Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)

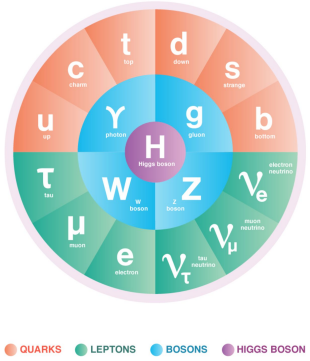
PI: Peter Elmer (Princeton), co-PIs: Brian Bockelman (Morgridge Institute), Gordon Watts (U.Washington) with UC-Berkeley, University of Chicago, University of Cincinnati, Cornell University, Indiana University, MIT, U.Michigan-Ann Arbor, U.Nebraska-Lincoln, New York University, Stanford University, UC-Santa Cruz, UC-San Diego, U.Illinois at Urbana-Champaign, U.Puerto Rico-Mayaguez and U.Wisconsin-Madison

http://iris-hep.org

IRIS-HEP was funded as of 1 September, 2018, and is currently ramping up activities

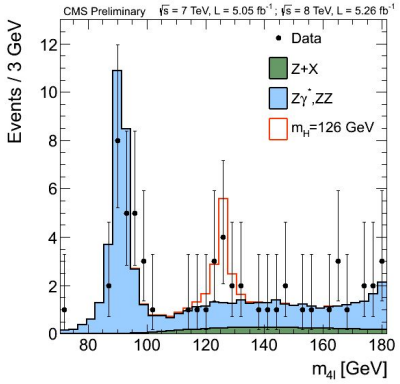
OAC-1836650

Science Driver: Discoveries beyond the Standard Model of Particle Physics

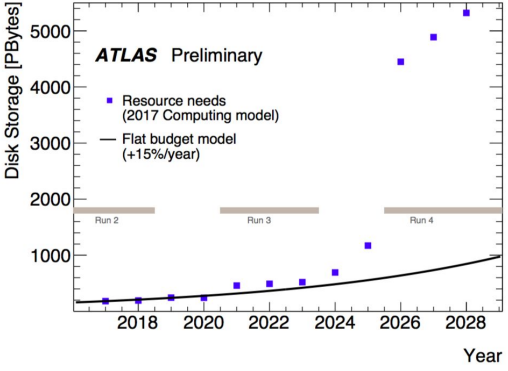


From "Building for Discovery - Strategic Plan for U.S. Particle Physics in the Global Context" - Report of the Particle Physics Project Prioritization Panel (P5):

1) Use the Higgs boson as a new tool for discovery
2) Pursue the physics associated with neutrino mass
3) Identify the new physics of dark matter
4) Understand cosmic acceleration: dark matter and inflation
5) Explore the unknown: new particles, interactions, and physical principles

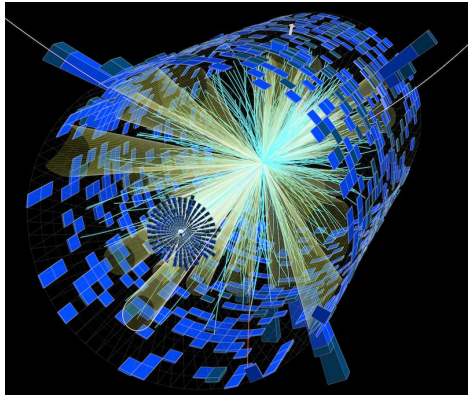


Computational and Data Science Challenges of the High Luminosity Large Hadron Collider (HL-LHC) and other HEP experiments in the 2020s

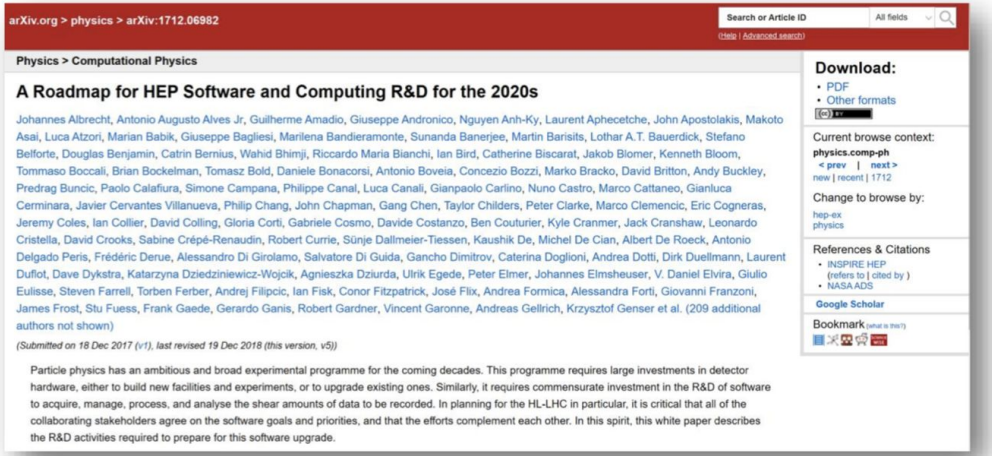


The HL-LHC will produce exabytes of science data per year, with increased complexity: an average of 200 overlapping proton-proton collisions per event.

During the HL-LHC era, the ATLAS and CMS experiments will record ~10 times as much data from ~100 times as many collisions as were used to discover the Higgs boson (and at twice the energy).



The Community White Paper Process (2016-2017)



Involved A Diverse

- Computing Management from the Experiments and Labs
- Individuals interested in the problems
- Members of other compute intensive scientific endeavors
- Members of Industry

Individual Papers on the arXiv:

Careers & Training, Conditions Data, DOMA, Data Analysis & Interpretation, Data and Software Preservation, Detector Simulation, Event/Data Processing Frameworks, Facilities and Distributed Computing, Machine Learning, Physics Generators, Security, Software Development, Deployment, Validation, Software Trigger and Event Reconstruction, Visualization

Community White Paper & the Strategic Plan

IRIS-HEP

Sustainable Software R&D objectives

1) Development of **innovative algorithms** for data reconstruction and triggering;

2) Development of highly performant **analysis systems** that reduce “time-to-insight” and maximize the HL-LHC physics potential; and

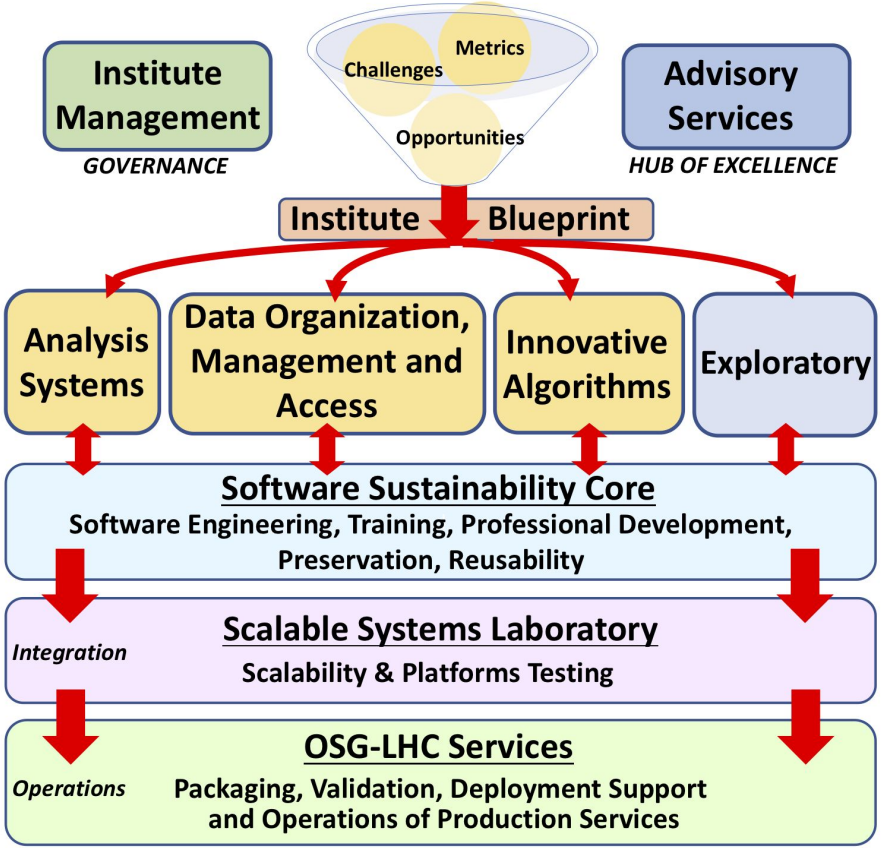
3) Development of **data organization, management and access systems** for the community’s upcoming Exabyte era.

4) Integration of software and scalability for use by **the LHC community on the Open Science Grid**, the Distributed High Throughput Computing infrastructure in the U.S.

Intellectual Hub for the HEP Community

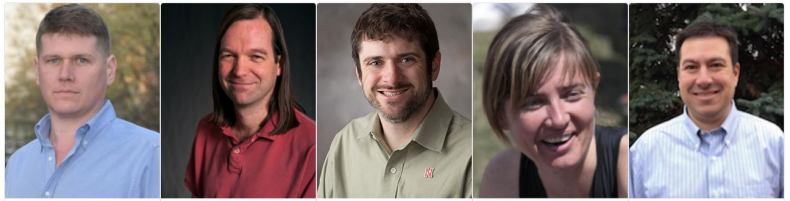
The plan for IRIS-HEP reflects a community vision developed by an international community process organized by the HEP Software Foundation (https://hepsoftwarefoundation.org). The S2I2-HEP conceptualization project (http://s2i2-hep.org) derived a Strategic Plan from the community roadmap which would leverage the strengths of, and could be executed by, the U.S. university community. This became a proposal to the National Science Foundation, which funded a 5 year project as of 1 September, 2018.

IRIS-HEP Structure and Executive Board



Executive Board

The IRIS-HEP Executive Board manages the day to day activities of the Institute.



Peter Elmer
Princeton University
Peter.Elmer@cern.ch
Institute PI and Executive Director

Gordon Watts
University of Washington
Institute co-PI and Deputy Executive Director

Brian Bockelman
Morgridge Institute
Institute co-PI and DOMA R&D Area Lead

Heather Gray
University of California, Berkeley
Innovative Algorithms Area co-Lead

David Lange
Princeton University
David.Lange@cern.ch
Innovative Algorithms Area co-Lead

Kyle Cranmer
New York University
Analysis Systems Area Lead

Sudhir Malik
University of Puerto Rico at Mayaguez
Training, Education and Outreach Coordinator

Mark Neubauer
University of Illinois at Urbana-Champaign
Blueprint Coordinator

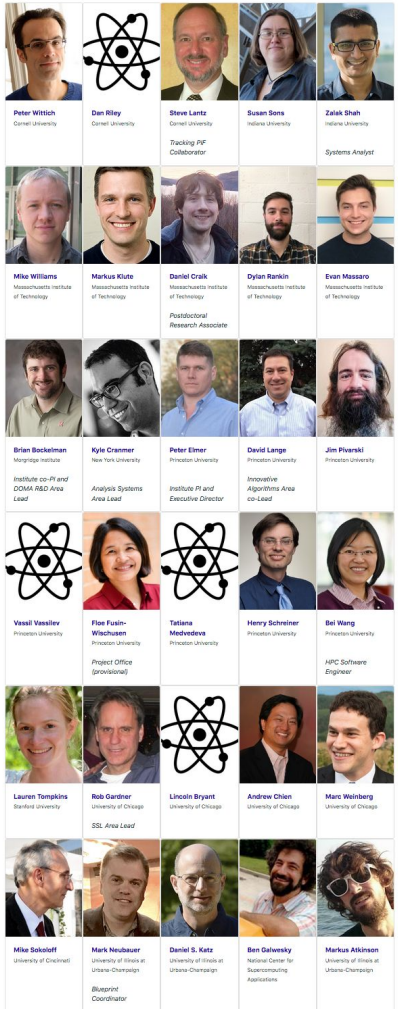
Rob Gardner
University of Chicago
SSL Area Lead

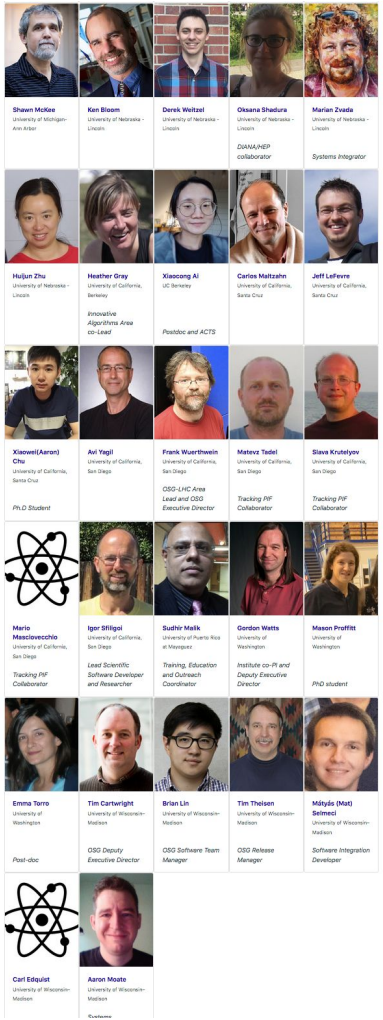
Frank Wuerthwein
University of California, San Diego
OSG-LHC Area Lead and OSG Executive Director

IRIS-HEP Team

http://iris-hep.org/about/team

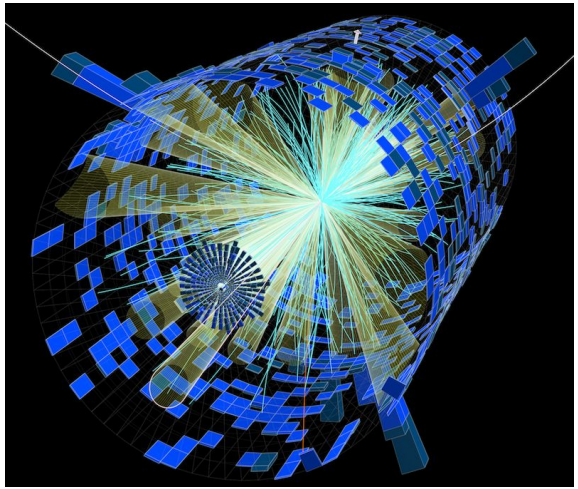
About 30 FTEs spread over a larger of people from 18 institutions

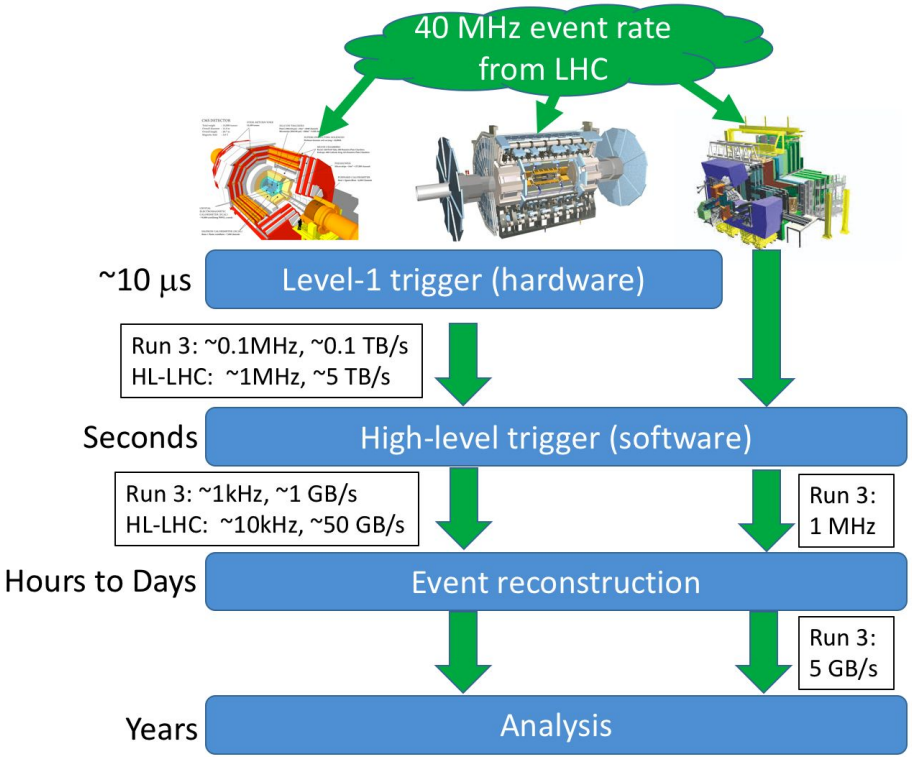




Innovative Algorithms - Trigger/Reconstruction

Algorithms for real-time processing of detector data in the software trigger and offline reconstruction are critical components of HEP's computing challenge.





Innovative Algorithms

These algorithms face a number of new challenges during HL-LHC:

- Upgraded accelerator capabilities, with more collisions per bunch crossing (“pile-up”)
- Detector upgrades, including new detector technologies and capabilities Increased event rates to be processed
- Emerging computing architectures

Specific R&D investments include collaboration with ACTS, continuation of the parallel Kalman Filter tracking project, work on HLS4ML, etc.

Data Organization, Management and Access (DOMA)

The DOMA focus area performs fundamental R&D related to the central challenges of organizing, managing, and providing access to exabytes of data from processing systems of various kinds.

- **Data Organization**: Improve how HEP data is serialized and stored.
- **Data Access**: Develop capabilities to deliver filtered and transformed event streams to users and analysis systems.
- **Data Management**: Improve and deploy distributed storage infrastructure spanning multiple physical sites. Improve inter-site transfer protocols and authorization

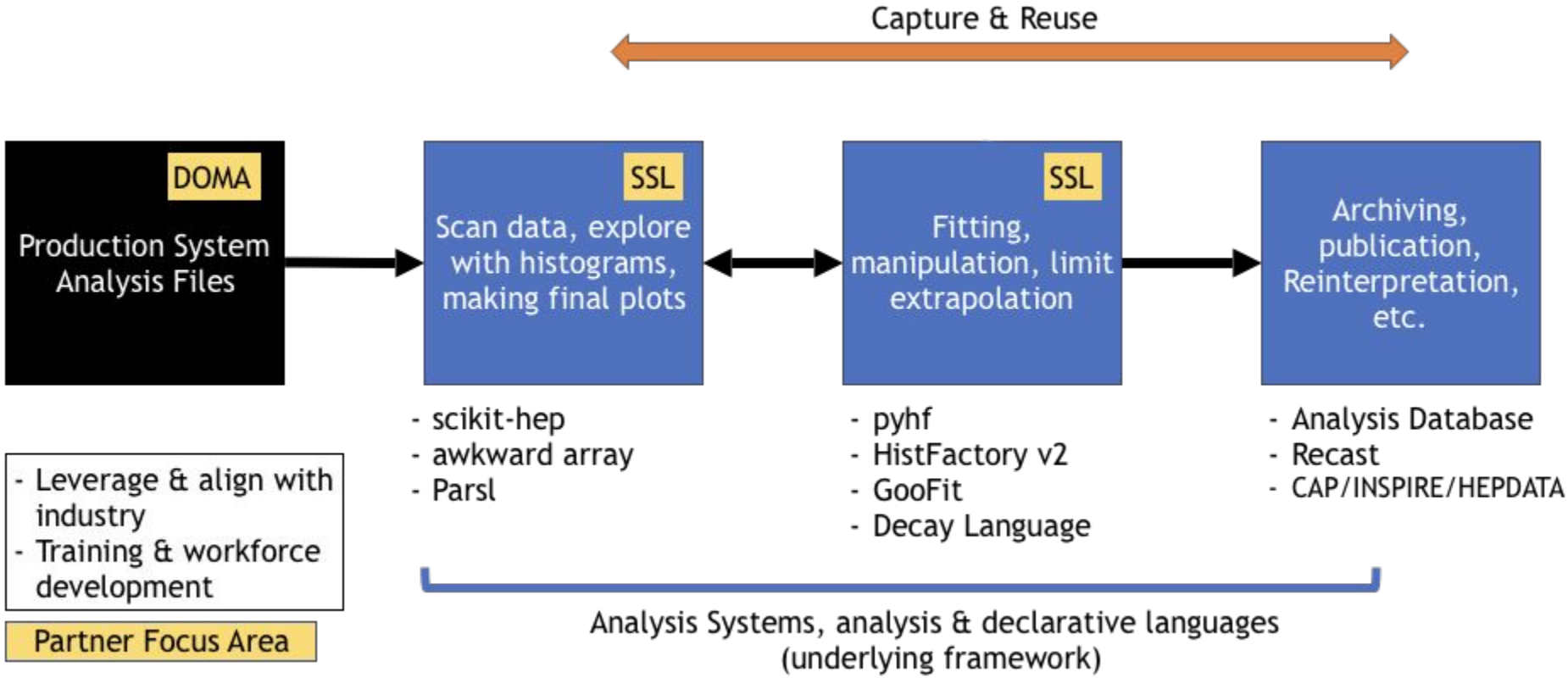
Analysis Systems R&D Goals

Develop sustainable analysis tools to extend the physics reach of the HL-LHC experiments by creating greater functionality, reducing time-to-insight, lowering the barriers for smaller teams, and streamlining analysis preservation, reproducibility, and reuse.

Compared to DOMA and Innovative Algorithms (which has more targeted reco/trigger goals), the Analysis Systems group is dealing with more “greenfield” area where there is a very heterogeneous set of use cases and relevant components.

The nature of IRIS-HEP Analysis Systems tasks is more exploratory and “big R” (R&d)

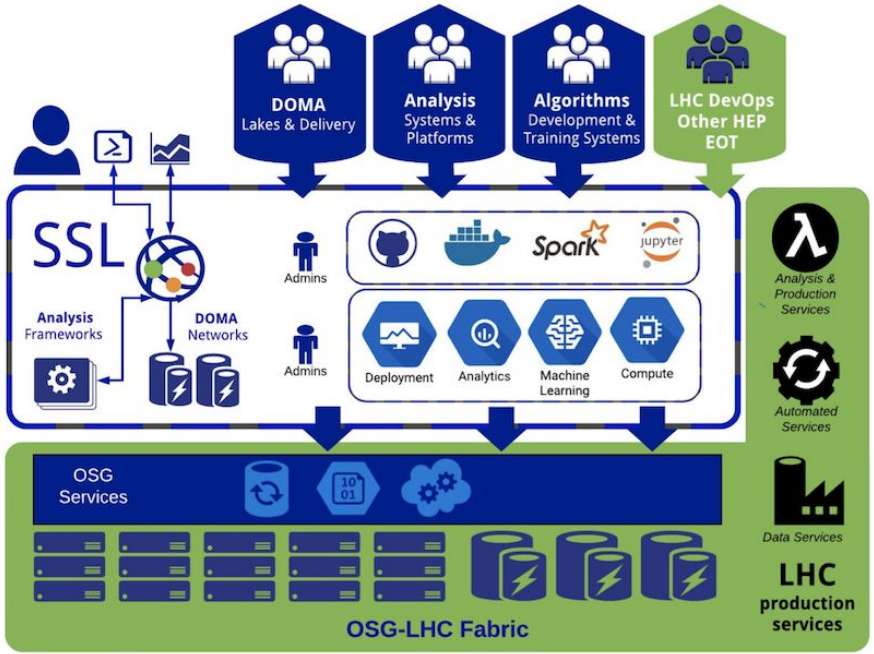
Analysis Systems Data Flow and Projects



Scalable Systems Laboratory (SSL)

Goal: Provide the Institute and the HL-LHC experiments with scalable platforms needed for development in context

- Provides access to infrastructure and environments
- Organizes software and resources for scalability testing
- Does foundational systems R&D on accelerated services
- Provides the integration path to the OSG-LHC production infrastructure

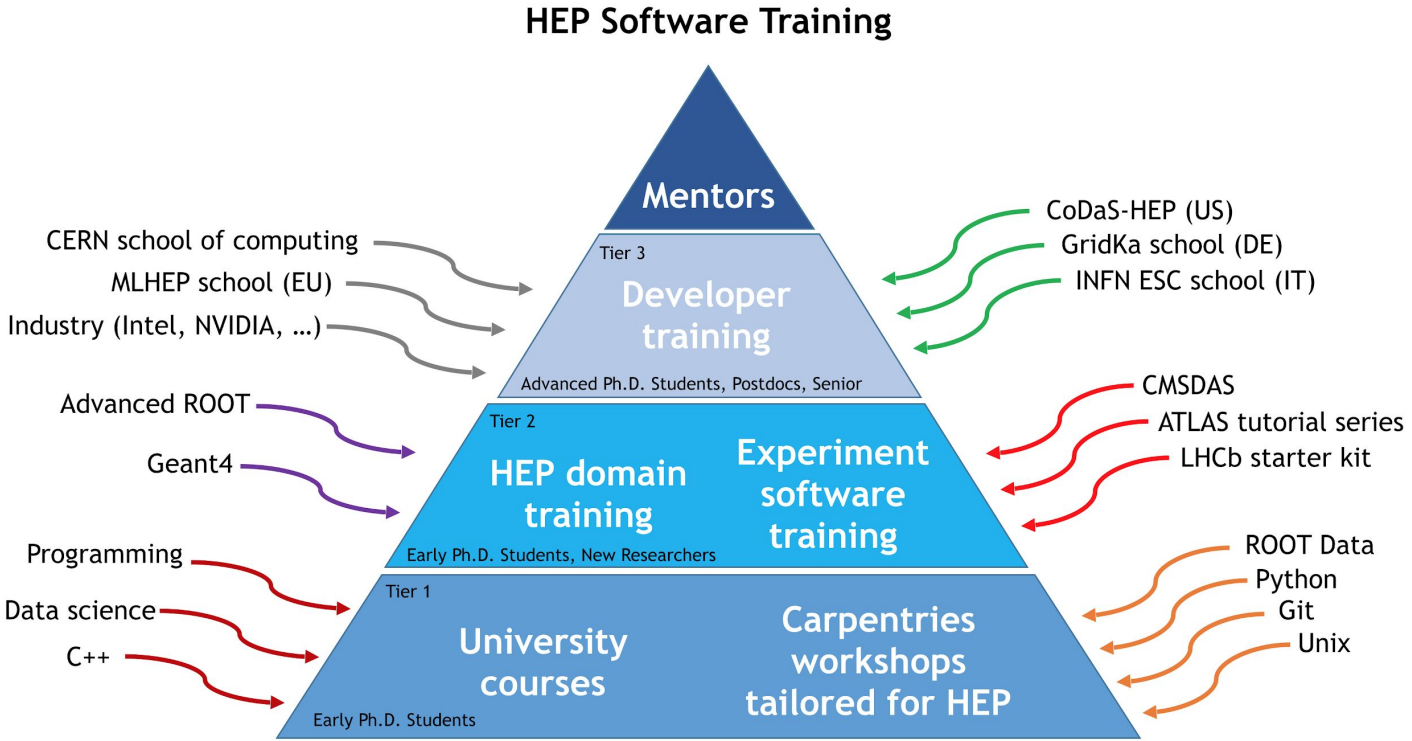


Open Science Grid (OSG) for the LHC

A total of 6 FTE across 11 people.
These people have worked together
and with the LHC program years.

(Slide from FKW)

Training and Education - Sustainability/Scalability



This is a general framework for training, but from the NSF we have funds from both IRIS-HEP (OAC-183665) and a separate project FIRST-HEP (OAC-1829707, OAC-1829729, http://first-hep.org) which can work towards implementing this model.

Training and Education - Upcoming events

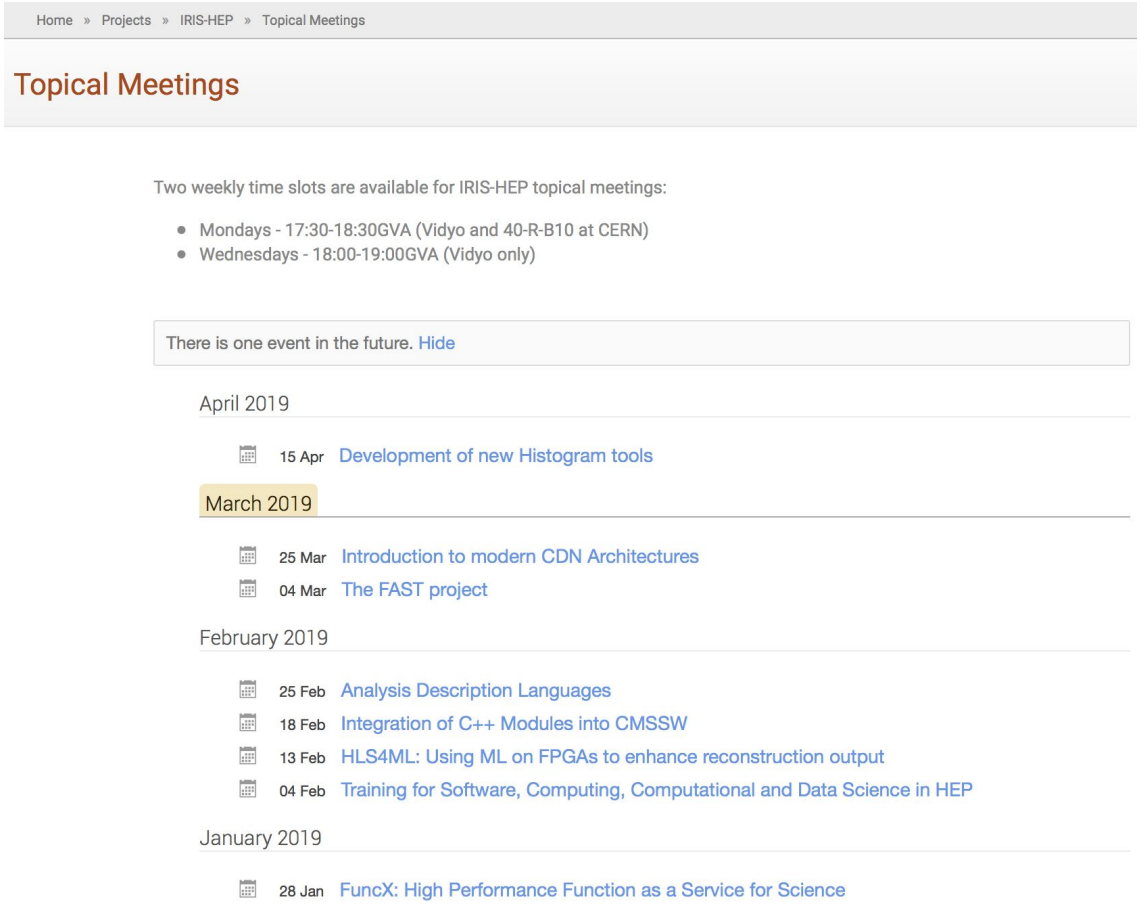
We are organizing a very first introductory level Software Carpentry (git/python/PyRoot/uproot) workshop at Fermilab (1-2 April 2019)

We are organize the CoDaS-HEP School (Princeton) - 22-26 July 2019 - tools, techniques and methods for Computational and Data Science for High Energy Physics (http://codas-hep.org)

With US-ATLAS colleagues we will be organizing an introductory “first test run” level Carpentry session for US-ATLAS at LBNL and CMS/ATLAS at CERN, the latter connecting to the international parts of the experiment, build on the experience of the FNAL workshop.

Invite LHCb/ALICE to showcase Starterkit on pre- and first analysis steps

IRIS-HEP Topical Meeting Series



https://indico.cern.ch/category/10570/

Meetings are announced on the announcements@iris-hep.org mailing list

Recorded videos are available in Youtube (see links on the individual agenda pages)

Connecting with IRIS-HEP

Website: http://iris-hep.org

Public announcement mailing list: announcements@iris-hep.org [Subscribe]

Topical meetings: https://indico.cern.ch/category/10570/

We will be organizing, co-organizing and hosting various events going forward, see the main webpage above.

Summary

HEP faces major challenges in the 2020s: Data, Compute, Staffing

The HSF executed an important community process that produced the CWP.

It was great to see this collaborative spirit continue here at HOW2019 and bodes well for us as a community to meet those challenges.

IRIS-HEP

We are focusing on 3 R&D areas from the CWP: Innovative Algorithms, Analysis Systems, and DOMA.

Plus training, a dedicated integration activity and continuity for the OSG services for the LHC.

We are just beginning our activities and are looking forward to collaborating with many of you in the coming years!