Opensource tools for Encouraging Sustainability in Diverse Multi-Library Projects

Stephen Thompson[1], Tom Couch[2], Thomas Dowrick[1], Matthew J Clarkson[1]

1 – Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), UCL

2 – Research Software Development Group, UCL

Project Vision

Modern software management tools enable the development of highly modular software libraries, allowing individual researchers and research software engineers to create compact libraries that can be reused in multiple projects by multiple users. This modularity results in researchers and developers working across many libraries, making it extremely hard to keep track of the sustainability of each library. Our vision is to develop a set a set of open-source tools to probe the sustainability of families of software libraries, enabling constant monitoring via an automated sustainability dashboard.

We will apply this to our own SciKit-Surgery family of libraries, however the project will develop tools that could be used by anyone to create a sustainability dashboard for a selected group of libraries. This could be applied in an organisation like UCL to add sustainability metrics to the UCL Research Software Dashboard enabling the recognition and celebration of software that meets the highest standards of sustainability, helping to embed sustainability across the community.

More generally, the tools will be available to any community, enabling better understanding of what exactly makes software sustainable.

Background

The SciKit-Surgery libraries are under active development by researchers within and beyond WEISS. From the outset we have attempted to follow best practice on software sustainability to create software that will be adopted and sustained by the research community. A key strategy has been to break the functionality up into a set of loosely coupled Python libraries that can be used independently of each other or in combination. This keeps the size of each library manageable for an individual researcher [1].

To date we have published over 35 libraries with contributions from at least 10 researchers within and beyond UCL. A key challenge for us now is how to build on the high level of sustainable software practice as the number of libraries grows. We began to address this in September 2021 by creating a which we won best poster for at SeptembRSE. The intent was to summarise the existing SciKit-Surgery libraries along with current values of metrics relevant to software sustainability. Metrics included code coverage, documentation, number of contributors, lines of code, and metrics from third party companies Code Climate and snyk.

This proposal will build on the implementation and deployment of the dashboard so that it becomes a useful resource for our community. Key issues to be addressed will include; research into what metrics are useful for monitoring software sustainability; automation of the deployment; application of the dashboard to existing projects to determine if there is any relationship between a libraries' sustainability and their adoption by the community.

We have engaged with the UK RSE community to determine whether there is a wider need for these tools, what metrics are of value, and what if any existing solutions exist. We found some examples of dashboards that pull together some relevant metrics, however they lacked the automation and generalisation that we are proposing. There was a clear need for dashboards of this type and some debate about the most relevant metrics, indicating that our proposed work is relevant and of value to the community.

The funding will be used to pay for the equivalent of 25 days of software engineering time, through UCL's Research Software Development Group. The software engineer will be responsible for researching what architecture should be used for the dashboard, what metrics should go on the dashboard and for deploying of the dashboard. WEISS will match the funding to supply a researcher who will examine how the chosen metrics correlate with sustainability. The researcher and software engineer will work jointly on a publication describing the open-source software developed and discussing software sustainability more generally.

Project Outline
We are asking for funding to support a software developer (Grade 8, SP 38) for 25 days (We are happy to have a flexible working pattern, e.g. it could be 1 day a week for 25 weeks, or full time for 5 weeks).
Days 1-5: Familiarisation with SciKit-Surgery. Further consultation with RSE community on dashboard requirements. Determine what information should go on the dashboard. Research similar distributed projects, and appropriate architectures for software dashboards,
Days 6-15: (Re)Implement software dashboard and set up hosted runner for regular updates of dashboard.
Days 16-25: Analyse SciKit-Surgery. Investigate sustainability of at least one other open-source community driven project (e.g. SciKit). Prepare publication for dissemination of results.

Summary of how the Project Meets the Funding Guidelines
Community Driven: We have already consulted with the UK RSE community and found an unmet need for our proposed work. We will continue to work with members of the RSE community and researchers to develop the architecture and metrics necessary for the global community research software community. The lessons we learn will be used to drive forward the sustainability of our SciKit-Surgery libraries. The SciKit-Surgery libraries themselves have a growing community and evidence of use and contribution beyond UCL.
Embedding Sustainability: It is not possible to embed new practices and improve something unless you can quantify and monitor it. The main output of this project will be a toolkit to continuously monitor sustainability across software projects. The toolkit will enable the identification of best practice and the development of an evidence base to demonstrate the importance of best practice on research impact.
Ability to Deliver: The project team brings over 30 years of experience in research software engineering and research, across industry and academia. The team has published well over 200 research papers and software libraries. We have carefully scoped the technical and research outputs to ensure that they are deliverable in the project timeline.
Benefit to UCL: This project will increase the impact and reputation of research at UCL and worldwide, by giving every software project the ability to measure improve sustainability. The project expands on the existing research software dashboard to enable the identification of projects following best practice. These projects can be used as exemplars to help embed software sustainability across UCL, increasing the impact of UCL's research outputs. The open-source software tools developed will also be UCL branded, increasing UCL's visibility and impact in research software development

1. Thompson, S., Dowrick, T., Ahmad, M. *et al.* SciKit-Surgery: compact libraries for surgical navigation. *Int J CARS* **15,** 1075–1084 (2020). https://doi.org/10.1007/s11548-020-02180-5