

# Review of DIANA project goals, metrics, broader impacts

Peter Elmer - Princeton University

16 December, 2016 - DIANA/HEP Team Meeting

In this meeting we need to evaluate our activities over the past year, and planned in the next months, relative to the overall project goals.

Now that the project is fully staffed and past the “bootstrap” phase, we should also put in place a public “project plan” to communicate to others.

We are not aiming to hash out specific technical details of projects in this meeting. Detailed technical discussions should be deferred to dedicated discussions or a DIANA topical meeting, as appropriate.

Going forward, we will have these project meetings more frequently. Today I am only outlining the most general goals and we can discuss at the end. We will probably need another in mid-January to go through things in greater detail.

**Capabilities:** Support the creation and maintenance of an innovative, integrated, reliable, sustainable and accessible software ecosystem providing new capabilities that advance and accelerate scientific inquiry and application at unprecedented complexity and scale.

**Research:** Support the foundational research necessary to continue to efficiently advance scientific software, responding to new technological, algorithmic, and scientific advances.

**Science:** Enable transformative, interdisciplinary, collaborative, science and engineering research and education through the use of advanced software and services.

**Education:** Empower the current and future diverse workforce of scientists and engineers equipped with essential skills to use and develop software. Further, ensure that the software and services are effectively used in both the research and education process realizing new opportunities for teaching and outreach.

**Policy:** Transform practice through new policies for software addressing challenges of academic culture, open dissemination and use, reproducibility and trust of data/models/ simulation, curation and sustainability, and that address issues of governance, citation, stewardship, and attribution of software authorship.

The NSF “Software Infrastructure for Sustained Innovation” (SI<sup>2</sup>) program implements this vision.

DIANA/HEP was funded as an “SSI” award, a “software framework”, see the original solicitation:

<https://www.nsf.gov/pubs/2014/nsf14520/nsf14520.htm>

The DIANA/HEP project (awards) had a start date of 1 May, 2015. This means that we must provide annual reports before that date each year.

A research and development agenda that leads to robust and sustainable software, that advances research capability for one or more areas of science

An engineering process used for the design, development, and release of the software, its deployments and associated outreach to the end user community, its interoperability with widely used tools by the community, and an evaluation plan that involves end users. Address issues of sustainability, manageability, usability, composability, and interoperability.

Tangible metrics, with end user involvement, to be used to measure the success of the software developed, and the steps necessary to take the software elements from prototype to broader use

How can the software be used by broader communities?

An outreach and education plan to allow additional end user groups to take advantage of our work.

How will the developed software be sustained beyond the lifetime of the project?

Building new collaborations is an additional objective of the DIANA/HEP project. The “vision for DIANA extends beyond the work the project team will do itself. We will collaborate with other physicists and data scientists from universities, national, and international labs as well as the private sector to catalyze a broad effort.”

Similarly “building community” around the software is critical and a primary metric for success.

We have in fact been doing both this year. We should continue to build both collaborations and communities. This should not be incidental or “passive”: activities in this direction (not just an occasional talk!) should be an integral part of the plan.

We proposed three interrelated areas of activity:

- ▶ Performance
- ▶ Interoperability
- ▶ Collaborative Analysis

**Performance:** we will greatly increase parallelism and eliminate CPU- and IO-bottlenecks to achieve higher processing rates necessary to efficiently and expeditiously analyze large volumes of data. We will address some key design and implementation issues from the early days of ROOT which impact not only performance, but also the manageability of the software.

**Interoperability:** we will reposition these key libraries to better interoperate with the larger scientific software ecosystem, transitioning the field to a more sustainable path where *new ideas and software developed elsewhere can be more easily used in particle physics* and our best products can be evaluated by other fields. We will create modular versions of the libraries that work both within the traditional ROOT framework as well as within other frameworks such as Hadoop MapReduce, Apache Spark, Mathematica, Python and R.

**Collaborative Analysis:** we will provide new tools that build on the concept and emerging practices in particle physics that data analysis is a collaborative activity, involving many individuals working within a given experiment, working in different experiments and even between the experimental and theory communities. This will involve directly integrating into the analysis tool suite the ability to capture elements of analysis workflows needed to satisfy best practices in data preservation, analysis archival, reproducibility, and open access.

The DIANA project was established with an Advisory Board (AB), see names and links to biographies on our DIANA team page:

<http://diana-hep.org/pages/team.html>

Our first meeting with them will be on 26 January, 2017, just after the HSF workshop at SDSC/UCSD

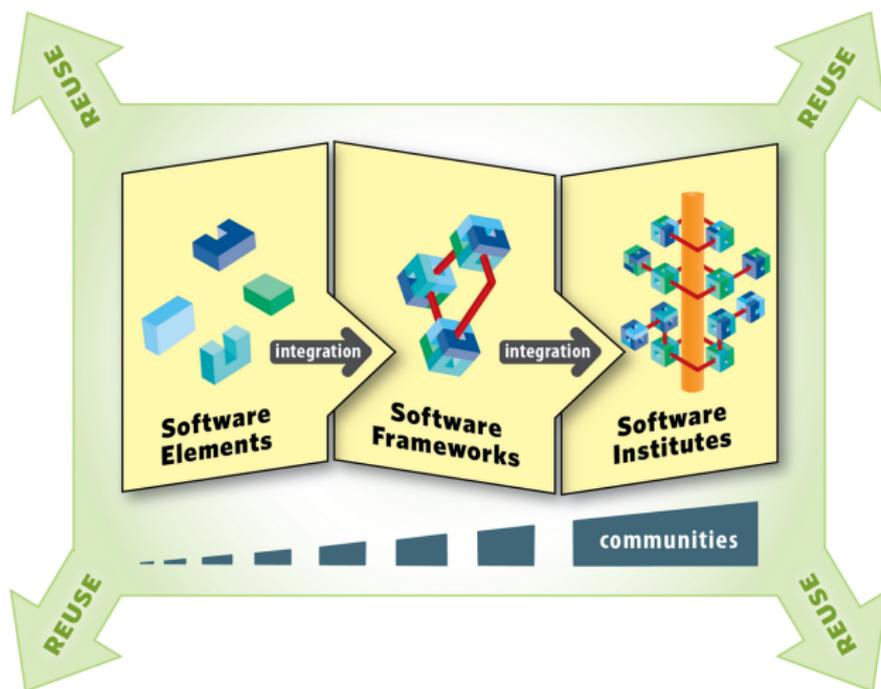
This will be our opportunity to explain not only our goals, but also the current plan for achieving those goals and the activities we are pursuing.

We will aim to meet with the DIANA AB twice/year.

We have had the idea of a DIANA blog since some time. The original idea was that we could have some content generated by DIANA team members and some “invited” content.

The current website has a placeholder for this, although it has only really been used for a few “news” items.

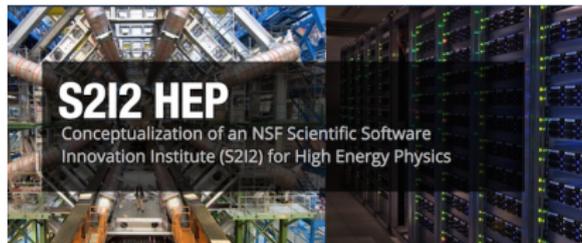
Should we pursue this? More generally should we aim for changes to the website?



In addition to SSI/Software-Framework projects (like DIANA), NSF has the concept of a "Software Institute" (S2I2)

NSF funded two S2I2 institutes in 2016: a Science Gateways institute (<http://sciencegateways.org>) and a Molecular Sciences Institute (<http://molssi.org>). The also funded a “conceptualization” (planning) project for a possible S2I2 institute for HEP (<http://s2i2-hep.org>).

[https://www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=189347](https://www.nsf.gov/news/news_summ.jsp?cntn_id=189347)



The DIANA project team should participate in the S2I2-HEP planning project.

- ▶ DIANA fellows (Mike's talk)
- ▶ Science Gateways S2I2
- ▶ Google Summer of Code (GSoC) will be coming up soon. Do we have projects and/or possible candidate students?
- ▶ DIANA topical meetings
- ▶ We should communicate clearly to the ROOT team how we fit into or add to the ROOT program of work:  
<https://docs.google.com/document/d/1nwyS50ff5vILVjvA7QdHWwYipJxJk5zI1fPw8M7NknA/edit>
- ▶ HEP analysis ecosystem workshop:  
[https://docs.google.com/document/d/1aAGCj\\_y9in\\_I-c9yYJ-XX3Qurf0PXH4tFoYmvuCY5tk/edit](https://docs.google.com/document/d/1aAGCj_y9in_I-c9yYJ-XX3Qurf0PXH4tFoYmvuCY5tk/edit)

- ▶ ROOT IO Weekly Meetings and Workshops
- ▶ ROOT team meeting  
(<https://indico.cern.ch/category/526/>)
- ▶ Intel Parallel Computing Center (IPCC project): Vassil Vassilev
- ▶ Conferences/Workshops:
  - ▶ ACAT 2017  
(<https://indico.cern.ch/event/567550/>)
  - ▶ others? non-HEP?
  - ▶ SI2 PI Meeting - 21-22 Feb, 2017 (Poster required!)
- ▶ iPad App, Masterclasses, training, Community Engagement Survey
- ▶ DASPOS, AAA, various NSF PIF projects
- ▶ email

In this presentation I've shown the “big picture” goals we have for the DIANA project, as well as the NSF SI2 program context.

After the presentations we should ask ourselves specifics about how our activities and plans are fully aligned with those goals, whether there are missing elements and discuss a common DIANA project plan for 2017.

We should probably meet again in mid-January to iterate on the full project plan for 2017 and prepare for the DIANA AB meeting in San Diego.