

Sharing data safely while preserving privacy

[https://doi.org/10.1016/S0140-6736\(19\)32603-0](https://doi.org/10.1016/S0140-6736(19)32603-0)

Analysing personal data is a privilege requiring researchers to safeguard data and to use data wisely. Safeguarding data means protecting the identity of individuals. Using data wisely means using, reusing, and sharing data to their maximum potential. More researchers should be given safe access to previously collected data from expensive clinical trials and laboratory or epidemiological studies. Journal editors therefore increasingly require a data sharing statement in published articles.[1]

Data sharing must, however, be safe, regulated, and in line with current legislation. Unfortunately, some journals encourage researchers to post their data openly in downloadable files, regardless of whether these are aggregated or individual level data. Such practice is potentially challenging to the privacy of individual participants.

Although sharing biological data such as serum values or RNA expression data in an Excel file on a journal website might seem innocuous, such open sharing requires extreme prudence from researchers and journal editors. Data on other variables might be necessary to replicate the original analyses (eg, age, gender, geographical location, socioeconomic status, lifestyle details, comorbidity, and disease). However, these data pose a potential risk of backwards identification of specific individuals. Increasingly sophisticated web crawler software can find common text or numbers across all published data and domains. The larger the number of variables, the greater the risk with modern technology.

For Europeans, such open data sharing is likely to violate the EU General Data Protection Regulation (GDPR) unless data are truly anonymous and therefore not covered by GDPR. However, the GDPR definition of anonymous data is stringent. To make pseudonymised individual data open and accessible, European researchers would need specific consent from participants. The informed consent of most studies currently being analysed do not state that the data will be made openly accessible, nor does it explain the risk of backwards identification. In addition, the rights for individuals provided by GDPR, such as the right to have data erased, must be properly handled.

Making pseudonymised individual data openly accessible, without careful attention to what is being shared, represents a potential threat to the privacy of the individuals whose data we all depend on. We therefore ask editors to reconsider their policy on making individual level data openly accessible. A data sharing statement should be mandatory and sufficient to satisfy the editors' requirements.

Data sharing through safe depositories with access control[2] is a step in the right direction, whereas making sensitive individual data publicly available on a journal's website is not.

We declare no competing interests. Where authors are identified as personnel of the International Agency for Research on Cancer or WHO, the authors alone are responsible for the views expressed in this Correspondence, and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer and WHO.

References

- [1] DB Taichman, P Sahni, A Pinborg, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *Lancet*, 389 (2017), pp. e12-e14.
- [2] D Mascalzoni, HB Bentzen, I Budin-Ljøsne, et al. Are requirements to deposit data in research repositories compatible with the European Union's General Data Protection Regulation? *Ann Intern Med*, 170 (2019), pp. 332-334.