

## As open as possible, as closed as necessary: how to find the right balance in sharing citizen science data for health?

Ria Wolkorte<sup>a\*</sup>, Lieke Heesink<sup>b</sup>, and Michelle M.A. Kip<sup>a</sup>

<sup>a</sup>*Department of Health Technology & Services Research, Technical Medical Centre, Faculty of Behavioural, Management and Social Sciences, University of Twente, Enschede, The Netherlands*

<sup>b</sup>*Biomedical Signals and Systems Group, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands*

E-mail: [r.wolkorte@utwente.nl](mailto:r.wolkorte@utwente.nl)

Citizen science should adhere as much as possible to the ideas of open science. On the topic of data sharing, the FAIR principles could be applied, making data as open as possible and as closed as necessary. However, it is unclear how open participants want to be with health-related data that is collected as part of a citizen science project. The current study is a first investigation of this topic with a multidisciplinary group of stakeholders and experts in a citizen science for health project. In two focus group discussions we discussed the views and desires regarding the sharing of data and the preferences for the levels of openness when sharing. The results of this study provide preliminary preferences regarding openness; people find sharing of data important yet not all are comfortable with making data completely open. Therefore, it was chosen to share metadata of our projects in repositories, and provide the full, anonymized or pseudonimized, dataset upon reasonable request. Future research needs to be done to confirm the current findings and to investigate the preferences of a broader group of participants in citizen science for health.

*Engaging Citizen Science Conference 2022 (CitSci2022)  
25-26 April 2022  
Aarhus University, Denmark*

\*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

<https://pos.sissa.it/>

## 1. Introduction

Citizen science inherently requires openness and transparency with regard to collecting and sharing data. This openness and transparency form one of the 10 principles for citizen science projects as defined by the European Citizen Science Association [1]. More specifically, it is argued that since data is collected in the public domain, these data and the results should also be shared publicly.

Sharing of results (i.e. aggregated data) can be accomplished through publications which offer open access formats such as scientific journals, general purpose repositories such as Zenodo [2], or freely accessible summaries and reports. Repositories allow the sharing of raw data, which in turn facilitates the reuse of these data for other research, and allow reproducing and confirming the findings of the original study.

In order to promote the sharing of data, citizen science projects should, where possible, adhere to the principles of FAIR (making data Findable, Accessible, Interoperable and Reusable) [3]. The basic notion of FAIR is "as open as possible, as closed as necessary", implying that the data does not necessarily have to be fully open to comply to the principles of FAIR. To accommodate this, repositories offer different levels of openness ranging from completely open, to placing data under embargo with access only upon reasonable request, to only publishing the metadata. This "metadata only" option implies that only the dataset is described, but the data itself are never shared.

In the domain of health and wellbeing the principles of openness and transparency sometimes collide with the principles of privacy. This issue is particularly present with qualitative data [4, 5]. For these situations, there is no one-size-fits-all approach in balancing openness against privacy.

The TOPFIT Citizenlab is a fieldlab in the Netherlands where researchers from the University of Twente work together with citizens, healthcare organisations, companies, and citizens. It focuses on projects to improve health and wellbeing through a citizen science approach. One of the projects involves a cooperation between researchers and people with rheumatoid arthritis and focuses on fatigue and factors associated with the fatigue experienced by people with rheumatoid arthritis. The project entailed interviews, surveys with open- and closed-ended questions, focus groups, and the longitudinal collection of data on a digital platform. The collected data concerned among others personal and health-related data and involves a mix of quantitative and qualitative data.

As there is currently no guidance available regarding how to manage such data, a crucial step was to understand the views of different stakeholders regarding the use of repositories to share the data collected during this project [6]. This was especially important as we believed that the participants should be in full control of their data.

To ensure that the data management strategy was broadly supported, we decided to investigate the views of the co-researchers (i.e. people with rheumatoid arthritis), citizen science researchers, ethicists, and data stewards.

## 2. Methods and results

Two focus group discussions were organised by the researchers. Both lasted 90 minutes and were held through the online video conferencing program Zoom. Before and at the start of each focus group participants received a short explanation of relevant concepts such as FAIR data, repositories, citizen science, and anonymization. Both focus groups were recorded with permission of the participants and the recording was transcribed verbatim using Amberscript [7]. Ethical approval was provided by the ethical review board of the faculty of Behavioural, Management and Social Sciences of the University of Twente (req. nr. 211340).

### 2.1 Monodisciplinary focus group discussion

The first focus group was attended by four people with rheumatoid arthritis. Participants discussed 1) the relevance of health-related data sharing, 2) the desired level of openness when sharing health-related data, 3) the importance of the type of data (i.e., qualitative vs quantitative), and 4) the information needs for the informed consent procedure. The views and preferences of participants were summarized and used as input for the second (i.e., a multidisciplinary) focus group discussion.

The main outcomes were that people with rheumatoid arthritis considered it important to share data from citizen science projects. The main reasons were that reuse of data is more efficient and saves both time and resources. However, they also felt that health data should not be openly available, even after anonymization. The consensus was that data should be anonymized and shared with restricted access, i.e. the data would only be made available upon reasonable request by the researchers, regardless of the type of data. Furthermore, participants should be made aware of the possibility that their data will be uploaded to a repository through the use of a simple and understandable informed consent procedure.

### 2.2 Multidisciplinary focus group discussion

The second focus group discussion was attended by two of the four people with rheumatoid arthritis who also joined the first focus group discussion, a data steward, an ethicist, and a citizen science researcher. The topics that were discussed were similar to the first focus group. Prior to sharing the outcomes of the first focus group with the participants, the data steward, ethicist, and researcher were asked to share their views. This was followed by general discussion on the topics of 1) the relevance of health-related data sharing, 2) the desired level of openness when sharing health-related data, 3) the importance of the type of data (i.e. qualitative vs quantitative), and 4) the information needs for the informed consent procedure.

The last part of the discussion concerned how to deal with the data (i.e., the transcript) resulting from this second focus group. In order to make an informed decision on providing consent to share the transcript with any degree of openness, we believed that people had to join the discussion first. This way, they would know 1) what kind of information was gathered, and 2) what the levels of openness meant. We therefore decided to ask for a strict version of consent prior to joining this discussion (i.e., the data would only be used for this specific research project). However, we also mentioned in the information letter that the data management of this focus group would be part of the discussion as well. The views and preferences of participants were summarized.

The main outcomes were similar to the first focus group discussion, i.e., research data should be made available upon reasonable request. This was also the participants' provisional choice for the transcript resulting from this focus group. To make a definitive decision, they asked if it was possible to see the anonymized transcript.

### 2.3 Follow-up

The researchers transcribed the focus group verbatim and wrote a public summary of the focus group results. Both were sent to all participants, allowing them to read the documents, notify the researchers about disagreements with the transcription of the text, and make an informed decision on the data sharing policy. Each participant sent an email with their preference to the researchers. Both researchers and participants agreed that the least open policy option that the group agreed upon would be applied to this dataset. This way, we would adhere to the principles of FAIR as much as possible.

All participants agreed with the level of anonymization of the dataset and agreed with the summary of the results. All participants agreed to making the dataset available upon request, i.e. publishing metadata in a repository and only making the full dataset available upon reasonable request from the main researcher. This researcher will judge whether a request is reasonable.

### 2.4 Roundtable dialogue

*The results were presented during the roundtable dialogue session of the Engaging Citizen Science Conference in Aarhus (25-26 April 2022).* The roundtable dialogues involved two separate rounds of discussions with 4-5 conference participants and the presenter (RW). RW briefly outlined the problem: 1) whether participants recognized the need for guidance for data management in health-related citizen science projects, 2) participants' views on openness of health-related citizen science data, and 3) suggestions for a guideline for FAIR data for citizen science for health. The problem was then discussed.

As health-related data is privacy sensitive, there was consensus among the participants that health-related data would benefit from specific guidance. It was acknowledged that other fields of citizen science may also deal with sensitive data (for instance, locations of endangered animals and/or plants), and the different research fields may inform each other. Furthermore, participants suggested that a single dataset might be shared with different levels of openness, i.e. aggregated data can be made open completely, whereas the raw anonymized dataset can be made available with restricted access, and perhaps part of the data (i.e., data from one focus group, data from one single person) can be made open with specific consent. All participants agreed that transparent data management is crucial and participants should be informed about this data management in a concise and understandable fashion.

## 3. Discussion and future steps

Although the principles of FAIR data are important in citizen science data management, the application to health-related data and qualitative data requires further elaboration and practical guidance. The guidance form should inform future projects in the field of health and

well-being on data management and sharing procedures, while taking ethical and privacy issues into account. These guidance forms should be developed in close cooperation with the participants/co-researchers and leave room for different applications depending on the specific context of each project. The current study informed first ideas on the data management policy for health-related data in citizen science projects at the TOPFIT Citizenlab. However, as the results are based on a small sample of participants, who were also very well informed, it is important to follow-up on the current results. This should provide more insight in the application of the current findings in a broader setting or with different target populations and topics.

## References

- [1] L. D. Robinson, J. L. Cawthray, S. E. West, A. Bonn & J. Ansine, Ten principles of citizen science, in: *Citizen Science: Innovation in Open Science, Society and Policy* (1 ed., pp. 27-40), eds. S. Hecker, M. Haklay, A. Bowser, Z. Makuch, J. Vogel, & A. Bonn, UCL Press (2018). <https://doi.org/10.14324/111.9781787352339>
- [2] Zenodo, <https://zenodo.org/>
- [3] M. Wilkinson, M. Dumontier, I. Aalbersberg, et al., The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3** (2016) 160018. <https://doi.org/10.1038/sdata.2016.18>
- [4] M. W. Ross, M. Y. Iguchi, & S. Panicker, Ethical aspects of data sharing and research participant protections, *American Psychologist* **73**, 2 (2018) 138–145. <https://doi.org/10.1037/amp0000240>
- [5] Katherine A. Tamminen, A. Bundon, B. Smith, M. H. McDonough, Z. A. Poucher & M. Atkinson, Considerations for making informed choices about engaging in open qualitative research, *Qualitative Research in Sport, Exercise and Health*, **13**, 5 (2021) 864-886. DOI: <https://www.tandfonline.com/doi/pdf/10.1080/2159676X.2021.1901138>
- [6] J. A. Cummings, & T. E. Day, But what do participants want? Comment on the “Data Sharing in Psychology” special section (2018), *American Psychologist*, **74**, 2 (2019) 245–247. <https://doi.org/10.1037/amp0000408>
- [7] Amberscript, <https://www.amberscript.com/en/>