



Packaging geoscience research for transparency and reproducibility: approaches and examples

Daniel Nüst

Institute for Geoinformatics (Opening Reproducible Research, o2r), University of Münster, Münster, Germany
(daniel.nuest@uni-muenster.de)

Scientific breakthroughs are built on previous research. With research relying increasingly on digital artefacts (data, software), the significance of these building blocks for science becomes more apparent. The metascience research field of Reproducible Research investigates the challenges and proposes practices for improving transparency, understandability, reviewability and ultimately also reusability of research outputs. In this work, we give an overview of the current state of the art in packaging research results.

The following concepts will be introduced to help geoscientists better select and apply the most suitable approach: (i) research compendia (<https://research-compendium.science/>), (ii) the Executable Research Compendium (ERC; Nüst et al., 2017a) and supporting infrastructure developed in the project Opening Reproducible Research (o2r, <https://o2r.info>), (iii) manual containerisation and virtual machines (Nüst et al., 2018a), (iv) containerising graphical user interfaces (x11docker), (v) virtual environments (cf. Marwick et al., 2018), and (vi) assisted containerisation (Chirigati et al., 2016; containerit; Jupyter et al., 2018). The approaches are illustrated with examples from geosciences and geographic information science: Geographic Object-based image analysis (GEOBIA; Knoth & Nüst, 2017), Citizen Science in environmental sensing (Nüst, 2018), geosimulation modelling (under development), metascience and surveys (Nüst et al., 2018b), and service-based environmental data science (Nüst & Schutzzeichel, 2016).

Putting your research in order ensures quality of work. It lays a foundation for collaborations and makes science's building blocks more accessible for reviewers, collaborators, interested readers, and students. The presented examples and concepts show that it is effective to adopt good habits, workflows and tools now to be prepared for future changes in policies of journals and funding agencies (cf. Editorial, 2018; Enserink, 2018).

References

Chirigati, F., Rampin, R., Shasha D., Freire, J., 2016. "ReproZip: Computational Reproducibility With Ease", In Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 2085-2088, 2016

containerit: <https://github.com/o2r-project/containerit/>

Editorial, 2018. "Does your code stand up to scrutiny?". *Nature*, 555(7695), 142–142. doi:10.1038/d41586-018-02741-4

Enserink, M., 2018. "European science funders ban grantees from publishing in paywalled journals". *Science*. doi:10.1126/science.aav3093

Jupyter et al., 2018. "Binder 2.0 - Reproducible, Interactive, Sharable Environments for Science at Scale." Proceedings of the 17th Python in Science Conference. doi:10.25080/Majora-4af1f417-011

Knoth, C., Nüst, D., 2017. "Reproducibility and Practical Adoption of GEOBIA with Open-Source Software in Docker Containers". *Remote Sens.* 9, no. 3: 290. doi:10.3390/rs903029

Marwick, B., Boettiger C., Mullen L., 2018. "Packaging data analytical work reproducibly using R (and friends)". *PeerJ Preprints* 6:e3192v2. doi:10.7287/peerj.preprints.3192v2

Nüst, D., Konkol, M., Schutzzeichel, M., Pebesma, E., Kray, C., Przibytzin, H. and Lorenz, J., 2017a. "Opening the Publication Process with Executable Research Compendia". D-Lib Mag. doi:10.1045/january2017-nuest

Nüst, D., Schutzzeichel, M., 2017b. "An Architecture for Reproducible Computational Geosciences". Zenodo. doi:10.5281/zenodo.1478543

Nüst, D., Boettiger, C., & Eddelbuettel, D., 2018a. "rocker/geospatial: a flexible runtime environment for geoscientific data analysis". Zenodo. doi:10.5281/zenodo.1216751

Nüst, D., 2018. "Open environmental data analysis with senseBox, openSenseMap, Jupyter Notebook, RStudio, and BinderHub". Zenodo. doi:10.5281/zenodo.1139929

Nüst, D., Granell C, Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., Cerutti, V., 2018b. "Reproducible research and GIScience: an evaluation using AGILE conference papers". PeerJ 6:e5072. doi:10.7717/peerj.5072

x11docker: <https://github.com/mviereck/x11docker/>