See, L, Sturn, T, MacFeely, S, Bowser, A, Campbell, J, Moorthy, I, Danylo, O, McCallum, I and **Fritz, S.** 2022b. Demonstrating the potential of Picture Pile as a citizen science tool for SDG monitoring. *Environmental Science & Policy*, 128: 81–93. DOI: https://doi.org/10.1016/j.envsci.2021.10.034

Fritz, S, See, L, Carlson, T, Haklay, M (Muki), Oliver, JL, Fraisl, D, Mondardini, R, Brocklehurst, M, Shanley, LA, Schade, S, Wehn, U, Abrate, T, Anstee, J, Arnold, S, Billot, M, Campbell, J, Espey, J, Gold, M, Hager, G, He, S, Hepburn, L, Hsu, A, Long, D, Masó, J, McCallum, I, Muniafu, M, Moorthy, I, Obersteiner, M, Parker, AJ, Weissplug, M and **West, S.** 2019. Citizen science and the United Nations Sustainable Development Goals. *Nature Sustainability*, 2(10): 922–930. DOI: https://doi.org/10.1038/s41893-019-0390-3

Haklay, M, Fraisl, D, Greshake Tzovaras, B, Hecker, S, Gold, M, Hager, G, Ceccaroni, L, Kieslinger, B, Wehn, U, Woods, S, Nold, C, Balázs, B, Mazzonetto, M, Ruefenacht, S, Shanley, LA, Wagenknecht, K, Motion, A, Sforzi, A, Riemenschneider, D, Dorler, D, Heigl, F, Schaefer, T, Lindner, A, Weißpflug, M, Mačiulienė, M and **Vohland, K.** 2021. Contours of citizen science: a vignette study. *Royal Society Open Science*, 8(8): 202108. DOI: https://doi.org/10.1098/rsos.202108

Higgins, K and **Cornforth, J.** 2015. *Civil Society and SDG Monitoring. Harnessing Civil Society and Citizen-generated data.* URL: http://civicus.org/images/SDG%20Monitoring%20-%20DataShift%20Background%20Note%20-%20Final.pdf.

HLPF. 2023. *Voluntary National Reviews.* Available at: https://hlpf.un.org/vnrs.

IAEG-SDGs. 2018. *Guidelines on Data Flows and Global Data Reporting for Sustainable Development Goals.* Available at: https://unstats.un.org/unsd/statcom/49th-session/documents/BG-Item-3a-IAEG-SDGs-DataFlowsGuidelines-E.pdf.

ISWGHS. 2019. *Achieving the Full Potential of Household Surveys in the SDG Era.* Available at: https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item4c-ISWGHS-E.pdf.

KNBS. 2022. *Kenya Statistical Quality Assurance Framework – KeSQAF.* Available at: Kenya Statistical Quality Assurance Framework – KeSQAF – Kenya National Bureau of Statistics (knbs.or.ke).

MacDonald, C. 2012. Understanding participatory action research: A qualitative resarch methodology option. *Canadian Journal of Action Research*, 13(2): 34–50. DOI: https://doi.org/10.33524/cjar.v13i2.37

MacFeely, S and **Nastav, B.** 2019. "You say you want a [data] revolution": A proposal to use unofficial statistics for the SDG Global Indicator Framework. *Statistical Journal of the IAOS*, 35(3): 309–327. DOI: https://doi.org/10.3233/SJI-180486

Olen, SM. 2022. Citizen science tackles plastics in Ghana. *Nature Sustainability*, 5(10): 814–815. DOI: https://doi.org/10.1038/s41893-022-00980-y

PARIS21, PSA, and PSRTI. 2020. *Use of CGD for SDG reporting in the Philippines: A case study.* Available at: https://paris21.org/sites/default/files/inline-files/PSA-report-FINAL.pdf (Last accessed 28 September 2022).

Price-Jones, V, Brown, PMJ, Adriaens, T, Tricarico, E, Farrow, RA, Cardoso, AC, Gervasini, E, Groom, Q, Reyserhove, L, Schade, S, Tsinaraki, C and **Marchante, E.** 2022. Eyes on the aliens: citizen science contributes to research, policy and management of biological invasions in Europe. *NeoBiota*, 78: 1–24. DOI: https://doi.org/10.3897/neobiota.78.81476

Proden, E. 2021a. Deliverable 5.1 Initial report on relevance and quality-related considerations of citizen-science generated data. Crowd4SDG Project. *Crowd4SDG Project.* Geneva, Switzerland: UNITAR. Available at: http://www.crowd4sdg.eu/ (Last accessed 28 September 2022).

Proden, E. 2021b. Deliverable 5.2 Data usability assessment and recommendations for SDGs GEAR Cycle 2. *Crowd4SDG Project.* Geneva, Switzerland: UNITAR. Available at: https://crowd4sdg.eu/wp-content/uploads/2021/10/D5.2-Data-usability-assessment.pdf (Last accessed 28 September 2022).

Proden, E, Bett, K, Chen, H, Duerto Valero, S, Fraisl, D, Gamez, G, MacFeely, S, Mondardini, R, See, L and **Min, Y.** 2022a. Citizen science data to track SDG progress: Low-hanging fruit for Governments and National Statistical Offices. *Policy Brief.* Available at: https://unitar.org/sites/default/files/media/publication/doc/2022%20Policy%20Brief%20NSOs%20%26%20CSD%204%20SDGs%20By%20Crowd4SDGs%20%26%20International_Organisations_0.pdf.

Proden, E, Zodeougin-Quist, A and **Imaralieva, M.** 2022b. Deliverable 5.3 Deliverable 5.3 Data usability assessment and recommendations for SDGs GEAR Cycle 3. *Crowd4SDG Project.* Geneva, Switzerland: UNITAR.

Republic of Ghana. 2019. *Ghana Statistical Service Act 2019 (ACT 1003).* https://statsghana.gov.gh/gssmain/storage/img/STATISTICAL%20SERVICE%20ACT,%202019%20(ACT%201003).pdf.

Shirk, JL, Ballard, HL, Wilderman, CC, Phillips, T, Wiggins, A, Jordan, R, McCallie, E, Minarchek, M, Lewenstein, BV, Krasny, ME and **Bonney, R.** 2012. Public participation in scientific research: A framework for deliberate design. *Ecology and Society,* 17(2): 29. DOI: https://doi.org/10.5751/ES-04705-170229

Sieber, RE and **Haklay, M.** 2015. The epistemology(s) of volunteered geographic information: a critique. *Geo: Geography and Environment,* 2(2): 122–136. DOI: https://doi.org/10.1002/geo2.10

UK Statistics Authority and Office for Statistics Regulation. 2023. *Code of Practice for Statistics.* Available at: https://code.statisticsauthority.gov.uk/.

UN. 2012. *Resolution adopted by the General Assembly on 27 July 2012. 66/288. The future we want.* Available at: https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_66_288.pdf.

UN. 2015. *Sustainable Development Goals.* Available at: https://unstats.un.org/sdgs/.

UN. 2020. *IAEG-SDGs Inter-agency and Expert Group on SDG Indicators.* Available at: https://unstats.un.org/sdgs/iaeg-sdgs/.

UN. 2023. *SDG indicators. United Nations Global SDG Database.* Available at: https://unstats.un.org/sdgs/indicators/database/.

UN OHCHR. 2018. *A Human Rights-Based Approach to Data. Leaving No One Behind in the 2030 Agenda for Sustainable Development.* OHCHR Guidance Note to Data Collection and Disaggregation. Geneva, Switzerland: United Nations. Available at: https://www.ohchr.org/Documents/Issues/HRIndicators/GuidanceNoteonApproachtoData.pdf (Last accessed 28 September 2022).

UNDESA. 2019. *United Nations National Quality Assurance Frameworks Manual for Official Statistics.* New York: United Nations. Available at: https://unstats.un.org/unsd/methodology/dataquality/references/1902216-UNNQAFManual-WEB.pdf. (Last accessed 28 September 2022).

UNEP. 2019. *Measuring Progress: Towards Achieving the Environmental Dimension of the SDGs.* Available at: https://wedocs.unep.org/handle/20.500.11822/27627.

UNEP. 2021a. *Measuring Progress: Environment and the SDGs.* Available at: https://www.unep.org/resources/publication/measuring-progress-environment-and-sdgs.

UNEP. 2021b. *Understanding the State of the Ocean: A Global Manual on Measuring SDG 14.1.1, SDG 14.2.1 and SDG 14.5.1.* https://wedocs.unep.org/bitstream/handle/20.500.11822/35086/USO.pdf?sequence=3&isAllowed=y.

UNFPOS. 2014. *General Assembly Resolution 68/261.* UN Fundamental Principles of Official Statistics. 29 January 2014.

UNSD. 2022. *The Sustainable Development Goals Report 2022.* Available at: https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf.

]u[🔓