Apache Taverna: Sustaining research software at the Apache Software Foundation

Document Version
Final published version

Citation for published version (APA):
Soiland-Reyes, S., Dunlop, I., Williams, A. R., & Apache Taverna Team (2015). *Apache Taverna: Sustaining research software at the Apache Software Foundation*. Abstract from Bioinformatics Open Source Conference, Dublin, Ireland. https://github.com/stain/bosc2015-apachetaverna/blob/master/soilandreyes_bosc.md

OPEN ACCESS

Title	Apache Taverna: Sustaining research software at the Apache Software Foundation
Authors	*Stian Soiland-Reyes*, Ian Dunlop, Alan R Williams, Apache Taverna team
Affiliation	Apache Software Foundation; eScience Lab, University of Manchester
Contact	http://taverna.incubator.apache.org/contact
URL	http://taverna.incubator.apache.org/
License	Apache License, Version 2.0

Apache Taverna is a *scientific workflow* system, encompassing a graphical workbench, command line tools, server and APIs. Although Taverna was conceived for bioinformatics, its user base also encompasses domains such as astronomy, digital preservation, biodiversity and virtual physiology. Taverna has been an open source project since 2003, developed by the myGrid consortium and originally led by the University of Manchester and EMBL-EBI. In October 2014, Taverna became an incubating project at the Apache Software Foundation.

Here we describe our motivation for changing the *governance* and *ownership* of the Taverna project, and reflect on our experience and challenges in transitioning a University-led research software activity to an open development process and building a wider developer community.

Research Software supports scientists and researchers; its development is usually funded for domain-specific projects and made publicly available as Open Source through code repositories (GitHub, BitBucket). By the time research software develops into a mature code base and gains a diverse user base, the initial funding and related projects may already be finished, yet ownership and control of the project typically remains with the original authors. Users and third-party developers have the *source code*, but are often not included in project decisions, and may not feel *ownership* to contribute code, documentation or support for others.

The original developers may eventually become involved in new projects that do not directly relate to the original project, and may lose focus as the user base changes. Thus it becomes critical to grow a sustainable and diverse *developer community*, and to build an *open governance* model that encourages engagement and commitment from everyone using the software. Efforts like the Software Sustainability Institute are crucial, as it helps guide research software engineers in best practices for making their research software open and maintainable. Ultimately the success of an open source project should lead to a change of its structure and management to widen its developer base.

For Taverna, we considered several options to reduce the lead role and responsibility that the University of Manchester had, and to move to a neutral ownership model where any interested developers could contribute to Taverna's development on an equal standing. One of the options was to create a Taverna Foundation, but this would have required legal administration and a dedicated budget. Another option was to assign copyright and management of the code to a well-established software foundation like the Software Freedom Conservancy, the GNU project at the Free Software Foundation, the Eclipse Foundation, the Outercurve Foundation which is backed by Microsoft, and the Apache Software Foundation. The different options come with implications for the project's way of working, licensing, community, politics, infrastructure and public impression - which must be evaluated for the particular research software project.

By choosing Apache as Taverna's new home, we put emphasis on the community building, a strong and neutral governance model, and clear intellectual property management, which we believe makes the project more approachable for new participants and commercial entities. While the initial months as an Apache Incubator involved a fair bit of administrative overhead, such as moving mailing lists, transitioning web sites and bug trackers, checking dependency licensing and verifying intellectual property ownership of the donated source code, we feel this effort is offset already by an increase in awareness and engagement in the community at large. Several new developers have been attracted to the project, with many proposals and new ideas, including four Google Summer of Code 2015 student applications.