Opening Reproducible Research

Daniel Nüst (1), Markus Konkol (1), Edzer Pebesma (1), Christian Kray (1), Stephanie Klötgen (2), Marc Schutzeichel (2), Jörg Lorenz (2), Holger Przibytzin (2), and Dirk Kussmann (2)

(1) Institute for Geoinformatics, University of Münster, Münster, Germany (daniel.nuest@uni-muenster.de), (2) University and Regional Library (ULB), University of Münster, Münster, Germany

Open access is not only a form of publishing such that research papers become available to the large public free of charge, it also refers to a trend in science that the act of doing research becomes more open and transparent. When science transforms to open access we not only mean access to papers, research data being collected, or data being generated, but also access to the data used and the procedures carried out in the research paper.

Increasingly, scientific results are generated by numerical manipulation of data that were already collected, and may involve simulation experiments that are completely carried out computationally. Reproducibility of research findings, the ability to repeat experimental procedures and confirm previously found results, is at the heart of the scientific method (Pebesma, Nüst and Bivand, 2012). As opposed to the collection of experimental data in labs or nature, computational experiments lend themselves very well for reproduction. Some of the reasons why scientists do not publish data and computational procedures that allow reproduction will be hard to change, e.g. privacy concerns in the data, fear for embarrassment or of losing a competitive advantage. Others reasons however involve technical aspects, and include the lack of standard procedures to publish such information and the lack of benefits after publishing them. We aim to resolve these two technical aspects.

We propose a system that supports the evolution of scientific publications from static papers into dynamic, executable research documents. The DFG-funded experimental project *Opening Reproducible Research* (ORR) aims for the main aspects of open access, by improving the exchange of, by facilitating productive access to, and by simplifying reuse of research results that are published over the Internet. Central to the project is a new form for creating and providing research results, the *executable research compendium* (ERC), which not only enables third parties to reproduce the original research and hence recreate the original research results (figures, tables), but also facilitates interaction with them as well as their recombination with new data or methods. Building on existing open standards and software, this project develops standards and tools for ERCs, and will demonstrate and evaluate these, focusing on the geosciences domains. The project goes beyond a technical solution for ERCs by evaluating the system from the perspectives of geoscience researchers as participants in a scientific publication process. It will focus on the statistical environment R, but also evaluate larger run time systems captured in virtual environments (Docker containers). ERCs are built upon and integrate well with both established day-to-day workflows of digital research and the scientific publication process. They make research accessible on different levels at any stage to anyone via open web platforms. Other scientists can transfer a compendium of software and tools to their own local environment and collaborate, while others make minimal changes and compare changed results in a web browser.

Building on recent advances in mainstream IT, ORR envisions a new architecture for storing, executing and interacting with the original analysis environment alongside the corresponding research data and text. ORR bridges the gap between long-term archives, practical geoscience researchers, as well as publication media.

Consequently, the project team seeks input and feedback from researchers working with geospatial data to ensure usable and useful open access publications as well as a publication process that minimizes effort while maximizing usability and re-usability.

References

Pebesma, E., D. Nüst, R. Bivand, 2012. The R software environment in reproducible geoscientific research. *Eos*, Transactions American Geophysical Union 93, vol 16, p. http://dx.doi.org/10.1029/2012EO160003163-164.

Opening Reproducible Research project description and website:
https://www.uni-muenster.de/forschungaz/project/9520?lang=en