Empowering Citizen Science Through Digital Technologies: New Participatory Approaches to Data Collection, Sharing, and Analysis

Luisa Pumplun
Software & Digital Business
Technical University of Darmstadt
Deutschland
luisa.pumplun@tu-darmstadt.de

Amina Wagner
Software & Digital Business
Technical University of Darmstadt
Deutschland
amina.wagner@tu-darmstadt.de

Tobias Störzinger
Philosophisches Seminar
University of Göttingen
Deutschland
tobias.stoerzinger@uni-goettingen.de

Mariska Fecho
Software & Digital Business
echnical University of Darmstadt
Deutschland
mariska.fecho@tu-darmstadt.de

Peter Buxmann
Software & Digital Business
Technical University of Darmstadt
Deutschland
peter.buxmann@tu-darmstadt.de

Catrin Misselhorn
Philosophisches Seminar
University of Göttingen
Deutschland
catrin.misselhorn@uni-goettingen.de

ABSTRACT

This workshop revolves around citizen science as an approach that allows people to collect data, share it with scientists, and actively participate in its analysis via modern digital technologies. In recent decades, Internet users have generated ever increasing amounts of personal data, e.g., through the use of mobile devices, wearables, or social networks. At the latest since the outbreak of the Covid 19 pandemic, it is clear that citizens are willing to donate this data for a good cause and in particular to promote science. Beyond donating data, citizens are increasingly involved in other phases of the research process, such as data analysis. This draws citizens and scientists closer together and blurs boundaries between the otherwise separated parties. Beyond the promise of larger data sets and more effective data analysis, however, citizen science has its downsides, especially when it comes to broad citizen participation through digital technologies. For example, key issues such as ensuring data privacy or quality are still unresolved. Therefore, the workshop will address the advantages and disadvantages of digitally enabled citizen science, the use of advanced digital technologies to engage people in research, and specifically data donation, which has been of particular interest in research and practice recently. We invite researchers to submit contributions in the area of digitally enabled citizen science to further develop this branch of research with respect to ethical, legal, and technical concerns, among others. We will also host a panel discussion with several experts from different disciplines on the special interest topic data donation to share ideas and discuss how this type of citizen science can be realized. We intend to use the results of this workshop as a starting point for further discussion, research, and steps towards participatory science.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Veröffentlicht durch die Gesellschaft für Informatik e.V.
in K. Marky, U. Grünefeld & T. Kosch (Hrsg.):
Mensch und Computer 2022 – Workshopband, 04.-07. September 2022, Darmstadt
© 2022 Copyright held by the owner/author(s).
https://doi.org/10.18420/muc2022-mci-ws05-111

1 INTRODUCTION TO WORKSHOP

According to [1, p. 467] "a citizen scientist is a volunteer who collects and/or processes data as part of a scientific enquiry". This definition results in three aspects for citizens to actively participate in science. That is, opportunities for citizens to (1) generate data, (2) actively share it with scientists, and (3) subsequently analyze the data. Citizen participation in science has grown steadily in recent decades. For instance, the German public health agency Robert Koch Institute recently asked citizens to disclose their health data in order to conduct research on the symptomatology of the novel Covid-19 virus, identify particular areas of spread, and take appropriate countermeasures [2]. But it is not only since the Covid-19 crisis that scientific institutions have called for participation in science and citizens have volunteered to serve scientific goals. In the crowdsourcing initiative Every Name Counts, more than 24,000 volunteers have helped digitize files of victims of National Socialism in the past several years [3], resulting in an invaluable basis for further reappraisal and historical research. In addition to generating and sharing data, citizens assist in analyzing the collected data. For example, in the Galaxy Zoo project, citizens can help better understand distant galaxies by evaluating images of them [4]. In all these projects, it is evident that science in general, and research on humans in particular, relies on the participation of citizens. Indeed, even completing a simple survey requires the involvement of people. Especially against the backdrop of datahungry analysis methods, such as analytics and machinelearning based artificial intelligence, data generated and shared by people is becoming increasingly important [5]. Consequently, as scientists who analyze data on a daily basis and use state-ofthe- art methods for this, we depend on the citizens to actively support us.

Citizen science can be significantly facilitated through the use of digital technologies. Technological approaches such as wearables,

social media, or crowdsourcing platforms promise to enable new forms of participation [6], making modern citizen science basically accessible to anyone owning a computer or smartphone [7]. For example, a platform called Open Humans allows Internet users to store and donate their personal information such as medical, social media, and location data to selected research projects [8]. Indeed, digital technologies enable data to be easily collected by people, shared quickly and efficiently with institutions around the world, and groups of citizens to join together and organize themselves to analyze data. Just as the fourth wall has been broken down in the theater to actively involve the audience, today's technology allows citizens to move beyond being mere spectators and increasingly engage in a process that was otherwise reserved for a scientific elite.

2 DIGITAL DATA DONATION AS A SPECIFIC FOCUS OF THE WORKSHOP

Researchers, especially from social science but also from computer science, depend on personal information being readily available at all times as they use this data to test theories, derive implications, propose countermeasures, and prepare innovations. At the same time, a lot of data is collected and aggregated by Internet users every day (e.g., via smartphones or wearables) that could serve as a valuable database for research. Recent surveys have shown that these Internet users are very willing to share their collected personal information for a good cause [9]. For example, [10] demnstrate the high willingness of Internet users to share their personal health data to combat viral diseases. So far, however, there is a lack of suggestions as to how this data could be donated, i.e., disclosed for a good cause without any consideration in return. Legal, ethical, organizational, societal, and technological aspects of the data donation process are still unresolved and need further expert discussion. Indeed, it is crucial for future research to make personal data accessible and usable without exploiting Internet users (i.e., donors) [11]. Therefore, data donation should be designed in a way that simultaneously enables individual privacy and benefits research. Other ethical challenges concern the possible implications and uses of scientific results and the ability of citizens to make autonomous and informed decisions about their contributions to them. Since these challenges call for a discussion between experts from different disciplines such as law, philosophy, medicine, among others, we will organize a panel discussion as a starting point for further research.

3 PROBLEMATIZATION AND DERIVED NEED FOR RESEARCH

Although digital technologies may help overcome the fourth wall between scientists and citizens, they also bring along challenges that have so far been addressed only rudimentarily. In particular, while digital technologies enable the sharing of individuals' data, this data is typically personal data that requires protection[10]. Accordingly, data privacy concerns play a major role, which are particularly evident in large-scale citizen participation that is only weakly coordinated by a central institution [6]. Leading scientists are therefore calling for new digital approaches to citizen science to be promoted without compromising citizens' privacy [11]. In addition, the question arises of how to ensure consistent data quality, even though large numbers of people participate in science via digital technologies (e.g., by typing in counted birds manually for recording the species population) [1, 7, 12, 13]. In addition, researchers should consider how to encourage people to participate in digitally-supported citizen science. Participation is by no means a matter of course; citizens expect feedback on their contribution, which must be communicated back to them through digital technologies [7].

This workshop will explore in more detail how these challenges can be overcome and what role digital technologies play in this effort. Topics of interest include but are not limited to:

- Design and development of new digital technologies that support citizen science (e.g., web scraping, machine learning-based approaches)
- Drivers and barriers to the usage of digital technologies for citizen science
- Real world examples of digitally-enabled citizen science projects, considering promises and pitfalls (e.g., descriptions of data analysis/performance, collected datasets)
- Financing digitally-enabled citizen science projects
- Digitally-enabled citizen science, privacy, and trust
- Overall societal impact of digitally-enabled citizen science on human welfare
- Measures to improve digitally-enabled citizen science and especially data quality (e.g., training, app design, approaches to automatic control of data quality)
- Historical impact of digital technologies on citizen science
- Effects of excluding non-digital natives from citizen science; aspects of fairness We particularly welcome submissions on digital data donation as a topic of special interest:
- Perils and promises of digital data donation
- Antecedents of digital data donation
- Data philanthropy
- Legal and ethical aspects of digital data donation
- Encouraging digital data donation while maintaining privacy

4 CONTRIBUTION TYPES, TARGET AUDIENCE, AND PLANNED WORKSHOP FORMAT

We welcome completed research papers that include theoretical and empirical studies and employ all types of methodological approaches, such as surveys, experiments, qualitative studies, simulations, or design science research. Submissions must be original and may not be published anywhere else. Papers for the workshop should be formulated in English, have a maximum length of eight pages, and be submitted in PDF format. Regarding the format of the submission, please follow the ACM Guidelines (see Link for more information). We are committed to a double-blind review process, meaning there should be no author information in the initial submission. Digitally-enabled citizen science is an interdisciplinary field of research that not only encompasses technical aspects, but focuses on people as well. Therefore, interested authors and listeners from various scientific disciplines are invited to participate in the workshop (e.g., researchers and practitioners, computer and social scientists). Accepted papers of authors will be published through

the 2022 proceedings of the Mensch und Computer Conference. Further, we will be hosting a panel discussion on the perils and promises of data donation; a topic of particular interest to digitally-enabled citizen science. The panel discussion will feature various specialists on the topic who will look at the prerequisites and effects of data donation from different perspectives (e.g., ethics, legal, medical, technical perspective).

REFERENCES

[1] E. Aceves-Bueno et al., "The Accuracy of Citizen Science Data: A Quantitative Review," Source Bull. Ecol. Soc. Am., vol. 98, no. 4, pp. 278–290, 2017, doi: 10.2307/90013289.

[2] Robert Koch Institute, "Corona-Datenspende-App 2.0," 2021. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Corona-Datenspende-allgemein.html (accessed Mar. 09, 2022).

[3] Arolsen Archives, "Help us build a digital memorial," 2022. https://enc.arolsen-archives.org/en/about-everynamecounts/ (accessed Mar. 09, 2022).

[4] Galaxy Zoo, "Few have witnessed what you're about to see," 2022. https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/ (accessed Mar. 09, 2022).

[5] M. Marabelli, S. Newell, and V. Handunge, "The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges," J. Strateg. Inf. Syst., vol. 30, no. 3, p. 101683, 2021, doi: 10.1016/j.jsis.2021.101683.

[6] J. Preece, "Citizen Science: New Research Challenges for Human–Computer Interaction," Int. J. Hum. Comput. Interact., vol. 32, no. 8, pp. 585–612, Aug. 2016, doi: 10.1080/10447318.2016.1194153.

[7] J. Silvertown, "A new dawn for citizen science," Trends Ecol. Evol., vol. 24, no. 9, pp. 467–471, 2009, doi: 10.1016/j.tree.2009.03.017.

[8] M. Bietz, K. Patrick, and C. Bloss, "Data donation as a model for citizen science health research," Citiz. Sci. Theory Pract., vol. 4, no. 1, pp. 1–11, 2019, doi: 10.5334/CSTP.178.

[9] Medical Informatics Initiative, "Substantial majority of Germans willing to provide data for medical research," 2019. https://www.medizininformatik-initiative.de/en/substantial-majoritygermans-willing-provide-data-medical-research (accessed Feb. 07, 2022).

[10] L. Pumplun, A. Wagner, C. M. Olt, A. Zöll, and P. Buxmann, "Acting egoistically in a crisis: How emotions shape data donations," 2022.

[11] M. Marabelli, S. Newell, and V. Handunge, "The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges," J. Strateg. Inf. Syst., vol. 30, no. 3, 2021, doi: 10.1016/j.jsis.2021.101683.

[12] B. Balázs, P. Mooney, E. Nováková, L. Bastin, and J. J. Arsanjani, "Data Quality in Citizen Science," in The Science of Citizen Science, K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, and K. Wagenknecht, Eds. Cham, Switzerland: Springer, 2021, pp. 139–157.

[13] R. Lukyanenko, A. Wiggins, and H. K. Rosser, "Citizen Science: An Information Quality Research Frontier," Inf. Syst. Front., vol. 22, no. 4, pp. 961–983, Aug. 2020, doi: 10.1007/s10796-019-09915-z.