Aspiration for Open Science in Accelerator & Beam Physics Modeling

(Letter of Interest to Snowmass21, Computational & Accelerator Frontiers)

A. Huebl*[1], J.-L. Vay[1], R. Lehe[1], C. Mayes[2], Y.-D. Tsai[3], A. Friedman[4], M. Thévenet[5], H. Vincenti[6], D. L. Bruhwiler[7], A. Sauers[3], N. M. Cook[7], S. J. Gessner[2], M. Bussmann[8,9], D. P. Grote[4], R. Schmitz[10], B. Cros[11], D. Sagan[12], S. R. Yoffe[13], A. L. Edelen[2], A. Ferran Pousa[5], G. Stark[14]

[1]*Lawrence Berkeley National Laboratory, Berkeley, California 94720 USA*
[2]*SLAC National Accelerator Laboratory, Menlo Park, California 94025 USA*
[3]*Fermi National Accelerator Laboratory (Fermilab), Batavia, Illinois 60510 USA*
[4]*Lawrence Livermore National Laboratory, Livermore, California 94550 USA*
[5]*Deutsches Elektronen Synchrotron (DESY), Hamburg, Hamburg 22607 Germany*
[6]*LIDYL, CEA-Université Paris-Saclay, CEA Saclay, 91 191 Gif-sur-Yvette, France*
[7]*RadiaSoft LLC, Boulder, Colorado 80301 USA*
[8]*Center for Advanced Systems Understanding (CASUS), Görlitz, Saxony 02826 Germany*
[9]*Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Dresden, Saxony 01328 Germany*
[10]*University of California, Santa Barbara, California 93106 USA*
[11]*CNRS, Université Paris Saclay, Orsay, 91400 France*
[12]*Cornell University, Ithaca, New York 14850 USA*
[13]*SUPA and University of Strathclyde, Glasgow G4 0NG, United Kingdom*
[14]*SCIPP, UC Santa Cruz, California, USA*

August 2020

Abstract

This LOI is a contribution to an ongoing discussion in the accelerator and beam physics modeling community. Unlike other large communities in the HEP field, groups in accelerator and beam physics modeling have not yet arrived at a consensus for open source as part of an overall open science approach for essential modeling tools used by the community. This document re-iterates the open science principles, and especially clarifies some misunderstandings with respect to open source practices. We further exemplify success stories in the community, industry and funding principles in and outside the US.

*axelhuebl@lbl.gov

1 Consensus and Status

Research roadmaps in accelerator and beam physics are funded with billions of dollars, with decade-long contributions from design to final experiments. Accelerator and beam modeling software is developed and advanced to tackle urgent and often highly specific scientific problems with advanced modeling capabilities. These software tools already enable a quick description of new problem sets and increase scientific productivity. Nonetheless, models for problems sets and the implementations that solve them need continuous effort to be extended and improved, so one can address more complex physics problems that arise internally in the research field. Over the course of these multi-decadal efforts, adoption to external changes are also necessary, such as rapidly evolving standards in computing, which are driven by industry and computer science partners [1, 2], both with regards to programming languages and computing hardware.

Accelerator and beam modeling software needs to be reliable and extensively tested in a way that allows advancing a project's functionality without fear to break complex, once-validated scenarios [3]. Performance and interoperability must be constantly improved to increase understanding (multi-physics problems) and optimization (machine learning). Last but not least, one must ensure a transfer of knowledge over generations of scientists in form of formal education and easy accessibility to the tools.

Communities in the computational sciences face similar challenges, and physics modeling groups would benefit from incorporating research results, best practices, participate actively in aforementioned bodies, and anticipate trends in the broader computer science and computational science community. In sustainable workflows, one wants to avoid heroic efforts relying on few individuals and provide an environment that is inclusive and thrives with contributions from various educational backgrounds. Physics groups that integrate early with external computer science, applied mathematics and computational physics efforts can share modular solutions [4], drive cross-domain visions and avoid missing out on solutions developed in related scientific domains.

Importantly, it is also consensus within the computer science and majority of computational science subdomains that computational results need to be reproducible and independently replicable [5]. Although performance metrics of scientific success often aim solely on publication numbers, a healthy scientific process actively fosters reproduction of scientific results and adoption/modification of successful scenarios for further studies.

For seamless integration, both University/lab groups and commercial providers build extensively on open source frameworks. Yet, when it comes to the implementation of physics algorithms in applications, one finds:

> *Computer simulations play an indispensable role in all accelerator areas. Currently, there are many simulation programs used for accelerator physics. There is, however, very little coordination and cooperation among the developers of these codes. Moreover there is very little effort currently being made to make these codes generally available to the accelerator community [...]*
>
> — HEPAP report*, 2015*

There are notable improvements in recent years, such as communities around novel, open source particle-in-cell codes for laser-plasma modeling and collaborations on standards [6]. Yet other projects are essentially walled gardens with varying access levels to simulation programs, their source code, documentation, support and usage rights. Challenges are also notable in computational publications that omit to openly archive analysis routines, source codes, inputs, and simulation data in sufficient detail to aid reproducibility and adoptions. This situation poses a significant risk and calls for an advancement of modeling practices that can adequately address the needs of decade-long basic science projects.

2 A full open science approach as part of the solution

As of today, many researchers in computer science and computational physics[1] arrive at the conclusion that the principles of open science are a preferable base to start addressing the aforementioned needs. Public funding agencies, including the National Science Foundation, the European Commission, and the Department of Energy,[2] often recommend those principles for a multitude of reasons [8], from sustainability of the scientific process itself, ensuring independent reproducibility as part of the scientific method, using public money for producing public scientific knowledge, and fostering collaborations without custom-made legal agreements in dynamic national and international collaborations.

[1] as part of the computational sciences (computational physics, computational biology, computational sociology, ...)
[2] recommended unless exceptions can be justified, e.g. for export controlled parts that can go in separate modules [4, 7]

Open science principles are well-defined [9, 10], by publishing open educational resources, open data, libre/open source codes [11, 12], developing and applying open methodology, open peer review, and providing documents via open access [13]. Note that all of the above guarantee freedoms of unlimited usage, adoption and redistribution of scientific knowledge (with author credit), among others. Simulation source codes in particular are numerical implementations of physical models essential to the modeling of complex problems and are as such as essential to modeling as the fundamental system of equations that describes a physical process. Open source software can also be commercialized, which is in fact one of the essential freedoms defined in open source.

Using open principles to build the basis for a field opens a way to truly independently scrutinize, understand and advance methods and re-use each others indispensable research results, which naturally includes scientific tools/codes, data, literature, etc. as part of scientific communications. Using open licenses provides clarity and introduces low organizational overhead, when compared with custom licensing approaches, and provides an efficient basis for reusability and sustainability. Yet, open modeling tools themselves are not self-developing, self-maintaining, self-modernizing, self-extending, self-supporting and the field needs dedicated funding and people for these tasks [14] - as a recent paper was titled *Open is not Enough* [15].

Discontinuations in modeling software are often caused by the challenge to timely adapt, transfer project knowledge, responsibilities and visibility/recognition before the original authors finish their PhD/change jobs/retire. This is a challenge in academic culture that can be addressed with open practices and open education [8]. It is important for the community of accelerator and beam physicists to continue developing computational tools that are sufficiently robust, benchmarked, tested, validated, documented and user-friendly that they can be used efficiently by researchers beyond the developers and their immediate collaborators. Ensuring sustainability of our research, we propose to focus further on the creation of open education and open documentation resources to explain methods, essential numerical parameters, and sometimes even empirically known-to-work constraints of the actual simulation frameworks used in daily research.

It is beneficial to step outside ones immediate expert group and open modeling software up for other audiences, explain them to the non-experts of the domain, make source code archives and their variations citable (and authors attributable), embrace open dissemination and adoption to new use-cases, and foster contributions (open and traceable). This trend has been even accelerated with the advance of machine learning and Exascale computing efforts in recent years, in which accelerator and beam modeling applications are not the final software product, but are integrated into further frameworks [16].

3 Advances Needed and Success Stories

There are numerous modeling success stories in academia as well as industry that embrace the open science and open source paradigms. In our domain, active communities formed organically such as SciKit-HEP, ROOT, Geant4, openPMD, LHC's open data portals [17, 18], novel plasma accelerator modeling codes, and many more, which adopt open science principles for better accelerator, detector, experimental data analysis, and HEP science as a whole. It was also open source codes in our community that spear-headed the adoption of modern architectures such as GPUs for production simulations by a significant time margin. Looking beyond our immediate HEP community, accelerator and beam modeling shares already many open source project dependencies with other computational science fields for visualization, user-interfaces, hardware acceleration, compilers, I/O, data science, visualization, software management, mathematical solvers, etc. Open source development is fundamental in Exascale and machine learning efforts, which produce cutting-edge computer science libraries that accelerator and beam modeling experts can readily reuse.

Industry can engage in scientific open source activities, even if it may be perceived as hindering a traditional software business model: synergistic financing through selling of many copies. Furthermore, companies use open source software extensively in their own products and one can list many successful commercial approaches related to scientific open source software [19]. Yet there are also indications that some central HPC accelerator and beam modeling tasks have little public demand outside of Academia [20].

In conclusion, for computational modeling of accelerators and beams, the codes that are used to produce results, the education & documentation that is needed to use them [14], their input and output data, analysis scripts, publications and presentations, and communications with users became an essential part of the scientific method. Further adoption of open science and open source practices by the community of accelerator and beam modeling can elevate complexities of the computational approach to science and should be strongly encouraged [8], except for limited and well-justified reasons [7].

References

[1] ISO working group JTC1/SC22/WG21. "ISO C++". In: *The International Organization for Standardization (ISO)* (1998). URL: https://isocpp.org.

[2] The Khronos Group industry consortium. "Open standards for graphics, media and parallel computation". In: *Khronos Group, Inc.* (2006). URL: https://www.khronos.org.

[3] Rémi Lehe et al. "Embracing modern software tools and user-friendly practices, when distributing scientific codes". In: *Snowmass21 LOI* (2020).

[4] Jean-Luc Vay et al. "A modular community ecosystem for multiphysics particle accelerator modeling and design". In: *Snowmass21 LOI* (2020).

[5] *REANA Reproducible research data analysis platform.* URL: http://reanahub.io.

[6] Axel Huebl et al. "Develop/integrate data standards and start-to-end workflows for Accelerator Physics". In: *Snowmass21 LOI* (2020).

[7] Ian Lee. *LLNL: Ambassadors of Code.* 2018. URL: https://str.llnl.gov/2018-01/lee.

[8] Cyberinfrastructure Framework for the 21st Century. "A Vision and Strategy for Software for Science, Engineering and Education". In: *National Science Foundation* (2012). URL: https://www.nsf.gov/pubs/2012/nsf12113/nsf12113.pdf.

[9] United Nations Educational, Scientific and Cultural Organization (UNESCO). *Open Science Movement.* URL: http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/open-science-movement/.

[10] FOSTER Plus (European Union funded open project in Horizon 2020 and beyond). *Open Science Definition.* URL: https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition.

[11] Open Source Initiative. *The Open Source Definition.* 2004. URL: https://opensource.org/docs/osd.

[12] Free Software Foundation. *The Free Software Definition.* 1990. URL: https://www.gnu.org/philosophy/free-sw.en.html.

[13] United Nations Educational, Scientific and Cultural Organization (UNESCO). *What is Open Access?* URL: https://en.unesco.org/open-access/what-open-access.

[14] Jean-Luc Vay et al. "Center(s) for accelerator and beam modeling". In: *Snowmass21 LOI* (2020).

[15] Xiaoli Chen et al. "Open is not enough". In: *Nature Physics* 15.2 (Nov. 2018), pp. 113–119. DOI: 10.1038/s41567-018-0342-2.

[16] Remi Lehe et al. "Machine learning and surrogates models for simulation-based optimization of accelerator design". In: *Snowmass21 LOI* (2020).

[17] *CERN Data Portal.* URL: http://opendata.cern.ch.

[18] *RECAST framework reinterpretation of an ATLAS Dark Matter Search constraining a model of a dark Higgs boson decaying to two b-quarks.* Tech. rep. ATL-PHYS-PUB-2019-032. Geneva: CERN, Aug. 2019. URL: https://cds.cern.ch/record/2686290.

[19] Examples of business opportunities are R&D software contract work extending open source modeling frameworks, support, hosting, consulting and services specialized on popular sub-aspects (GUIs for antenna/RF design, shielding, etc.) that are successfully provided by companies, both in business-to-business and public-private partnerships. Scientifically relevant examples of US companies that produce open source software that is used by accelerator and beam modeling scientist are, from startups over non-profits to multi-billion USD companies: Continuum Analytics, Kitware, NumFOCUS, RadiaSoft, Red Hat, to name a few.

[20] Task Force on Accelerator R&D commissioned by Jim Siegrist. "Office of High Energy Physics Accelerator R&D Task Force Report". In: *Office of Science* (2012). URL: http://www.acceleratorsamerica.org/report/accelerator_task_force_report.pdf.