Publication and citation of scientific software with persistent identifiers

Martin Hammitzsch and Jens Klump

GFZ German Research Centre for Geosciences, CeGIT Centre for GeoInformation Technology, Potsdam, Germany
(martin.hammitzsch@gfz-potsdam.de)

Scientific progress depends increasingly on the use and development of software and the combination of diverse software components. Scientific software, as well as the underlying hardware platforms, also evolve quickly. But when using different versions of software and operating systems the processing of complex data can lead to significantly different results, and thus is in conflict with the principle of reproducibility in natural sciences. Incomplete documentation and ambiguous quality of the software hamper the understanding, reuse and the reproducibility of generated results. In addition, available software code repositories are developed for industrial and open source community needs and do not meet the demands of the scientific community for the sustainable use of software in sciences.

Furthermore, software development in general is not perceived as a scientific achievement, similar to the situation of research data years ago. However, the development of software accounts for an increasingly prominent space in research, especially in natural sciences software has become an indispensable commodity. In its complexity the software developed within research ranges from scripts for data processing, programs of increasing complexity to extensive program packages and system images, e.g. for use in cloud infrastructures. This software, its quality and its handling highly influence the quality of research results obtained and their traceability.

As a consequence of this, the scientific community is in active discussion on how to overcome the related problems, to find and implement solutions serving researchers' needs regarding software used in a scientific context. As for research data, answers to a variety of related questions and a common understanding of handling scientific software with defined processes have to be developed jointly. Amongst others, these processes have to cover issues regarding quality assurance, versioning and documentation, traceability, reproducibility and reusability, archiving and the use of persistent identifiers, metrics for evaluation and validation, measuring of productivity and impact as well as the dissemination and recognition of scientific achievement. Furthermore, open access and the use and interplay of software publication, data publication and traditional paper publication have to be considered.

As a contribution to improve the publication of scientific software, the project SciForge [1], funded by the German Research Foundation (DFG), addresses these issues. A network of interested groups and individuals in different research areas contributes to the project to recognize, create, and act upon opportunities for the development of concepts establishing defined processes and a reference platform. Established and working mechanisms, such as the Digital Object Identifier (DOI) based on the Handle System, will be an integral part of the concept. By this way established processes and existing frameworks are extended by new possibilities to publish software and recognize its scientific contributions and achievements.

[1] www.gfz-potsdam.de/sciforge