RESEARCH · RESEARCH INTEGRITY

Research Data Management Plans and Long-Term Knowledge Preservation

What a research data management plan contains, why funders require one, how the FAIR principles guide it, and how to preserve data so that knowledge remains usable for decades.

ModQu research article · 878 words · 8 references (APA 7th edition)

Data outlast the projects that create them, but only if someone plans for it. Files saved on a laptop, in an obsolete format, without notes on what each variable means, are effectively lost within a few years. A **data management plan (DMP)** is a short document, written at the start of a project, that sets out how data will be collected, organised, documented, stored, shared and preserved. Many funders now require one. This article explains what a good plan contains and how long-term preservation works.

Why plans matter

The evidence on what happens without planning is stark. Vines et al. (2014) requested the data behind biology papers of different ages and found that the chance of obtaining them fell steeply as papers grew older, mostly because the data had been lost or the authors could not be found. Wicherts et al. (2006) found that most psychologists did not share their data when asked, even though their journals' policies required it. Tenopir et al. (2011) found that researchers often lacked the time, funding and infrastructure to manage data well. A plan helps by making these needs visible and budgeted from the start.

What a plan contains

Michener (2015) set out simple rules for a good data management plan, and most funder templates cover similar ground:

- **Data description.** What types and volumes of data will be collected or reused, and in what formats.
- **Documentation and metadata.** How variables, methods and file structures will be described so that others can understand the data.
- **Storage and security.** Where data will be kept during the project, how they will be backed up and how sensitive data will be protected.
- **Ethics and legal issues.** Consent, confidentiality, data protection and ownership, and how these affect sharing.
- **Sharing and access.** Which data will be shared, when, where and under what licence or conditions.
- **Preservation.** Which data will be kept long term, in which repository and for how long.
- **Responsibilities and resources.** Who is responsible for each task and what it will cost.

Corti et al. (2019) stress that consent forms and data management must be designed together: if participants are not told at the outset that anonymised data may be shared, sharing later may not be possible.

Common mistakes

Several mistakes recur. Plans are written only to satisfy a funder and then ignored. File names and folder structures are improvised, so that no one can tell which version of a data set was used for the final analysis. Variables are renamed or recoded by hand without a record. Backups are kept in the same place as the originals. And consent forms promise to destroy data that the funder expects to be shared. Treating the plan as a working document, reviewed as the project develops, avoids most of these problems.

The FAIR principles

Wilkinson et al. (2016) proposed that research data should be **Findable**, **Accessible**, **Interoperable** and **Reusable**. In practice, this means giving data persistent identifiers, describing them with rich metadata, using open and widely supported formats, and attaching clear licences and documentation. FAIR does not require that all data be open; sensitive data can be FAIR if their existence and the conditions of access are clearly described.

Long-term preservation

Preservation is more than storage. Formats become obsolete, storage media decay and the knowledge needed to interpret data disappears as people move on. The **Open Archival Information System (OAIS)** reference model, an international standard for digital archives, describes the functions a repository must perform to keep information usable over time: taking in data with adequate documentation, storing it securely, monitoring formats and migrating them when needed, and providing access to future users (Consultative Committee for Space Data Systems, 2012).

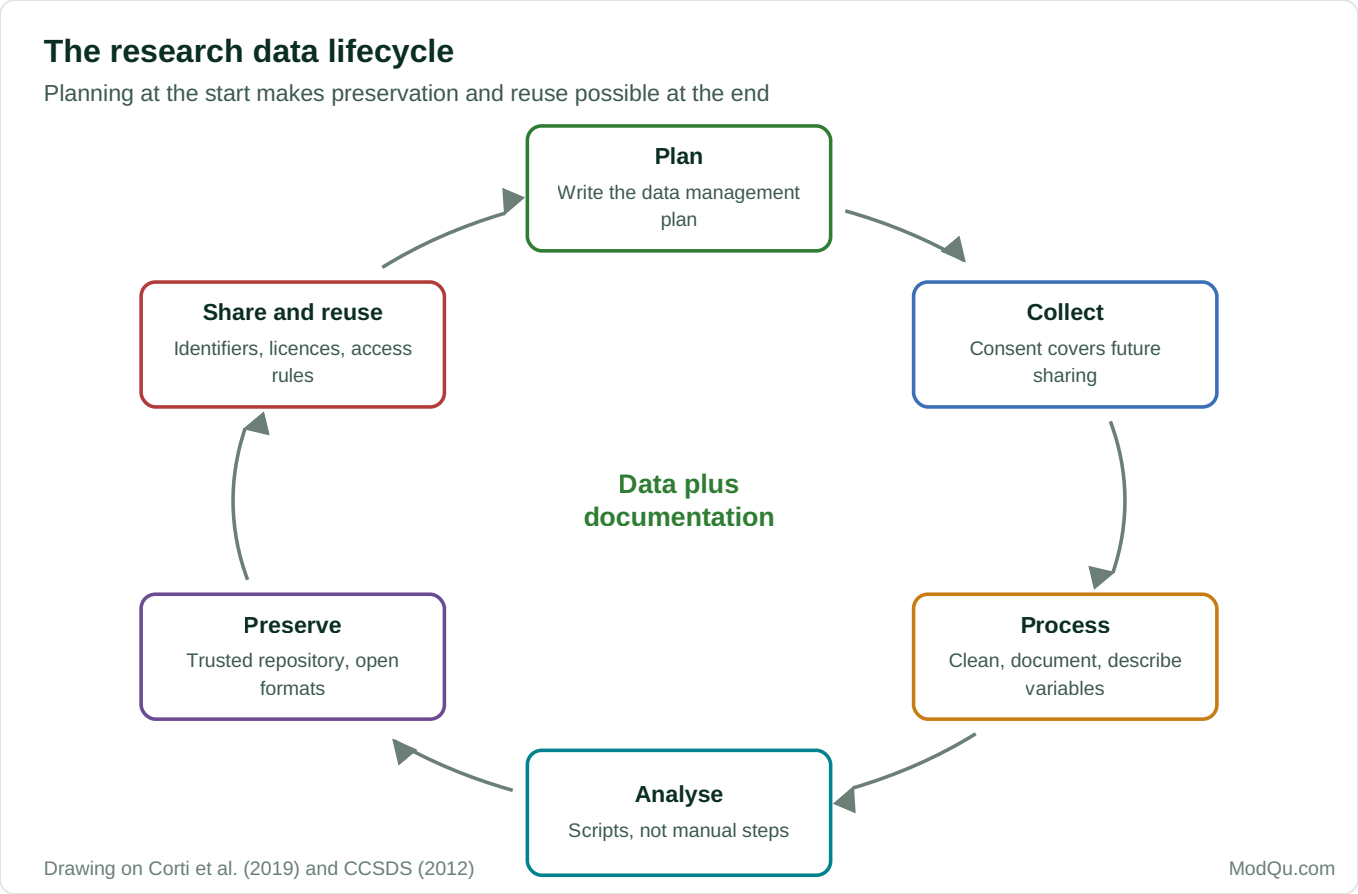


Figure 1. The research data lifecycle. Diagram by ModQu, drawing on Corti et al. (2019) and the OAIS reference model (CCSDS, 2012).

Borgman (2015) cautioned that data have value only in relation to the knowledge and context needed to use them. Preserving numbers without the questionnaires, codebooks and descriptions of how they were produced preserves little. Good preservation therefore keeps data, documentation and code together.

For most researchers, the practical choice is to deposit data in a trusted repository, such as a national data archive, a discipline-specific repository or an institutional repository, rather than to attempt long-term preservation themselves. Such repositories provide persistent identifiers, backup, format migration and managed access to sensitive data.

Plans in business research

Commercial research organisations face the same problems, often more acutely, because projects end quickly and staff move on. A simple internal plan for each project, recording where data are stored, what each file contains, how long data must be kept under contract and data protection law, and when they will be deleted, protects both the organisation and the people whose data it holds. It also makes past studies usable for future comparison, turning one-off projects into a cumulative body of knowledge.

Conclusion

A data management plan is a small investment at the start of a project that determines whether its data remain usable at the end and long after. Planning for documentation, storage, sharing and preservation, guided by the FAIR principles and supported by trusted repositories, turns data from a private by-product into a lasting public resource.

References

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.

Consultative Committee for Space Data Systems. (2012). *Reference model for an Open Archival Information System (OAIS)* (CCSDS 650.0-M-2). CCSDS.

Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2019). *Managing and sharing research data: A guide to good practice* (2nd ed.). SAGE.

Michener, W. K. (2015). Ten simple rules for creating a good data management plan. *PLoS Computational Biology, 11*(10), Article e1004525. https://doi.org/10.1371/journal.pcbi.1004525

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE, 6*(6), Article e21101. https://doi.org/10.1371/journal.pone.0021101

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J.-S.,
Renaut, S., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology, 24*(1), 94–97. https://doi.org/10.1016/j.cub.2013.11.014

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*(7), 726–728. https://doi.org/10.1037/0003-066X.61.7.726

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, Article 160018. https://doi.org/10.1038/sdata.2016.18