Using Docker in environmental research

Alexander Dolich, Mirko Mälicke, Ashish Manoj J, Jan Wienhöfer, and Erwin Zehe
Karlsruhe Institute of Technology, Institute of Water and River Basin Management, Chair of Hydrology, Germany
(alexander.dolich@kit.edu)

The virtual research environment V-FOR-WaTer provides functionalities to store and access hydrological and other environmental data from various sources and disciplines. We propose a framework to run containerized tools within the V-FOR-WaTer toolbox, that is intended to solve the problem of combining software or scripts developed in different programming languages.

The framework is used to manage Docker containers, which can contain software like tools for data analysis or environmental modeling. Alongside the well-known advantages of containerization, such as development speed and efficiency, isolation from the local system, dependency management and portability, the usage of containers also ensures a high degree of reproducibility.

Given a scientific context, containers are especially useful to combine scripts in different languages following different development paradigms. To do so, we developed a framework-agnostic container specification which standardizes inputs and outputs from and to containers to ease the development of new tools. As of now we also provide templates for tools developed in Python, R, Octave and NodeJS.

We present an exemplary workflow for the CATFLOW hydrological model. Data from the V-FOR-WaTer environment is loaded using a Python tool and preprocessed with an existing R tool. After running the FORTRAN model, existing tools in Python, R and MATLAB are used for analysis and presentation of results. When executing the workflow, the user does not need to be familiar with the different programming languages of individual tools, since the containerized tools are self-contained by definition.