• The primary goal of DIANA/HEP is to develop state-of-the-art tools for experiments which acquire, reduce, and analyze petabytes of data.

• DIANA is not a piece of software itself, but a collaborative project to improve and extend analysis tools as sustainable infrastructure for the community.

• DIANA is a 4 year project, 6-7 FTE spread over 4 universities (Princeton, NYU, U.Cincinnati, U.Nebraska-Lincoln)
DIANA/HEP is part of the NSF SI2 program

• Not just software development, but part of a larger set of strategic goals:

- **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.

• Need to build only software, but also better structures for collaboration, career paths, education, etc.
The SI2 program includes four classes of awards:

1. **Scientific Software Elements (SSE)**: SSE awards are Software Elements. They target small groups that will create and deploy robust software elements for which there is a demonstrated need that will advance one or more significant areas of science and engineering.

2. **Scientific Software Integration (SSI)**: SSI awards are Software Frameworks. They target larger, interdisciplinary teams organized around the development and application of common software infrastructure aimed at solving common research problems. SSI awards will result in sustainable community software frameworks serving a diverse community.

3. **Scientific Software Innovation Institutes (S2I2)**: S2I2 awards are Software Institutes. They focus on the establishment of long-term hubs of excellence in software infrastructure and technologies that will serve a research community of substantial size and disciplinary breadth.

4. **Reuse**: In addition, SI2 provides support through a variety of mechanisms (including co-funding and supplements) to proposals from other programs that include, as an explicit outcome, reuse of software. Proposals that integrate with previously developed software, either by reference or inclusion, are encouraged. Proposals developing new software with an explicitly open design for reuse may also be considered. The purpose of the Reuse class is to stimulate connections within the broader software ecosystem. The class of reuse awards is currently being developed.

[DIANA is an SSI](http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504817)
DIANA is about (cross-experiment) analysis tools. Grant runs 2015-2019. We have broad areas of activity and goals:

- **performance**: ROOT I/O, vectorization, …
- **interoperability**: scientific python ecosystem, R, hadoop, spark, …
- **collaborative tools & reproducibility**: RooFit workspace, HEPdata, CAP

**Approach:**

- Specific focus is meant to be coordinated with needs of experiments.
- Of course ROOT sits at the center of the analysis tools ecosystem in HEP, thus are collaborating directly with ROOT team (and others).

As part of the NSF's Software Infrastructure for Sustained Innovation (SI2) program, DIANA is concerned with the overarching goal of transforming innovations in research and education into sustained software resources that are an integral part of the cyberinfrastructure.
Collaborative Analyses
Establish infrastructure for a higher-level of collaborative analysis, building on the successful patterns used for the Higgs boson discovery and enabling a deeper communication between the theoretical community and the experimental community.

Reproducible Analyses
Streamline efforts associated to reproducibility, analysis preservation, and data preservation by making these native concepts in the tools.

Interoperability
Improve the interoperability of HEP tools with the larger scientific software ecosystem, incorporating best practices and algorithms from other disciplines into HEP.

Faster Processing
Increase the CPU and IO performance needed to reduce the iteration time so crucial to exploring new ideas.

Better Software
Develop software to effectively exploit emerging many- and multi-core hardware. Promote the concept of software as a research product.

Training
Provide training for students in all of our core research topics.

Design by Eamonn Maguire, CERN fellow, HEPdata developer
PROJECT TEAM

Peter Elmer (Lead PI) - Princeton, CMS

Brian P. Bockelman (PI) - University of Nebraska-Lincoln, CMS

Kyle Cranmer (PI) - NYU, ATLAS

Michael D. Sokoloff (PI) - Cincinnati, LHCb

Jinyang Li (Senior Personnel) - New York University, Computer Science Department

David Lange - Princeton, CMS co-coordinator Offline Software and Computing

Gilles Louppe - NYU, ATLAS, Machine Learning PhD (former CERN fellow), scikit-learn developer

Jim Pivarski - Princeton, CMS, ROOT interoperability with Hadoop, Spark, etc.

Eduardo Rodrigues - Cincinnati, LHCb, coordinator analysis tools; tracking, trigger, and fitting

Zhe Zhang - Nebraska, Comp. Sci PhD student, improving ROOT IO performance

Chien-Chin Huang - NYU, Comp. Sci. PhD student, RooFit data parallelism, Theano, TensorFlow, etc.

Note: Gilles Louppe started fall 2015, others just started
DIANA FELLOWS

Each year, 4 DIANA Graduate Fellows will each spend 3 months intensively developing tools in conjunction with collaborating institutions.

- call for applications will go out soon

Similarly, a DIANA Undergraduate Fellow will work 10 - 12 weeks during the summer, either developing or using data-intensive tools.
DIANA topical meetings

• A forum for presentations and discussion about analysis techniques and analysis tools, of relevance to the broader HEP community

• These meetings are meant to explore near and long term possibilities, ideas and collaborations. We hope to engage people from a number of experiments and from beyond HEP.

• In the steady state we expect approx. 2 meetings per month. If you have ideas, please contact us or bring them up in the meetings.
DIANA topical meetings (Monday 17:30 GVA)

The **DIANA/HEP project** focuses on improving performance, interoperability, and collaborative tools through modifications and additions to ROOT and other packages broadly used by the HEP community.

**June 2016**
- 06 Jun  **DIANA Meeting - HDF5 File Format (TBC)**

**May 2016**
- 23 May  **DIANA Meeting - Python/ROOT interoperability (TBC)**
- 16 May  **DIANA Meeting - MC generation and numerical integration (TBC)**
- 02 May  **DIANA Meeting - the "Big Data" Ecosystem (TBC)**

**April 2016**
- 25 Apr  **DIANA Meeting - Histogram primitives and map-reduce (TBC)**
- 18 Apr  **DIANA Meeting - HEPData (TBC)**
- 11 Apr  **DIANA Meeting - Bayesian Optimisation**

Google group:  diana-hep@googlegroups.com
GITHUB ORGANIZATION

https://github.com/diana-hep