The value of open data
A meta-research(er) perspective

AGU/CHORUS Forum: How open is open data and software?
April 19, 2023

Kathleen Gregory
University of Vienna
Scholarly Communications Lab, University of Ottawa
Today’s talk
Focus data sharing/reuse

Three questions (‘provocations’)
Three responses
So what can be done?
Provocation 1:

Does ‘openness’ depend on context and demand flexibility?
No need to wait for an answer:

A resounding yes!
We have to be very careful sharing the raw data, otherwise we can be sued by the lab or by the university...In terms of the clinical data (in Portugal), the situation is different...it is not that formal as in the United States. It is quite frequent to share clinical data with other researchers.
Why do you not reuse data?

Weighted percentages of respondents to this question (n=466). Items with a significant disciplinary difference in purple. Multiple responses possible.
Recognize the spectrum of research

- Flexible ‘snap-in’ policies
- Sharing ‘openly’ or reusing data are not always relevant or possible
Provocation 2:
What is our responsibility around attribution and credit?
Wait a minute...
Are researchers already ‘responsibly’ attributing data use?
Citing data

Majority of data re-users report citing!

Variety in practice

**What data objects do you cite or reference?**

- A data paper: 22.7% Never, 17.2% Rarely, 16.8% Sometime, 11.5% Often, 22.1% Always
- An article or publication analyzing the data: 3.4% Never, 12.5% Rarely, 21.9% Sometime, 55.8% Often
- A collection of related data: 19.1% Never, 17.7% Rarely, 22.0% Sometime, 12.5% Often, 16.8% Always
- The source of the data: 6.1% Never, 15.5% Rarely, 13.6% Sometime, 15.3% Often, 55.4% Always
- The data themselves: 9.7% Never, 13.1% Rarely, 15.3% Sometime, 16.8% Often, 41.5% Always

Weighted percentages of respondents to this question (n=2,026). Significant disciplinary differences marked with asterisk.
Motivations

Reasons reflect ideal (responsible?) research practice

Why do you *cite* data?

- Show intellectual debt to the data creator/data provider: 77.0%
- Help others to locate and access the data you used: 75.9%
- Support the validity of your claims: 71.4%
- Indicate that you have used the data in some way: 65.7%
- Reward the data creators/data providers: 54.6%
- Because you were advised to cite data (e.g. by a journal editor): 8.4%
- Other: 3.1%

Weighted percentages of respondents to this question (n=2,026). Items with a significant disciplinary difference in purple. Multiple responses possible.
So what?
Build on existing practices

Data citation practices
- Encourage standardized data citation
- But meet researchers where they are
Provocation 3: What is the value of data management and data sharing?
The value of *Slow (Reflexive) Science*

Wait a minute – isn’t faster science the goal?
This takes work.

Detailed, tedious

Lots of questions – also an opportunity

https://zenodo.org/communities/meaningfuldatacounts/
Data management takes time and work.

Existing steps (e.g. DMPs) as a check-in not a checkbox.
Thank you.

Questions/Discussion?

Kathleen Gregory
@gregory_km | kathleengregory@mastodon.social
kathleen.gregory@univie.ac.at