Using Docker for reproducible workflows

Mirko Mälicke, Alexander Dolich, Ashish Manoj Jaseetha, Balazs Bischof, and Lucas Reid
Karlsruhe Institute of Technology (KIT), Institute of Water and Environment, Hydrology, Karlsruhe, Germany
(mirko.maelicke@kit.edu)

We propose a framework-agnostic specification for contextualizing Docker containers in environmental research. Given a scientific context, containers are especially useful to combine scripts in different languages following different development paradigms.

The specification standardizes inputs and outputs from and to containers to ease the development of new tools, retrace results and add a provenance context to scientific workflows. As of now we also provide templates for the implementation of new tools developed in Python, R, Octave and NodeJS, two different server applications to run the containers in a local or remote setting and a Python client to seamlessly include containers into existing workflows. A flutter template is in development, which can be used as a basis to build use-case specific applications for Windows, Linux, Mac, the Web, Android and iOS.

We present the specification itself, with a focus on ways of contributing, to align the specification with as many geoscientific use-cases as possible in the future. In addition a few insights into current implementations are given, namely the role of the compliant pre-processing tools in the generation of the CAMELS-DE dataset, as well as result presentation for a Machine learning application for predicting soil moisture. Both applications are presented at EGU as well. We use these examples to demonstrate how the framework can increase the reproducibility of associated workflows.