



GUIDANCE

The truth about open data: As open as possible, as closed as necessary

The British Psychological Society is committed to supporting and facilitating open research and its underlying principles. A core aim of the open science movement is to make science more reliable and to increase trust in research through openness and transparency. Sharing research data – ‘open data’ – so that it can be freely used, repurposed and redistributed is integral to open research.

Data sharing is beneficial because it can facilitate the discovery of new knowledge, as the outputs of previous research can be utilised in new research. It enhances transparency and reduces the potential for bias and improves data management practices, which ultimately improves the robustness of the scientific evidence base¹.

Interestingly – researchers think that other researchers are fearful that alternative analyses might expose poor research skills, invalid conclusions and therefore negative scrutiny². In contrast, fear of misinterpretation of data or of being scooped topped their own fears³.

Many researchers do not see how sharing their data could advance science and fear the opposite – that data in the hands of naïve or malicious analysts could lead to poor science. However, the issue of who is harmed by sharing data needs to be balanced against who is harmed by NOT sharing the data⁴.

SO WHAT ARE THE MOST COMMONLY PERCEIVED CHALLENGES TO DATA SHARING?

Being scooped (other researchers publishing results before the primary researcher has had the opportunity to)

Too much effort to present the data in an organised way

The loss of control over intellectual property/not wanting to ‘give away’ data

Not receiving credit or acknowledgement for your data

¹ Hrynaszkiewicz et al., 2013.

² Branney et al., 2019; Houtkoop et al., 2015.

³ Branney et al., 2019; Houtkoop et al., 2015.

⁴ Martone et al., 2019.

The detection of errors or flaws in the data and the consequent negative impact on reputation

The rejection of conclusions as a result of alternative analysis of the data

The misinterpretation of the data by secondary users

Compromises to the identities of participants

Failure to redact identifying information⁵.

SO WHAT IS THE TRUTH ABOUT OPEN DATA?

CHALLENGE #1: BEING SCOOPED/GIVING AWAY DATA

Truth: Any uses of research data must be appropriately cited and acknowledged. The best practice set out in the BPS (2020a) *Statement on Authorship and Publication Credit* should be adopted.

Moreover, researchers who are the original data creators must have the opportunity for primary use of the data for an appropriate and well-defined period. Data should be shared according to a clear timetable for release/partial release.

Embargoed data sharing allows time for the author to re-use their own data as well as de-identify and annotate data appropriately. The latter may improve the quality of the original work. The researcher can also register their protocol for future intentions to publish (and ask those whom they share the data with to check this before proceeding with their own analysis to the original team).

Publishing their data puts a date and time stamp on the researchers intellectual property. Many online platforms have version control which can evidence the date of registration of the work, and the dates of any amendments to documents even before the information is made public. The researcher can also assign a digital object identifier (DOI) to ensure their work can be referred to and cited.

The researcher can also retain copyright of their own work by setting a creative commons licence. The author can specify citation of the work appropriately on the reuse in the format that they require which acknowledges the data source and request that they check the existing protocols to minimise duplication of effort. More information is creativecommons.org on different CC licences.

Mandatory data sharing policies may also improve practices but only when supported by resources and incentives. We also recognise that the type of data to be shared, in what format and with what metadata, will vary depending on the methodology used and the data generated. One size does not fit all and any mandatory requirements for open data must reflect this (BPS, 2020b).

⁵ Banks et al., 2018; Houtkoop et al., 2015; Martone, 2019; Stuart et al., 2018.

Training in the responsible conduct of research should focus more explicitly on privacy concerns including data privacy. Authors who collect raw data that could never be shared could be trained to use de-identification and aggregation methods, or the creation of synthetic datasets (whose properties mirror those of the original data, but can be safely shared without fear or participants being identified) that might still allow for useful data to be shared.

CHALLENGE #2: TOO MUCH EFFORT

Truth: Data management plans can help researchers plan for the sharing of their data. Data management plans can help researchers plan for the sharing of their data and should be part of research planning from the outset. In this way, data collected already has a clear plan for sharing or a justification for not sharing. This plan for the sharing of data should also include the consent process with participants as well as data analysis, coding and storage. Meyer (2018) provides an excellent overview for practical tips for ethical data sharing.

CHALLENGE #3: DETECTION OF ERRORS/FLAWS IN THE DATA WHICH COULD IN TURN LEAD TO THREATS/HARASSMENT/TROLLING/REPUTATIONAL DAMAGE

Truth: Mistakes and errors will happen, we are all human. Open data, and indeed open research more generally, aims to provide researchers with the tools to detect and fix errors as they occur. Therefore, whilst the prospect of other researchers finding errors in your work may feel threatening, particularly to early career researchers, over time this can 'normalise the humanness of research'⁶ and improve the robustness of the scientific evidence base.

CHALLENGE #4: ALTERNATIVE CONCLUSIONS ARISING FROM NEW ANALYSIS WHICH COULD RESULT IN A REJECTION OF THE ORIGINAL CONCLUSIONS AND LEAD TO REPUTATIONAL DAMAGE

Truth: This is a concern across all science. Replication is a key element of scientific development and advancement. Unless the new analysis is reckless and malicious, for the most part it may prompt debate, further studies and evidence for the original findings.

CHALLENGE #5: COMPROMISING PARTICIPANTS/LOSS IN ACCESS TO DATA

Truth: As outlined above, not all data has to be shared. We recognise and advocate the principle of 'as open as possible; as closed as necessary'⁷; that not all data needs to be open and that there is a need to maintain confidentiality, protect individuals' privacy and the consent given, as well as manage security and risk. Nevertheless, restrictions to data must be justifiable and justified (BPS, 2020a)⁸. The key to protecting research participants is planning for responsible data sharing at the beginning of a project.

LOOKING TO THE FUTURE

We acknowledge that the effective adoption of open data requires the appropriate infrastructure, resources (software, codes and models/frameworks), training and investment. Infrastructure is especially important to ensure the long-term sustainability and protection of data⁹. We support and will actively engage in the development of more accessible resources and appropriate guidance to enable researchers at all levels to adopt open data practices.

⁶ Kathawalla et al., 2020, p.21.

⁷ European Commission. Directorate-General for Research & Innovation. *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*. Version 3.0. 26 July 2016.

⁸ <https://www.bps.org.uk/news-and-policy/open-data-position-statement>

REFERENCES

- Banks, G.C., Field, J.G., Oswald, F.L. et al. (2018). Answers to 18 questions about open science Practices. *Journal of Business and Psychology*, <https://doi.org/10.1007/s10869-018-9547-8>
- Branney, P., Reid, K., Frost, N. et al. (2019). A context-consent meta-framework for designing open (qualitative) data studies. *Qualitative Research in Psychology*, *16*(3), 483–502. <https://doi.org/10/gf429z>
- British Psychological Society (2020a). *Statement on Authorship and Publication Credit*. Leicester: BPS.
- British Psychological Society (2020b). *Position statement on open data*. Leicester: BPS.
- Houtkoop, B.L., Chambers, C., Macleod, M. et al. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 70–85.
- Kathawalla, U.K., Silverstein, P. & Syed, M. (2020). *Easing into open science: A guide for graduate students and their advisors (WWW Document)*. doi:10.31234/osf.io/vzjdp
- Martone, M.E., Garcia-Castro, A. & VandenBos, G.R. (2019). Data sharing in psychology. *American Psychologist*, *73*(2), 111–125.
- Meyer, M.N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, *1*(1), 131–144.
- Stuart, D., Baynes, G., Hrynaskiewicz, I. et al. (2018): *Whitepaper: Practical challenges for researchers in data sharing*. <https://doi.org/10.6084/m9.figshare.5975011.v1>.



the british
psychological society
promoting excellence in psychology