



Perspectives on Citizen Science Data Quality

Robert R. Downs^{1*}, Hampapuram K. Ramapriyan^{2,3}, Ge Peng⁴ and Yaxing Wei⁵

¹ NASA Socioeconomic Data and Applications Center, Center for International Earth Science Information Network, The Earth Institute, Columbia University, Palisades, NY, United States, ² Science Systems and Applications, Inc., Lanham, MD, United States, ³ Earth Science Data and Information System Project, Goddard Space Flight Center, NASA, Greenbelt, MD, United States, ⁴ Earth System Science Center/NASA Marshall Space Flight Center (MSFC) Interagency Implementation and Advanced Concepts Team (IMPACT), The University of Alabama in Huntsville, Huntsville, AL, United States, ⁵ Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States

Information about data quality helps potential data users to determine whether and how data can be used and enables the analysis and interpretation of such data. Providing data quality information improves opportunities for data reuse by increasing the trustworthiness of the data. Recognizing the need for improving the quality of citizen science data, we describe quality assessment and quality control (QA/QC) issues for these data and offer perspectives on aspects of improving or ensuring citizen science data quality and for conducting research on related issues.

Keywords: citizen science, data quality, information quality, citizen science data, citizen science methods

OPEN ACCESS

Edited by:

Sven Schade,
European Commission, Italy

Reviewed by:

Rob Stevenson,
University of Massachusetts Boston,
United States
David Neil Bonter,
Cornell University, United States

*Correspondence:

Robert R. Downs
rdowns@ciesin.columbia.edu

Specialty section:

This article was submitted to
Climate Risk Management,
a section of the journal
Frontiers in Climate

Received: 04 November 2020

Accepted: 15 March 2021

Published: 09 April 2021

Citation:

Downs RR, Ramapriyan HK, Peng G
and Wei Y (2021) Perspectives on
Citizen Science Data Quality.
Front. Clim. 3:615032.
doi: 10.3389/fclim.2021.615032

INTRODUCTION

Citizen science (CS) is recognized as having broad potential benefits to society. Citizen science projects are providing unique and sometimes fundamental scientific insights and offer a wide variety of scientific outcomes (Pettibone et al., 2017; Paul et al., 2018; Wiggins et al., 2018; Bautista-Puig et al., 2019; Miller et al., 2019; van Etten et al., 2019). Citizen science also offers opportunities for efficiently collecting data that otherwise might not be obtainable in a practical manner (Li et al., 2019; Van Eupen et al., 2021). Citizen science data (CSD) provides valuable environmental measurements and observations that can be used independently and in conjunction with other data products and services to improve research and decision making capabilities (Robinson et al., 2018; Poisson et al., 2020). Especially given the increased opportunity to supplement traditional scientific data with CSD, it is essential that the CSD be as trustworthy and of known quality as other scientific data (Swanson et al., 2016; Aceves-Bueno et al., 2017; Budde et al., 2017; Burgess et al., 2017; Kallimanis et al., 2017; Steger et al., 2017; Sandahl and Tøttrup, 2020). Information about the quality of CSD builds trust, provides opportunities for potential users to discover CSD that are appropriate for their purposes, and enables users to determine whether and how the data can be used to meet their objectives (Alabri and Hunter, 2010; Hunter et al., 2013; Freitag et al., 2016; Lukyanenko et al., 2016; Stevenson, 2018; Anhalt-Depies et al., 2019). The quality of CSD also can influence the analysis and interpretation of the data (Kelling et al., 2015; Clare et al., 2019). Quality information is important for scientific data, including CSD (Roman et al., 2017; Gharaibeh et al., 2019). Citizen science data contributes to many scientific endeavors that are important for environmental science and for the well-being of society, including sustainable development, humanitarian efforts, and disaster prevention and response (Hicks et al., 2019; Fraisl et al., 2020). Providing data quality information can improve opportunities for CS to contribute to important societal efforts and to the reuse of CSD (Kosmala et al., 2016; Hecker et al., 2019; Shanley et al., 2019).

While CS initiatives offer possibilities for obtaining observations and gathering data that supplement traditional data collection on important environmental issues, there is healthy

skepticism about the quality of CSD (Brown and Williams, 2019; Cross, 2019). Fritz et al. (2019) indicate that uncertainty regarding quality of the data is a major barrier to the use of CSD, despite their value for the United Nations Sustainable Development Goals (SDGs). They also provide examples of several activities where steps have been taken to ensure that CSD are of high (and known) quality. Earp and Liconti (2020) describe the disparity between benefits of using marine CSD for research and perceptions of quality. Incompatible design of CS studies and inconsistencies in nomenclature also can affect data quality, resulting in challenges for integrating data from different CS programs (Campbell et al., 2020). User interfaces of digital tools provided to participants also can affect CSD quality (Sharma et al., 2019; Torre et al., 2019). Studying CSD management practices, Bowser et al. concluded: “While significant quality assurance/quality control (QA/QC) checks are taken across the data lifecycle, these are not always documented in a standardized way” (Bowser et al., 2020, p. 12). Recognizing a perceived bias among scientists regarding the use of CSD, Albus et al. (2019) reviewed comparison studies that were conducted on volunteer and professional data collection efforts for large-scale water quality projects, concluding that more comparison studies are needed and that such studies should include accuracy, while controlling for variations among the datasets that are compared.

Considering such concerns about the quality of CSD, as well as other data, and how data quality can affect data and their use, the Earth Science Information Partners (ESIP) Information Quality Cluster (IQC) is attempting to provide recommendations on practices to help ensure or improve CSD quality and build trust for CSD in the scientific community. This manuscript aims to lay out ESIP IQC’s perspectives on the existing challenges and important aspects of CSD quality that should be tackled by the community in the near future.

In section ESIP Information Quality Cluster, activities of the ESIP Information Quality Cluster, relevant to CSD, are introduced along with four quality dimensions that occur throughout the data lifecycle. Section Challenges and Approaches for Improving CSD Quality introduces challenges, directions, and approaches for improving the quality of CSD. The first subsection offers a brief overview of opportunities for improving CSD quality during the recruitment, selection, self-selection, and training of CS volunteers. The second subsection describes selected issues that pertain to transparency of information about QA/QC practices during the production of CSD. The third subsection describes the importance of documenting CSD quality. The fourth subsection describes the importance of and need for establishing rubrics for evaluating CSD quality levels. Section Discussion concludes the paper with a discussion of these CSD quality issues and offers recommendations for progressively improving the quality of CSD.

ESIP INFORMATION QUALITY CLUSTER

The ESIP IQC studies and promotes the awareness of data and information quality (Ramapriyan et al., 2017). Like other ESIP Collaboration Areas (ESIP, 2020), the IQC reflects perspectives

of various partner organizations that contribute to the collection, curation, dissemination, and interdisciplinary use of Earth science data. Information Quality Cluster activities include regular meetings, workshops, conference sessions, white papers, and journal publications. Information Quality Cluster activities also leverage the work of the NASA Earth Science Data System Working Group (ESDSWG) on Data Quality, which was active during 2014–2019 and completed its recommendations to the NASA Earth Science Data and Information System Project (NASA, 2020a). The IQC also organized sessions on CS during recent ESIP meetings. Directly related to data quality concerns for CS and other types of studies, the IQC recently began developing guidelines for documenting and enabling the sharing and reuse of data quality information (Peng et al., 2020). The strength of the IQC is in its membership, consisting of experts in data and information quality from various organizations and disciplines, and promoting collaboration among them and resulting in synergy for developing recommendations with broad applicability.

CHALLENGES AND APPROACHES FOR IMPROVING CSD QUALITY

Applying CSD can be problematic if researchers and other users are not aware of data quality issues that could affect their analyses, contributions, or operational uses. However, there are several challenges for improving CSD quality. Assessing CSD quality can be extremely difficult due to heterogeneous observers and methods and lack of information about such methods. In particular, data bias, errors, uncertainty, and ethical issues pose challenges that should be assessed regularly as part of CS research projects. These and other challenges that occur throughout the data lifecycle are being investigated in an effort to improve the quality of CSD.

Taking a lifecycle approach can help CSD investigators to consider data quality issues and improve the information about data quality that is recorded and provided to users along with the data. The term, data lifecycle, has been defined variously with different levels of detail by different groups. For example, at a very high level, the NOAA Environmental Data Management Framework shows three types of activities—Planning and Production, Data Management, and Usage—in that order, but with feedback from each to the previous type of activity (NOAA, 2013). The US Geological Survey (USGS) defines a science data lifecycle model consisting of the following activities: “Plan, Acquire, Process, Analyze, Preserve and Publish/Share” (Henkel et al., 2015), with cross-cutting activities including “Describe (including metadata and documentation), Manage Quality, and Backup and Secure” (Henkel et al., 2015), thus emphasizing that management of quality cuts across all parts of the lifecycle (Faundeen et al., 2013). Strasser et al. (2012, p. 3) define a data lifecycle with eight components: “Plan, Collect, Assure, Describe, Preserve, Discover, Integrate, and Analyze.” Ramapriyan et al. (2017) consider information quality (i.e., quality of information about data quality) throughout the entire lifecycle to be four-dimensional. These dimensions, also referred

to as aspects of information quality, are: 1. Scientific quality, 2. Product quality, 3. Stewardship quality, and 4. Service quality. Activities that focus on these four dimensions can be regarded as constituting four stages in the lifecycle. The specific activities of the four stages and their mappings to the four dimensions are: “1. Define, develop, and validate; 2. Produce, assess, and deliver (to an archive or data distributor); 3. Maintain, preserve, and disseminate; and 4. Enable data use, provide data services and user support” (Ramapriyan et al., 2017). **Figure 1** depicts data lifecycle stages with each of these activities represented within the four quality dimensions.

Regardless of the terminology used and the level of detail into which the data lifecycle is subdivided, it is important that characterizing and documenting data quality is considered within each stage of the lifecycle. For convenience of discussion, the terms, stages 1–4, as defined, above, in terms of the four quality dimensions, are used in sections Recruitment, Selection, Self-Selection, and Training of CSD Contributors, Transparency in Information about QA/QC Practices during the Data Production Process, Documenting Data Quality to Facilitate Discovery and Reuse, and Establishing Rubrics for Evaluating Quality Levels of CSD to indicate when the recommended actions need to be taken during CSD projects.

Information about the quality of data, including CSD, should be recorded throughout the data lifecycle to improve data for potential use and reuse. Effective planning is critical to the success of a CS project (Freitag et al., 2016) and improved data stewardship (Peng et al., 2018). Considering data quality during the earliest stages of the data project can improve planning and enable the research team to identify issues that could affect data quality later during the project. A framework for data quality issues to be considered while planning and designing CSD research is offered by Wiggins et al. (2011) for applying data quality and validation methods throughout the research

process. In particular, when planning the CSD project, the questions and techniques identified by Kosmala et al. (2016) provide a good starting point for investigators and also provide considerations that can be assessed by evaluators and users of CSD. Such planning would be applicable to CS projects that involve a small number of volunteers as well as to large-scale projects, such as those that were the focus of the study conducted by Albus et al. (2019). A white paper has been developed by NASA’s Citizen Science Data Working Group, for the benefit of researchers desiring to incorporate CS and crowdsourcing into their projects (NASA, 2020b). While this white paper is targeted for NASA-funded researchers in the Citizen Science for Earth Science Program, the discussion in the paper is relevant to a much broader audience. Many aspects of CSD management are addressed in this white paper, including a significant amount of detail describing how information about data quality should be handled.

The ESIP IQC recognizes some of the challenges in and potential approaches to addressing these data quality issues that are pertinent to CSD. These are discussed in more detail within the following subsections.

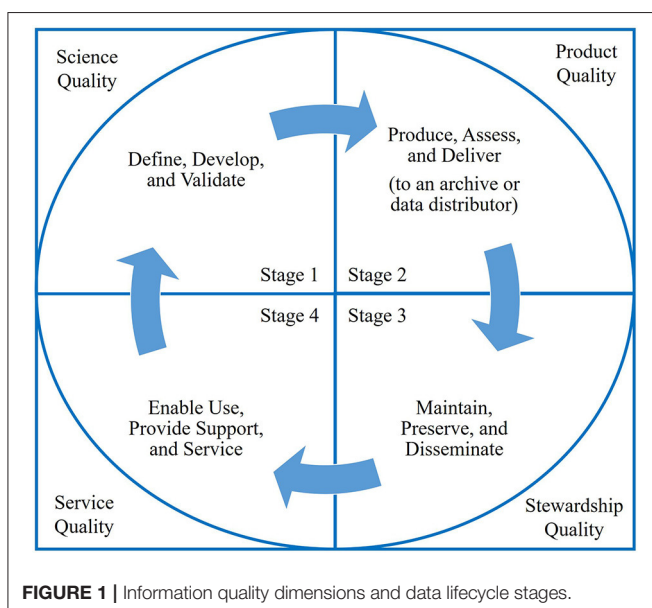
Recruitment, Selection, Self-Selection, and Training of CSD Contributors

Bias, errors, uncertainty, and ethical issues can be addressed through well-designed and documented procedures and proper training by providing volunteers with instructions and written procedures for fieldwork. For studies that involve large numbers of volunteers in additional aspects of the research process besides data collection, training of volunteers contributes to QA (Wilderman and Monismith, 2016). Investigators should consider sources of potential bias when recruiting CS participants and, including recognizing the potential for errors, the proper use of instruments, and techniques for reducing and flagging data uncertainty. Developing a data collection instrument and recruiting volunteers to use the instrument in the field provides opportunities to identify enhancements that can improve the quality of data collected by future volunteers (Compas and Wade, 2018). When engaging volunteers, protecting indigenous people and privacy also must be considered (Bowser et al., 2017; Carroll et al., 2019; Global Indigenous Data Alliance, 2019). Human research subject protections further reduce risks (Resnik, 2019). The NASA Earth Science Data Systems CSD Working Group also offers guidance on these and other relevant issues (NASA, 2020b).

Citizen science data quality efforts for recruitment, selection, self-selection, and training should be initiated during stage 1 (science quality focus) of the data lifecycle, when defining, developing, and validating CSD. These activities also should be pursued during subsequent stages.

Transparency in Information About QA/QC Practices During the Data Production Process

Uncorrected errors, missing data, and undocumented corrections and modifications could influence findings resulting from the analysis of CSD. Such lack of transparency could result



in lost time when exploring whether to use the data. Identified usage limitations should be recorded and, when possible, addressed during research design. Similarly, appropriate uses of data should be identified to reduce the potential for misuse. Verification procedures should be planned and conducted to ensure correctness of data values. Completeness should be ensured by reducing the potential for missing values.

Deploying automated verification and parsing to address data quality issues also could reduce the potential for human errors. However, human oversight is recommended to avoid potential pitfalls of fully-automated systems, such as underestimating extremes. In addition, increasing transparency about pitfalls that have compromised the quality of CSD can avoid a cycle of repeating failures in CS research (Balázs et al., 2021). Enabling volunteers to contribute to transparent validation of observations also contributes to the improvement of CSD quality and to the motivation of contributors (Bonnet et al., 2020).

Considering that CSD is produced largely from voluntary contributions, it is also critical to be transparent about other aspects of CSD that can facilitate use, especially when designating CSD as open data. Providing simple language that enables users to understand their intellectual property rights for using CSD facilitates their use as open data. Ideally, such language should describe permissive intellectual property rights that eliminate restrictions on the use of the data and the documentation (Anhalt-Depies et al., 2019).

Facilitating transparency of information about QA/QC practices should be completed as part of stage 1 (focus on science quality) and stage 2 (focus on product quality) of the data lifecycle. Such transparency also should be facilitated during subsequent stages.

Documenting Data Quality to Facilitate Discovery and Reuse

Describing the quality of CSD in documentation and metadata improves its potential for use and improves capabilities for assessing whether data are appropriate for reuse by those who did not participate in the original study that collected the data. Furthermore, describing data quality can improve the interoperability and integration of CSD with other data. Documentation of CSD also should describe provenance for collection, validation, curation, dissemination, and use of the data. As data originators, the roles and responsibilities of investigators and volunteer observers for ensuring and documenting the scientific quality of data should be defined (e.g., Peng et al., 2016).

Relevant guidance on practices for managing data also delineate the importance of documenting data quality. These include the FAIR Principles (Wilkinson et al., 2016), the Group on Earth Observations System of Systems (GEOSS) Data Management Principles (Group on Earth Observations, 2016), the TRUST Principles for Digital Repositories (Lin et al., 2020), and data maturity models (Peng et al., 2019).

Data quality documentation should be conducted throughout all four stages of the data lifecycle. The development of data quality documentation should be initiated early during stage 1,

delivered to a repository during stage 2, disseminated along with the data during stage 3, and used to support use of the data in stage 4.

Establishing Rubrics for Evaluating Quality Levels of CSD

To enable and maximize the reuse of CSD in environmental research and other areas, easy-to-understand quality levels that address the specific needs of target user communities, e.g. researchers, decision supporters, and the general public, on CSD will be important. Establishing rubrics to evaluate CSD quality information against such quality levels will be consequential. For example, Balázs et al. (2021) recommend communicating data quality goals to volunteers and providing accessible training materials, guidance, and understandable instructions for data collection to improve the quality of CSD. Tredick et al. (2017) developed a rubric for evaluating CS programs. This structured rubric acknowledges the importance of CSD management, quality assurance, and information integrity to the success of a CS program. The BiodivERsA Citizen Science Toolkit For Biodiversity Scientists (Goudeseune et al., 2020) also described the evaluation of output, including data quality, as one of the ten key principles for successful CS. Vocabularies for CSD quality levels, which link to the needs of diverse user communities and rubrics to assess CSD against such vocabularies, are important next steps to maximize the scientific and societal benefits of CS programs.

Rubrics for information quality levels of CSD apply to the dimensions across all stages of the data lifecycle. However, it should be noted that the development of rubrics should be initiated very early during stage 1, and that such rubrics will support users during stage 4.

DISCUSSION

Enabling the use of CSD offers opportunities for new research projects to investigate issues while avoiding costly or redundant data collection. To allow for broad use of CSD, data QA/QC should be performed, and information about QA/QC procedures should be captured and conveyed to users. Since improving CSD quality offers opportunities for additional uses, data quality efforts should begin during project conceptualization and planning, continuing throughout the data lifecycle, to enable data reuse. Efforts to improve the quality of CSD should begin during stage 1, when science quality activities are performed and quality information is prepared when defining, developing, and validating the data. Citizen science data quality efforts should continue with stage 2, so that product quality information is prepared, assessed, and delivered along with the data to a repository for dissemination. Citizen science data quality information should be maintained, preserved, and disseminated with the data to ensure stewardship quality during stage 3. Providing quality information along with the data to provide service quality during stage 4 enables and supports the use of CSD.

Furthermore, documenting CSD quality can improve trust in CS within the scientific community and reflects ethical approaches to conducting CS. When preparing CSD for use, investigators should describe data quality in the metadata and data documentation, as well as in data papers and publications. Documentation should differentiate between various quality issues to avoid confusing potential users.

Consequently, we recommend employing a systematic approach for ensuring CSD quality. Future research should consider implications of data quality throughout the data lifecycle and data quality as it pertains to collecting CSD.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

RD, HR, GP, and YW contributed to conception and design of the manuscript and wrote the first draft and sections of the manuscript. All the authors reviewed and revised the draft with beneficial edits, and approved the submitted version.

REFERENCES

- Aceves-Bueno, E., Adeleye, A. S., Feraud, M., Huang, Y., Tao, M., Yang, Y., et al. (2017). The accuracy of citizen science data: a quantitative review. *Bull. Ecol. Soc. Amer.* 98, 278–290. doi: 10.1002/bes.2.1336
- Alabri, A., and Hunter, J. (2010). “Enhancing the quality and trust of citizen science data,” in *2010 IEEE Sixth International Conference on e-Science* (Washington, DC: IEEE), 81–88. doi: 10.1109/eScience.2010.33
- Albus, K., Thompson, R., and Mitchell, F. (2019). Usability of existing volunteer water monitoring data: what can the literature tell us? *Citiz. Sci. Theory Pract.* 4:28. doi: 10.5334/cstp.222
- Anhalt-Depies, C., Stenglein, J. L., Zuckerberg, B., Townsend, P. A., and Rissman, A. R. (2019). Tradeoffs and tools for data quality, privacy, transparency, and trust in citizen science. *Biol. Conserv.* 238:108195. doi: 10.1016/j.biocon.2019.108195
- Balázs, B., Mooney, P., Nováková, E., Bastin, L., and Arsanjani, J. J. (2021). “Data quality in citizen science,” in *The Science of Citizen Science*, eds K. Vohland, A. Land-Zandra, R. Lemmens, J. Perello, M. Ponti, R. Samson, and K. Wagenknecht (Switzerland: Springer), 139–157. Available online at: <https://www.springer.com/gp/book/9783030582777>
- Bautista-Puig, N., De Filippo, D., Mauleón, E., and Sanz-Casado, E. (2019). Scientific landscape of citizen science publications: dynamics, content and presence in social media. *Publications* 7:12. doi: 10.3390/publications7010012
- Bonnet, P., Joly, A., Faton, J. M., Brown, S., Kimiti, D., Deneu, B., et al. (2020). How citizen scientists contribute to monitor protected areas thanks to automatic plant identification tools. *Ecol. Solut. Evid.* 1:e12023. doi: 10.1002/2688-8319.12023
- Bowser, A., Cooper, C., de Sherbinin, A., Wiggins, A., Brenton, P., Chuang, T. R., et al. (2020). Still in need of norms: the state of the data in citizen science. *Citiz. Sci. Theory Pract.* 5:1. doi: 10.5334/cstp.303
- Bowser, A., Shilton, K., Preece, J., and Warrick, E. (2017). “Accounting for privacy in citizen science: Ethical research in a context of openness,” in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, OR), 2124–2136. doi: 10.1145/2998181.2998305

FUNDING

RD was supported by the National Aeronautics and Space Administration (NASA) under Contract 80GSFC18C0111 for operation of the NASA Socioeconomic Data and Applications Center (SEDAC). HR was supported under NASA Contract 80GSFC20C044 with Science Systems and Applications, Inc. GP was supported in part by NOAA under Cooperative Agreement NA19NES4320002 and by NASA under Cooperative Agreement NNM11AA01A. YW was supported by NASA under Interagency Agreement 80GSFC19T0039.

ACKNOWLEDGMENTS

This article reflects perspectives of the authors, who are members of the ESIP Information Quality Cluster (IQC) leadership team and appreciate the insight received from discussions among IQC members and from invited presentations on the CS programs at the U.S. agency level, including those at NASA and NOAA. The authors also appreciate the thoughtful comments and recommendations provided by the reviewers. The views expressed in the article do not represent the position of ESIP, its sponsors, the authors’ employers, or their sponsors.

- Brown, E. D., and Williams, B. K. (2019). The potential for citizen science to produce reliable and useful information in ecology. *Conserv. Biol.* 33, 561–569. doi: 10.1111/cobi.13223
- Budde, M., Schankin, A., Hoffmann, J., Danz, M., Riedel, T., and Beigl, M. (2017). Participatory sensing or participatory nonsense? Mitigating the effect of human error on data quality in citizen science. *Proc. ACM Interact. Mob. Wear. Ubiquit. Technol.* 1, 1–23. doi: 10.1145/3131900
- Burgess, H. K., DeBey, L. B., Froehlich, H. E., Schmidt, N., Theobald, E. J., Ettinger, A. K., et al. (2017). The science of citizen science: exploring barriers to use as a primary research tool. *Biol. Conserv.* 208, 113–120. doi: 10.1016/j.biocon.2016.05.014
- Campbell, D. L., Thessen, A. E., and Ries, L. (2020). A novel curation system to facilitate data integration across regional citizen science survey programs. *PeerJ* 8:e9219. doi: 10.7717/peerj.9219
- Carroll, S. R., Rodriguez-Lonebear, D., and Martinez, A. (2019). Indigenous data governance: strategies from United States native nations. *Data Sci. J.* 18:31. doi: 10.5334/dsj-2019-031
- Clare, J. D., Townsend, P. A., Anhalt-Depies, C., Locke, C., Stenglein, J. L., Frett, S., et al. (2019). Making inference with messy (citizen science) data: when are data accurate enough and how can they be improved? *Ecol. Appl.* 29:e01849. doi: 10.1002/eap.1849
- Compas, E. D., and Wade, S. (2018). Testing the waters: a demonstration of a novel water quality mapping system for citizen science groups. *Citiz. Sci. Theory Pract.* 3:6. doi: 10.5334/cstp.124
- Cross, I. D. (2019). ‘Changing behaviour, changing investment, changing operations’: using citizen science to inform the management of an urban river. *Area* 51:1–10. doi: 10.1111/area.12597
- Earp, H. S., and Liconti, A. (2020). “Science for the future: the use of citizen science in marine research and conservation,” in *YOU MARES 9 - The Oceans: Our Research, Our Future*, eds S. Jungblut, V. Liebich, and M. Bode-Dalby (Cham: Springer) 1–19. doi: 10.1007/978-3-030-20389-4_1
- ESIP (2020). *Collaboration Areas*. Available online at: <https://www.esipfed.org/get-involved/collaborate> (accessed September 17, 2020)
- Faundeen, J. L., Burley, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., et al. (2013). *The United States Geological Survey Science Data*

- Lifecycle Model*. U.S. Geological Survey Open-File Report 2013–1265, p. 4. doi: 10.3133/ofr20131265
- Fraisl, D., Campbell, J., See, L., Wehn, U., Wardlaw, J., Gold, M., et al. (2020). Mapping citizen science contributions to the UN sustainable development goals. *Sustain. Sci.* 15, 1735–1751. doi: 10.1007/s11625-020-00833-7
- Freitag, A., Meyer, R., and Whiteman, L. (2016). Strategies employed by citizen science programs to increase the credibility of their data. *Citiz. Sci. Theory Pract.* 1:2. doi: 10.5334/cstp.6
- Fritz, S., See, L., Carlson, T., Haklay, M. M., Oliver, J. L., Fraisl, D., et al. (2019). Citizen science and the United Nations sustainable development goals. *Nat. Sustain.* 2, 922–930. doi: 10.1038/s41893-019-0390-3
- Gharaibeh, N., Oti, I., Meyer, M., Hendricks, M., and Van Zandt, S. (2019). Potential of citizen science for enhancing infrastructure monitoring data and decision-support models for local communities. *Risk Anal.* 39, 1–7. doi: 10.1111/risa.13256
- Global Indigenous Data Alliance (2019). *CARE Principles for Indigenous Data Governance*. GIDA. Available online at: <https://www.gida-global.org/care> (accessed October 6, 2020).
- Goudeseune, L., Eggermont, H., Groom, Q., Le Roux, X., Paleco, C., Roy, H. E., et al. (2020). *BiodivERsA Citizen Science Toolkit For Biodiversity Scientists*. BiodivERsA Report, p. 44. doi: 10.5281/zenodo.3979343
- Group on Earth Observations (2016). *GEOSS Data Management Principles*. Available online at: http://earthobservations.org/open_eo_data.php# (accessed September 17, 2020).
- Hecker, S., Wicke, N., Haklay, M., and Bonn, A. (2019). How does policy conceptualise citizen science? A qualitative content analysis of international policy documents. *Citiz. Sci. Theory Pract.* 4:32. doi: 10.5334/cstp.230
- Henkel, H. S., Hutchison, V. B., Langseth, M. L., Thibodeaux, C. J., and Zolly, L. (2015). *USGS Data Management Training Modules—USGS Science Data Lifecycle*. U.S. Geological Survey. doi: 10.5066/F7RJ4GGJ
- Hicks, A., Barclay, J., Chilvers, J., Armijos, M. T., Oven, K., Simmons, P., et al. (2019). Global mapping of citizen science projects for disaster risk reduction. *Front. Earth Sci.* 7:226. doi: 10.3389/feart.2019.00226
- Hunter, J., Alabri, A., and van Ingen, C. (2013). Assessing the quality and trustworthiness of citizen science data. *Concurr. Comput. Pract. Exp.* 25, 454–466. doi: 10.1002/cpe.2923
- Kallimanis, A. S., Panitsa, M., and Dimopoulos, P. (2017). Quality of non-expert citizen science data collected for habitat type conservation status assessment in Natura 2000 protected areas. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-09316-9
- Kelling, S., Fink, D., La Sorte, F. A., Johnston, A., Bruns, N. E., and Hochachka, W. M. (2015). Taking a 'Big Data' approach to data quality in a citizen science project. *Ambio* 44, 601–611. doi: 10.1007/s13280-015-0710-4
- Kosmala, M., Wiggins, A., Swanson, A., and Simmons, B. (2016). Assessing data quality in citizen science. *Front. Ecol. Environ.* 14, 551–560. doi: 10.1002/fee.1436
- Li, E., Parker, S. S., Pauly, G. B., Randall, J. M., Brown, B. V., and Cohen, B. S. (2019). An urban biodiversity assessment framework that combines an urban habitat classification scheme and citizen science data. *Front. Ecol. Environ.* 7:277. doi: 10.3389/fevo.2019.00277
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giarretta, D., et al. (2020). The TRUST Principles for digital repositories. *Sci. Data* 7, 1–5. doi: 10.1038/s41597-020-0486-7
- Lukyanenko, R., Parsons, J., and Wiersma, Y. F. (2016). Emerging problems of data quality in citizen science. *Conserv. Biol.* 30, 447–449. doi: 10.1111/cobi.12706
- Miller, E. T., Leighton, G. M., Freeman, B. G., Lees, A. C., and Ligon, R. A. (2019). Ecological and geographical overlap drive plumage evolution and mimicry in woodpeckers. *Nat. Commun.* 10, 1–10. doi: 10.1038/s41467-019-09721-w
- NASA (2020a). *ESDIS Standards Office Standards and Practices*. Available online at: <https://earthdata.nasa.gov/esdis/eso/standards-and-references#data-quality> (accessed September 17, 2020).
- NASA (2020b). *ESDS Citizen Science Data Working Group White Paper, Version 1.0-24*. Available online at: <https://cdn.earthdata.nasa.gov/conduit/upload/14273/CSDWG-White-Paper.pdf> (accessed October 5, 2020).
- NOAA (2013). *NOAA Environmental Data Management Framework*. Available online at: https://nosc.noaa.gov/EDMC/documents/NOAA_EDM_Framework_v1.0.pdf (accessed February 5, 2021).
- Paul, J. D., Buytaert, W., Allen, S., Ballesteros-Cánovas, J. A., Bhusal, J., Cieslik, K., et al. (2018). Citizen science for hydrological risk reduction and resilience building. *Wiley Interdiscipl. Rev. Water* 5:e1262. doi: 10.1002/wat2.1262
- Peng, G., Lacagnina, C., Downs, R. R., Ivanova, I., Moroni, D. F., Ramapriyan, H., et al. (2020). *Laying the Groundwork for Developing International Community Guidelines to Share and Reuse Digital Data Quality Information – Case Statement, Workshop Summary Report, and Path Forward*. Open Science Foundation (OSF) Preprints. Available online at: <https://osf.io/75b92/> (accessed February 5, 2021).
- Peng, G., Milan, A., Ritchey, N. A., Partee, I. I., R. P., Zinn, S., et al. (2019). Practical application of a data stewardship maturity matrix for the NOAA OneStop Project. *Data Sci. J.* 18, 1–18. doi: 10.5334/dsj-2019-041
- Peng, G., Privette, J. L., Tilmes, C., Bristol, S., Maycock, T., Bates, J. J., et al. (2018). A conceptual enterprise framework for managing scientific data stewardship. *Data Sci. J.* 17:15. doi: 10.5334/dsj-2018-015
- Peng, G., Ritchey, N. A., Casey, K. S., Kearns, E. J., Privette, J. L., Saunders, D., et al. (2016). Scientific stewardship in the Open Data and Big Data era - roles and responsibilities of stewards and other major product stakeholders. *D. Lib Mag.* 22. doi: 10.1045/may2016-peng
- Pettibone, L., Vohland, K., and Ziegler, D. (2017). Understanding the (inter) disciplinary and institutional diversity of citizen science: a survey of current practice in Germany and Austria. *PLoS ONE* 12:e0178778. doi: 10.1371/journal.pone.0178778
- Poisson, A. C., McCullough, I. M., Cheruvilil, K. S., Elliott, K. C., Latimore, J. A., and Soranno, P. A. (2020). Quantifying the contribution of citizen science to broad-scale ecological databases. *Front. Ecol. Environ.* 18, 19–26. doi: 10.1002/fee.2128
- Ramapriyan, H., Peng, G., Moroni, D., and Shie, C.-L. (2017). Ensuring and improving information quality for earth science data and products. *D. Lib Mag.* 23. doi: 10.1045/july2017-ramapriyan Available online at: <http://www.dlib.org/dlib/july17/07contents.html>
- Resnik, D. B. (2019). Citizen scientists as human subjects: ethical issues. *Citiz. Sci. Theory Pract.* 4:11. doi: 10.5334/cstp.150
- Robinson, O. J., Ruiz-Gutiérrez, V., Fink, D., Meese, R. J., Holyoak, M., and Cooch, E. G. (2018). Using citizen science data in integrated population models to inform conservation. *Biol. Conserv.* 227, 361–368. doi: 10.1016/j.biocon.2018.10.002
- Roman, L. A., Scharenbroch, B. C., Östberg, J. P., Mueller, L. S., Henning, J. G., Koeser, A. K., et al. (2017). Data quality in citizen science urban tree inventories. *Urban Forest. Urban Green.* 22, 124–135. doi: 10.1016/j.ufug.2017.02.001
- Sandahl, A., and Tøttrup, A. P. (2020). Marine citizen science: recent developments and future recommendations. *Citiz. Sci. Theory Pract.* 5:24. doi: 10.5334/cstp.270
- Shanley, L. A., Parker, A., Schade, S., and Bonn, A. (2019). Policy perspectives on citizen science and crowdsourcing. *Citiz. Sci. Theory Pract.* 4:30. doi: 10.5334/cstp.293
- Sharma, N., Sam, G., Colucci-Gray, L., Siddharthan, A., and van der Wal, R. (2019). From citizen science to citizen action: analysing the potential for a digital platform to cultivate attachments to nature. *J. Sci. Commun.* 18, 1–35. doi: 10.7717/peerj.5965
- Steger, C., Butt, B., and Hooten, M. B. (2017). Safari Science: assessing the reliability of citizen science data for wildlife surveys. *J. Appl. Ecol.* 54, 2053–2062. doi: 10.1111/1365-2664.12921
- Stevenson, R. (2018). A three-pronged strategy to improve trust in biodiversity data produced by citizen science programs. *Biodivers. Inform. Sci. Stand.* 2:e25838. doi: 10.3897/biss.2.25838
- Strasser, C., Cook, R., Michener, W., and Budden, A. (2012). *Primer on Data Management: What You Always Wanted to Know, But Were Afraid to Ask*. Available online at: http://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf (accessed February 5, 2021).
- Swanson, A., Kosmala, M., Lintott, C., and Packer, C. (2016). A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conserv. Biol.* 30, 520–531. doi: 10.1111/cobi.12695
- Torre, M., Nakayama, S., Tolbert, T. J., and Porfiri, M. (2019). Producing knowledge by admitting ignorance: enhancing data quality through an "I don't know" option in citizen science. *PLoS ONE* 14:e0211907. doi: 10.1371/journal.pone.0211907

- Tredick, C. A., Lewison, R. L., Deutschman, D. H., Hunt, T. A., Gordon, K. L., and Von Hendy, P. (2017). A rubric to evaluate citizen-science programs for long-term ecological monitoring. *BioScience* 67, 834–844. doi: 10.1093/biosci/bix090
- van Etten, J., de Sousa, K., Aguilar, A., Barrios, M., Coto, A., Dell'Acqua, M., et al. (2019). Crop variety management for climate adaptation supported by citizen science. *Proc. Natl. Acad. Sci. U.S.A.* 116, 4194–4199. doi: 10.1073/pnas.1813720116
- Van Eupen, C., Maes, D., Herremans, M., Swinnen, K. R., Somers, B., and Luca, S. (2021). The impact of data quality filtering of opportunistic citizen science data on species distribution model performance. *Ecol. Modell.* 444:109453. doi: 10.1016/j.ecolmodel.2021.109453
- Wiggins, A., Bonney, R., LeBuhn, G., Parrish, J. K., and Weltzin, J. F. (2018). A science products inventory for citizen-science planning and evaluation. *BioScience* 68, 436–444. doi: 10.1093/biosci/biy028
- Wiggins, A., Newman, G., Stevenson, R. D., and Crowston, K. (2011). “Mechanisms for data quality and validation in citizen science,” in *2011 IEEE Seventh International Conference on e-Science Workshops* (Washington, DC: IEEE), 14–19. doi: 10.1109/eScienceW.2011.27
- Wilderman, C. C., and Monismith, J. (2016). Monitoring marcellus: a case study of a collaborative volunteer monitoring project to document the impact of unconventional shale gas extraction on small streams. *Citiz. Sci. Theory Pract.* 1:7. doi: 10.5334/cstp.20
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18
- Conflict of Interest:** HR is employed by the company Science Systems and Applications, Inc.
- The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2021 Downs, Ramapriyan, Peng and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.