



Standing on the shoulders of software developers

Last week's SciencePAD workshop, held at CERN, examined ways to ensure scientific software stays citable and sustainable over the long term.

Left



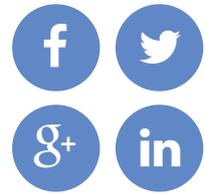
Alberto di Meglio, project director of the European Middleware Initiative and leader of its SciencePAD activities.

Posted on FEB 6
2013 11:57AM



Andrew Purcell
European editor

Share this story

[↻ Republish](#)

Tags

uncared for, software decays. Like a grand, old building, it may fall into ruin if abandoned. But, as with any treasured monument, there are those who will fight to preserve scientific software; those who seek to ensure that the hard-won gains of today's researchers are not frivolously lost for the researchers of tomorrow.

On Wednesday 30 January, around 40 such 'e-preservationists' converged on CERN for [the SciencePAD Persistent Identifier's Workshop \(SPID2013\)](#), the aim of which was to investigate ways of improving collection, storage, and preservation of information concerning the software used in scientific research. Of course, the motivation for doing so is not mere nostalgia, but a desire to ensure that software developed can be reused by researchers in the future and that the scientific results generated through the use of specific software remain reproducible for years to come.

Alberto di Meglio, project director of [the European Middleware Initiative](#) and leader of its [SciencePAD activities](#), stresses the importance of online software repositories in achieving these goals. "It is important that the software used to help produce scientific publications is properly cited," he says. "The ultimate goal is for scientists to be able to locate software in a repository and get citation information which they can use in their own publications."

"One solution is to write a publication about the software you've developed," suggests Martin Fenner of [ORCID](#). This way, he argues, researchers would be able to cite software used for research within existing academic publishing structures. Another approach would be to tag software with persistent identifiers, long-lasting reference codes - akin to the digital object identifiers (DOIs) assigned to research papers themselves.

However, this isn't just about software developers trying to get the recognition they feel they deserve; rather, it's about ensuring that research groups don't duplicate one another's efforts and that scientists are able to successfully build on the work of their peers. Rudolf Dimper of the [European Synchrotron Radiation Facility](#) says that online software repositories have a vital role to play in building the trust necessary to achieve these goals: "If software is well maintained and it is well documented, researchers then actually trust these programs to do their own data analysis and will be less tempted to redesign similar or even identical programs themselves."

Neil Chue Hong, director of the [Software Sustainability Institute](#), says that many of the issues discussed at the workshop could be solved by providing scientists with a better understanding of computational science. "We need to work on skills and capabilities," he says. "A lot of researchers do not have the basics that they need to know about computational science in the same way that we

hope all scientists have been taught the basics of statistics."

Inspired by [Tim Berners-Lee's 5 stars of linked, open data](#), Chue Hong goes on to propose a 5-star ratings scheme for software. His categories are as follows:

* **Existence:** there is accurate metadata that defines the software

** **Availability:** the software can be accessed and run

*** **Openness:** the software has an open, permissible license

**** **Linked:** related data, dependencies and papers are referenced

***** **Assured:** the software provides ways of demonstrating "correctness"

However, even putting a system such as this in place isn't enough; the system of incentives for researchers needs to be fundamentally changed, suggests Chue Hong. He cites work by the [ImpactStory](#) organization to incorporate non-traditional outputs, such as developing scientific software, into measuring academic reputation and impact. "One of the things they've recently developed is a way of linking to researchers' [GitHub](#) profiles," says Chue Hong. "This allows you to create different types of alternative metrics about your impact on the community." (GitHub is an online hosting service for software development

projects - a sort of social networking site for programmers who want to share code.)

Björn Brembs of **DataCite** agrees with Chue Hong on the need to broaden the range of activities deemed important in influencing the academic reputation of researchers. "Today, we produce three things and only one of these is publications," he says. "In principle, we produce software (both to generate and validate the data), data, and publications." He argues that all three of these things ought to be taken into account when academic institutions judge the 'success' of their researchers.

Brembs goes on to say that the current state of affairs in the academic research community, with regards to software sustainability, is a mess. "There's no software archiving to speak of," he says. "It's like the web in 1995."

You can read the original version of this article on the CERN website, [here](#).

**Join the
conversation**

Contribute



Do you have story ideas or something to contribute? **Let us know!**

OUR UNDERWRITERS

Thank to you our underwriters, who have supported us since the transition from International Science Grid This Week (iSGTW) into Science Node in 2015. We are incredibly grateful.

[View all underwriters](#)

CATEGORIES

[Advanced computing](#)
[Research networks](#)
[Big data](#)
[Tech trends](#)
[Community building](#)

CONTACT

Science Node

Email:

editors@sciencenode.org

Website:

sciencenode.org



Copyright © 2022 Science Node™ | [Privacy Notice](#) | [Sitemap](#)

Disclaimer: While Science Node™ does its best to provide complete and up-to-date information, it does not warrant that the information is error-free and disclaims all liability with respect to results from the use of the information.