



# TRAINING MATERIALS for Responsible Open Science

## Collection of cases

**Authors:** Signe Mežinska, Ivars Neiders

**Contributors:** Jaana Eigi, Kadri Simm

**Reviewers:** Rosemarie Bernabe, Sandra Bendiscioli, Stephanos Cherouvis, Ilaria Anna Colussi, Keziah Chanyisa Khayadi Dash, Su Nee Goh, Eva Hnátková, Margarita Poškutė, Vivian Mbanya, Lilian Kwamboka Mocheche, Mari-Liisa Parder, Vana Stavridi.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under GA No 101006430

### Table of Contents

---

Table of Contents.....	2
Introduction.....	4
Cases grouped on content.....	5
CASE 1: Conflict of interests .....	6
CASE 2: Informed consent to open publishing of a dataset.....	8
CASE 3: Citizen science and privacy of data.....	9
CASE 4: Data quality in citizen social science.....	11
CASE 5: Research involving vulnerable groups .....	13
CASE 6: Using sensitive social media data for open science .....	14
CASE 7: Open access biodiversity data .....	16
CASE 8: Open data and risk of looting in archaeology.....	18
CASE 9: Using sensitive data from social media.....	19
CASE 10: Data quality in citizen science: climate research.....	21
CASE 11: Data sharing and reanalysis in medicine .....	23
CASE 12: Sharing sensitive qualitative data .....	25
CASE 13: Data privacy in water sciences.....	27
CASE 14: Open geospatial data in agriculture research .....	29
CASE 15: Refusal to share raw data.....	30
CASE 16: Low-resourced research environments as a barrier to openness.....	31
CASE 17: Open data in genome research .....	33
CASE 18: Refusal to share neuroscience data .....	35
CASE 19: Should scientists share data in climate science?.....	37
CASE 20: Collaborative authorship in digital humanities .....	39
CASE 21: Recognizing citizen scientists in scientific publications .....	41
CASE 22: Authorship and contributorship in citizen science.....	43
CASE 23: Trade-offs in publishing preprints.....	45
CASE 24: Open post-publication peer review.....	47

CASE 25: Open peer review.....	49
CASE 26: Publication of clinical case in open access journal.....	51
CASE 27: Retraction due to the non-compliance with the journal's data policy.....	52
CASE 28: Predatory publishing – the dark side of open science? .....	54
CASE 29: Should all correction notices be open access? .....	56
CASE 30: Accessing pirated papers for research.....	57
CASE 31: Informing the general public about research.....	59
CASE 32: Inequities and potential of exploitation in open science.....	60

### Introduction

---

This is a case collection for the ROSiE Training Materials for Responsible Open Science. Many cases included in this material are used in the ROSiE training materials – handouts and their use are described in instructions for trainers; however, if trainers want to use additional cases or address a topic that is not directly addressed in the training materials, they might consult this collection of cases. Also, this material can be used independently from the training materials for any course addressing ethical issues in open science and citizen science. Additionally, for six of the cases there are animations available on the [ROSiE Knowledge Hub](#).

After each case, there are questions for discussion, as well as supplementary readings that may be used by a trainer or assigned as required or optional readings for trainees. The case collection includes an index (p. 5) where cases are grouped according to field of science, stage of research and topic.

There are many approaches how to discuss a case study in ethics (see, for example, the case deliberation methods compiled by the team of the EnTIRE project<sup>1</sup>). Some approaches are suggested in ROSiE instructions for trainers and handouts. We encourage trainers to choose an approach tailored to the needs of a specific group of trainees.

Trainers are encouraged to discuss cases both according to the field of science and interdisciplinary. In groups including trainees from different fields of science, it might be beneficial to form small groups according to disciplinary lines. The small groups representing particular fields of science might be asked to discuss a case from their point of view. In a plenary discussion following the group work, the different perspectives may be compared and analysed. It might help, first, to see the differences and then reach a consensus. Another methodological approach for interdisciplinary discussions is “pairing the cases”. For example, in situations where there is a group consisting of medical scientists and engineers, it might be useful to introduce two cases, one medical case and one engineering case, and ask trainees to discuss these cases in mixed groups. It might also be beneficial to involve several trainers with different backgrounds to foster a dialogue between different fields of science.

---

1

<https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c3a7e938&appId=PPGMS>

## Cases grouped on content

### Field of science

Social sciences	<a href="#">2</a> <a href="#">5</a> <a href="#">6</a> <a href="#">9</a> <a href="#">12</a>
Humanities	<a href="#">8</a> <a href="#">9</a> <a href="#">12</a> <a href="#">20</a>
Health and life sciences	<a href="#">2</a> <a href="#">7</a> <a href="#">11</a> <a href="#">17</a> <a href="#">18</a> <a href="#">22</a> <a href="#">23</a> <a href="#">24</a> <a href="#">26</a>
Natural sciences	<a href="#">10</a> <a href="#">13</a> <a href="#">14</a> <a href="#">19</a> <a href="#">27</a>
Citizen science	<a href="#">1</a> <a href="#">3</a> <a href="#">4</a> <a href="#">7</a> <a href="#">10</a> <a href="#">21</a> <a href="#">22</a>
Interdisciplinary	<a href="#">15</a> <a href="#">16</a> <a href="#">25</a> <a href="#">28</a> <a href="#">29</a> <a href="#">30</a> <a href="#">31</a> <a href="#">32</a>

### Stage of research

Data collection	<a href="#">1</a> <a href="#">3</a> <a href="#">4</a>
Data sharing, using and reanalysis	<a href="#">2</a> <a href="#">5</a> <a href="#">6</a> <a href="#">7</a> <a href="#">8</a> <a href="#">9</a> <a href="#">11</a> <a href="#">12</a> <a href="#">14</a> <a href="#">15</a> <a href="#">16</a> <a href="#">17</a> <a href="#">18</a> <a href="#">19</a>
Open publishing	<a href="#">20</a> <a href="#">23</a> <a href="#">24</a> <a href="#">25</a> <a href="#">26</a> <a href="#">27</a> <a href="#">28</a> <a href="#">29</a> <a href="#">30</a>
Dissemination	<a href="#">31</a>

### Topic

Authorship	<a href="#">20</a> <a href="#">21</a> <a href="#">22</a>
Conflicts of interests	<a href="#">1</a>
Climate research	<a href="#">10</a> <a href="#">19</a>
Epistemic injustice	<a href="#">16</a> <a href="#">30</a>
Genomic research	<a href="#">17</a>
Low- and middle-income countries	<a href="#">16</a> <a href="#">30</a> <a href="#">32</a>
Predatory publishing	<a href="#">28</a>
Preprints	<a href="#">23</a>
Peer-review	<a href="#">23</a> <a href="#">24</a> <a href="#">25</a>
Privacy	<a href="#">2</a> <a href="#">3</a> <a href="#">4</a> <a href="#">6</a> <a href="#">9</a> <a href="#">12</a> <a href="#">13</a> <a href="#">14</a> <a href="#">17</a> <a href="#">26</a>
Data quality	<a href="#">4</a> <a href="#">10</a>
Intellectual property	<a href="#">4</a> <a href="#">18</a>
Informed consent	<a href="#">2</a> <a href="#">6</a> <a href="#">12</a> <a href="#">17</a> <a href="#">26</a>
Vulnerable groups	<a href="#">5</a> <a href="#">6</a> <a href="#">12</a> <a href="#">17</a>
Anonymization, anonymous data	<a href="#">2</a> <a href="#">5</a> <a href="#">6</a> <a href="#">9</a> <a href="#">12</a> <a href="#">17</a>
Social media	<a href="#">6</a> <a href="#">9</a> <a href="#">13</a> <a href="#">31</a>
Risk of harm	<a href="#">2</a> <a href="#">7</a> <a href="#">8</a> <a href="#">9</a> <a href="#">11</a> <a href="#">12</a> <a href="#">13</a> <a href="#">14</a> <a href="#">17</a> <a href="#">19</a>
Retraction of publication	<a href="#">26</a> <a href="#">27</a>

### CASE 1: Conflict of interests

SOURCE: Macey, G. P. et al. (2014). Air concentrations of volatile compounds near oil and gas production: a community-based exploratory study. *Environmental Health*, 13(1), 1-18. <https://doi.org/10.1186/1476-069X-13-82>

A community-based environmental research study was published in 2014, focused on examining air quality near oil and gas production sites in several parts of the United States. Residents living close to these hydraulic fracturing sites were experiencing a range of symptoms such as headaches, dizziness, light-headedness, running nose, nausea, and sore throat. Air samples were collected by the community members who played a vital role in the study. Volunteers were asked to document visible emissions, odours and sounds deriving from the hydraulic fracturing sites, activity on-site and acute health symptoms in humans. Before data collection, community members completed a training program and received written instructions on how to obtain accurate data.

When analysing the collected samples researchers identified unique chemical mixtures at each location, including significant concentrations of four volatile organic compounds: benzene, formaldehyde, hexane, and hydrogen sulphide. The study revealed that concentrations of several volatile organic compounds at the research sites significantly exceed health-based risk levels, and the existing regulations may not be sufficient to reduce human health risks.

This research study initiated a discussion about different possible conflicts of interest. In the paper, the authors stated that they had no conflicting financial interests; however, questions were raised about the involvement of community members driven by a particular agenda and motivated by their interests. Additionally, five of the authors disclosed that they are employed by non-profit organizations committed to reducing exposure to toxic chemicals. Some critics suggested that it would also be important to consider that some industry-sponsored studies have presented differing conclusions regarding the impact of fracking on air quality.



Animation of this case is available on the [ROSIE Knowledge Hub](#).

#### Questions for discussion:

- 1) Do you think there is a conflict of interest in this case? Why yes or no? If yes, should it be disclosed in a publication?
- 2) It is widely believed that when researchers publish the results of their research, they should disclose their financial conflicts of interest. However, this case suggests that this might be a too narrow way of how a conflict of interests should be understood in contemporary research and in the context of citizen science. If so, what other conflicts of interest should be disclosed?
- 3) Do conflicts of interest in citizen science differ from conflicts of interest in science in general? If yes, what is the difference?

### Supplementary readings:

1. Aytug, Z. G., Rothstein, H. R., Kern, M. C., & Zhu, Z. (2019). Is there social consensus regarding researcher conflicts of interest? *Ethics & Behavior*, 29(2), 101-140. <https://doi.org/10.1080/10508422.2017.1402683>
2. COPE Council (2016). COPE Discussion Document: Handling competing interests. <https://doi.org/10.24318/ElTeSLhp>
3. COPE Council (2021). COPE Flowcharts and infographics: Undisclosed conflict of interest in a published article. <https://doi.org/10.24318/cope.2019.2.7>
4. Resnik, D. B., Konecny, B., & Kissling, G. E. (2017). Conflict of interest and funding disclosure policies of environmental, occupational, and public health journals. *Journal of occupational and environmental medicine*, 59(1), 28. <https://doi.org/10.1097/JOM.0000000000000910>
5. Resnik, D. B., Elliott, K. C., & Miller, A. K. (2015). A framework for addressing ethical issues in citizen science. *Environmental Science & Policy*, 54, 475-481. <https://doi.org/10.1016/j.envsci.2015.05.008>
6. The Embassy of Good Science: "[Conflict of interests](#)", "[Intellectual conflicts of interest](#)"

### CASE 2: Informed consent to open publishing of a dataset

AUTHOR: Signe Mežinska

A team of researchers is conducting a study to investigate the impact of prolonged screen time on adolescent mental health. The research includes surveys, psychological assessments and collecting data via a smartphone tracking application. The researchers aim to publish their findings in an open access journal and include a link to the dataset collected during the study. The dataset will be accessible to other researchers (and in general – to everyone) in an open access data repository. Participants are recruited from local schools, and their data will be included in this dataset in an anonymized form.

#### Questions for discussion:

- 1) What information on the open publication of the dataset should be included in informed consent/assent forms?
- 2) Is it important for research participants or their parents to understand that the data will be available in open access?
- 3) How to ensure proper anonymisation of data in this case? Is it possible?
- 4) What are the risks and benefits of making datasets like this openly accessible?

#### Supplementary readings:

1. Florea, D., & Florea, S. (2020). Big Data and the ethical implications of data privacy in higher education research. *Sustainability*, 12(20), 8744. <https://doi.org/10.3390/su12208744>
2. Kreuter, F., Haas, G. C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, 38(5), 533-549. <https://doi.org/10.1177/0894439318816389>
3. Solymosi, R., Buil-Gil, D., Ceccato, V., Kim, E., & Jansson, U. (2023). Privacy challenges in geodata and open data. *Area*. <https://doi.org/10.1111/area.12888>



### CASE 3: Citizen science and privacy of data

SOURCE: Anhalt-Depies, C. et al. (2019). Tradeoffs and tools for data quality, privacy, transparency, and trust in citizen science. *Biological Conservation*, 238, 108195. <https://doi.org/10.1016/j.biocon.2019.108195>

Snapshot Wisconsin is a USA-based citizen science project utilizing a network of trail cameras to monitor wildlife. The project was initiated in 2016 by the Wisconsin Department of Natural Resources (DNR). The main aims of the project are collecting essential data to aid in making decisions about wildlife management and fostering a greater connection between the general public and the agency. Citizens interested in taking part in the Snapshot Wisconsin project have the option to register as hosts for photo cameras in privately owned properties. After receiving the necessary training and equipment from the Wisconsin DNR, volunteers are asked to upload photos a few times per year. Afterwards, the volunteers can check out the photos of the animals they have captured through their online profiles and assist in identifying the species present in the images. Any photos that remain unidentified are then shared on a collaborative online platform ([snapshotwisconsin.org](https://snapshotwisconsin.org)), where people from all around the world can help identify the animals.

By 2018, the cameras hosted by the Wisconsin volunteers had snapped over 20 million photos, and more than 5800 individuals had registered to participate in the collaborative online animal identification effort. However, from the start of the project, there were concerns raised about the possibility of accidentally capturing images of humans. To decrease this risk, Wisconsin DNR issued guidelines for hosts of cameras on how to avoid areas used by humans. It was also decided that hosts would not see the photos until they were uploaded to the agency and subjected to a proprietary decryption procedure to remove human images. This approach was introduced to prevent privacy violations.

At the same time, the idea of the project was to involve the public more effectively in wildlife management and enhance transparency in wildlife monitoring. Thus, for ensuring project success it is very important to provide photos and data back to volunteers. Nonetheless, granting volunteers unrestricted access to all photos and data before uploading could lead to privacy violations and hinder ensuring a comprehensive dataset. Also, volunteers might lack the motivation to promptly upload photos if their primary interest was discovering the wildlife on their property. This situation presented the Snapshot Wisconsin project with a dilemma: striking a balance between privacy, data quality and open data sharing with volunteers.

#### Questions for discussion:

- 1) What are the main privacy concerns raised by Snapshot Wisconsin and other similar citizen science projects?

- 2) What policies and measures you as a researcher would implement to mitigate the privacy concerns? Prepare your proposal of the measures and provide a justification for it.
- 3) Should citizen scientists follow the same ethics and privacy requirements as 'traditional' scientists? Why yes or no?

### Supplementary readings:

1. Evans, B. J. (2020). The perils of parity: should citizen science and traditional research follow the same ethical and privacy principles? *The Journal of Law, Medicine & Ethics*, 48(1\_suppl), 74-81.  
<https://doi.org/10.1177/1073110520917031>
2. Scheibner, J., Jobin, A., & Vayena, E. (2021). Ethical issues with using Internet of Things devices in citizen science research: a scoping review. *Frontiers in Environmental Science*, 9, 629649. <https://doi.org/10.3389/fenvs.2021.629649>
3. The Embassy of Good Science: "[Privacy in research](#)"

### CASE 4: Data quality in citizen social science

SOURCE: Heiss, R., & Matthes, J. (2017). Citizen science in the social sciences: A call for more evidence. *GAIA-Ecological Perspectives for Science and Society*, 26(1), 22-26.  
<https://doi.org/10.14512/gaia.26.1.7>

In the citizen science project called [Young Adults' Political Experience Sampling \(YAPES\)](#), school students from Austria took part in independent data collection to gather their political experience every day and send the data to a research group by email or WhatsApp. They were asked to photograph, document and comment on everything that was politically interesting and important to them. The scientists used data to identify the political issues that young people face in their everyday lives. The project focused on various aspects, including the places where political engagement occurs, the channels through which political information is received and the content of political communication among young people. Typically, researchers have limited access to this kind of data, often relying on surveys or a small number of qualitative interviews. The citizen science approach enables the collection of large amounts of real-life data. Despite being initially designed as a small-scale pilot project, YAPES managed to involve 254 volunteers who gathered a total of 1768 observations.

However, there are several challenges to implementing citizen science in social sciences projects like YAPES, including quality of data and ethical considerations. Ensuring data quality is a fundamental issue in citizen science, but it may be even more important in social science research. Instead of purely objective measurements facilitated by technical devices that are common in natural science projects, social science research heavily relies on human observation. Activities like observing, counting, documenting, and photographing are inherently subjective due to the observer's perception. Moreover, the measurement process can be biased and affected by the individual characteristics of volunteers, such as their political views etc. The ethical concerns are related, for example, to data sharing and privacy, especially when citizens use their smartphones to collect and submit data which could include sensitive information about themselves or others they observe. For instance, sharing data concerning an individual's political beliefs or interpersonal behaviour might raise greater ethical considerations than sharing data about a local species of insects in a natural sciences project.

#### Questions for discussion:

- 1) What are the challenges for data quality that researchers might face in a study like YAPES? What are the reasons for those challenges? What can be done to mitigate these problems? Please, develop a list of proposals that the researchers should implement in this case to ensure data quality!
- 2) In your view, what are other considerable ethical challenges for social scientists collaborating with citizen scientists? How to address these challenges?

### Supplementary readings:

1. Chesser, S., Porter, M. M., & Tuckett, A. G. (2020). Cultivating citizen science for all: ethical considerations for research projects involving diverse and marginalized populations. *International Journal of Social Research Methodology*, 23(5), 497-508. <https://doi.org/10.1080/13645579.2019.1704355>
2. Riesch, H., & Potter, C. (2014). Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public understanding of science*, 23(1), 107-120. <https://doi.org/10.1177/0963662513497324>
3. Resnik, D. B., Elliott, K. C., & Miller, A. K. (2015). A framework for addressing ethical issues in citizen science. *Environmental Science & Policy*, 54, 475-481. <https://doi.org/10.1016/j.envsci.2015.05.008>

### CASE 5: Research involving vulnerable groups

AUTHOR: Laura Guntrum

Johanna is a peace and conflict researcher who conducts qualitative empirical research on a group of activists under a totalitarian regime. Johanna is concerned over the possible consequences of openly sharing the collected data because it is not and cannot be sufficiently anonymized to prevent the identification of informants. Sharing the data might have detrimental consequences for the study informants because they could be identified by the regime officials who then might harm them. However, to proceed with her research Johanna has to share the data with her collaborators in another country.

#### Questions for discussion:

- 1) What are the ethical issues associated with this kind of research project and what are the challenges of doing this research within the framework of open science?
- 2) Why in this case the data cannot be sufficiently anonymized? What kind of ethical implications follow from that?
- 3) What should Johanna do in this situation?
- 4) What information should be included in the informed consent form, especially regarding open sharing of research data?

#### Supplementary readings:

1. DuBois, J. M., Strait, M., & Walsh, H. (2018). Is it time to share qualitative research data? *Qualitative Psychology*, 5(3), 380–393.  
<https://doi.org/10.1037/qup0000076>
2. VandeVusse, A., Mueller, J., & Karcher, S. (2022). Qualitative Data Sharing: Participant Understanding, Motivation, and Consent. *Qualitative Health Research*, 32(1), 182-191. <https://doi.org/10.1177/10497323211054058>
3. The Embassy of Good Science: "[Privacy in research](#)"

### CASE 6: Using sensitive social media data for open science

SOURCE: O'Callaghan, E., & Douglas, H. M. (2021). # MeToo online disclosures: A survivor-informed approach to open science practices and ethical use of social media data. *Psychology of Women Quarterly*, 45(4), 505-525.  
<https://doi.org/10.1177/03616843211039175>

Back in 2006, Tarana Burke kicked off the "MeToo" movement, aiming primarily to bring attention to the experiences of women who had suffered from sexual abuse and to highlight the unique impact of sexual abuse on persons of colour. In 2017, the online #MeToo movement, centred on sexual violence, took off in a big way and people from all over the world began openly sharing their personal stories of experiencing sexual violence and abuse, using the #MeToo hashtag. Since the inception of the #MeToo movement, researchers have explored related topics in academia, including research on online accounts of sexual violence on social media.

With the ability to extract data from social media platforms or partner with third-party organizations to access specific data collections, researchers now have the potential to access large amounts of data online.

While compiling extensive databases of statements from persons who have experienced sexual violence, as in the case of the #MeToo movement, questions arise about measures taken to ensure anonymity and providing a rationale for the quantity of collected social media data. Also, there are questions about whether these databases should be made freely accessible in open science repositories, and if yes, should the access be free or given only upon request and research ethics committee approval?

Currently, more and more journals require publishing data in open access repositories for their published papers. This dynamic introduces a dilemma, pitting researchers' efforts to protect persons' identities and privacy against the imperative to adhere to open science principles.

#### Questions for discussion:

- 1) The posts gathered from social media are publicly available. If so, does it mean, that there are no restrictions on how this data can be gathered, shared and reused for research purposes?
- 2) Might there be any need to get consent from the social media account owners to use the data for research? What role if any is played by the fact, that the data contains disclosures of sexual violence? On what conditions the data can be used?
- 3) How should the data be anonymized or pseudonymised in this case?
- 4) What are the specific ethical concerns related to archiving, open sharing and reuse of social media data?

### Supplementary readings:

1. Campbell, R., Goodman-Williams, R., & Javorka, M. (2019). A trauma-informed approach to sexual violence research ethics and open science. *Journal of interpersonal violence*, 34(23-24), 4765-4793. <https://doi.org/10.1177/0886260519871530>
2. DuBois, J. M., Strait, M., & Walsh, H. (2018). Is it time to share qualitative research data? *Qualitative Psychology*, 5(3), 380–393. <https://doi.org/10.1037/qup0000076>
3. Gerrard, Y. (2021). What's in a (pseudo) name? Ethical conundrums for the principles of anonymisation in social media research. *Qualitative Research*, 21(5), 686-702. <https://doi.org/10.1177/1468794120922070>
4. Mason, S., & Singh, L. (2022). Reporting and discoverability of “Tweets” quoted in published scholarship: current practice and ethical implications. *Research Ethics*, 18(2), 93-113. <https://doi.org/10.1177/17470161221076948>
5. VandeVusse, A., Mueller, J., & Karcher, S. (2022). Qualitative Data Sharing: Participant Understanding, Motivation, and Consent. *Qualitative Health Research*, 32(1), 182-191. <https://doi.org/10.1177/10497323211054058>
6. The Embassy of Good Science: "[Privacy in research](#)"

### CASE 7: Open access biodiversity data

SOURCE: Quinn, A. (2021). Transparency and secrecy in citizen science: Lessons from herping. *Studies in History and Philosophy of Science Part A*, 85, 208-217.  
<https://doi.org/10.1016/j.shpsa.2020.10.010>

eBird is an online platform for posting observations of birds that was launched in 2002 by the Cornell Lab of Ornithology. Now eBird is among the world's largest biodiversity-related science projects with more than 100 million bird sightings contributed annually by eBirders around the world. In 2008, the success of eBird inspired three students to develop a platform for natural history observations of any organism - iNaturalist. Now iNaturalist is maintained by the California Academy of Sciences and the National Geographical Society. As of 2020, iNaturalist hosted 42 427 731 observations contributed by 1 149 886 observers.

Both platforms bring substantial epistemic benefits by contributing data for natural sciences and life sciences research, as well as for educational purposes. As A. Quinn points out: "Citizens contributing observations in a casual manner can rapidly expand the scale of the dataset far beyond what is possible using traditional research methods. Moreover, it is not even possible to predict what kind of knowledge can be pulled from long-term, huge datasets." Besides that, the platforms encourage many people to experience nature. However, the popularity of the platforms also may create a danger for many species. Smartphone photos uploaded by volunteers contain location coordinates, and, for example, iNaturalist by default makes observation locations visible to all users. While users can choose an option to hide observation locations when uploading data, it is easy to forget to enable this function and there still is a potential for overrides. Many users also are unaware of the risks tied to sharing location information. While iNaturalist automatically hides location data for some species in need of conservation, users are primarily accountable for hiding location data.

The main concern is that the information might be used by poachers. For many species, the more severe threat is the destruction or degradation of habitat. A record of an interesting species on iNaturalist might attract many people who might go to look for the species. But as many people are ignorant of what interventions can destroy habitat, such visits might turn out to be fatal for the animals. This concern is especially raised by the community of herpers as herps (amphibians and reptiles) are very sensitive to any changes in their habitat.

#### Questions for discussion:

- 1) How to responsibly implement the principle 'as open as possible and as closed as necessary' regarding biodiversity data?
- 2) Platforms like iNaturalist promote citizens' interest in the natural world by enabling participation. However, by doing that they also increase interest in finding rare species, which in turn might pose a risk to habitats that are



necessary for their survival. Are there any ways to avoid or at least minimize the risks created by these platforms?

- 3) Should there be stricter ethical guidelines for sharing location-specific data of rare or sensitive species on platforms like iNaturalist?

### Supplementary readings:

1. Cooke, S. J. et al. (2017). Troubling issues at the frontier of animal tracking for conservation and management. *Conservation Biology*, 31(5), 1205–1207.  
<https://doi.org/10.1111/cobi.12895>
2. Quinn, A. (2021). Transparency and secrecy in citizen science: Lessons from herping. *Studies in History and Philosophy of Science Part A*, 85, 208–217.  
<https://doi.org/10.1016/j.shpsa.2020.10.010>
3. Soroye, P. et al. (2022). The risks and rewards of community science for threatened species monitoring. *Conservation Science and Practice*, 4(9), e12788.  
<https://doi.org/10.1111/csp2.12788>
4. Tulloch, A. I. T. et al. (2018). A decision tree for assessing the risks and benefits of publishing biodiversity data. *Nature Ecology & Evolution*, 2(8), Article 8.  
<https://doi.org/10.1038/s41559-018-0608-1>
5. <https://www.inaturalist.org/pages/about>
6. <https://ebird.org/about>

### CASE 8: Open data and risk of looting in archaeology

SOURCE: Frank, R. D., Kriesberg, A., Yakel, E., & Faniel, I. M. (2015). Looting hoards of gold and poaching spotted owls: Data confidentiality among archaeologists & zoologists. *Proceedings of the Association for Information Science and Technology*, 52(1), 1-10.  
<https://doi.org/10.1002/pr2.2015.145052010037>

Archaeologists have faced ethical challenges inherent in their research since their discipline's inception. One major concern is historical site looting - unauthorized and often illicit removal of artefacts, objects, and valuable items from archaeological sites. This destructive practice harms sites, fuels illegal antiquities markets, and places human remains at risk. Ethical guidelines for archaeologists issued by professional organizations (e.g., guidelines by the Society for American Archaeology) emphasize archaeologist's duty to protect sites from potential looting. This requirement means also careful consideration of risk before open data sharing and publication in archaeology. Digital archaeology repositories, like Open Context, have addressed this risk by issuing data publication guidelines highlighting researchers' responsibility to assess the necessity to remove or restrict access to location data before submitting datasets to a repository.

In the research interviews that were conducted during the study by Frank et al., archaeologists voiced a particular concern. They feared that sharing location details as open data could lead to site looting and subsequent blame placed on them. An archaeologist recounted a situation where she reconsidered her publication plans due to the anti-looting regulations and the potential harm to the site. She pondered whether safeguarding a site from looters was her duty when the host country already had protective laws in effect and whether she could face legal accountability if any damage occurred to the site.

#### Questions for discussion:

- 1) How to responsibly implement the principle 'as open as possible and as closed as necessary' regarding archaeological research data?
- 2) Who should make decisions about whether and how to manage access to sensitive archaeological data? Who should decide whether data is sensitive?
- 3) What are the responsibilities of individual researchers when depositing datasets into a digital repository?

#### Supplementary readings:

1. Huggett, J. (2015). Digital Haystacks: Open Data and the Transformation of Archaeological Knowledge. *Open Source Archaeology: Ethics and Practice*, 6-29.  
<http://dx.doi.org/10.17613/yfss-zt74>
2. Smith, C. (2020). Ethics and best practices for mapping archaeological sites. *Advances in Archaeological Practice*, 8(2), 162-173.  
<https://doi.org/10.1017/aap.2020.9>

### CASE 9: Using sensitive data from social media

SOURCE: Suomela, T., Chee, F., Berendt, B., & Rockwell, G. (2019). Applying an ethics of care to internet research: Gamergate and digital humanities. *Digital Studies/Le Champ Numérique*, 9(1). <http://doi.org/10.16995/dscn.302>

"The Gamergate controversy erupted during the summer of 2014 and quickly engulfed the online gaming community in an intense debate about what and who belonged in gaming culture. The conflict quickly escalated into another battlefield in the culture wars involving gender, identity, and political beliefs [...]. Overt harassment of women and others who were critical of aspects of gaming culture quickly became a dominant feature of online forums, Twitter, and other media in which discussion about gaming culture occurred. Rape and death threats were sent to prominent game developers and journalists, some of whom were also doxed (had personally identifiable information such as home addresses disseminated online). This harassment and other extreme expressions of hatred and intolerance quickly eclipsed the alleged ethical issues in gaming culture that supposedly was the impetus for Gamergate."

Research done by Suomela et al. included building a primary data archive comprising data collected online on the Gamergate controversy and including numerous instances of hate speech. They describe the ethical challenges as follows:

"Privacy and reputation were the two biggest harms to research subjects that we discussed and attempted to mitigate in the Gamergate project. The privacy of our subjects was protected in two overlapping ways. First, the results of the research were reported only in aggregate forms, and second, the sources for direct quotes were not identified [...]. Neither method of privacy protection can completely guarantee that people will not be identified because the activity we collected and analyzed occurred in online forums like Twitter that anyone can search. A determined person could still recover the original source of a quote by searching Twitter or the internet, so the results could not be completely anonymized. The question of privacy is highly fraught when it comes to research about topics such as Gamergate, which depends on the internet as the primary medium for communication. Any quote from a publicly accessible website could potentially be re-identified after a research study has been completed."

#### Questions for discussion:

- 1) The posts gathered from social media are publicly available. If so, does it mean, that there are no restrictions on how this data can be gathered, shared and reused for research purposes?
- 2) How should the data be anonymized or pseudonymised? Are the methods suggested by the authors sufficient to protect the privacy of research subjects?
- 3) What are the specific ethical concerns related to archiving, open sharing and reuse of social media data?

- 4) Might there be any need to get consent from the social media account owners to use the data for research? Are they research subjects? On what conditions the data may be used?

### Supplementary readings:

1. Fox, J., Pearce, K. E., Massanari, A. L., Riles, J. M., Szulc, Ł., Ranjit, Y. S., ... & L. Gonzales, A. (2021). Open science, closed doors? Countering marginalization through an agenda for ethical, inclusive research in communication. *Journal of Communication*, 71(5), 764-784. <https://doi.org/10.1093/joc/jqab029>
2. Gerrard, Y. (2021). What's in a (pseudo) name? Ethical conundrums for the principles of anonymisation in social media research. *Qualitative Research*, 21(5), 686-702. <https://doi.org/10.1177/1468794120922070>
3. Mason, S., & Singh, L. (2022). Reporting and discoverability of "Tweets" quoted in published scholarship: current practice and ethical implications. *Research Ethics*, 18(2), 93-113. <https://doi.org/10.1177/17470161221076948>

### CASE 10: Data quality in citizen science: climate research

SOURCE: Herodotou, C., Scanlon, E., & Sharples, M. (2021). Methods of promoting learning and data quality in citizen and Community Science. *Frontiers in Climate*, 53. <https://doi.org/10.3389/fclim.2021.614567>

"The "Heatwave: Are you coping?" investigation has been designed in collaboration with the Royal Meteorological Society and support from the BBC Weather (see <https://nquire.org.uk/mission/heatwave-are-you-coping/contribute>). The mission was an outcome of a workshop with citizens and organizations interested in weather issues, which was organized by the Open University UK, as part of the UKRI funded project EduCS: EDUcating Citizens and organizations in Citizen Science methodologies. Workshop attendees were asked to brainstorm, vote, and rank ideas for research investigations they would like to design using nQuire. How comfortable people feel in extreme weather conditions was one of the two most popular investigations (alongside the impact of climate change). The investigation with more than 1200 responses, was launched on the 7th of August 2020, during which England experienced a heatwave and was ended in September 2020. The purpose of the mission was to explore how people's experiences of hot weather may differ depending on where they live and work, and how people are able to adapt their routines to heat. Citizens were asked to take their first temperature recording around 3–4 pm, when maximum daily temperatures are normally observed. The rationale behind the mission was to collect data about how different people are affected by extreme weather conditions and how working and living conditions could be improved. Results could, for example, help people plan for heatwaves in the future. In terms of the learning benefits for citizens, the mission was an opportunity to learn about what forecast temperatures mean in practice, how to make and record measurements, and how to increase personal comfort in a heatwave.

Citizen Science temperature measurements have the unique value of providing data about air temperature on scales smaller than those measured by the official meteorological service, and such data could be possibly used in weather monitoring or even forecasting [...]. Yet, the quality of weather data collected is a major challenge and a source of bias, often related to possible overheating of the thermometer by, for example, not being shielded. This was an issue raised and discussed during the workshop, with weather scientists expressing concerns about the quality of data collected and whether amateur scientists could actually offer reliable recordings."

#### Questions for discussion:

- 1) What are the challenges for data quality that researchers might face in the case above? What are the reasons for those challenges? What can be done to mitigate these problems?

## Supplementary readings:

1. Balázs, B., Mooney, P., Nováková, E., Bastin, L., Jokar Arsanjani, J. (2021). Data Quality in Citizen Science. In: Vohland, K., et al. *The Science of Citizen Science*. Springer. [https://doi.org/10.1007/978-3-030-58278-4\\_8](https://doi.org/10.1007/978-3-030-58278-4_8).
2. Haklay, M. (2021). Why is it so difficult to integrate citizen science into practice? *Citizen Science and Public Policy Making*, 108. <https://discovery.ucl.ac.uk/id/eprint/10130136>

### CASE 11: Data sharing and reanalysis in medicine

SOURCE: LeNoury, J., Nardo, J. M., Healy, D. et al. (2015). Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ*, 51:h4320. <https://doi.org/10.1136/bmj.h4320>

In 2015 a team of scientists published a paper “Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence” in the British Medical Journal. The study aimed to reanalyse SmithKline Beecham’s Study 329 published by Keller et al. in 2001. The primary objective of Study 329 was to compare the efficacy and safety of paroxetine and imipramine with placebo in the treatment of adolescents with unipolar major depression. The conclusion by Keller et al. was that paroxetine is well tolerated and effective for major depression in adolescents.

In the reanalysis of data, researchers found out that neither paroxetine nor imipramine showed efficacy for major depression in adolescents. Moreover, with both drugs, there was an increase in harm. The researchers identified several potential barriers to accurate reporting of harms in the original study:

- “Use of an idiosyncratic coding system
- Failure to transcribe all adverse events from clinical record to adverse event database
- Filtering data on adverse events through statistical techniques
- Restriction of reporting event to that occurred above a given frequency in any one group
- Coding event under different heading for different patients (dilution)
- Grouping of adverse events
- Insufficient consideration of severity
- Coding of relatedness to study medication
- Masking effects of concomitant drugs
- Ignoring effects of drug withdrawal”

In the conclusions of the reanalysis, the authors wrote: “Access to primary data from trials has important implications for both clinical practice and research, including that published conclusions about efficacy and safety should not be read as authoritative. The reanalysis of Study 329 illustrates the necessity of making primary trial data and protocols available to increase the rigour of the evidence base.”

#### Questions for discussion:

- 1) What is the role and significance of open data in scientific research? What are the benefits and risks of reanalysis of open data sets?
- 2) If you would perform a reanalysis of an openly accessible data set and discover similar problems, what would/should you do?
- 3) Should the original publication of the study be retracted in this case?

### Supplementary readings:

1. *The original study:*  
Keller, M. B., Ryan, N. D., Strober, M., Klein, R. G., Kutcher, S. P., Birmaher, B., ... & McCafferty, J. P. (2001). Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(7), 762-772.  
<https://doi.org/10.1097/00004583-200107000-00010>
2. Bauchner, H., Golub, R. M., & Fontanarosa, P. B. (2016). Data sharing: an ethical and scientific imperative. *Jama*, 315(12), 1238-1240.  
<https://doi.org/10.1001/jama.2016.2420>
3. Faria, M., Spoljaric, S., & Caruso, F. (2022). Reanalysis: the forgotten sibling of reproducibility and replicability. *Nature Reviews Methods Primers*, 2(1), 1-2.  
<https://doi.org/10.1038/s43586-022-00103-z>
4. Neutra, R. R., Cohen, A., Fletcher, T., Michaels, D., Richter, E. D., & Soskolne, C. L. (2006). Toward guidelines for the ethical reanalysis and reinterpretation of another's research. *Epidemiology*, 17(3), 335-338.  
<https://doi.org/10.1097/01.ede.0000209464.97895.bf>



### CASE 12: Sharing sensitive qualitative data

**AUTHOR:** Kadri Simm

The Russia-Ukraine war that began in 2022 caused a refugee crisis in Europe, with millions of Ukrainians escaping war zones and settling, at least temporarily, in various countries. In Estonia, scholars working on life stories wish to collect narratives from recent Ukrainian refugees and publish them on an open access project platform. Aside from the academic value of this material, the project can also be seen as a way of recording and safekeeping the Ukrainians' tragic experiences for both them and a wider public. Some participants would like to openly publish their stories of war and escape even under their own names so that their experience can inform the public about what happened and allow scientists to analyse their experiences. However, while the war in Ukraine continues, there is also a simultaneous information war raging, and the refugees' relatives may still be fighting in the former. Further, some refugees might have witnessed war crimes and the possibility exists that these stories could later be used as evidence in a court of law. Further, among refugees, there are children and adolescents whose stories form part of those told by their family members.



**Animation of this case is available on the [ROSiE Knowledge Hub](#).**

#### Questions for discussion:

- 1) How to ensure the quality of informed consent and voluntariness of participation in this case?
- 2) Participants may be willing to publish their stories of war and escape (even using their real names) so that their experience can inform the public about what happened. Yet there is a war going on, also an information war and relatives of the refugees might still be fighting the actual war. What about the potential misuse of these stories? What kind of harm might this facilitate for the refugees and their relatives? Are there ways to minimize risks?
- 3) What are the criteria for publishing such life stories as open data?
- 4) Some refugees might have witnessed war crimes. Can these stories later be used as evidence in a court of law? Should the researchers inform participants about this possibility? How to deal with other "incidental findings" (stories of exploitation etc.)?

#### Supplementary readings:

1. Campbell, R., Goodman-Williams, R., & Javorka, M. (2019). A trauma-informed approach to sexual violence research ethics and open science. *Journal of interpersonal violence*, 34(23-24), 4765-4793.  
<https://doi.org/10.1177/0886260519871530>

2. DuBois, J. M., Strait, M., & Walsh, H. (2018). Is it time to share qualitative research data? *Qualitative Psychology*, 5(3), 380–393.  
<https://doi.org/10.1037/qup0000076>
3. VandeVusse, A., Mueller, J., & Karcher, S. (2022). Qualitative Data Sharing: Participant Understanding, Motivation, and Consent. *Qualitative Health Research*, 32(1), 182-191. <https://doi.org/10.1177/10497323211054058>

### CASE 13: Data privacy in water sciences

Zipper, S. C. et al. (2019). Balancing open science and data privacy in the water sciences. *Water Resources Research*, 55(7), 5202-5211.

<https://doi.org/10.1029/2019WR025080>

Open science practices, like sharing data, research results and code, are providing new opportunities to scientists working in the field of water science, e.g., by enhancing data availability and reproducibility. However, as physical and social science domains are merging in fields like socio-hydrology, researchers may inadvertently compromise privacy and security when sharing sensitive information in open access. High-resolution spatial data, including satellite data, hydrological model outputs, and other geospatial datasets, are widely used in hydrologic sciences. This type of data can be sensitive, despite not meeting traditional human subject research definitions. For example, 30% of farmers in Iowa, US in a survey stated that collecting geospatial data on private property compromises privacy. Also, other groups of data, like water or electricity consumption, are potentially sensitive. The possibility of commercializing the data amplifies these concerns by adding issues of data ownership.

Privacy and other risks may increase when researchers lack cultural understanding and sensitivity. There have been cases when powerful groups or companies have misused open data “at the expense of the intended beneficiaries of the shared data”. For example, the digitization of land records in Karnataka, India which was intended to democratize information access, instead empowered wealthy landowners. These concerns may be particularly important in environmental justice contexts, especially when working with indigenous communities and historically disadvantaged groups.

#### Questions for discussion:

- 1) How sharing the different types of data mentioned in the case description might violate the privacy and security of individuals or communities?
- 2) Do you agree with the authors’ statement that: “Natural scientists have little guidance to deal with privacy concerns for open science, which are inherent in socio-environmental research”?
- 3) What should the scientists do to protect data privacy and security?

#### Supplementary readings:

1. Blatt, A. J. (2015). The benefits and risks of volunteered geographic information. *Journal of Map & Geography Libraries*, 11(1), 99-104.  
<https://doi.org/10.1080/15420353.2015.1009609>
2. Richardson, D. B., Kwan, M. P., Alter, G., & McKendry, J. E. (2015). Replication of scientific research: addressing geoprivacy, confidentiality, and data sharing challenges in geospatial research. *Annals of GIS*, 21(2), 101-110.  
<https://doi.org/10.1080/19475683.2015.1027792>

3. Solymosi, R., Buil-Gil, D., Ceccato, V., Kim, E., & Jansson, U. (2023). Privacy challenges in geodata and open data. *Area*. <https://doi.org/10.1111/area.12888>

## CASE 14: Open geospatial data in agriculture research

Prince Czarnecki, J. M., & Jones, M. A. (2022). The problem with open geospatial data for on-farm research. *Agricultural & Environmental Letters*, 7(1), e20062.

<https://doi.org/10.1002/ael2.20062>

On-farm research in agriculture may involve collection of spatially referenced farm data, such as fertilizer application, plant populations, and yield, which can be traced back to individual properties and private collaborators. At the same time, this type of research lacks comprehensive approaches for de-identifying geospatial data. Commonly used techniques for general geospatial data, like random perturbation and temporal cloaking, are not well-suited for farm data. Shifting point locations and altering time stamps may distort statistical measures of the data without ensuring adequate privacy. Another approach involves removing geospatial references and rescaling points to a spatially correct grid. While this helps to keep spatial relationships, it hinders contextual analyses as features with geographic concurrence cannot be identified and it is not possible to use, e.g. climate data for analysis.

In the situation where publishers and funders require researchers to share research data in open access, researchers in the field of on-farm research face challenges to ensure open access and at the same time safeguard the privacy of farm owners. Some researchers state that insisting on the publication of the data may make the involvement of collaborators more complicated, and the requirement of de-identification of data may make on-farm research more challenging.

### Questions for discussion:

- 1) How open sharing the geospatial data mentioned in the case description might violate the privacy and security of individuals or communities?
- 2) What should the scientists do to protect privacy of their collaborators?

### Supplementary readings:

1. Blatt, A. J. (2015). The benefits and risks of volunteered geographic information. *Journal of Map & Geography Libraries*, 11(1), 99-104.  
<https://doi.org/10.1080/15420353.2015.1009609>
2. Richardson, D. B., Kwan, M. P., Alter, G., & McKendry, J. E. (2015). Replication of scientific research: addressing geoprivacy, confidentiality, and data sharing challenges in geospatial research. *Annals of GIS*, 21(2), 101-110.  
<https://doi.org/10.1080/19475683.2015.1027792>
3. Solymosi, R., Buil-Gil, D., Ceccato, V., Kim, E., & Jansson, U. (2023). Privacy challenges in geodata and open data. *Area*. <https://doi.org/10.1111/area.12888>

### CASE 15: Refusal to share raw data

**AUTHOR:** Fernando Marmolejo-Ramos

Researcher X asks researcher Y to share the raw data that Y has collected and used in a publicly funded research project. A research report on the study has already been published. Y refuses to share the raw data with X. He says that no one has ever come to them with such a request and therefore they see no reason to start sharing their data now. As a response, X elaborates their reasons for their request for data sharing. Firstly, X would wish to gain access to raw data because X would like to re-analyse it with a new statistical technique. Secondly, X appeals to the open science policy of the funding organization Z that funded the study at hand. Z requires, as a pre-condition for getting funding, that the data collected and analysed be shared with other researchers. As a response to this, Y points out that according to the same policy, research grants below 100K do not obligate researchers to share their data, even though it is highly recommended. Therefore, Y refuses to share the raw data with X. However, Y does not comment on X's idea to re-analyse the raw data.

#### Questions for discussion:

- 1) It is clear, that it is not mandatory for researcher Y to share the data with X, however, is there a moral duty to do this? If so, should the policy of data sharing be changed to oblige researchers to share the data? What would be the practical and moral implications of such a change?
- 2) Researcher Y did not give any compelling reason not to share the data. What might be the possible reasons for refusal to share the data?

#### Supplementary readings:

1. Pampel, H., Dallmeier-Tiessen, S. (2014). Open Research Data: From Vision to Practice. In: Bartling, S., Friesike, S. (eds.) *Opening Science*. Springer.  
[https://doi.org/10.1007/978-3-319-00026-8\\_14](https://doi.org/10.1007/978-3-319-00026-8_14)
2. Zuiderwijk, A., Shinde, R., & Jeng, W. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PloS One*, 15(9), e0239283.  
<https://doi.org/10.1371/journal.pone.0239283>

### CASE 16: Low-resourced research environments as a barrier to openness

SOURCE: Rappert, B., & Bezuidenhout, L. (2016). Data sharing in low-resourced research environments. *Prometheus*, 34(3-4), 207-224.

<https://doi.org/10.1080/08109028.2017.1325142>

Scientists have many reasons to share their data and research results in open access. Empirical studies have shown that data sharing practices are often motivated by the potentially increased impact of the research, intellectual credit received through improved visibility, greater efficiency through the reuse of data, discovery of novel research questions and directions, promotion of scientific integrity and replication, and facilitation of collaboration. However, the existing differences and inequalities in research infrastructures and environments globally mean that researchers from low- and middle-income countries are struggling with a very different set of challenges as regards data sharing and openness. Issues like unreliable internet access, out-of-date software and hardware, severe lack of research funding and thus little ability to pay for open access publications affect motivations for sharing and openness. Often fears of being scooped (losing control of their data) or exploited are voiced because, based on access to open data, researchers in high-income countries have the resources to push for results much faster. Thus, scientists from low- and middle-income countries are arguing that they are often not able to contribute to and benefit from such a “gift economy” as open science that ultimately harms them.

#### Questions for discussion:

- 1) Discuss the potential challenges that researchers from low- and middle-income countries can face when practising open science. Which challenges are most difficult to solve?
- 2) What are the instances where open science practices could specifically benefit from researchers from low- and middle-income countries?
- 3) If you were a collaborator in an international network of researchers, what actions could you take to support researchers from low- and middle-income countries in practicing open science?

#### Supplementary readings:

1. Bezuidenhout, L., Leonelli, S., Kelly, A., & Rappert, B. (2016). “\$100 is not much to you”: open access and neglected accessibilities for data-driven science in Africa. *Critical Public Health*, 27(1), 39-49.  
<https://doi.org/10.1080/09581596.2016.1252832>
2. Bull, S., & Bhagwandin, N. (2020). The ethics of data sharing and biobanking in health research. *Wellcome Open Research*, 5.  
<https://doi.org/10.12688/wellcomeopenres.16351.1>

3. Zeitlyn, D. (2003). Gift economies in the development of open source software: anthropological reflections. *Research Policy*, 32(7), 1287-1291.  
[https://doi.org/10.1016/S0048-7333\(03\)00053-2](https://doi.org/10.1016/S0048-7333(03)00053-2)



## CASE 17: Open data in genome research

SOURCE: Callaway, E. (2013). HeLa publication brews bioethical storm. *Nature*, 1, 12689. <https://doi.org/10.1038/nature.2013.12689>

HeLa cell line is the first immortalized human cell line that is derived from the cancer cells of African American woman Henrietta Lacks (1920-1951). The cells were obtained during Lacks's treatment in 1951. It was done without her informed consent, in line with the practice that existed at that time. Neither Lacks nor her family members were informed about the collection of cells and their use in research. The Lacks family was not aware of the cell line's existence until 1975.

In 2013, Lars Steinmetz and his group at the European Molecular Biology Laboratory in Heidelberg published the genome of the HeLa cell line in open access. Steinmetz and his colleagues saw it as a helpful resource for their own research and for the countless other scientists studying the cell line. The descendants of Henrietta Lacks and many bioethicists on the other hand criticized the decision to publish the genome. They pointed out that the cells were acquired without consent and publishing of the genome may provide information about some genetic traits of surviving family members.

In reaction to the objections, Steinmetz removed the genomic data from public databases. "We were surprised, we did not expect this to happen at all," he said. "We wanted to respect the wishes of the family, and we didn't intend to cause them any anxiety by the publication of our research."

### Questions for discussion:

- 1) What was wrong with the publishing of the genomic data of HeLa cells? Are there any conditions on which it would be right to publish the genome of the HeLa cell line? If yes, then under what conditions?
- 2) In many ways HeLa cells present a unique dilemma - the donors of most other human cell lines are anonymous, but in the case of HeLa many people know that HeLa is derived from the cells of Henrietta Lacks. However, recent work has shown, that anonymized participants in large genomics projects can be identified by cross-referencing their genomes with genealogy databases. If that is the case, how worried we should be about the anonymity of other cell lines? What implications does that have for sharing genomic data of other cell lines?

### Supplementary readings:

1. Akyüz, K., Goisauf, M., Chassang, G., Kozera, Ł., Mežinska, S., Tzortzatou-Nanopoulou, O., & Mayrhofer, M. T. (2023). Post-identifiability in changing sociotechnological genomic data environments. *BioSocieties*, 1-28. <https://doi.org/10.1057/s41292-023-00299-7>

2. Berger, B., & Cho, H. (2019). Emerging technologies towards enhancing privacy in genomic data sharing. *Genome biology*, 20(1), 1-3.  
<https://doi.org/10.1186/s13059-019-1741-0>
3. Eisen, M. (2013). The Immortal Consenting of Henrietta Lacks.  
<https://www.michaeleisen.org/blog/?p=1341>
4. Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., & Erlich, Y. (2013). Identifying personal genomes by surname inference. *Science*, 339(6117), 321-324.  
<https://doi.org/10.1126/science.1229566>

### CASE 18: Refusal to share neuroscience data

SOURCE: Barron, D. (2018). How freely should scientists share their data? *Scientific American Blog Network*. <https://blogs.scientificamerican.com/observations/how-freely-should-scientists-share-their-data/>

Jack Gallant is a cognitive neuroscientist at the University of California, Berkeley who works on brain decoding technology. He has showed that based only on brain activity it is possible to reconstruct images of movies people are watching. Gallant's work has made him a prominent neuroscientist who runs a successful lab.

In 2018 Gallant was promoting open science on his Twitter account. He argued that giving away free code is pointless if it only works within an expensive software system. The next day a theoretical physicist Manilo De Domenico tweeted in reply to Gallant: "Nice advice. But what about data? We keep trying to ask access to data in your Nature 2016, but we received not a single reply, yet". Gallant replied: "The original authors are still writing further primary research papers on these data so they haven't been released yet but we expect to be able to do that very soon." Another Twitter user Andre Brown pointed out that "We still want exclusivity to publish more papers' isn't a great excuse. Did you note data restrictions in the manuscript?" and referred to Nature's policy that, on publication, authors should make their data, code and protocols "promptly" and publicly available. Therefore, it appeared that Gallant had violated *Nature's* policy and fundamental principles of open science. De Domenico further complained that Gallant's paper has given him several ideas that he would like to test but not having access to Gallant's data he is not able to do that. To this Gallant answered: "And why do you assume that your project is better than the ones that we are continuing with these data? My students and postdocs are an awesome group of people, the stuff they have in the pipeline is great! But I can't afford for them to be scooped." Gallant then affirmed his commitment to open science, that he had shared many datasets in the past and then provided further explanation of why he has not yet shared this particular data set. He pointed out that complex data takes time to understand, and his team wanted to work on data more before releasing it. Also, he argued that his lab has competed for and won the grant to collect the data and then worked to collect it, they should be able to work on it first before others do it. Many academics on Twitter were not happy about Gallant's answer. They called it a "nonsense excuse", "scandalous", etc. Someone on *Nature's* website wrote that "Jack Gallant refuses to share the data (in violation with Nature's Journal Policy and with his NSF grants)." Some called to boycott Gallant and to retract his paper.

#### Questions for discussion:

- 1) Who is right in this debate? Are the objections to Gallant's position justified? What do you think about Gallant's reasons for not sharing the data set? Did it violate *Nature's* policy? Did he violate the principles of open science?

- 2) Who owns the data? Do scientists have duties to share the data? How are those duties justified?

### Supplementary readings:

1. Availability of Data. Nature portfolio. <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#availability-of-data>
2. Data sharing and the future of science. *Nature Communications* 9, 2817 (2018). <https://doi.org/10.1038/s41467-018-05227-z>
3. Gewin, V. (2016). Data sharing: An open mind on open data. *Nature*, 529(7584), 117-119. <https://doi.org/10.1038/nj7584-117a>
4. Staunton, C., Barragán, C. A., Canali, S., Ho, C., Leonelli, S., Mayernik, M., ... & Wonkham, A. (2021). Open science, data sharing and solidarity: who benefits? *History and Philosophy of the Life Sciences*, 43(4), 115. <https://doi.org/10.1007/s40656-021-00468-6>

### CASE 19: Should scientists share data in climate science?

SOURCE: McAllister, J. W. (2012). Climate science controversies and the demand for access to empirical data. *Philosophy of Science*, 79(5), 871-880.

<https://doi.org/10.1086/667871>

In recent years, critics of climate science have persistently sought access to raw data from the Climatic Research Unit at the University of East Anglia. Their efforts, often invoking the UK's Freedom of Information Act 2000, aimed to uncover evidence contradicting the scientific consensus on anthropogenic climate change. Climate scientists, viewing these requests as a campaign to waste their time and undermine their research unfairly, became increasingly sceptical. Once the correspondence between scientists was made public, critics cited selected messages in order to support their claims of a conspiracy among climate scientists to hinder data access and prevent external scrutiny of their work.

Subsequent inquiries devoted considerable attention to the issue of raw data access. Several reports highlighted climate scientists' reluctance to release data into the public domain and emphasized the importance of sharing scientific data with fellow researchers and the general public. For instance, the report by the UK House of Commons Science and Technology Committee quoted a response from Phil Jones of the Climatic Research Unit to Warwick Hughes, who had requested access to the raw data held by the unit: "Even if the World Meteorological Organization agrees, I will still not pass on the data. We have 25 or so years invested in the work. Why should I make the data available to you, when your aim is to try and find something wrong with it?" The report critically remarked that this response appeared unreasonable and stated that transparency and full disclosure of data and methods are fundamental to scientific integrity.

Further arguments put forth by Jones and his colleagues included assertions that releasing all the data was unnecessary and impractical: parts were already accessible through other sources such as the Global Historical Climatology Network in the United States, commercial agreements restricted the publication of certain data, most scientists preferred working with adjusted data rather than raw data, and the Climatic Research Unit did not have a specific obligation to provide raw data to the general public. While the committee appeared to acknowledge some of these points and sympathized with Jones' frustration in handling data requests driven by motives to undermine his work, the report concluded that the Climatic Research Unit should have been more transparent with the raw data and followed a more open approach to data availability.



Animation of this case is available on the [ROSiE Knowledge Hub](#).

#### Questions for discussion:

- 1) Who is right in this debate? Is the contested and politicized nature of some research fields a legitimate argument not to share raw data?
- 2) Why might scientists have reservations about sharing the data?

## Supplementary readings:

1. McAllister, J. W. (2012). Climate science controversies and the demand for access to empirical data. *Philosophy of Science*, 79(5), 871-880.  
<https://doi.org/10.1086/667871>
2. Zuiderwijk, A., Shinde, R., & Jeng, W. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PloS One*, 15(9), e0239283.  
<https://doi.org/10.1371/journal.pone.0239283>

### CASE 20: Collaborative authorship in digital humanities

SOURCE: Spiro, L. (2009). Collaborative Authorship in the Humanities.  
<https://digitalscholarship.wordpress.com/2009/04/21/collaborative-authorship-in-the-humanities/>

Lisa Spiro in her blog post on collaborative authorship in humanities wrote: “Recently I heard the editors of a history journal and a literature journal say that they rarely published articles written by more than one author—perhaps a couple every few years. Around the same time, I was looking over a recent issue of *Literary and Linguistic Computing* and noticed that it included several jointly-authored articles. This got me wondering: is collaborative authorship more common in digital humanities than in “traditional” humanities?

“Collaboration” is often associated with “digital humanities.” Building digital collections, creating software, devising new analytical methods, and authoring multimodal scholarship typically cannot be accomplished by a solo scholar; rather, digital humanities projects require contributions from people with content knowledge, technical skills, design skills, project management experience, metadata expertise, etc. [..]

Of course, collaboration poses some significant challenges, such as divvying up and managing work, negotiating conflicts, finding funding for complex projects, assigning credit, etc. But as Lisa Ede and Andrea A. Lunsford point out, collaborative authorship can lead to a “widening of scholarly possibilities.” In talking to humanities scholars (particularly those in global humanities), I’ve noticed genuine enthusiasm about collaborative work that allows scholars to engage in community, consider alternative perspectives, and undertake ambitious projects that require diverse skills and/or knowledge.”

#### Questions for discussion:

- 1) What are the advantages and disadvantages of collaborative research and publishing in digital humanities? What might be the role of collaborative research in the context of open science?
- 2) How to recognize the contribution of each co-author in case of collaborative authorship?
- 3) What ethical problems might arise in the context of collaborative authorship? How to prevent and solve these problems?

#### Supplementary readings:

1. McCarty, W. (2016). Collaborative research in the digital humanities. In *Collaborative Research in the Digital Humanities* (pp. 13-22). Routledge.  
<https://doi.org/10.4324/9781315572659>
2. COPE Council (2003). How to Handle Authorship Disputes: A Guide for New Researchers. <https://doi.org/10.24318/cope.2018.1.1>

## 3. The Embassy of Good Science: "[Authorship criteria](#)"



### CASE 21: Recognizing citizen scientists in scientific publications

SOURCE: Ward-Fear, G., Pauly, G. B., Vendetti, J. E., & Shine, R. (2020). Authorship protocols must change to credit citizen scientists. *Trends in Ecology & Evolution*, 35(3), 187-190. <https://doi.org/10.1016/j.tree.2019.10.007>

In Australia, native apex predators face fatal poisoning when they consume toxic invasive species of cane toads. To try to reduce the harm, a group of researchers investigated a novel conservation intervention aiming at discouraging native fauna from ingesting cane toads. The researchers collaborated with a group known as Balanggarra Rangers representing the indigenous traditional owners of the region. The research team consisted of both scientists and the Rangers, and the latter's role in the research was substantial.

However, the researchers soon realized that acknowledging the contribution of the Balanggarra people is a very difficult task. The scientists strongly believed that selecting only a few citizen scientists for authorship would have been both arbitrary and culturally insensitive. After negotiating with editors and editorial staff they succeeded in adding the 'Balanggarra Rangers' as group co-authors on two scientific papers. That was an expression of appreciation for the scientific value of traditional knowledge and skills, as well as respect for the collective cultural identity. Unfortunately, the group name of co-authors was often misleadingly abbreviated in citations as 'B. Rangers', because of using citation management software. In other scientific publications, the researchers were unable to include the Balanggarra Rangers as co-authors, because in some journals group authors are not allowed and other journals require an official academic affiliation for all co-authors. This experience has raised questions on proper and respectful ways of recognizing the contribution of citizen scientists in scientific publications.

*The International Committee of Medical Journal Editors (ICMJE) recommends that **authorship** should be based on the following four criteria: "(1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved."*

*According to these criteria, an individual can be considered as an author only on the condition if he/she fulfils all four conditions mentioned above.*

*Persons who have contributed to the paper but whose contribution does not justify authorship may be considered as contributors and their role should be described in the **contributorship** statement or acknowledgements.*



Animation of this case is available on the [ROSiE Knowledge Hub](#).

### Questions for discussion:

- 1) The case raises some issues about authorship, group co-authorship and contributorship in the context of citizen science. What are these issues? Can the ICMJE criteria be used to solve them? If yes, what that solution might be? If not, how to solve the issues raised by the case above?
- 2) Is the concept of group co-authorship helpful to solve the problems raised by this case? Why yes or no?
- 3) Do you have other suggestions for recognizing the contribution of citizen scientists in scientific publications?

### Supplementary readings:

1. Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature*, 508(7496), 312-313.  
<https://doi.org/10.1038/508312a>
2. COPE Council (2003). How to Handle Authorship Disputes: A Guide for New Researchers. <https://doi.org/10.24318/cope.2018.1.1>
3. ICMJE. Defining the role of authors and contributors. <https://bit.ly/N7uoq3>
4. Smith, E., Bélisle-Pipon, J. C., & Resnik, D. (2019). Patients as research partners; how to value their perceptions, contribution and labor? *Citizen science: theory and practice*, 4(1). <https://doi.org/10.5334/cstp.184>
5. The Embassy of Good Science: "[Authorship criteria](#)"
6. Vasilevsky, N. A. et al. (2021). Is authorship sufficient for today's collaborative research? A call for contributor roles. *Accountability in Research*, 28(1), 23-43.  
<https://doi.org/10.1080/08989621.2020.1779591>

### CASE 22: Authorship and contributorship in citizen science

SOURCE: Ward-Fear, G., Pauly, G. B., Vendetti, J. E., & Shine, R. (2020). Authorship protocols must change to credit citizen scientists. *Trends in Ecology & Evolution*, 35(3), 187-190. <https://doi.org/10.1016/j.tree.2019.10.007>

In 2015, to map the distribution of gastropods in Southern California, the Natural History Museum of Los Angeles County initiated a citizen science project named SLIME (Snails and Slugs Living in Metropolitan Environments) on the online platform iNaturalist (<https://www.inaturalist.org/projects/slime>). Given the challenges of urban sprawl and restricted private property access in the Greater Los Angeles Area, citizen science proved to be highly effective in generating species occurrence data. By late 2019, SLIME reached around 2200 contributors and 14000 observations. The project's findings were summarized in two publications that showed the initial evidence of several introduced gastropod species in California and in the USA.

In the first publication, four citizen scientists who provided specimens were listed as co-authors alongside five academic researchers. These citizen scientists provided feedback to the manuscript, and the editor of The Journal of Natural History accepted their co-authorship. The second article, published in The American Malacological Bulletin, lists three researchers and two groups of citizen scientists: a 14-person SLIME contributor team and a family that hosted a Malaise insect trap for snail collection. The information about group co-authors reads "and citizen science participants in SLIME and BioSCAN". The names of the individuals and the family appear in the author affiliations. Although the group co-author credit is visible in the published article, it is omitted from the article's citation in BioOne and Google Scholar but is included in Web of Science.

*The International Committee of Medical Journal Editors (ICMJE) recommends that **authorship** should be based on the following four criteria: "(1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved."*

*According to these criteria, an individual can be considered as an author only on the condition if he/she fulfils all four conditions mentioned above.*

*Persons who have contributed to the paper but whose contribution does not justify authorship may be considered as contributors and their role should be described in the **contributorship** statement or acknowledgements.*

#### Questions for discussion:

- 1) The case raises some issues about authorship and contributorship in the context of citizen science. What are these issues? Can the ICMJE criteria be used

to solve them? If yes, what that solution might be? If not, how to solve the issues raised by the case above?

- 2) Would the concept of contributorship be helpful to solve the problems raised by this case? Why yes or no?
- 3) Do you have other suggestions for recognizing the contribution of citizen scientists in scientific publications?

### Supplementary readings:

1. Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature*, 508(7496), 312-313.  
<https://doi.org/10.1038/508312a>
2. COPE Council (2003). How to Handle Authorship Disputes: A Guide for New Researchers. <https://doi.org/10.24318/cope.2018.1.1>
3. ICMJE. Defining the role of authors and contributors. <https://bit.ly/N7uoq3>
4. Smith, E., Bélisle-Pipon, J. C., & Resnik, D. (2019). Patients as research partners; how to value their perceptions, contribution and labor? *Citizen science: theory and practice*, 4(1). <https://doi.org/10.5334/cstp.184>
5. The Embassy of Good Science: "[Authorship criteria](#)"
6. Vasilevsky, N. A. et al. (2021). Is authorship sufficient for today's collaborative research? A call for contributor roles. *Accountability in Research*, 28(1), 23-43.  
<https://doi.org/10.1080/08989621.2020.1779591>

### CASE 23: Trade-offs in publishing preprints

SOURCE: Molldrem, S., Hussain, M. I., & Smith, A. K. (2021). Open science, COVID-19, and the news: Exploring controversies in the circulation of early SARS-CoV-2 genomic epidemiology research. *Global Public Health*, 1-14.

<https://doi.org/10.1080/17441692.2021.1896766>

In 2020, a team of genomic epidemiologists led by Bette Korber posted a preprint titled "Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2" on the bioRxiv repository. A preprint is a version of a research paper that is posted and shared publicly by researchers before it has undergone formal peer review and been published in a scientific journal. Preprints are usually shared on preprint servers, which are online platforms specifically designed for this purpose.

The story regarding the preprint was immediately featured in the *Los Angeles Times*. Molldrem et al. in their analysis of the events write about this publication: "Originally, the headline was 'A mutant coronavirus has emerged, even more contagious than the original, study says'. Within hours, the headline was changed to the less-sensational 'Scientists say a now-dominant strain of the coronavirus appears more contagious than original'; the following day, it was modified to read 'Scientists say a now-dominant strain of the coronavirus could be more contagious than original.'"

Vartabedian, the journalist who wrote the article, conducted interviews with members of the research team and several experts. He even cited first author's Bette Korber Facebook post:

'we see a mutated form of the virus very rapidly emerging, and over the month of March becoming the dominant pandemic form...When viruses with this mutation enter a population, they rapidly begin to take over the local epidemic, thus they are more transmissible.'

Afterwards Vartabedian wrote, including a quote from the preprint:

'D614G is increasing in frequency at an alarming rate, indicating a fitness advantage relative to the original Wuhan strain that enables more rapid spread,' the study said. Still unknown is whether this mutant virus could account for regional variations in how hard COVID-19 is hitting different parts of the world.'

Molldrem et al. in their analysis of the events write: "Vartabedian's inclusion of the qualifying phrase '*Still unknown is whether*' before suggesting the possibility of a SARS-CoV-2 evolutionary trajectory toward greater transmissibility by region is deceptive rhetoric. It presents a hypothetical, putting forward the possibility of an alternative outcome without making it sufficiently clear that there was not yet evidence to say that the alternative outcome (in this case, SARS-CoV-2 evolution toward greater transmissibility by region) was actually occurring."

Molldrem et al. state that this case shows several problems. First, it validates concerns within the scientific community that preprints frequently receive media coverage similar to fully peer-reviewed papers. Furthermore, it demonstrates that preprints have a potential to influence scientific landscape and society. Lastly, it illustrates how a

journalist's interpretation can overly sensationalize and misrepresent preliminary conclusions, especially in situations where expert opinions are uncertain, divergent, and perplexing.

### Questions for discussion:

- 1) What are the potential benefits of publishing preprints? What are the potential risks?
- 2) How to minimise the potential risks caused by publishing preprints?
- 3) How to minimize risks of misrepresentation of preprints in media?

### Supplementary readings:

1. COPE Council (2018). COPE Discussion Document: Preprints.  
<https://doi.org/10.24318/R4WByao2>
2. Elmore, S. A. (2018). Preprints: what role do these have in communicating scientific results? *Toxicologic pathology*, 46(4), 364-365.  
<https://doi.org/10.1177%2F0192623318767322>
3. Ravinetto, R. et al. (2021). Preprints in times of COVID19: the time is ripe for agreeing on terminology and good practices. *BMC Medical Ethics*, 22(1), 1-5.  
<https://doi.org/10.1186/s12910-021-00667-7>
4. Sheldon, T. (2018). Preprints could promote confusion and distortion. *Nature*, 559(7715), 445–445. <https://doi.org/10.1038/d41586-018-05789-4>

### CASE 24: Open post-publication peer review

SOURCE: Molldrem, S., Hussain, M. I., & Smith, A. K. (2021). Open science, COVID-19, and the news: Exploring controversies in the circulation of early SARS-CoV-2 genomic epidemiology research. *Global Public Health*, 1-14.

<https://doi.org/10.1080/17441692.2021.1896766>

In 2020, a group of scientists led by Xiaolu Tang published the paper 'On the origin and continuing evolution of SARS-CoV-2'. The manuscript was one of the earliest genomic epidemiology studies to be released following the initial COVID-19 outbreak in Wuhan. Remarkably, the paper underwent the processes of submission, review, and publication within just four days.

The article claimed that "there were two dominant 'types' of the novel coronavirus: 'S' and 'L.' The authors characterised 'L' as more 'aggressive,' stating that it had 'potentially higher transmission and/or replication rates.' In addition to working from a small number of sequences from many jurisdictions, the authors used questionable methodologies to make assertions about the evolution of 'L' from 'S' as well as transmission directionality within their sample. This involved re-identifying two cases – though not by name. The authors used demographic data from the GISAID entries and cited a January 2020 press release from the U.S. Centers for Disease Control and Prevention and a news report from Australia. The authors described how they cross-referenced information in those documents with the sequences to make inferences about the travel history of particular entries. They then extrapolated from this to make claims about patterns of global SARS-CoV-2 viral mutation."

These findings were widely picked up by news media. Immediately following the publication of the paper, a response by MacLean and colleagues emerged on the Virological.org website. Their paper was titled 'Addressing the Claims in "On the origin and continuing evolution of SARS-CoV-2"'. Virological.org is an open-source platform used by genomic epidemiologists. The platform is a "discussion forum for analysis and interpretation of virus molecular evolution and epidemiology". MacLean et al. critically evaluated the sample size and methodology employed by Tang et al. They contended that the authors had not successfully differentiated between two distinct SARS-CoV-2 types; instead, they had mistakenly attributed significance to harmless mutations which, while possibly informative in an epidemiological context, held no relevance to the virus's severity or transmissibility. Subsequent discussion unfolded on Virological.org, encompassing contributions from other scientists, as well as several coauthors of the initial Tang et al. publication.

#### Questions for discussion:

- 1) The case illustrates risks created by an accelerated peer-review process (four days in this case). What are the risks and benefits of an accelerated peer review



process? Are the risks justified in this case? Are there any ways to minimize risks?

- 2) The case is an example of how open science practices affect the traditional model of peer review and response. In your view, what are the benefits and risks of open post-publication peer review, e.g., by using platforms like [virological.org](https://virological.org), [f1000.com](https://f1000.com) or [pubpeer.com](https://pubpeer.com)?

### Supplementary readings:

1. *The original study:*  
Tang, X. et al. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review*, 7(6), 1012-1023. <https://doi.org/10.1093/nsr/nwaa036>
2. *Response :*  
MacLean, O. A. et al. (2020). Response to “On the origin and continuing evolution of SARS-CoV-2”. <https://virological.org/t/response-to-on-the-origin-and-continuing-evolution-of-sars-cov-2/418>
3. Bagdasarian, N., Cross, G. B., & Fisher, D. (2020). Rapid publications risk the integrity of science in the era of COVID-19. *BMC Medicine*, 18: 192  
<https://doi.org/10.1186/s12916-020-01650-6>
4. Besancon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P. Smout, C., Billy, E., Deforet, M., Leyrat, C. (2021). Open science saves lives: lessons from the COVID-19 pandemic. *BMC Medical Research Methodology*, 21: 117.  
<https://doi.org/10.1186/s12874-021-01304-y>
5. Harms, P. D., & Credé, M. (2020). Bringing the review process into the 21st century: Post-publication peer review. *Industrial and Organizational Psychology*, 13(1), 51-53. <https://doi.org/10.1017/iop.2020.13> "
6. The Embassy of Good Science: "[Post-publication peer review](#)"



### CASE 25: Open peer review

SOURCE: Fox, J., Pearce, K. E., Massanari, A. L., Riles, J. M., Szulc, Ł., Ranjit, Y. S., ... & L. Gonzales, A. (2021). Open science, closed doors? Countering marginalization through an agenda for ethical, inclusive research in communication. *Journal of Communication*, 71(5), 764-784. <https://doi.org/10.1093/joc/jqab029>

Open peer review has been seen as an important aspect of a more transparent, more open science. In traditional peer review, the identities of authors and reviewers are usually kept confidential. The open peer review introduces openness in different ways: the authors and reviewers may know each other's identity, the reviews may be published along with articles and their different versions, the comments to the articles may be open etc.:

- **open identity** "makes authors and reviewers known to each other",
- "in **open reports**, peer reviews are published alongside articles",
- **open pre-review** is, for example, "a crowdsourced platform where any scholar could review a manuscript before publication and a cumulative score would be displayed",
- in **open final-version commenting** "the public can comment on published articles, and authors are expected to engage with commenters to promote public communication about science".

While this can certainly affect the quality and the tone of the reviews, as well as offer opportunities to acknowledge the effort of reviewers, some scholars have drawn attention to the possibility that there are certain risks involved in the open-peer review, especially for marginalized researchers and research. Open reviewing might result in self-censorship for fear of retaliation or discrimination (felt especially by young, marginalized researchers). Open commenting could, in worst-case scenarios, turn into a witch-hunt.

#### Questions for discussion:

- 1) Imagine that you are a young scientist asked to review a well-established researcher's paper in an open peer review process. What are the challenges as well as opportunities involved?
- 2) How the potential dangers of open peer review could be handled in a way that best protects the researchers (both authors and reviewers)?

#### Supplementary readings:

1. Ferber, A. L. (2018). "Are you willing to die for this work?" Public targeted online harassment in higher education: SWS presidential address. *Gender & Society*, 32(3), 301-320. <https://doi.org/10.1177/0891243218766831>

2. Ross-Hellauer, T., Deppe, A., & Schmidt, B. (2017). Survey on open peer review: Attitudes and experience amongst editors, authors and reviewers. *PloS One*, 12(12), e0189311. <https://doi.org/10.1371/journal.pone.0189311>
3. Tenorio-Fornés, Á., Tirador, E. P., Sánchez-Ruiz, A. A., & Hassan, S. (2021). Decentralizing science: Towards an interoperable open peer review ecosystem using blockchain. *Information Processing & Management*, 58(6), 102724. <https://doi.org/10.1016/j.ipm.2021.102724>
4. The Embassy of Good Science: "[Open peer review - transparent way of gatekeeping science](#)"

### CASE 26: Publication of clinical case in open access journal

SOURCE: Retraction Watch (2016). Family decries publication of child's picture in open access journal. <https://retractionwatch.com/2016/11/07/family-decries-publication-of-childs-picture-in-open-access-journal/>

In 2012 the *Indian Journal of Dermatology, Venerology and Leprology* published a paper "[Delleman syndrome: Report of a case in an adolescent boy](#)" detailing a 14-year-old boy's case. The syndrome causes cysts and malformations in the eyes, brain, and skin. The paper included the boy's photo. Researchers obtained parental consent before publication, but the paper was later retracted as the parents realized the open access nature of the paper and withdrew the consent.

According to Mabel Nocito the study's first and corresponding author, the parents gave oral informed consent, and it was a legally sound approach at the time: "Up until 2014 the Argentinian law allowed free publication of portrait images for scientific purposes which is no longer the case. An oral permission was enough." At the same time, Nocito explained that the parents "did not understand fully what the term open access journal meant and even as the article was written in 2011, nowadays there's a more generalized possibility of linking, liking and sharing images in different social media." The paper was retracted. Reflecting on the consequences of the retraction, Nocito pointed out: "In the case of rare syndromes like Delleman, where case reports are the way to find out more about the disease, it is a setback but more and more we need to consider the patient's rights and feelings towards their conditions."

#### Questions for discussion:

- 1) It is mentioned in the case description, that the parents did not understand, what the term 'open access' meant and that their decision to withdraw the consent for publishing the picture was motivated by their realization that the picture is freely available. How should researchers explain open access of data or publications to research participants?
- 2) What (if anything) can be done to avoid the problem depicted in the case description?

#### Supplementary readings:

1. Roguljić, M., Šimunović, D., Poklepović Peričić, T., Viđak, M., Utrobičić, A., Marušić, M., & Marušić, A. (2022). Publishing Identifiable Patient Photographs in Scientific Journals: Scoping Review of Policies and Practices. *Journal of Medical Internet Research*, 24(8), e37594. <https://doi.org/10.2196/37594>
2. The Embassy of Good Science: "[Privacy in research](#)"

### CASE 27: Retraction due to the non-compliance with the journal's data policy

SOURCE: Retraction Watch (2019). PLOS ONE retracts perfume study when data don't pass the sniff test. <https://retractionwatch.com/2019/09/26/plos-one-retracts-perfume-study-when-data-dont-pass-the-sniff-test/>

In 2019, *PLOS ONE* published a paper titled "Social success of perfumes" by Vaiva Vasiliauskaite and Tim S. Evans of the Theoretical Physics Group and Centre for Complexity Science at Imperial College London. In the paper's abstract, the authors pointed out that they had studied "data on perfumes and their odour descriptors – notes - to understand how note compositions, called accords, influence successful fragrance formulas".

However, the paper was soon retracted by the editors of *PLOS ONE* due to concerns about the reproducibility of the study and noncompliance with the journal's data availability policy. The editors explained their concerns, writing that after publishing several questions were raised over the dataset used in the study. They noted that further inquiry revealed that it "was obtained from a third-party commercial entity" whose identity could not be disclosed "due to a nondisclosure agreement" and that the authors could not share the raw data or disclose information about how the data were collected and processed. Although the authors posted anonymized summary data on Figshare, the reported methods are insufficient "to enable other researchers to reproduce the study". The data the authors provided do not meet *PLOS ONE*'s requirements as described in the journal's data availability policy. According to the editors, the authors admitted to them that "they cannot reproduce the analyses using another public dataset as no comparable dataset is currently available".

The authors of the paper wrote to *Retraction Watch* and explained that the data is owned by a third party and that in order to use it, they had "to agree to very tight restrictions". For instance, they pointed out that at the point of communication with *Retraction Watch*, they no longer had access to the original data. Therefore, they were "very well aware of the restrictions when writing the paper". As they wanted to be as open as possible, they made as much of the data available as they could, and this data has been accessible in the repository listed in the paper's references. Before the publication of the paper, they explained the situation to the referees and to the journal. After publication, the journal reviewed the situation again and "at that point decided that the paper did not comply with their open data policy".



Animation of this case is available on the [ROSIE Knowledge Hub](#).

#### Questions for discussion:

- 1) Who is right in this debate? Are commercial interests and protection of intellectual property legitimate arguments not to share raw data?

2) Why might scientists have reservations about sharing their data?

### Supplementary readings:

1. Editorial (2018). Data sharing and the future of science. *Nature Communications*, 9, 2817-2817. <https://doi.org/10.1038/s41467-018-05227-z>
2. Staunton, C., Barragán, C. A., Canali, S., Ho, C., Leonelli, S., Mayernik, M., ... & Wonkham, A. (2021). Open science, data sharing and solidarity: who benefits? *History and Philosophy of the Life Sciences*, 43(4), 1-8. <https://doi.org/10.1007/s40656-021-00468-6>

## CASE 28: Predatory publishing – the dark side of open science?

SOURCE: De Rijcke, S., & Stöckelová, T. (2020). Predatory publishing and the imperative of international productivity: Feeding off and feeding up the dominant. In: De Rijcke, S., & Stöckelová, T. (eds.) *Gaming the Metrics: Misconduct and Manipulation in Academic Research*, pp. 101-110. <https://doi.org/10.7551/mitpress/11087.003.0010>

In 2015, a significant discussion emerged in an Eastern European country, centred around the publication and evaluation of research. The catalyst for this debate was the case of a highly productive junior researcher affiliated with one of the country's most prestigious universities. At first glance, this scientist appeared to be a successful scholar, boasting an extensive record of international publications, collaborations, and co-authorships – precisely in line with the prevailing research standards established in the country. However, upon closer examination, the situation revealed a rather different story.

This researcher earned his PhD in 2007 and afterwards claimed authorship, co-authorship or co-editorship of 17 scientific monographs between 2011 and 2013, as well as over 80 journal articles from 2006 to 2015. While the sheer volume of his output was impressive, several aspects of his output raised concerns:

- 1) the author also acted as an editor in chief, editorial board member, and even publisher of some of the journals in which he published,
- 2) some of the journals in which he published were included in Jeffrey Beall's database of predatory journals,
- 3) one of his co-authors appeared to be a fictional character with fake affiliations of prestigious universities.

Amid this controversy, some whistleblowers submitted a complaint to the University's Ethics Commission. The researcher's contract was subsequently terminated; however, at the same time, the contract of the main whistleblower was not renewed.

### Questions for discussion:

- 1) What ethical problems do you see in the activities of this researcher?
- 2) Why do academics publish their research in predatory journals or books published by predatory publishers? What are the negative consequences of such a practice?
- 3) What policies might minimise predatory publishing practices? Can the open science framework mitigate the incentives to violate the norms of academic integrity in this and similar cases?
- 4) How to recognise a predatory publisher? How should scientists treat publications in predatory journals, e.g., should research studies published in predatory journals be included in systematic reviews?

## Supplementary readings:

1. Bartholomew, R. E. (2014). Science for sale: The rise of predatory journals. *Journal of the Royal Society of Medicine*, 107(10), 384–385. <https://doi.org/10.1177/0141076814548526>
2. Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, 489(7415), 179. <https://doi.org/10.1038/489179a>
3. Beall, J. (2015). Criteria for determining predatory open access publishers. <https://beallslist.net/wp-content/uploads/2019/12/criteria-2015.pdf>
4. COPE Council (2019). COPE Discussion Document: Predatory Publishing. <https://doi.org/10.24318/cope.2019.3.6>
5. Kurt, S. (2018). Why do authors publish in predatory journals? *Learned Publishing*, 31(2), 141-147. <https://doi.org/10.1002/leap.1150>
6. The Embassy of Good Science: "[Predatory publishing](#)"

### CASE 29: Should all correction notices be open access?

SOURCE: Retraction Watch (2014). Should all correction notices be open access?

<https://retractionwatch.com/2014/02/04/should-all-correction-notices-be-open-access/>

A correction note in a scientific journal is a formal statement issued by the journal's editorial team to address errors or inaccuracies that have been identified in a previously published article. When errors are discovered in a published article, either by the authors, readers, or the editorial team itself, a correction note is typically issued. It serves to rectify mistakes or clarify information, ensuring the accuracy and integrity of the scientific literature.

A chemistry blogger See Arr Oh was irritated when he found out that a correction in *Organometallics* an American Chemical Society's journal was behind the paywall. To read the correction note one had to pay 35 USD to American Chemical Society. He wrote a blog post "Why Aren't All Correction Articles Free?" where he argues that correction articles should be freely available. Some journals offer correction articles as open access. The Committee on Publication Ethics recommends that all retraction notices should be open access but does not have a stance on whether corrections should be.

#### Questions for discussion:

- 1) Should all correction notices be open access? Are there any good reasons for them not to be open access?
- 2) Some corrections — the ones often referred to as errata — are quite minor. Should, for example, misspelt names be subject to the same rules as retractions?

#### Supplementary readings:

1. Oh, S.A. (2014). Why Aren't Correction Articles Free?  
<http://justlikecooking.blogspot.com/2014/01/why-arent-correction-articles-free.html>



### CASE 30: Accessing pirated papers for research

**AUTHORS:** Ivars Neiders & Signe Mežinska

Johanna, a prominent researcher at a European university, works at a research institution. Although its academic library has subscribed to several crucial journals in Johanna's field of research, there are still some relevant sources missing due to budgetary constraints. To access these unavailable papers, Johanna turns to Sci-Hub, a website described as a "shadow library" on Wikipedia, which provides free access to millions of research papers and books, regardless of copyright restrictions.

Alexandra Elbakyan created Sci-Hub in 2011. The website is widely used, serving approximately 400 000 requests per day in 2019, as the site itself reported. According to Sci-Hub's website, their database contained 88 343 822 research articles and books as of 2 June 2022. Some advocates of open access have praised Elbakyan's efforts. The prominent natural science journal *Nature* picked her as one of its Top 10 people in science who mattered in 2016<sup>2</sup>. According to the journal: "Few people support the fact that she acted illegally, but many see Sci-Hub as advancing the cause of the open-access movement, which holds that papers should be made [legally] free to read and reuse". Michael Eisen, a biologist at the University of California, expressed his admiration to *Nature*, stating, "What she did is nothing short of awesome. Lack of access to the scientific literature is a massive injustice, and she fixed it with one fell swoop". However, many publishing companies have sued Elbakyan, arguing that her actions are illegal.

Johanna shares Eisen's sentiment. When faced with arguments that Sci-Hub violates copyrights, Johanna highlights her belief that the current academic publishing model is morally flawed. She argues that since most research is funded by the public it is unfair for people to have to pay to access the results. According to Johanna, this creates a situation where individuals are essentially paying twice for the same information. Critics acknowledge that Johanna's point has some validity but argue that it does not justify either Johanna's use of pirated papers or Elbakyan's activities.



Animation of this case is available on the [ROSIE Knowledge Hub](#).

#### Questions for discussion:

- 1) Evaluate the argument provided by Johanna in her defence of using Sci-Hub! Can you add any other reasons why she might be right?
- 2) What reasons can you mention for Johanna being wrong? Evaluate the arguments pro and contra Johanna's view on the issue!

<sup>2</sup> Van Noorden, R. (2016). Alexandra Elbakyan: Paper pirate. *Nature*, 540, 512.  
<https://doi.org/10.1038/540507a>

- 3) What makes Sci-Hub so popular? Using Sci-Hub violates copyrights, but it seems clear that a simple ban on using it does not work. What would be the best solution to this problem at the policy level?

### Supplementary readings:

1. Bender, M. 'It's a Moral Imperative:' Archivists made a directory of 5000 Coronavirus studies to bypass paywalls. *Vice*, February 3, 2020.  
<https://www.vice.com/en/article/z3b3v5/archivists-are-bypassing-paywalls-to-share-studies-about-coronaviruses>
2. Monbiot, G. Scientific publishing is a rip-off. We fund the research - it should be free. *The Guardian*. September 13, 2018.  
<https://www.theguardian.com/commentisfree/2018/sep/13/scientific-publishing-rip-off-taxpayers-fund-research>.
3. Plan S (2018). Open Access is Foundational to the Scientific Enterprise.  
<https://www.coalition-s.org/why-plan-s/>
4. Van Noorden, R. (2016). Alexandra Elbakyan: Paper pirate. *Nature*, 540, 512.  
<https://doi.org/10.1038/540507a>
5. Vogel, G., & Kupferschmidt, K. (2017). A bold open-access push in Germany could change the future of academic publishing. *Science*, 23.  
<https://doi.org/10.1126/science.aap7562>

### CASE 31: Informing the general public about research

AUTHOR: Kadri Simm

Adam is a successful researcher at a big European research institution. He is a biologist and he studies yeasts. The communication department at Adam's institution and the sponsor of Adam's research project expect that Adam will not only do the research but also popularize the research results to the wider public via different public events and social media. Adam on the other hand considers it a waste of his time and thinks that the most important duty of a scientist is to conduct research.

#### Questions for discussion:

- 1) Please, evaluate Adam's claim, that the only duty of scientists is to conduct scientific research! Is this a defensible position? What arguments can be made in its support? What might be wrong with Adam's view?
- 2) It might be argued that researchers must inform the wider public about their research. However, does that include posting research results on social media (Twitter and Facebook)? Can research institutions and the sponsors of research demand this from the researchers?

#### Supplementary readings:

1. Davies, S. R., Franks, S., Roche, J., Schmidt, A. L., Wells, R., & Zollo, F. (2021). The landscape of European science communication. *Journal of Science Communication*, 20(3). <https://doi.org/10.22323/2.20030201>
2. Kessler, S. H., Schäfer, M. S., Johann, D., & Rauhut, H. (2022). Mapping mental models of science communication: How academics in Germany, Austria and Switzerland understand and practice science communication. *Public Understanding of Science*, 31(6), 711-731. <https://doi.org/10.1177/0963662521106574>

### CASE 32: Inequities and potential of exploitation in open science

SOURCE: Ewuoso, C., Cordeiro-Rodrigues, L., Wonkam, A., & de Vries, J. (2022). Addressing exploitation and inequities in open science: A relational perspective. *Developing World Bioethics*. <https://doi.org/10.1111/dewb.12378>

Open collaboration in science sometimes triggers concerns about exploitation, both in terms of control over resources and professional equity. For example, scholars may worry that open science practices, including sharing data, methods, codes etc., might lead to various forms of exploitation and inequity, e.g., researchers might not be adequately acknowledged or credited for the resources that they have shared, that they may be scooped by other scientists who manage to publish faster and so on.

Cornelius Ewuoso and his co-authors in their article express worries from the point of view of the African scientific community that the potential of exploitation may disadvantageously affect under-resourced scholars, particularly those from developing countries. Two main concerns revolve around exploitation in open science. Firstly, relinquishing control and ownership of shared methods, data, or codes raises questions about their use. Researchers fear losing control over these resources, which could leave them vulnerable. Retaining control, however, can facilitate non-exploitative collaborations and ensure resources are available on researchers' terms. Secondly, there is a concern about professional vulnerability and equity. Researchers who share resources may not receive adequate acknowledgement or rewards for their efforts. There's also fear of being scooped by others who can publish quickly. Scholars worry that well-resourced researchers might exploit shared methods and data, potentially sidelining others.

Ewuoso et al. suggest: "[...] it appears reasonable that individuals who have shared their materials should be recognised by researchers who use them. This could be by way of giving credit to those who have shared. Sharing will cease to be just if individuals are not recognised for their work, scooped by research parasites, or undercut by others who can publish quickly on any subject. It could equally be by way of acknowledgement in the published work or co-authorship. We also think open datasets, codes and methods in repositories may be recognised as works of scholarship in their own right, reflecting the capabilities and inventiveness of those who created and shared them."

#### Questions for discussion:

- 1) Do you agree that practising open science might lead to exploitation? Why yes or no?
- 2) What should be done to reduce risk of exploitation in the context of open science?
- 3) How to protect intellectual property when practicing open science? How the authors of open datasets, codes, methods etc. be acknowledged? Are the existing practices of acknowledgement working well?

### Supplementary readings:

1. Bull, S., & Bhagwandin, N. (2020). The ethics of data sharing and biobanking in health research. *Wellcome Open Research*, 5.  
<https://doi.org/10.12688/wellcomeopenres.16351.1>
2. Ross-Hellauer, T., Reichmann, S., Cole, N. L., Fessler, A., Klebel, T., & Pontika, N. (2022). Dynamics of cumulative advantage and threats to equity in open science: a scoping review. *Royal Society Open Science*, 9(1), 211032.  
<https://doi.org/10.1098/rsos.211032>
3. Zeitlyn, D. (2003). Gift economies in the development of open source software: anthropological reflections. *Research Policy*, 32(7), 1287-1291.  
[https://doi.org/10.1016/S0048-7333\(03\)00053-2](https://doi.org/10.1016/S0048-7333(03)00053-2)