



Getting it right: implementing data protection in citizen science research

As citizen science projects increase in number, scope and relevance to policy, the conduct of the underlying research is likely to come under greater scrutiny, particularly if any resulting research findings are to be regarded as legitimate. Among the most important aspects of research ethics scrutiny is the handling of data. The ever-increasing attention to the data protection rights of citizens can be seen across research domains, from health to environmental impacts. Careful attention to data protection is essential to the sustainability of citizen science for practical reasons as well. Complying with data protection laws is not optional. It is a legal requirement. Failure to meet the requirements can result in substantial fines, which most citizen science projects can ill afford. Moreover, the scientific, political and legal viability of citizen science hinges on getting data protection right. This article sets out some of the key data protection considerations that commonly arise in conducting a citizen science project.

Keywords

data protection; citizen science; right to data protection; privacy in research; GDPR

ROBIN PIERCE

Associate Professor
Tilburg Institute for
Law
Technology, Markets
and Society

MARIANA EVRAM

Student Assistant
Tilburg University

Introduction

As citizen science projects increase in number,¹ scope and relevance to policy, the conduct of the underlying science is likely to come under greater scrutiny. Moreover, the viability of citizen science hinges not only on the quality and relevance of the results that it produces, but equally as much on the process through which the results were obtained. One of the most important areas is the handling of data, particularly in view of the increased attention to the data protection rights of citizens. This article sets out some of the key data protection considerations that commonly arise in conducting a citizen science project. We have selected a few issues at different stages of a citizen science project to highlight that data protection is an ongoing process – one that starts in the planning, before the project begins, and continues even after the project ends. The scientific, political and legal viability of citizen science hinges significantly on getting data protection right. From principles to practice, this article points to the path to achieving responsible handling of data by citizen scientists.

Among the purported benefits of citizen scientists is that it empowers laypersons to take an active role in knowledge production, thereby 'democratizing' science, a traditionally closed activity reserved for professionals. However, with this more inclusive approach to science comes additional responsibilities. Depending on the type of citizen science, whether bottom-up (with questions, goals and methods originating from a community of laypersons) or institution-led projects that recruit laypersons to collect or contribute data, or some combination in between, citizen scientists may find themselves with some or all of the responsibilities of an institutional principal investigator.² In a sense, the citizen scientist takes on the dual role of 'participant' and 'researcher' when taking an active role in directing the research. They are now cast as both 'protected participant' and 'researcher', responsible for providing those protections. Among the most serious responsibilities that they must take on is the proper handling of data.

2 This article aims to act as a starting point for thinking about the data protection aspects of a citizen science project. The selected issues aim to demonstrate how data protection principles and regulations translate into real-world applications. Of course, there is no substitute for consultation with a data protection expert about the needs and considerations for a particular project in a particular jurisdiction. However, citizen scientists benefit themselves and their projects by having some understanding of what goes into setting up a data-responsible project.

'citizen scientists benefit ... by having some understanding of what goes into setting up a data-responsible project'

Planning – before recruiting participants

How do I decide what data to collect?

To a researcher, the notion that 'more data is better' seems natural. However, the General Data Protection Regulation (GDPR) sets forth the principle of 'data minimization', which provides an important guide for this question. Data minimization states that data collection should be 'adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed'.³ Therefore, when deciding what data to collect, it is essential to identify which data you absolutely need for your research and why. Some data about participants may be necessary to answer the question, e.g. gender, ethnicity, age, if the study is seeking to understand or identify patterns or problems that may have policy implications related to these attributes as they pertain to the subject under study. If not, the principle of data minimization should be applied.

'it is essential to identify which data you absolutely need for your research and why'

Are there different categories of data?

Yes. The general aim of data protection law is to offer broad protection to individuals by protecting their personal data (any information relating to an identifiable person⁴). Therefore, whether data can identify you becomes a critical consideration. If the information is likely to identify you, then it falls into the category of 'personal data' and will need to be protected. Information such as a name, e-mail address, or even less direct attributes, such as the physical, psychological, genetic, mental, economic, cultural or social identity, may constitute personal data.⁵ A citizen science project may collect many types of data. For example, the number of pedestrians crossing a bridge during rush hour is non-personal data, while the names or addresses of people filling out a survey about crossing the bridge is personal data. If a project needs to correspond with participants, provide training, or keep track of inputs from a person over time, it will involve the processing of the personal data of those participants and data protection provisions will apply.

A common practice is to 'code' the identity of a participant, using a pseudonym, number, or other way of labelling the data, but also allowing for reidentification with access to a key that links the pseudonym to the real name. For example, writing in a survey 'Participant AB12495' instead of 'John Smith' while keeping a separate key may offer a greater level of protection of John Smith's identity, but only if the key is stored and handled in a secure manner. This process, called pseudonymization, still involves the handling of personal data and data protection laws apply. Additionally, there are many other applicable provisions, including providing the means for persons to access their data, and in the case of inaccuracies, to rectify it.⁶ Thus, citizen scientists, in these dual roles as researcher-participant, must decide which data is necessary to the project, whether this data is personal (identifiable), and how it will be collected and stored securely.

Is there some data we should not collect or data that must be handled differently?

There are 'special categories' of personal data that should be handled differently. 'Sensitive data' are special because they can reveal information that can make the person vulnerable or can be used against them in some way. The law regards this data as needing a heightened level of protection. For example, 'sensitive data' includes political opinions, religion, racial or ethnic origin, trade union membership and data concerning the health, sex life or sexual

3 orientation of the individual.⁷ Therefore, the citizen scientist, as researcher-participant, should decide whether sensitive data is necessary to the study and, if so, how to ensure adequate protection.

If you find that the project needs to collect 'sensitive data', you must obtain the *explicit consent* of the participant either via a written statement or e-mail. Again, determining whether you need to collect 'sensitive data' should be governed by the principle of data minimization.

Who is responsible for what happens with the data?

Identifying responsible persons for any project is a key part of planning. For citizen science projects that are not primarily run by institutions, the citizen researchers essentially become the responsible parties. Where institutions and citizens enter into a research partnership, written agreements about responsibilities may be appropriate.⁸ Data flow presents many challenges pertaining to collection, access and storage, for example. Designating someone to orchestrate and manage this process is highly recommended. There are two roles regarding data protection that citizen scientists should be aware of at the outset of a project – 'data controller' and 'data processor'. The data controller is the ultimately responsible person(s) or organization charged with complying with data protection legal obligations.⁹ This responsibility lasts throughout the project¹⁰ and includes identifying the purpose of collection and authorizing who can process or handle the data, as well as ensuring compliance during the project and designating duration and manner of storage at the end. This responsibility may be shared by all partners who co-determine purpose and contribute to decision-making. The data processor, on the other hand, refers to anyone who processes data on behalf of the project. Thus, each team in a citizen science project, for example, may be a data processor if they collect and analyse data.

'whether you need to collect "sensitive data" should be governed by the principle of data minimization'

Recruitment

What do prospective participants need to know about what happens with their data?

One of the most important aspects of setting up a research project is obtaining informed consent from the participants. Essential information that must be in the consent includes the *purpose* for which data is being processed, data controller contact information and duration of data processing and storage, among other things. Thus, in order to give proper informed consent, the participants must be provided with information about what will happen with their data.¹¹

Study and analysis

Can we circulate the data among different project teams in different countries?

Some citizen science projects have multiple project partners located in different countries. As a result, project partners may wish to transfer data across different countries for purposes of analysis. How feasible this transfer is will depend on many factors, including whether the data is anonymized or identifiable and which countries are involved.

For example, projects whose partners are all located in the European Union may be less complicated because each country is subject to the GDPR. If partners are located outside of the EU, there will be additional responsibilities and requirements before the data can be shared to ensure 'adequate protection', either by having been issued with an adequacy decision by the European Commission or by implementing a set of Standard Contractual Clauses (SCC) (publicly available on the European Commission's website).¹² Some citizen science projects may try to minimize complicating factors in international transfers either by using anonymous data or by conducting all processing within the country.

Are there special considerations for the storage of data?

Setting a clear time frame for the storage of the data is important because the longer the data is kept, the higher the risk that it could be leaked or misused in ways that will harm the rights or interests of participants. The principle of 'storage limitation' in the GDPR can clarify how to approach this issue. It states that data should be 'kept in a form which permits identification of data subjects for *no longer than is necessary for the purposes for which the personal data are processed*'.¹³

Naturally, the more sensitive the data categories stored, e.g. health data, the higher the security measures ought to be. Hence, to prevent unauthorized access, loss or alteration of the data, technical and organizational measures must be used.¹⁴ It is important to note that citizen scientists remain responsible for the handling and storage of personal data even after the project is over. At the end of the project, the data should either be deleted or anonymized. Data protection law does not apply to anonymized data.

'citizen scientists remain responsible for the handling and storage of personal data even after the project is over'

What do we need to do if there is a security incident/data breach?

Even with the best efforts, a data breach can occur and the project must be prepared to deal with this quickly and appropriately. In traditional institution-led research, designated institutional personnel will be responsible for notifying the appropriate authorities if a data breach occurs. However, in citizen science projects where responsibility is shared, it is helpful to designate a specific person to take these actions since the project will be responsible for quickly notifying the regional supervisory authority and, under certain circumstances, any individuals whose data was breached.¹⁵

It is useful to be aware that multiple types of incidents could lead to a data breach, such as:

- the loss/theft of a laptop, which included citizen scientists' data
- a misdirected e-mail
- unauthorized access (e.g. someone from your organization who was not authorized to handle citizen science data).

'After an individual withdraws their consent ... that person's data can no longer be used'

What do we do (with the data) if someone withdraws from the study?

Consider a situation in which a few months into your project, a few citizen scientists decide that they no longer wish to participate. The project should demonstrate in a reasonable and timely way that a participant can withdraw.¹⁶ For example, an opt-out link at the bottom of the relevant webpage can provide this option. After an individual withdraws their consent, generally, that person's data can no longer be used for the study from the date of the withdrawal.

Concluding thoughts

It should be clear that data protection is not something to be left to the lawyers (although we strongly recommend consultation with a data protection expert). Data protection is an integral part of a citizen science project. This short excursion through a few common questions about how data protection is an ever-present part of citizen science aims to alert those involved in citizen science to the centrality of this body of rights and interests. Moreover, by becoming familiar with the data protection principles of accountability, accuracy, data minimization, purpose limitation, lawfulness, transparency, integrity and confidentiality and storage limitation set out in Article 5 of the GDPR, citizen scientists can learn how to think in terms of what data-responsible citizen science looks like. The questions and answers are intended to illustrate how some of these principles operate in practice. Data protection is a very detailed field, but the more you are aware of what data protection means for your project, the better you are situated to seek out and implement what the law requires.

'data protection is not something to be left to the lawyers'

Acknowledgements

The authors wish to acknowledge grant funding ID 873169 from the H2020 project, "Citizen Scientists Investigating Cookies and GDPR Compliance". The authors also wish to thank Ana-Maria Hriscu for her valuable input to this paper.

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: <http://www.uksg.org/publications#aa>.

Competing interests

The authors have declared no competing interests.

References

1. Lea Den Broeder et al., "Citizen Science for public health," *Health Promotion International*, 33, no. 3, (2018): 505–514, DOI: <https://doi.org/10.1093/heapro/daw086> (accessed 11 November 2021).
2. Alena Buyx et al., "Every participant is a PI. Citizen science and participatory governance in population studies," *International Journal of Epidemiology*, 46, no. 2, (April 2017): 377–384, DOI: <https://doi.org/10.1093/ije/dyw204> (accessed 11 November 2021).
3. "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," European Union, April 27, 2016, Article 5(c), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> (accessed 12 November 2021).
4. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Recital 26.
5. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Article 4.1.
6. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Article 15–16.
7. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Article 9.
8. David B. Resnik, Kevin C. Elliott, and Aubrey K. Miller, "A framework for addressing ethical issues in citizen science," *Environmental Science & Policy*, 54, (2015): 475–481, DOI: <https://doi.org/10.1016/j.envsci.2015.05.008> (accessed 11 November 2021).
9. European Parliament and Council of European Union (2016) Regulation (EU) 2016/679. Article 4.7, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN> (accessed 23 November 2021).
10. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Article 24.
11. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Article 7.
12. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Chapter 5.
13. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Article 5(e).
14. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Article 32.
15. European Union, "REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Article 33
16. "Guidelines 05/2020 on consent under Regulation 2016/679," Article 7.3, European Data Protection Board, adopted May 4, 2020, [edpb_guidelines_202005_consent_en.pdf \(europa.eu\)](https://www.edpb.europa.eu/our-work-and-activities/guidelines-recommendations-and-best-practices/guidelines-202005-consent-en.pdf), (accessed 11 November 2021).

Article copyright: © 2022 Robin Pierce and Mariana Evram. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author:

Robin Pierce

Associate Professor

Tilburg Institute for Law

Technology, Markets and Society, NL

E-mail: R.L.Pierce@tilburguniversity.edu

ORCID ID: <https://orcid.org/0000-0002-4799-241X>

To cite this article:

Pierce R and Evram M, "Getting it right: implementing data protection in citizen science research," *Insights*, 2022, 35: 2, 1–6; DOI: <https://doi.org/10.1629/uksg.538>

Submitted on 09 November 2020

Accepted on 22 September 2021

Published on 19 January 2022

Published by UKSG in association with Ubiquity Press.