

Reproducibility in Geospatial Research

Position paper for ASU SPARC R&R workshop, February 2019

Daniel Nüst, Institute for Geoinformatics, University of Münster, <https://o2r.info>

The challenges of contemporary science led to widespread calls for more transparency and reproducibility (cf. e.g. Munafò et al., 2017). It is open for discussion whether science really faces a reproducibility crisis (cf. Baker, 2016 and Fanelli, 2018), but the challenges of computer-based research sparked numerous articles (e.g. Sandve et al., 2013, Wilson et al., 2017, Hardwicke et al., 2018, Gil et al., 2016). These technical challenges drive the development of infrastructures and tools to assist researchers in writing articles effectively and efficiently, publishing data and code underlying computations, and communicating their findings in a reproducible way (e.g. Jupyter et al., 2018, ReproZip, 2016). Extending on Claerbout's claim "*an article about a computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.*" (Donoho, 2010), authors, readers and reviewers need to learn to "package" and to consume more than just articles, even for their own benefit (Markowetz, 2015). This includes learning how to create and inspect **research compendia** (cf. research-compendium.science; Nüst, 2018) building on current technology, e.g. using containerisation or virtual environments. Recent studies showed that the geosciences and GIScience also have troubles with reproducibility (Konkol et al., 2018, Nüst et al., 2018, Ostermann & Granell, 2017). Hence, more and more journals adopt policies on sharing data and code, and the Association of Geographic Information Laboratories in Europe (AGILE) supports an initiative to develop new author and reviewer guidelines for reproducible conference publications.

Computational reproducibility, i.e. getting the same output data using the author's data and method/code, is one side of the coin. Replication, i.e. coming to the same conclusions using different tools or inputs, is the other one. It is arguably an even stronger goal. To achieve replicability, there need to be changes in **academic reward systems, scholarly institutions, publication practices**, and an open research culture (Nosek et al., 2015). The distinction between the terms should be discussed widely within the geospatial sciences to reach a consensus about their meaning and application in a community of practice. These discussions should be led in the light of prevailing issues of a common definition of reproducibility (Barba, 2018). Why not support the approach by Philip Stark (2018) who recently coined a new term, preproducibility, to embrace openness and helpfulness? "*An experiment or analysis is preproducible if it has been described in adequate detail for others to undertake it. Preproducibility is a prerequisite for reproducibility [...]*" (Stark, 2018).

Along with reproducibility/replicability often comes a promotion of **Open Science practices and Open Scholarship** on an individual level. Connecting the terms meanings' and open practices with the existing best practices in community will require critical yet open minds and readiness to change. One could say, geospatial sciences have been lucky not to have faced their own "crisis" so far. In my opinion, the potential impact to people's lives by geospatial analyses does not rank behind domains such as life sciences or psychology, where replication and reproducibility are more widely discussed. I see the lack of an easily usable open solution for processing of large remote sensing data (cf. openEO), a widespread use of proprietary software, and unspecific article submission guidelines as starting points for questioning the state of the art in GI Science/geosciences research. The preservation of computer-based research requires a higher awareness for our virtual laboratories. Scripted workflows using open source software are possible today, and they are crucial for successful future scientific work. Therefore students should be taught the tools and mindset of Open Science and teachers should set examples.

Scientific breakthroughs build on previous research. Open Science enables everyone to read (Open Access), contribute to (Open Peer Review), extend (Open Data, Open Source), improve (Open Methodology, preprints, pre-registration), and teach (Open Educational Resources) the latest findings. It is essential to answer societal challenges of today. Changes towards openness and transparency are sometimes driven from within a community of practice, but also imposed from the outside, e.g. by funders or journals. Outside requirements as well as the fast pace of technological progress pose challenges to researchers of all career stages who have to adapt their habits and workflows, and continuously update their toolboxes. It's not too late for the GI Science/geospatial sciences to react. Building upon a common understanding of terms,

organisations and domain leaders can promote an **open research culture**. This means, for example, senior researchers are open to question and change their own habits, and teach their students the skills they need in the future, but also to live by these standards working as editors, reviewers, or evaluation committee members. In Nüst et al. (2018), we present further steps for individuals, labs, institutes, and scholarly societies that help scientists to improve the reproducibility of their research. It would be most welcome if geoscience journals publish editorials on the issue of reproducibility to raise awareness. Journal editors as the leaders of research communities should reevaluate and consider updating author and reviewer guidelines towards replicable and reproducible research

References

- Baker, M.: 1,500 Scientists Lift the Lid on Reproducibility, *Nature*, 533(7604), 452–454, doi:10.1038/533452a, 2016.
- Barba, L. A.: Terminologies for Reproducible Research, arXiv:1802.03311 [Cs], February 9, 2018.
- David L. Donoho.: An invitation to reproducible computational research, *Biostatistics*, Volume 11, Issue 3, 1 July 2010, Pages 385–388, doi:10.1093/biostatistics/kxq028.
- Fanelli, D.: Opinion: Is Science Really Facing a Reproducibility Crisis, and Do We Need It To?, and do we need it to?, *Proceedings of the National Academy of Sciences*, 115(11), 2628–2631, doi:10.1073/pnas.1708272114, 2018.
- Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., Karlstrom, L., Lee, H., Mills, H. J., Oh, J.-H., Pierce, S. A., Pope, A., Tzeng, M. W., Villamizar, S. R. and Yu, X.: Towards the Geoscience Paper of the Future: Best Practices for Documenting and Sharing Research from Data to Software to Provenance, *Earth and Space Science*, 3(10), 388–415, doi:10.1002/2015ea000136, 2016.
- Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B. and Frank, M. C.: Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*, *Royal Society Open Science*, 5(8), 180448, doi:10.1098/rsos.180448, 2018.
- Konkol, M. and Kray, C.: In-depth examination of spatiotemporal figures in open reproducible research, *Cartography and Geographic Information Science*, 1–16, doi:10.1080/15230406.2018.1512421, 2018.
- Konkol, M., Kray, C. and Pfeiffer, M.: Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study, *International Journal of Geographical Information Science*, 1–22, doi:10.1080/13658816.2018.1508687, 2018.
- Markowetz, F.: Five selfish reasons to work reproducibly, *Genome Biology*, 16(1), doi:10.1186/s13059-015-0850-7, 2015.
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J. and Ioannidis, J. P. A.: A manifesto for reproducible science, *Nature Human Behaviour*, 1(1), 21, doi:10.1038/s41562-016-0021, 2017.
- Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R. and Cerutti, V.: Reproducible research and GIScience: an evaluation using AGILE conference papers, *PeerJ*, 6, e5072, doi:10.7717/peerj.5072, 2018.
- Nüst, D., Boettiger, C., and Marwick, B.: How to read a research compendium, arXiv:1806.09525 [cs.GL], June 11, 2018.
- Ostermann, F. O. and Granell, C.: Advancing Science with VGI: Reproducibility and Replicability of Recent Studies using VGI, *Transactions in GIS*, 21(2), 224–237, doi:10.1111/tgis.12195, 2017.
- ReproZip: Computational Reproducibility With Ease, F. Chirigati, R. Rampin, D. Shasha, and J. Freire. In *Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pp. 2085–2088, 2016
- Sandve, G. K., Nekrutenko, A., Taylor, J. and Hovig, E.: Ten Simple Rules for Reproducible Computational Research, edited by P. E. Bourne, *PLoS Computational Biology*, 9(10), e1003285, doi:10.1371/journal.pcbi.1003285, 2013.
- Stark, P. B.: Before reproducibility must come preproducibility, *Nature*, 557(7707), 613–613, doi:10.1038/d41586-018-05256-0, 2018.
- Jupyter, P., Bussonnier, M., Forde, J., Freeman, J., Granger, B., Head, T., Holdgraf, C., Kelley, K.,

- Nalvarte, G., Osheroﬀ, A., Pacer, M., Panda, Y., Perez, F., Ragan-Kelley, B. and Willing, C.: Binder 2.0 - Reproducible, interactive, sharable environments for science at scale, in Proceedings of the 17th Python in Science Conference, SciPy., 2018.
- Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L. and Teal, T. K.: Good enough practices in scientiﬁc computing, edited by F. Ouellette, PLOS Computational Biology, 13(6), e1005510, doi:10.1371/journal.pcbi.1005510, 2017.