

Department of Earth Sciences
Institute of Geophysics
ETH Zürich NO H23
Sonneggstrasse 5
8092 Zürich
Switzerland

October 6, 2016

Dear rOpenSci Project,

I am excited to apply for the position of Postdoctoral Scholar, focusing on open source software development to promote reproducible research.

My recent work has given me great exposure to the worlds of scientific computing, high performance computing, scalable algorithms, and “modern” (that is, highly hierarchical and/or heterogeneous) compute architecture. Clearly these fields, and in general the ascendant “third pillar” that is computation, hold great promise. However, the rapidly-changing nature of the fields, combined with the pressures of academic reality, which prioritize quantity and novelty over quality, have conspired to produce a scientifically unsatisfactory situation. Studies involving complex computing machinery, particularly that which is expensive or cumbersome to use, are often extremely difficult to interpret, reproduce, or extend. My attached proposal details some ideas to try to address these issues.

The attached example of my work, “Pipelined, Flexible Krylov Subspace Methods,” relates to some of these issues. It concerns algorithms whose usefulness is only directly demonstrable on a supercomputer (in this case, Piz Daint, a cray XC30 at the Swiss National Supercomputing Center). It relies on the behavior of a complex, largely proprietary communication stack, and as such results can only be interpreted by running a statistically meaningful number of experiments. To generate the results in the paper, I spent a large amount of time writing custom scripts to launch and analyze jobs, going back and forth to try to benchmark the MPI-3 features in the cray libraries, and running experiments on thousands of cores. While this was probably worth my time trying to write such a paper, this would be an undue burden to impose on someone doing applications research, or even algorithms research not centered directly involving these performance issues.

My interest in open source scientific software began in earnest during my internship with the PETSc team at Argonne National Lab. I have used and contributed to that package ever since. I am very interested in the “full stack” aspects of computational science; I consider myself to be a very flexible and creative person, and working across disciplines is one of the most appealing aspects of my job. I have greatly enjoyed attending linear algebra, earth sciences, and

supercomputing conferences in the last couple of years, and believe that one of my greatest strengths is my potential to collaborate with diverse scientific communities (and implement our ideas in code). I have a good deal of teaching experience (I was a TA for applied mathematics courses for most terms during my PhD, and have designed and taught two courses and several lectures since then¹). While I have not used the Open Science Framework, I am very comfortable with software design and the tools involved; in particular, I use git daily and have given lectures on its usage.

I look forward to any further communication regarding this position. I have increasingly come to believe that the interests of science would be greatly advanced with investment in the tools of reproducible research (and I am glad to see that the Supercomputing conference and others are also making this more of a priority). The introduction of extremely robust, standardized, and transparent-to-use tools is a way to bring scientific rigor to the communities without imposing undue burden on already-busy people.

Sincerely,

A handwritten signature in black ink, consisting of a series of fluid, connected loops and strokes.

Patrick Sanan

¹patricksanan.com/teaching