Data Sharing for Greater Scientific Transparency

January 06, 2021

Openness and transparency are core values of the scientific process. Sharing research data "can be perceived as a signal of commitment to transparency and of confidence in the integrity of the data and analyses," wrote APS Fellow D. Stephen Lindsay, former editor of *Psychological Science*, in 2017. Accessible data can also foster critical reanalyses that shed new light on findings and facilitate meta-analytic work, he added.

Major funding agencies often require researchers to share the data provided by their research. The National Institutes of Health (NIH) policy on data sharing, for example, states that "data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data." An updated NIH policy for data management and sharing, effective in 2023, will require NIH-funded researchers to submit a plan outlining how their data will be managed and shared before beginning their research.

The European Research Council (ERC) has also expressed its support for the so-called FAIR data principles of findability, accessibility, interoperability, and reusability. In information provided to grantees in 2019, the ERC suggested that "the next step in the development of open science is making research data publicly available when possible."

The ERC recommends that its funded researchers share their data in depositories that:

- store the data safely (e.g., prevent access from unauthorized users);
- make sure the data remain findable (i.e., use identifiers and links that are permanently and uniquely attached to each data set), accessible, and reusable;
- describe the data in a standard way, so their usability is maximized; and
- add a license stating who can access and reuse the data.

The OSF

One of the most commonly used platforms for data sharing is the OSF (formerly known as the Open Science Framework), a not-for-profit depository developed and maintained by the Center for Open Science. Founded by Brian Nosek and Jeffrey Spies, the Center for Open Science's mission is to increase the openness, integrity, and reproducibility of research.

OSF Guide

The OSF provides several guides with instructions on everything from creating and managing projects to ensuring the security and privacy of data. In a 2018 article in *Advances in Methods and Practices in Psychological Science*, Courtney K. Soderberg wrote a step-by-step guide for quickly using OSF to share data. The OSF user interface has since changed; download an updated quick-use guide below.

JAN-FEB_OBS_2021-MethodsDownload

The OSF is a free, open-source collaboration tool that lets researchers share their data for private collaborations with specific researchers or disseminate their entire projects to the public. Brian Nosek explained in the March, 2014 edition of the *Observer* that the "OSF helps individuals and research teams organize, archive, document, and share their research materials and data. Users have accounts and create projects…. The OSF logs actions and retains version histories of the wikis and files so that the history of the research process is recoverable."

Ethical data sharing

In a 2018 tutorial published in *Advances in Methods and Practices in Psychological Science*, Michelle N. Meyer (Geisinger Health System) described do's and don'ts for ethically sharing data from research involving humans.

Sharing future data

If you plan to share data that you have yet to collect, consider these measures in both your consent form and your institutional review board (IRB) submission:

- In your consent form, state that the data will not be destroyed and might be shared.

- Don't promise that research analyses of the collected data will be limited to certain topics.
- Get consent to retain and share data. Disclose who will have access to the data (e.g., other researchers at the same institution, researchers at other institutions, government agencies, the general public, commercial entities) and the purposes for which the data may be reused (e.g., reanalysis and replication, new analyses).
- Incorporate data-retention and data-sharing clauses into IRB templates.
- Consider the risks of reidentification.
- Choose a data repository judiciously, given that different repositories have different guidelines and privacy settings and might be more suitable for some data sets than others (see Meyer, 2018).

Sharing collected data

Sharing data that you've already collected poses two risks to participants. First, they may become subject to harms including privacy loss and discrimination, and second, their data may end up being used for research purposes to which they would not have consented. Meyer (2018) proposed that sharing previously collected data is more acceptable under the following conditions:

Open Science Badges

Open Data: For making publicly available the study data that other researchers would need to reproduce the reported results.

Open Materials: For making publicly available the materials and methods that other researchers would need to reproduce the experiments leading to the reported results.

Preregistered: For having disclosed a plan for the experimental design and analysis (i.e., specification of the variables and the statistical analyses that will be conducted) before the research was conducted and having followed that preregistered plan.

- The original consent form did not include a promise not to share data.
- Sharing the data is unlikely to cause significant harm to participants.
- The shared data are not individually identified and are not likely to be relinked to individuals.
- The shared data are accessible only under restricted conditions, protected by agreements prohibiting reidentification.
- Sharing is limited to research purposes that fall within the scope of the research described in the

original consent form.

Sharing "public" data

During the COVID-19 pandemic, many researchers have been forced to find alternatives to collecting data in their laboratories, often opting to use publicly available data to attempt to answer their research questions. Sources of publicly available data include social media platforms such as Twitter, Facebook, and Instagram, along with dating websites such as OkCupid.

The use and sharing of publicly available data raise specific ethical questions, given that "participants" are usually not aware of their participation, did not provide informed consent, did not have an option to opt out of the research, and did not intend for their data to be shared or used in research. Moreover, efforts to eliminate connections between data and participants' identities might not completely prevent the reidentification of data.

Meyer (2018) illustrated some of these issues by describing a study in which the researchers joined a closed online community to retrieve data. "The fact that users were willing to disclose personal information to fellow members of a particular community, for a particular purpose… does not mean that they would have agreed to share the same information with researchers, much less with the public, and much less in a permanent data repository" (p. 142).

Employing good research practices and seeking institutional approval before retrieving and sharing publicly available data might help to avoid ethical violations. Josh VanArsdall (Elmhurst College), in the APS webinar Online Research: Tools and Techniques, encouraged researchers to keep in mind that most online data were not intended to be used in research and may not fully represent the people behind the data. Be careful not to misrepresent data, he added, and always seek IRB approval.

References

Lindsay, D. S. (2017). Sharing data and materials in psychological science. *Psychological Science*, *28*(6), 699–702.

Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, *1*(1), 131–144.

Soderberg, C. K. (2018). Using OSF to share data: A step-by-step guide. *Advances in Methods and Practices in Psychological Science*, *1*(1), 115–120.