IROSiE

Training Materials for Responsible Open Science

Case study 1

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 politicised 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 2

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?

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

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