

1 Institute Role

1.1 Software Lifecycle

A Sustainable Software Innovation Institute (S^2I^2) for High Energy Physics should serve as an intellectual hub for computing and software for the HL-LHC era.

In practice, the direct resources funded by the institute will not be enough by itself to design, build and sustain the cyberinfrastructure required for the challenges of the HL-LHC. What the institute can do is serve as the primary engine for innovation in collaboration with the wider HEP and CS research communities. It should leverage and amplify the impact of effort from those communities.

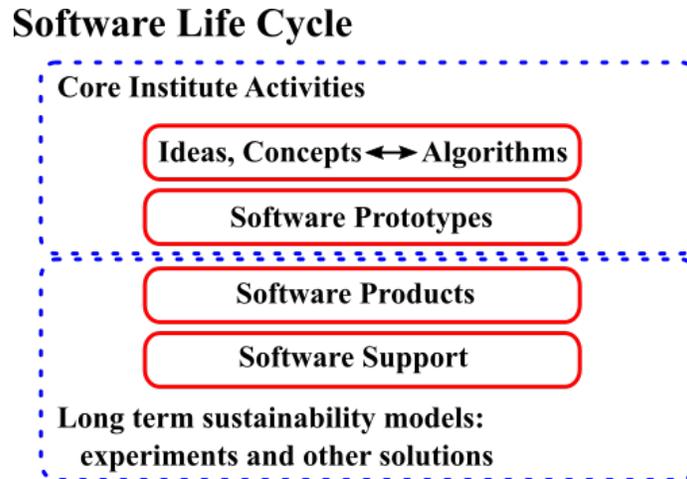


Figure 1: Roles of the Institute

The Institute will focus its resources on developing innovative ideas through the prototype stage and along the path to become software products used by the wider community. It will also provide technical support to the experiments and others to develop sustainability and support models for the software products developed. In its role as an intellectual hub for HEP software innovation, it provide advice and guidance broadly on software development within the HEP ecosystem. For example, a new idea or direction under consideration by an experiment could be critically evaluated by the Institute in terms of its essence, novelty, sustainability and impact which would then provide written recommendations for the proposed activity. This will be achieved through a critical mass of experts in scientific software development inside and outside of HEP and the computer science community who partner with the Institute.

- The institute can have different levels of ownership.
- Community software to be developed (and owned) by the institute needs a sustainability model

1.2 Notes

Need to think about role of computing models, “architecture” of software; life-cycle costs; how to design algorithms which are agnostic with respect to hardware; also, algorithms and software which are agnostic with respect to languages and operating systems.

The “in-house” activities might/should include studying and designing algorithms and building prototypes of software, evaluating externally developed software, comparing performance of prototype algorithms within one hardware/software environment, comparing single software prototypes on multiple platforms.

Where do we support “Fellows”? Should they be associated with the individual centers, or somehow affiliated with the “in-house” activities? How about sabbatical support? What are the advantages of having fellows and affiliates co-located at one institution? Disadvantages?

2 Institute Organizational Structure and Evolutionary Process

2.1 Institute Management and Governance

The management and governance structures chosen for the Institute should answer to the following questions:

1. **Goals:** What are the goals of the Institute?
2. **Interactions:** Who are the primary clients/beneficiaries of the Institute? How are their interests represented? How can the Institute align its priorities with those of the LHC experiments?
3. **Operations:** How does the Institute execute its plan with the resources it directly controls? How does the Institute leverage and collaborate with other organizations? How does the Institute maintain transparency?
4. **Metrics:** How is the impact of the Institute evaluated? And by whom?
5. **Evolution:** What are the processes by which the Institutes areas of focus and activities evolve?

PI/co-PIs: as on the proposal, with project responsibilities as defined by NSF.

Executive Board: the Executive Board will manage the day to day activities of the Institute. It will consist of the PI, coPIs and the managers of the areas of focus. A weekly meeting will be used to manage the general activities of the institute and make shorter term plans. In many cases, a liaison from other organizations (e.g. the US LHC Ops programs) would be invited as an “observer” to weekly Executive Board meetings in order to facilitate transparency and collaboration (e.g. on shared services or resources).

Focus Areas: A number of Focus Areas will be defined for the institute at any given point in time. These areas will represent the main priorities of the institute in terms of activities aimed at developing the software infrastructure to achieve the mission of the Institute. Each focus area will have a written set of goals for the year and corresponding institute resources. The number and size of these will depend on funding available and resources needed to achieve the goals, however a typical number will be 3-5. The active focus areas will be reviewed together with the Advisory Panel once/year and decisions will be taken on updating the list of areas and their yearly goals, with input from the Steering Board.

Area Manager(s): each Area Manager will manage the day to day activities within a focus area. It is for the moment undefined whether there will be an Area Manager plus a deputy, co-managers or a single manager. An appropriate mix of HEP, Computer Science and representation from different experiments will be a goal. Area Managers should be available to serve at least 20% time.

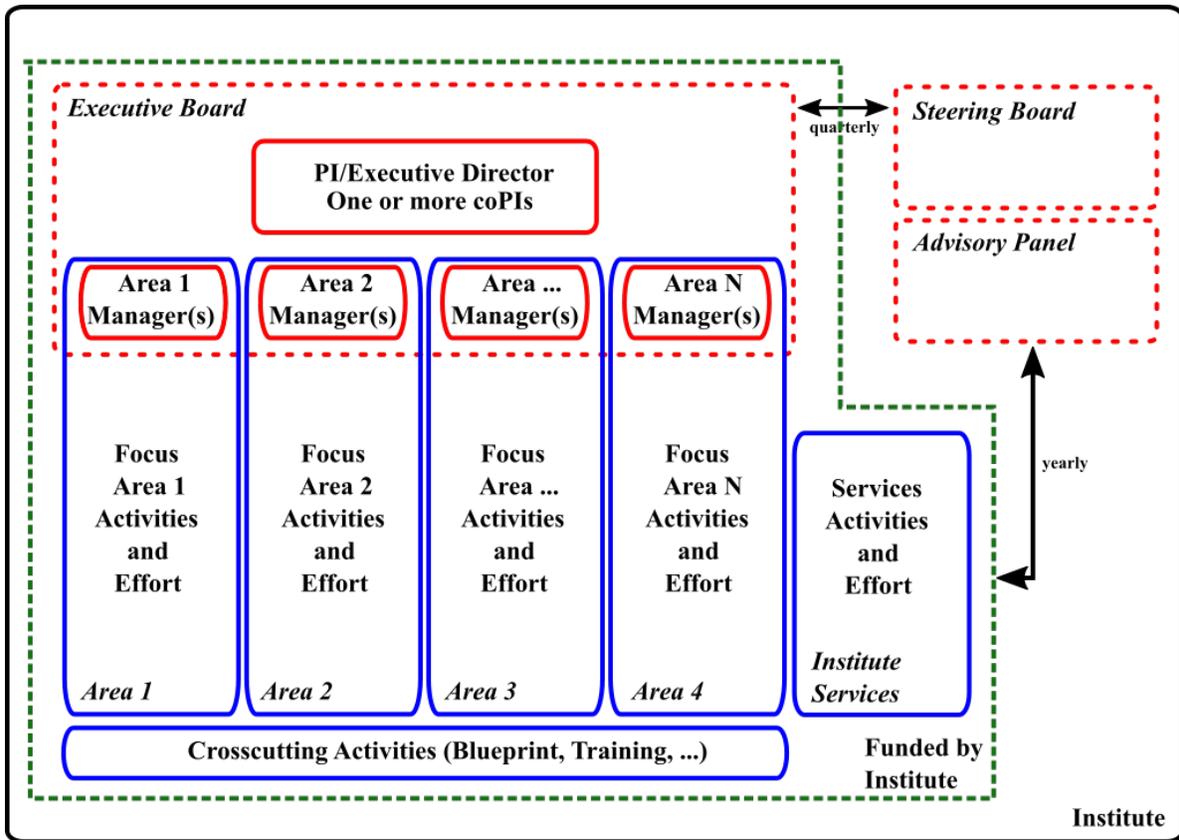


Figure 2: Institute Management and Governance

Steering Board: a Steering Board will be defined to meet with the executive team approximately quarterly to review the large scale priorities and strategy of the institute. (Areas of focus will also be reviewed, but less frequently.) The steering board will consist of two representatives for each participating experiment, plus representatives of CERN, FNAL, etc. Members of the Steering Board will be proposed by their respective organizations and accepted by the Executive Director in consultation with the Executive Board.

Executive Director: an Executive Director will manage the overall activities of the institute and its interactions with external entities. In general day-to-day decisions will be taken by achieving consensus in the Executive Board and strategy and priority decisions based on advice and recommendations by the Steering and Executive Boards. In cases where consensus cannot be reached, the Executive Director will take a final decision. It would also be prudent for the Institute to have a Deputy Director who is able to assume the duties during periods of unavailability of the Executive Director.

Advisory Panel: an Advisory Panel will be convened to conduct an internal review of the project once per year. The members of the panel will be selected by the PI/co-PIs with input from the Steering Board. The panel will include experts not otherwise involved with the institute in the areas of physics, computational physics, sustainable software development and computer science.

2.2 Relationship to the Community

The Institute will partner with a number of other entities, as shown in Figure 3.

2.3 Institute Elements

The Institute will have a number of internal elements, as shown in Figure 3. There are the N focus areas previously mentioned, in addition to an exploratory element where new ideas outside of the strict confines of the focus areas but a with strong potential can be researched and developed. Such activities are critical for the vitality of the Institute and could inform the Institute Blueprint as it evolves to meet the challenges of the HL-LHC era.

2.4 External Interactions of the Institute

- **HEP Researchers (University, Lab, International):**
- **LHC Experiments:**
- **LHC Ops Programs:**
- **Computer Science Community:**
- **External Software Providers:**
- **OSG-LHC:**

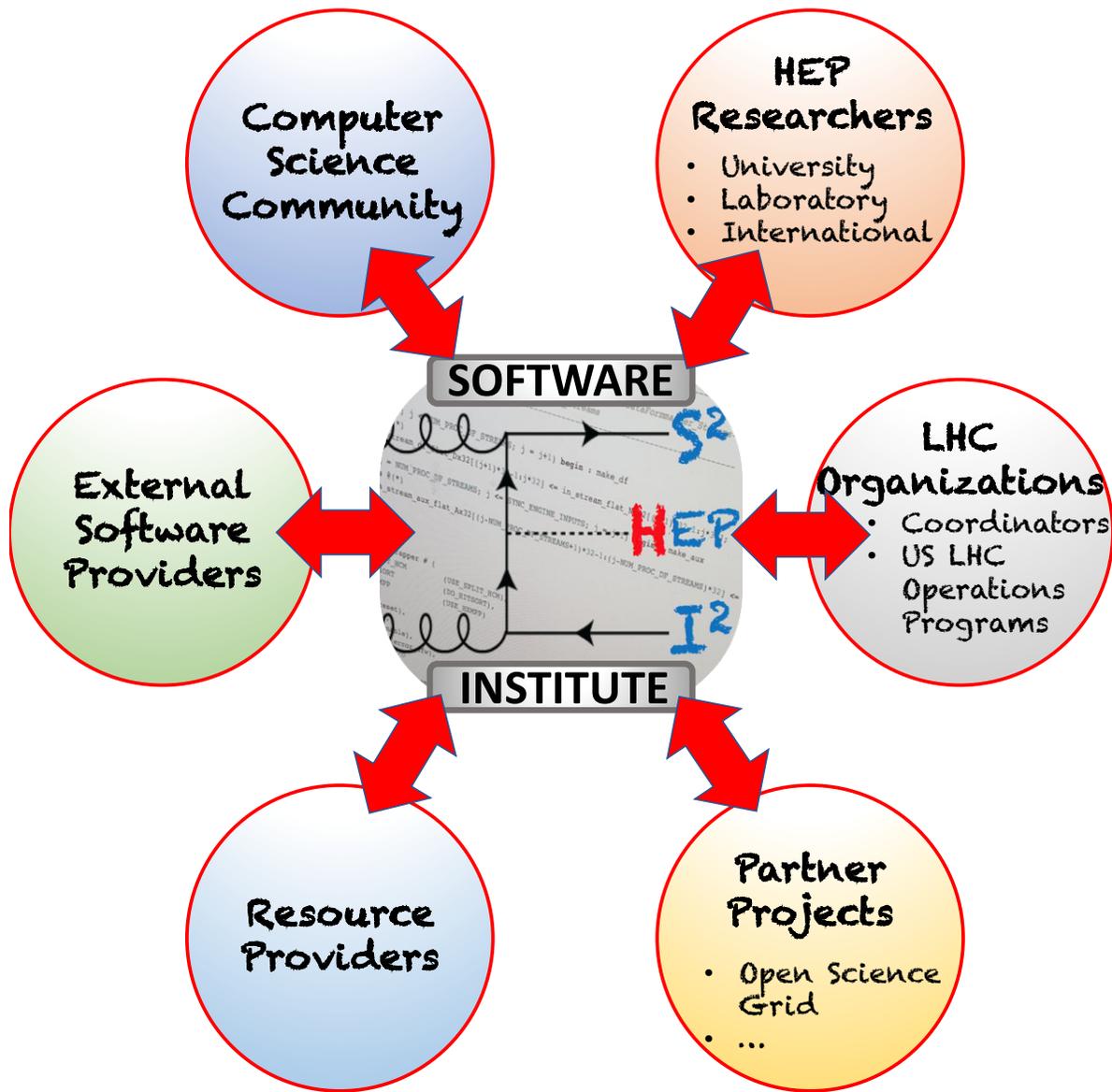


Figure 3: Relationship of the Institute to other entities

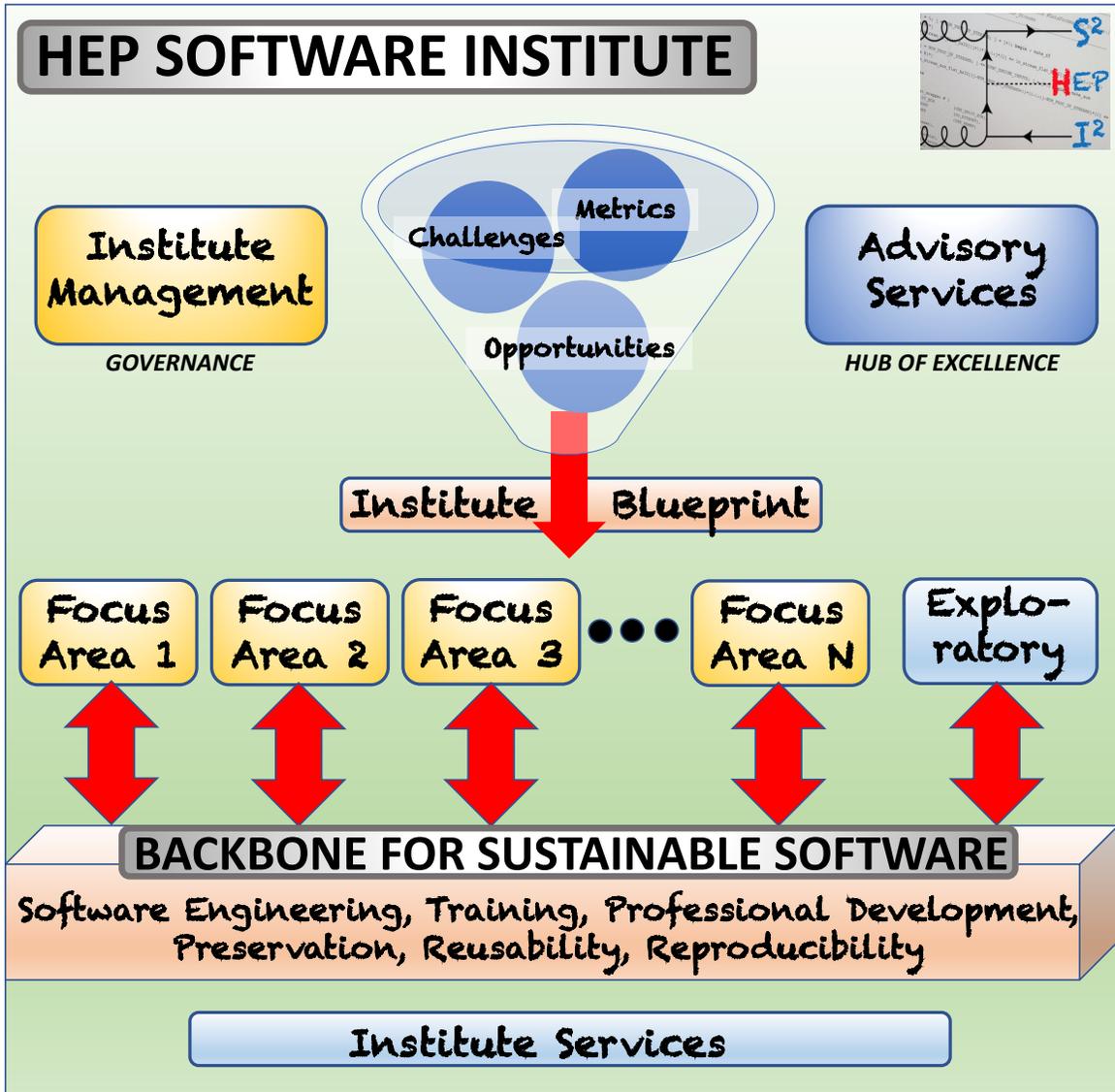


Figure 4: Institute elements

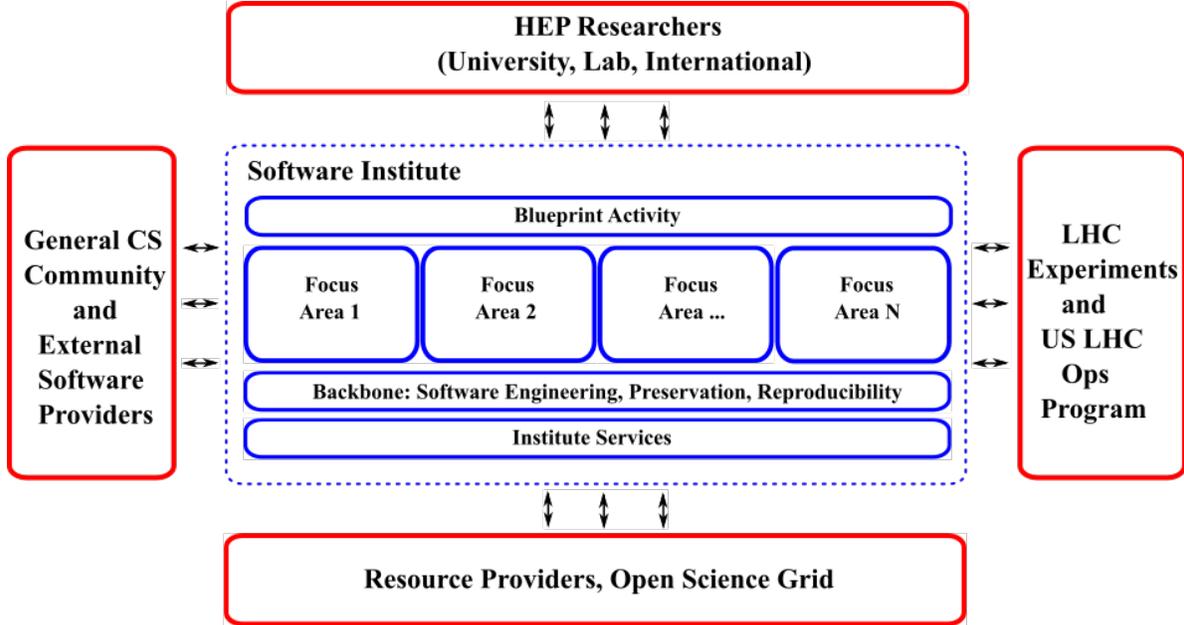


Figure 5: Relationship of the Institute to other entities

3 Strategic Areas for Initial Investment

3.1 Rationale for choices and prioritization of a university-based S^2I^2

The primary goal of the S^2I^2 -HEP conceptualization project is to prepare a Strategic Plan for a potential NSF Scientific Software Innovation Institute (S^2I^2) to develop software for experiments taking data in the “High-Luminosity Large Hadron Collider” (HL-LHC) era in the 2020s. We have been working with the HEP Software Foundation to prepare a larger HEP Community White Paper (CWP) describing a global roadmap for HEP Software and Computing R&D for the 2020s.

The S^2I^2 -HEP Strategic Plan should identify the specific areas of focus and initial priorities for an NSF-funded Institute funded in the U.S. universities, as well as the corresponding challenges and why addressing those challenges will be transformative. Equally importantly, the Strategic Plan should identify areas where the Institute will not lead efforts, and those where it might participate with lower priority.

The following questions were used as the rationale to identify priority focus areas relevant for a U.S. university-based Institute:

1. **Impact - Physics:** Will efforts in this area enable new approaches to computing and software that maximize, and could potentially radically extend, the physics reach of the detectors?
2. **Impact - Resources:** Will efforts in this area achieve required improvements in software efficiency, scalability and performance and make use of the advances in CPU, storage and network technologies?
3. **Impact - Sustainability:** Will efforts in this area guarantee the long term sustainability of the software through the lifetime of the HL-LHC?
4. **Interest/Expertise:** Does the U.S. university community have a strong interest and expertise in the area?

5. **Leadership:** Are the proposed focus areas complementary to efforts funded by the US-LHC Ops programs, DOE or international entities?
6. **Value:** Is there potential to provide value to more than one LHC experiment and to the wider HEP community?
7. **Research/Innovation:** Are there opportunities for combining research and innovation as part of partnerships between the HEP and Computer Science communities?

CWP working group areas:

- Data Analysis and Interpretation
- Machine Learning
- Software Trigger and Event Reconstruction
- Data Access, Organization and Management
- Workflow and Resource Management
- Data and Software Preservation
- Careers, Staffing and Training
- Visualization
- Detector Simulation
- Various Aspects of Technical Evolution (Software Tools, Hardware, Networking)
- Data Acquisition Software
- Conditions Database
- Physics Generators
- Computing Models, Facilities and Distributed Computing
- Software Development, Deployment and Validation/Verification
- Event Processing Frameworks

4 Addendum - S^2I^2 Strategic Plan Elements

Other points from the solicitation NSF 15-553 [1]?

“The product of a conceptualization award will be a strategic plan for enabling science and education through a sustained software infrastructure that will be freely available to the community, and will address the following elements:”

- the science community and the specific grand challenge research questions that the S^2I^2 will support;
- specific software elements and frameworks that are relevant to the community, the sustainability challenges that need to be addressed, and why addressing these challenges will be transformative;
- appropriate software architectures and lifecycle processes, development, testing and deployment methodologies, validation and verification processes, end usability and interface considerations, and required infrastructure and technologies;
- the required organizational, personnel and management structures and operational processes;
- the requirements and necessary mechanisms for human resource development, including integration of education and training, mentoring of students, postdoctoral fellows as well as software professionals, and proactively addressing diversity and broadening participation;
- potential approaches for long-term sustainability of the software institute as well as the software; and
- potential risks including risks associated with establishment and execution, necessary infrastructure and associated technologies, community engagement, and long-term sustainability.

“The strategic plan resulting from the conceptualization phase is expected to serve as the conceptual design upon which a subsequent S^2I^2 Implementation proposal could be based.”

The S^2I^2 implementation proposals asked for the following things in the (20 page) Project Description:

- The overall rationale for the envisioned institute, its mission, and its goals.
- A set of software issues and needs and software sustainability challenges faced by a particular, well-defined yet broad community (that is clearly identified in the proposal) that can best be addressed by an institute of the type proposed, a compelling case these are the most important issues faced by the community, and that these issues are truly important.
- A clear and compelling plan of activities that shows how the proposed institute will address these issues and needs by involving (and leveraging) the community, including its software developers, in a way that will benefit the entire community.
- If there are other NSF-funded activities that might appear to overlap the institute’s activities, a discussion clarifying how the funding of each activity will be independent and non-overlapping.
- Metrics of how success will be measured, that include at least impact on the developer and user communities.
- Evidence that the people involved in planning and setting up the institute have the organizational, scientific, technical, and sociocultural skills to undertake such a task, and that they are trusted and respected by the community as a whole.
- Evidence of a high degree of community buy in that a) these are the urgent/critical needs and b) this institute is the way to address them.

- A plan for management of the institute, including 1) the specific roles of the PI, co-PIs, other senior personnel and paid consultants at all institutions involved, 2) how the project will be managed across institutions and disciplines, 3) identification of the specific coordination mechanisms that will enable cross-institution and/or cross-discipline scientific integration, and 4) pointers to the budget line items that support these management and coordination mechanisms.
- A steering committee composed of leading members of the targeted community that will assume key roles in the leadership and/or management of the institute. A brief biography of the members of the steering committee and their role in the conceptualization process should be included.
- A plan for how the institute activities will continue and/or the value of the institutes products will be preserved after the award, particularly if it does not receive additional funds from NSF.

References

- [1] NSF 15-553. <https://www.nsf.gov/pubs/2015/nsf15553/nsf15553.htm>.