

A carrot, not a stick

Konrad Hinsen konrad.hinsen@cnr.fr

Centre de Biophysique Moléculaire (UPR4301 CNRS), Rue Charles Sadron, 45071 Orléans, France

Your editorial "A problem shared is a problem halved" raises several important and difficult questions concerning the publication of digital scientific knowledge. I would like to comment on one specific aspect: sharing scientific software.

Most scientists view software primarily as a tool for doing research, and from this point of view it is reasonable to consider it a potentially commercial product, much like lab equipment. However, contrary to lab equipment, scientific software is also executable theory. With few exceptions, scientific software implements scientific models, and these models are often not fully documented anywhere else than in the software's source code [1]. Thus, using proprietary software sometimes means applying unpublished models, a practice that is in contradiction with the principles of scientific research.

Another important aspect that is often overlooked is that observations – experiments – and models – theory – do not have the same status in science. Observations provide inherently imperfect information about nature. Models are mental constructions that are perfectly well known, even though their relation to reality is inherently approximate. Laboratory equipment is part of the observation branch. Like samples, it can never be replicated exactly, and only be known approximately. Two different labs never have the exact same equipment, and therefore it is not a fundamental problem that some details of its construction are commercial secrets. Models, on the other hand, lose much of their power if only approximate descriptions are available. This is especially true for complex computational models, in which small changes can lead to unforeseeably large differences in predictions.

There are numerous fields where computational research is primarily performed using community-developed Open Source software, which shows that science can function without commercial software. It is also worth pointing out that there is no problem with delegating the development and maintenance of Open Source software to software professionals, as long as they are paid for providing services rather than for licensing a product. My conclusion is that we should ban commercial software from scientific research.

Compared to this fundamental question, the issue of documenting and preserving software environments and library dependencies is a minor technical issue. The problems are well understood and we know how to solve them, given sufficient resources. However, until now, the incentives for doing computational work transparently have been insufficient to ensure the allocation of sufficient resources to improving our software environments. If scientific journals slowly but steadily increase their requirements for the reproducibility of computational results, this will ultimately happen.

[1] K. Hinsen. Computational science: shifting the focus from tools to models [version 2; referees: 2 approved]. *F1000Research* 2014, 3:101 <https://doi.org/10.12688/f1000research.3978.2>