

Reproducing the computational environments of experiments

September 6 2017



A view from inside the MDC's data center, with racks of high performance computers. Credit: MDC

Experiments increasingly rely on high-performance computing software. Differences in software environments can cause problems when those experiments need to be reproduced—so scientists at the MDC in Berlin are seeking a solution.

Reproducing experiments and results is a cornerstone of science, but researchers acknowledge that actually achieving this feat can be tricky. Specific experimental setups are usually the result of a lab's painstaking work, and are increasingly expensive in today's [environment](#) of high-throughput methods. The fact that complex, customized sets of [software](#) are frequently involved in the analysis and interpretation of data makes it even more difficult to achieve true reproducibility.

Guix—a free program that is used to fully reproduce computational environments—might be part of the solution, says Ludovic Courtès of Inria of the French National Institute for computer science and applied mathematics in Bordeaux. To implement it, he has collaborated with Ricardo Wurmus of the platform for bioinformatics and modeling at the the MDC's Berlin Institute of Medical Systems Biology (BIMSB), scientists from the Utrecht University Medical Center and a growing group of international colleagues.

Capturing complete computational environments

Science authorities insist that researchers share source code and support reproducibility. "The ability to reproduce an experiment depends—among other things—on the ability to reproduce the [software environment](#)," Courtès says. "That poses particular difficulties in the

many cases requiring high-performance computing (HPC) environments."

Guix is an outgrowth of a project called GNU launched almost 40 years ago at MIT in the U.S. It makes up for some deficits of earlier efforts and addresses several challenges: Users are no longer dependent on software package management by system administrators, empowering them to fully customize the environment to their needs. It also solves problems that arise when scientists draw on "container solutions," which Courtès compares to receiving a brand-new computer with everything pre-installed. "That works until you make a small modification in the experiment to test a new hypothesis—which often happens in the world of research."

The advantage of Guix is that it characterizes software environments in unambiguous terms, similar to a mathematical function. It completely describes all its relations and thus can reproduce them bit-for-bit. This way, Guix facilitates both reproducibility and customizability.

Adapting Guix to scientists' needs

Guix was not originally designed for the high-performance computing environments required by today's experiments. So scientists at the MDC, Inria and the partner institutes are building functions that permit Guix to be used on a computing cluster, to implement reproducible workflows. They are also adding packages that were developed at each site.

"Before Guix, the installation of scientific software was necessarily ad-hoc," Wurmus says. "Groups would build their own software, statically link it into existing systems, and hope that it would never have to change—because managing software environments was virtually impossible. Now not only can we manage a single environment per group in a reliable fashion, but we use Guix at all levels: of the group, user,

workflow and so on."

The project is scheduled to last two years, at which time its initiators hope to have met the software reproducibility needs of their institutions. "The wider objective," Courtès says, "is to convince others who rely on high-performance computing that Guix represents a major advance toward a fundamental goal in science."

More information: guix-hpc.bordeaux.inria.fr/

Provided by Max Delbrück Center for Molecular Medicine

Citation: Reproducing the computational environments of experiments (2017, September 6)
retrieved 25 April 2024 from <https://phys.org/news/2017-09-environments.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.