

Everybody talks about research data management — But where is the science?

Till Biskup

Physikalisch-Technische Bundesanstalt (PTB), FB 7.1 Radiometrie mit Synchrotronstrahlung

Newton, when seeing further, stood on the shoulder of giants,[1] not on the piles of his predecessors' cruffy research data. Nearly 300 years later, Richard Feynman coined the term "cargo cult science" for research that appears scientific but has no scholarly contribution nor impact.[2] With all the current hype about research data management, we tend to forget that science is about gaining knowledge, not accumulating research data to train large-language models and other AI tools. Data is not insight, but its prerequisite at best. And sharing data is a highly non-trivial concept resting on many implicit assumptions often not fulfilled.[3] We present a number of strategies and tools that, if used competently, will help us to improve the quality of our research and ultimately contribute to science and scholarship. First come tools for capturing all relevant information during data acquisition,[4] and electronic lab notebooks.[5] A framework for scientific data analysis that provides a gap-less and complete protocol of each step and relieves the user of actual programming [6] is a huge step forward. This is complemented by a larger (local) infrastructure consisting of persistent and unique IDs (PID, UID), a repository for "warm" research data, lab management, and knowledge base.[7] All these strategies and tools focus on the individual scientists, as only they can potentially ensure the urgently required quality of data and results that underpin scientific insight. Eventually, we need to teach [8, 9] young researchers early on what science is all about and why properly handling research data is a prerequisite for scholarly contribution. "At stake is the future of scholarship." [3]

References

- [1] I. Newton, letter to Robert Hooke, February 5th, 1676
- [2] R. P. Feynman, Cargo cult science, *Eng. Sci.* 1974, **37**(7), 10.
- [3] C. L. Borgman, *Big Data, Little Data, No Data*. MIT Press, Cambridge MA 2015.
- [4] B. Paulus, T. Biskup, Towards more reproducible and FAIRer research data: documenting provenance during data acquisition using the Infofile format, *Digit. Discov.* 2023, **2**, 234.
- [5] M. Schröder, T. Biskup, LabInform ELN: A lightweight and flexible electronic laboratory notebook for academic research based on the open-source software DokuWiki, *ChemRxiv* 2023, doi:10.26434/chemrxiv-2023-2tvct.
- [6] J. Popp, T. Biskup, ASpecD: A modular framework for the analysis of spectroscopic data focussing on reproducibility and good scientific practice, *Chem. Meth.* 2022, **2**, e202100097
- [7] T. Biskup, LabInform: A modular laboratory information system built from open source components, *ChemRxiv* 2022, doi:10.26434/chemrxiv-2022-vz360
- [8] <https://www.till-biskup.de/de/lehre/programmierkonzepte/>.
- [9] <https://www.till-biskup.de/de/lehre/forschungsdatenmanagement/>.