HEP Software Consortium

Lothar Bauerdick, Paolo Calafiura, Peter Elmer, Liz Sexton-Kennedy,
Panagiotis Spentzouris, Torre Wenaus, Avi Yagil

May 28, 2014

1 Objective

HEP computing is facing many challenges ahead, including the High Luminosity LHC program in Europe, the Intensity Frontier (IF) program in the US and the evolution of computing hardware technology. Significant resources are required to maintain and further develop the scientific computing infrastructure needed by the current and future HEP programs. Data analysis campaigns last for many years (decades!) and involve hundreds of developers and experimenters. The data volumes and processing needs for the LHC continue to increase while the many IF experiments will be in various phases of design and operations in the next decade. In addition, the evolution of the technical landscape calls for major software re-engineering. Both the physics and the technology *require more sophisticated software tools.*

At the same time, overall HEP funding has been declining in the past few years. *Multiple independent efforts to build this software will prove both more costly and ultimately produce less sophisticated and less sustainable software.* It is thus desirable to take advantage of commonality in needs of experiments and leverage expertise across all programs and projects. While some organizational structures exist within individual labs, between experiments at a given lab or in specific software domains, no general framework exists for collaborative software efforts across the entire HEP field.

In this document we describe a proposal for an HEP Software Consortium (HSC) whose aim is to *foster the development of a high quality, innovative, efficient and sustainable software ecosystem of general utility for the HEP community.*

2 Software Consortium

This HEP Software Consortium (HSC) would aim to identify areas that benefit from collaboration between individual projects and to facilitate the sharing/pooling of expertise and resources to solve common problems. It will organize forums to discuss best practice and software development issues and propose better software engineering solutions. Because many of our tools have broader applicability for science outside HEP, the existence of such a project could help generate new funding opportunities outside our traditional funding "sandbox" (funding agencies with narrow HEP focus) by increasing the visibility and appeal of our scientific software activities.

There is a plethora of mature and successful HEP software projects, each with different models of organization, established processes, programs, and sponsors. The HSC will provide an effective mechanism for these projects to connect and collaborate, while maintaining their separate and independent entities. We envision the HSC as a partnership of projects, which explores the common needs and interests of the participants, identifies community goals and deliverables, and facilitates community contributions. Consortium participants will come initially from HEP scientific software projects (e.g. Root, Geant4, GENIE, xrootd, etc), HEP experiments, and university and laboratory scientific software and computer science groups. In the future, if the HSC enables and fosters outreach to other communities beyond HEP, participation could be expanded to similar groups from non-HEP domains.

The Consortium will involve all partners in the design, development and possibly deployment process to ensure that the HSC will function and evolve consistently while meeting the needs of the individual partners. The Consortium will sponsor activities (workshops, meetings, discussion forums) to facilitate the development of software engineering standards and software architecture guidelines by identifying common needs and encouraging common solutions. The participating institutions and partner projects will contribute to the Consortium by providing input and expertise in defining common goals and developing such principles and guidelines. In addition, they will contribute to the HSC software infrastructure by developing and deploying software within the common development and testing guidelines and following the HSC software architecture principles. This will include adapting existing software tools to the HSC standards, when possible. The partner software projects will maintain their independent organization frameworks and have ownership of the software they are developing and deploying utilizing the common standards. The HSC will publish the code that complies with the standards and organize peer-reviews for the software that is developed under its process and guidelines. Specific examples of HSC technical activities could include:

- sharing technical expertise in architecture and design issues,
- the development of standards for component interfaces and layout, to enable independent development and easy component integration,
- facilitating the development of non-domain specific components (geometry representation, data representation) that can be used by many projects,
- developing standards and guidelines for testing releasing and documenting software,
- proposing solutions for distributed collaborative environment and common infrastructure such as build tools, testing and validation suites,
- maintaining a list of HEP software products, and
- providing information as to whether the software packages follow architectural and engineering recommendations developed in the HSC.

Note that the initiative (and buy-in) for these activities must come from motivated individuals, software projects, institutions and/or experiments within the HSC. It is not intended to be a "top-down" process driven entirely by the chair of one or another governance entity,

but a structured community process. In order for this program of work to be successful, the Consortium will also adopt an "open-source" software model and rely on agile development methods.

3 Governance

From the discussions at the workshop at CERN on 3-4 April, 2014, it is clear that the community in general prefers a lightweight organization, whose successes will be driven in a "bottoms-up" fashion. Indeed fostering a software ecosystem is a different type of activity from a detector construction project or a traditional HEP experimental collaboration. We describe here the functional entities of such a lightweight "governance" that would provide the basic mechanisms for the consortium to achieve the overall aim of fostering the development of a high quality, innovative, efficient and sustainable software ecosystem of general utility for the HEP community.

Software Engineering Board (SEB): The activities of the SEB provide the primary "software ecosystem" functionality of the Consortium. The SEB provides expertise and advise on software architecture and engineering issues, organizes and contributes to peer-reviews, discusses and proposes best practices and develops guidelines. The members of the board play a leading role in the activities of the Technical Domain Forums (described below) by organizing and chairing the Forum discussions and meetings, and discussing the outcome of the Forum activities at the Board meetings. In essence, it should play a role similar to the LHC "Architects Forum", with clearer mechanisms for interacting with major stakeholders (Consortium Council below). The membership of this board is *inclusive and open*. In general each software project and/or major user group (e.g. experiments) simply self-identifies a representative to participate. The members of the Board elect a fixed term chair to moderate discussions and foster consensus. The ultimate success of the HSC should be judged on the synergies, initiatives and activities enabling a stronger software ecosystem which would not have existed in the absence of the HSC.

Technical Domain Forums: provide coordination and facilitate communication between the participating projects and the consortium on specific topics. One such example today is the "Concurrency Forum". These could be standing forums on topics of general interest or temporary working groups on specific topics or initiatives. To achieve their goals they may organize workshops; maintain wikis, blogs, etc. The SEB may create new technical forums as needed and pre-existing entities effectively playing this role can simply self-identify to the chair of the SEB. The chair of the SEB will maintain a list of such forums, their scope and objectives.

Consortium Council: The role of the council is to enable the work of the Software Engineering Board. It consists of representatives of institutions providing resources to the software projects and major experiments and/or user groups; the stakeholders board. (We believe that the healthiest starting point for the HSC is that it *not* have resources of its own, independent from these stakeholders. Eventually, the HSC could evolve to also incorporate an HSC centric project with its own resources, if the need arises.) The Council holds infrequent meetings (at least once a year) to identify common areas of interest and discuss common goals and overall direction based on the input from the Advisory Board and the results of the

activities of the Software Engineering Board. The Council will also elect a fixed term chair.

Advisory Board: The role of the Advisory Board is to provide feedback and recommendations on scientific and technical priorities and needs to the Consortium Council. It consists of individuals from the HEP community at large, other scientific communities, and software industry experts. They are *not* chosen as representatives of particular user groups, experiments, institutions or software projects, but rather for their general expertise. Proposals for membership in the Advisory Board are solicited from the Consortium Council. The Advisory Board may be asked to perform general reviews, or simply provide feedback directly to the Council, depending on the specific needs in any given time period.

4 Incentives

As the HSC will not direct resources of its own, at least initially, the primary incentives that the HSC can provide to software developers are *recognition* and *visibility*. The key question is how these incentives can be deployed to reach the goal of creating a high quality, innovative, efficient and sustainable software ecosystem of general utility for the HEP community. The secondary question will then be how to increase the *value* of these incentives with the community and with external entities like funding agencies.

It is not possible to answer these questions fully for the purposes of this document. The community needs to work out specific methods over time. We provide some initial ideas here, but emphasize that if the HSC is to succeed, discussions must not focus only on the specific technical activities such as the examples listed in Sec. 2, but also on the use of such incentives in pursuit of the overall goal. The SEB and Consortium Council chairs should help keep this aspect of the discussion in view, but ultimately it is the software projects and the institutions supporting them that must find value in, and find methods to add value to, the incentives.

To give a concrete example, we note that a recent study [1] of software packages which are widely used in HEP noted the following common characteristics relevant for the HSC:

1. Clearly defined individual or individuals exist as champion(s) with a strong sense of ownership for the software package and its success
2. The software is created in the context of an experiment or driven by people who are also users
3. Distribution via known mechanisms enables wider use
4. *Collaborations* between individuals, institutions or experiments are formed early to facilitate development
5. Adherence to or development of useful or recognized (de-facto or documented) standards enables wider use

We can examine how such incentives might serve to encourage more software projects to adopt these characteristics.

While the HSC itself cannot create sofware project champions as such, it can provide standard mechanisms for recognizing and making visibile both software projects and their

champions. This should be especially beneficial to R&D efforts related to new technologies, since they involve a lot of uncertainty and risk for participating developers. At the most basic level this could be done via a community catalog of software packages (similar to today's HepForge, for example), but it could eventually evolve towards a HEP software build, testing and distribution. (Note that today's LCG AA software distribution has elements of this, but additional flexibility is needed to allow different experiments groups to use and deploy different portions and versions of software in the ecosystem.) The simple participation of software projects in the HSC will already help identify champions. The visibility and recognition from successful integration of packages into a community build and test system is an incentive as a step towards facilitating further adoption by new user groups, especially for new software projects. It is simultaneously a means for evaluation of community software standards. The HSC can also provide independent recognition of adoption of software packages by experiments and user groups, where such statistics can clarify the importance of individual software packages to the field (to funding agencies, etc.)

Similarly, as broader collaborations foster stronger and more sustainable software, the HSC can provide a forum for existing software projects to publicize themselves *as software projects*, describing their goals and their needs to possible collaborators. This could be facilitated through the Techincal Domain Forums. Eventually new collaborations could formed be via *quid pro quo* mechanisms between institutions or (preferably) via pursuit of new funding. Here the HSC uses the fact that it can provide visibility to encourage collaborations (explicitly in the "marketplace" sense of bringing together buyers and sellers).

5 Summary

A HEP Software Consortium will provide a community-wide framework for leveraging and coordinating ongoing HEP software activities and facilitating the creation of an *ecosystem* of software activities. The proposed HSC will:

- be based on a primarily "bottoms-up" structure
- provide the means to recognize and encourage the common aspects of successful projects, facilitate collaboration between partner projects, and facilitate adoption of successful software engineering practices,
- promote sharing of software architecture assistance and expertise,
- help coordinate exploration of new technologies and provide expertise in their implementation,
- and eventually enable contact with other scientific fields, providing access to either resources or software tools and expertise.

References

[1] P. Elmer 2014 Case Studies of Scientific Software Collaborations in High Energy Physics and Beyond (In preparation, draft available at `https://indico.cern.ch/event/297652/session/0/contribution/2/material/0/0.pdf`)