IRoSiE

Training Materials for Responsible Open Science

Case study 1

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?

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

IRoSiE

Training Materials for Responsible Open Science

Case study 2

Should scientists share the 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.

Questions for discussion:

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

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

IRoSiE

Training Materials for Responsible Open Science

Case study 3

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 reused for research purposes?
2) What role if any is played by the fact, that the data contains disclosures of sexual violence? On what conditions the data can be reused?
3) How should the data be anonymized or pseudonymised in this case?
4) What are the specific ethical concerns related to reuse of social media data?

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