9 Building Partnerships

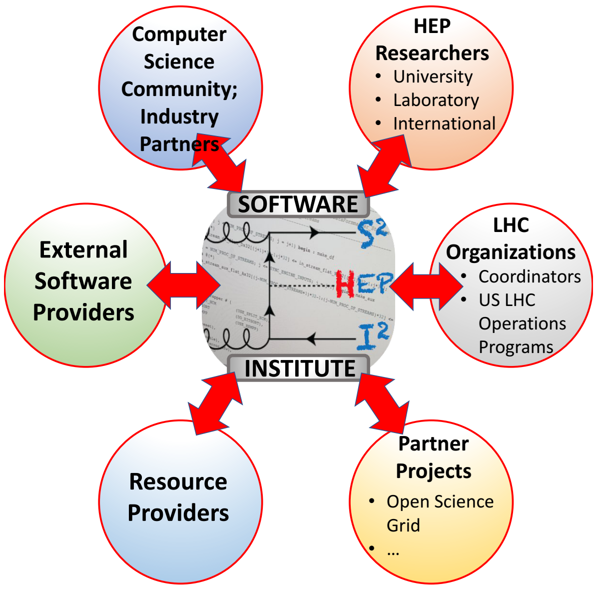

Figure 9: Partners of the Institute.

9.1 Partners

Various roles envisioned for the Institute in Section 6 will require collaborations and partnerships with external entities, as illustrated in Figure 9. These include:

HEP Researchers (University, Lab, International): LHC researchers are the primary repository of expertise related to all of the domain-specific software to be developed and deployed; they also define many of the goals for domain-specific implementations of more general types of software such as workflow management. Areas in which collaboration with HEP researchers will be especially close include technical aspects of the detectors, their performances, algorithms for reconstruction and simulation, analysis techniques and, ultimately, the potential physics reach of the experiments. These researchers will define the detailed and evolving physics goals of the experiments. They will participate in many of the roles described in Section 8, and some will be co-funded by the Institute. In addition, the Institute should identify, engage and build collaborations with other HEP researchers whose interests and expertise align with the Institute's focus areas.

LHC Experiments: The LHC experiments, and especially the US-ATLAS and US-CMS participants, are key partners. In large measure, the success of an Institute will be judged in terms of its impact on their Computing Technical Design Reports (CTDRs), to be submitted in 2020, and its impact on software deployed (at least experimentally) during LHC's Run 3. The experiments will play leading roles in understanding and defining software requirements and how the pieces fit together into coherent computing models with maximum impact on cost/resources, physics reach, and sustainability. As described in Section 8, representatives of the experiments will participate explicitly in the Institute Steering Board to help provide big-picture guidance and oversight. In terms of daily work, the engagement will be deeper. Many people directly supported by the Institute will be collaborators on the LHC experiments, and some will have complementary roles in the physics or software & computing organizations of their experiments. Building on these natural connections will provide visibility for Institute activities within the LHC experiments, foster collaboration across experiments, and provide a feedback mechanism from the experiments to the Institute at the level of individual researchers. The experiments will also be integral to developing sustainability paths for software products they deploy that emerge from R&D performed by the Institute; therefore, they must be partners starting early in the software lifecycle.

US LHC Operations Programs: As described in Section 6, the Institute will be an R&D engine in the earlier phases of the software life cycle. The Operations Programs will be one of the primary partners within the U.S. for integration activities, testing "at-scale" on real facilities, and eventual deployment. In addition they will provide a long run sustainability path for some elements of the software products. Ultimately, much of the software emerging from Institute efforts will be used by the LHC Operations Programs or run in facilities they operate. The Institute will address many of the issues that the Operations Programs expect to encounter in the HL-LHC era. Thus, the Institute must have, within the U.S., a close relationship to the Operations Programs. Their representatives will serve as members of the Steering Board, and they will be invited to participate in Executive Board meetings as observers.

Computer Science (CS) Community: During the S^2I^2-HEP conceptualization process we ran two workshops focused on how the HEP and CS communities could work to their mutual benefit in the context of an Institute, and, also, more generally. We identified some specific areas for collaboration, and others where the work in one field can inform the other. Several joint efforts have started as results of these conversations. More importantly, we discussed the challenges, as well as the opportunities, in such collaborations, and established a framework for continued exchanges. For example, several of the computer scientists explicitly defined their interests as mapping our specific problems to more abstract problems whose solutions would interest them. They distinguished this from providing software engineering solutions to specific problems. We anticipate direct CS participation in preparing a proposal if there is a solicitation, and collaboration in Institute R&D projects if it comes to fruition. Continued dialogue and engagement with the CS community will help assure the success of the Institute. This may take the form of targeted workshops focused on specific software and computing issues in HEP and their relevance for CS, or involvement of CS researchers in blueprint activities (see below). It may also take the form of joint exploratory projects. Identified topics of common interest include: Science Practices & Policies, Sociology and Community Issues; Machine Learning; Software Life Cycle; Software Engineering; Parallelism and Performance on modern processor architectures, Software/Data/Workflow Preservation & Reproducibility, Scalable Platforms; Data Organization, Management and Access; Data Storage; Data Intensive Analysis Tools and Techniques; Visualization; Data Streaming; Training and Education; and Professional Development and Advancement. We also expect that one or two members of the CS and Cyberinfrastructure communities will serve on the Institute Advisory Panel, as described in Section 8, to provide a broad view of CS research.

External Software Providers: The LHC experiments depend on numerous software packages developed by external providers, both within HEP and from the wider open source software community. For the non-HEP software packages, the HEP community interactions are often a bit diffuse and unorganized. The Institute could play a role in developing the collaborations with these software providers, as needed, including engaging them for relevant planning, discussions regarding interoperability with other HEP packages, and software packaging and performance issues. For non-HEP packages the Institute can also play a role in introducing key developers of these external software packages to the HEP community (through seminars or sponsored visits to work at HEP institutions) and raising the visibility of HEP use cases in the external development communities. Examples of the latter activity can be found in the "topical meetings" being organized by the

DIANA-HEP project [69, 70].

Open Science Grid: The strength of the Open Science Grid project is its fabric of services that allows the integration of an at-scale globally distributed computing infrastructure for HTC that is fundamentally elastic in nature, and can scale out across many different types of hardware, software, and business models. It is the natural partner for the Institute on all aspects of "productizing" prototypes, or testing prototypes at scale. For example, the OSG already supports machine learning environments across a range of hardware and software environments. New environments could be added in support of the ML focus area. It is also a natural partner to facilitate discussions with IT infrastructure providers, and deployment experts, e.g. in the context of the DOMA and Data Analysis Systems focus areas.

DOE and the National Labs: The R&D roadmap outlined in the Community White Paper [11] is much broader than what will be possible within an Institute. The DOE labs will necessarily engage in related R&D activities both for the HL-LHC and for the broader U.S. HEP program in the 2020s. Many DOE lab personnel participated in both the CWP and S^2I^2-HEP processes. In addition, a dedicated workshop was held in November, 2017, to discuss how S^2I^2- and DOE-funded efforts related to HL-LHC upgrade software R&D might be aligned to provide the best coherence (see Appendix B). Collaborations between university personnel and national labortory personnel will be critical, as will be collaborations with foreign partners. In particular, the HEP Center for Computational Excellence (HEP-CCE) [71], a DOE cross-cutting initiative focused on preparations for effectively utilizing DOE's future high performance computing (HPC) facilities, and the R&D projects funded as part of DOE's SciDAC program are critical elements of the HL-LHC software upgrade effort. While S^2I^2 R&D efforts will tend to be complementary, the Institute will establish contacts with all of these projects and will use the blueprint process (described below) to establish a common vision of how the various efforts align into a coherent set of U.S. activities.

CERN: As the host lab for the LHC experiments, CERN must be an important collaborator for the Institute. Two entities within CERN are involved with software and computing activities. The IT department is focused on computing infrastructure and hosts CERN openlab (for partnerships with industry, see below). The Software (SFT) group in the CERN Physics Department develops and supports critical software application libraries relevant for both the LHC experiments and the HEP community at large, most notably the ROOT analysis framework and the Geant4 Monte Carlo detector simulation package. There are currently many ongoing collaborations between the experiments and U.S. projects and institutions with the CERN software efforts. CERN staff from these organizations were heavily involved the CWP process. The Institute will naturally build on these existing relationships with CERN. A representative of CERN will be invited to serve as a member of the Institute Steering Board, as described in Section 8.

The HEP Software Foundation (HSF): The HSF was established in 2015 to facilitate international coordination and common efforts in high energy physics (HEP) software and computing. Although a relatively new entity, it has already demonstrated its value. Especially relevant for the S^2I^2 conceptualization project, it organized the broader roadmap process leading to the parallel preparation of the Community White Paper. This was a collaboration with our conceptualization project, and we expect that the Institute will naturally partner with the HSF in future roadmap activities. Similarly, it will work under the HSF umbrella to sponsor relevant workshops and coordinate community efforts to share information and code.

Industry: Partnerships with industry are particularly important. They allow R&D activities to be informed by technology developments in the wider world and, through dedicated projects, to inform and provide feedback to industry on their products. HEP has a long history of such collaborations in many technological areas including software and computing. Prior experience indicates that involving industry partners in actual collaborative projects is far more effective than simply inviting them for occasional one-way presentations or training sessions. There are a number of projects underway today with industry partners. Examples include collaboration with Intel like the Big Data Reduction Facility [72], through an Intel Parallel Computing Center [73], with Google [74, 75] and AWS [74–76] for cloud computing, etc. A variety of areas will be of interest going forward, including processor, storage and networking technologies, tools for data management at the Exabyte scale, machine learning and data analytics, computing facilities infrastructure and management, cloud computing and software development tools and support for software performance. In 2001 CERN created a framework for such public-private partnerships with industry called CERN openlab [77]. Initially this was used to build projects between CERN staff and industry on HEP projects, however in recent years the framework has been broadened to include other research institutions and scientific disciplines. Fermilab has recently joined CERN openlab collaboration and Princeton University is currently finishing the process to join. Others may follow. CERN openlab can also be leveraged by the Institute to build partnerships with industry and to make them maximally effective. This can be done in addition to direct partnerships with industry.

9.2 The Blueprint Process

To facilitate the development of effective collaborations with the various partners described above, the Institute should proactively engage and bring together key personnel for small "blueprint" workshops on specific aspects of the full R&D effort. During these blueprint workshops the various partners will not only inform each other about the status and goals of various projects, but actively *articulate and document* a common vision for how the various activities fit together into a coherent R&D picture. The scope of each blueprint workshop should be sized in a pragmatic fashion to allow for convergence on the common vision, and some of the key personnel involved should have the means of realigning efforts within the individual projects if necessary. The ensemble of these small blueprint workshops will be the process by which the Institute can establish its role within the full HL-LHC R&D effort. The blueprint process will also be the mechanism by which the Institute and its various partners can drive the evolution of the R&D efforts over time, as shown in Figure 10.

Following the discussions at the November, 2017, S^2I^2-DOE workshop on HL-LHC R&D, we expect that jointly sponsored blueprint activities between NSF and DOE activities relevant for HL-LHC, the US LHC Operations Programs and resource providers like OSG will likely be possible. All parties felt strongly that an active blueprint process would contribute significantly to the coherence of the combined U.S. efforts. The Institute could also play a leading role to bring other parties into specific blueprint activities, where a formal joint sponsorship is less likely to be possible. This may include specific HEP and CS researchers, other relevant national R&D efforts (non-HEP, non-DOE, other NSF), international efforts and other external software providers, as required for the specific blueprint topic.

Blueprint activities will likely happen 3-4 times per year, typically with a focus on a different specific topics each time. The topics will be chosen based on recognizing areas where a common vision is required for the coordination between partners. Input from the Institute managment, the Institute Steering Board and the management of various partner projects and key personnel will be explicitly solicited to identify potential blueprint activities. The Institute will take an active role in organizing blueprint activities by itself and jointly with its partners based on this input. From year to year specific topics may be revisited.

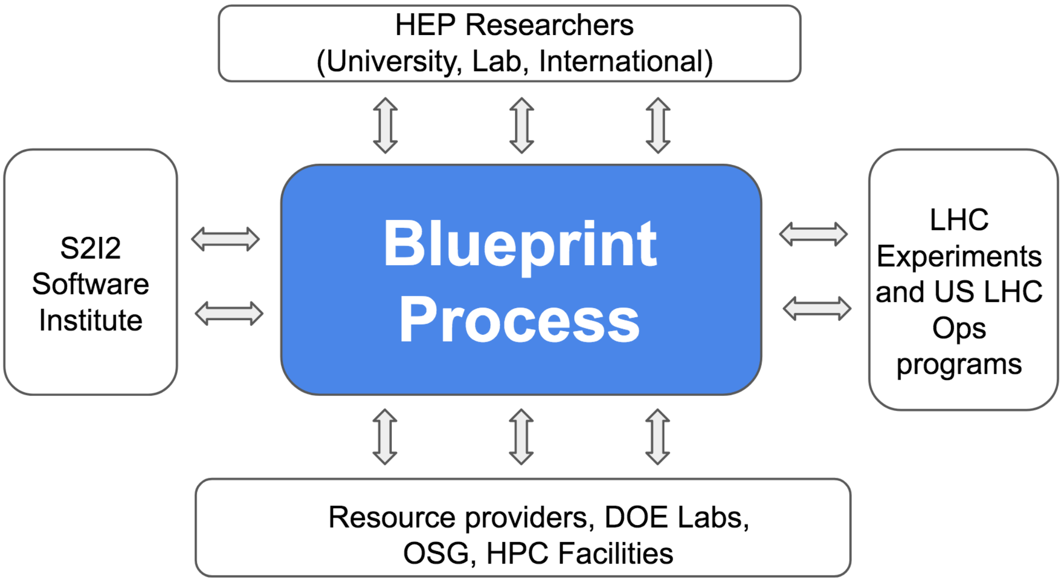

Figure 10: The Blueprint Process will be a primary means of developing a common vision with the major partners.