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, 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

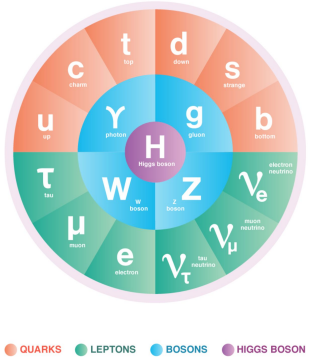
IRIS-HEP was funded as of 1 September, 2018, and is currently ramping up activities

OAC-1836650

And Introduction: My Background

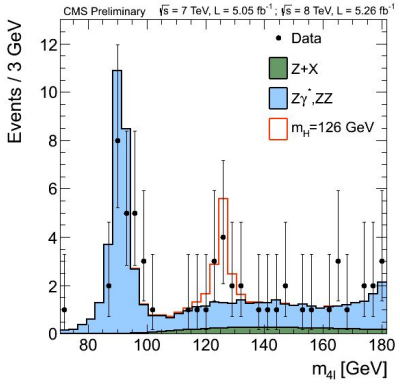
- PhD in experimental particle physics (Wisconsin, SLW)
- HEP Experiments: ALEPH (CERN), BaBar (SLAC), CMS (CERN)
- Princeton staff since 2001, based in Geneva, Switzerland
- I've had a variety of software/computing roles in BaBar and CMS software and computing over the years. These have been mostly on the development side, but (in particular in BaBar) occasionally in operations
- I played a major role in computing model updates in BaBar (2003-2004) and CMS (2005-2007), as well as subsequent evolution in CMS
- Things I have not done in the past: manage facilities myself or direct involvement in OSG!

Science Driver: Discoveries beyond the Standard Model of Particle Physics

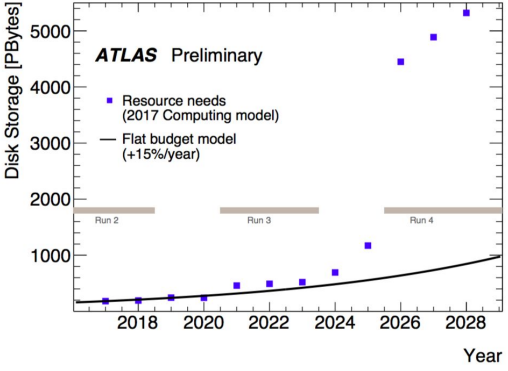


From "Building for Discovery - Strategic Plan for U.S. Particle Physics in the Global Context" - Report of the Particle Physics Project Prioritization Panel (P5):

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

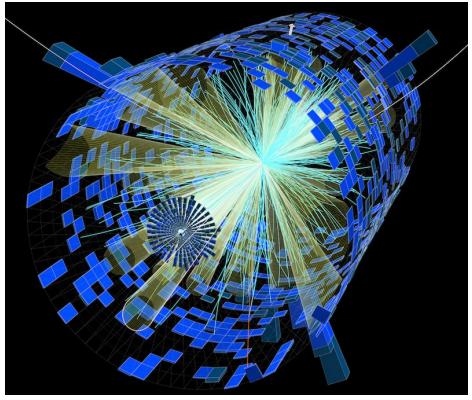


Computational and Data Science Challenges of the High Luminosity Large Hadron Collider (HL-LHC) and other HEP experiments in the 2020s



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).



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 the S2I2-HEP conceptualization project (http://s2i2-hep.org) and an international community process organized by the HEP Software Foundation (https://hepsoftwarefoundation.org).

IRIS-HEP will function as an intellectual hub for the national and international HEP community, through training, community workshops and the development of wider collaborations with the larger computer and data science communities.

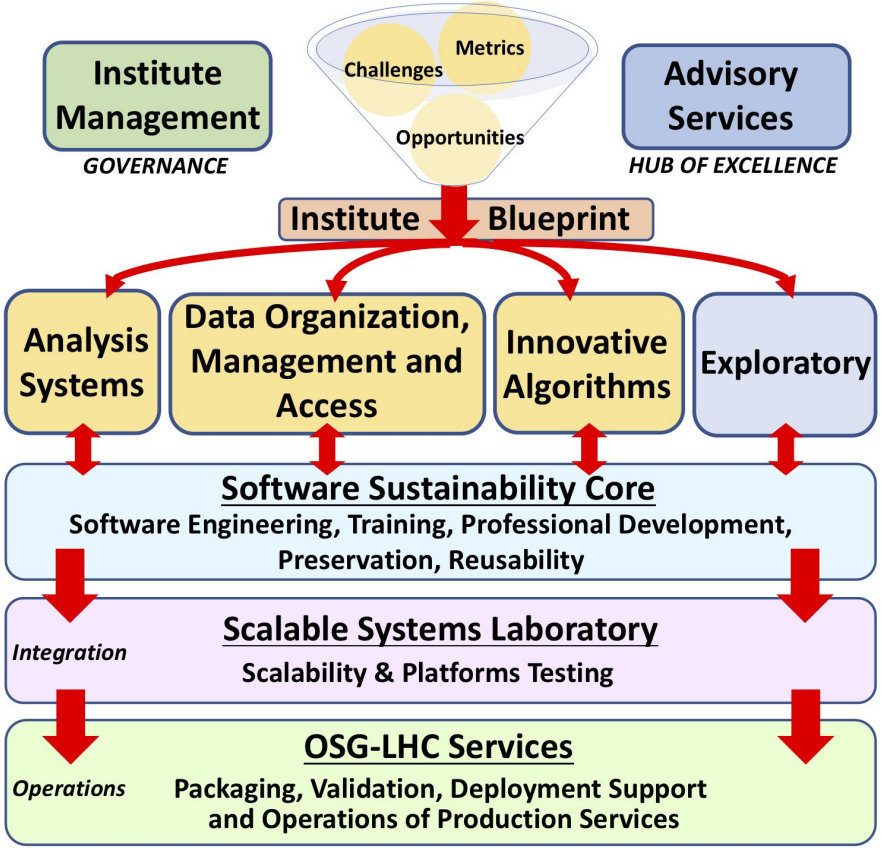
Strategic Plan and the Community Roadmap

IRIS-HEP implements the roadmap described in the S2I2-HEP Strategic plan (https://arxiv.org/abs/1712.06592), with an R&D plan which aims to:

- enable new approaches to computing and software that can radically extend the physics reach of the detectors; and
- achieve improvements in software efficiency, scalability, and performance, and to make use of the advances in CPU, storage, and and network technologies;
- ensure the long term sustainability of the software through the lifetime of the HL-LHC.

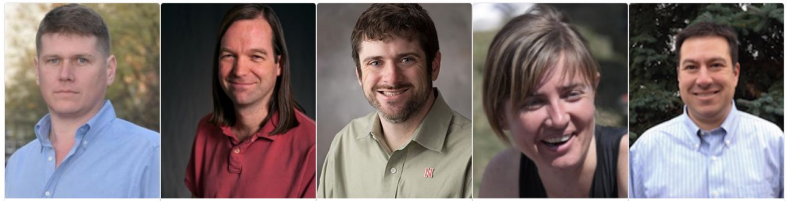
The S2I2-HEP Strategic Plan is a subset of the community roadmap described in the Community White Paper (https://arxiv.org/abs/1712.06982).

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

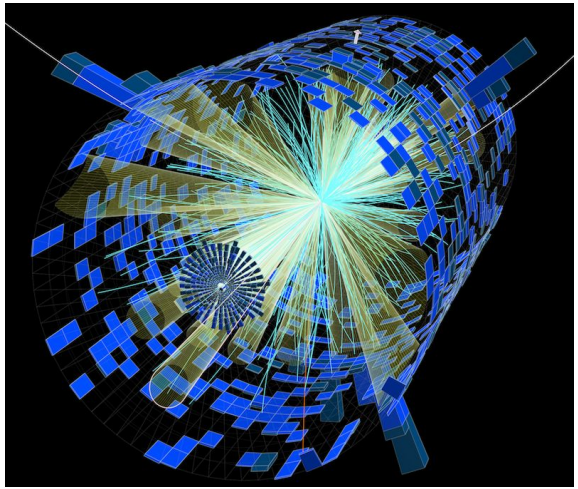
Mark Neubauer
University of Illinois at Urbana-Champaign
Blueprint Coordinator

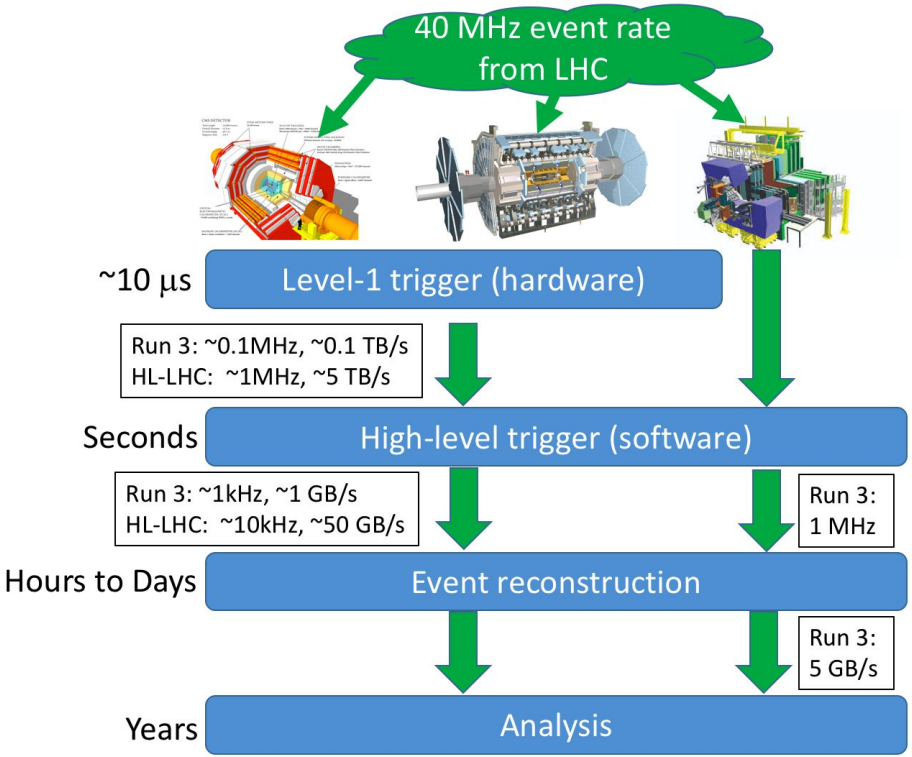
Rob Gardner
University of Chicago
SSL Area Lead

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

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

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.

- **Modeling**: Better understand LHC data usage patterns as a DOMA is organized into three areas:
- **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

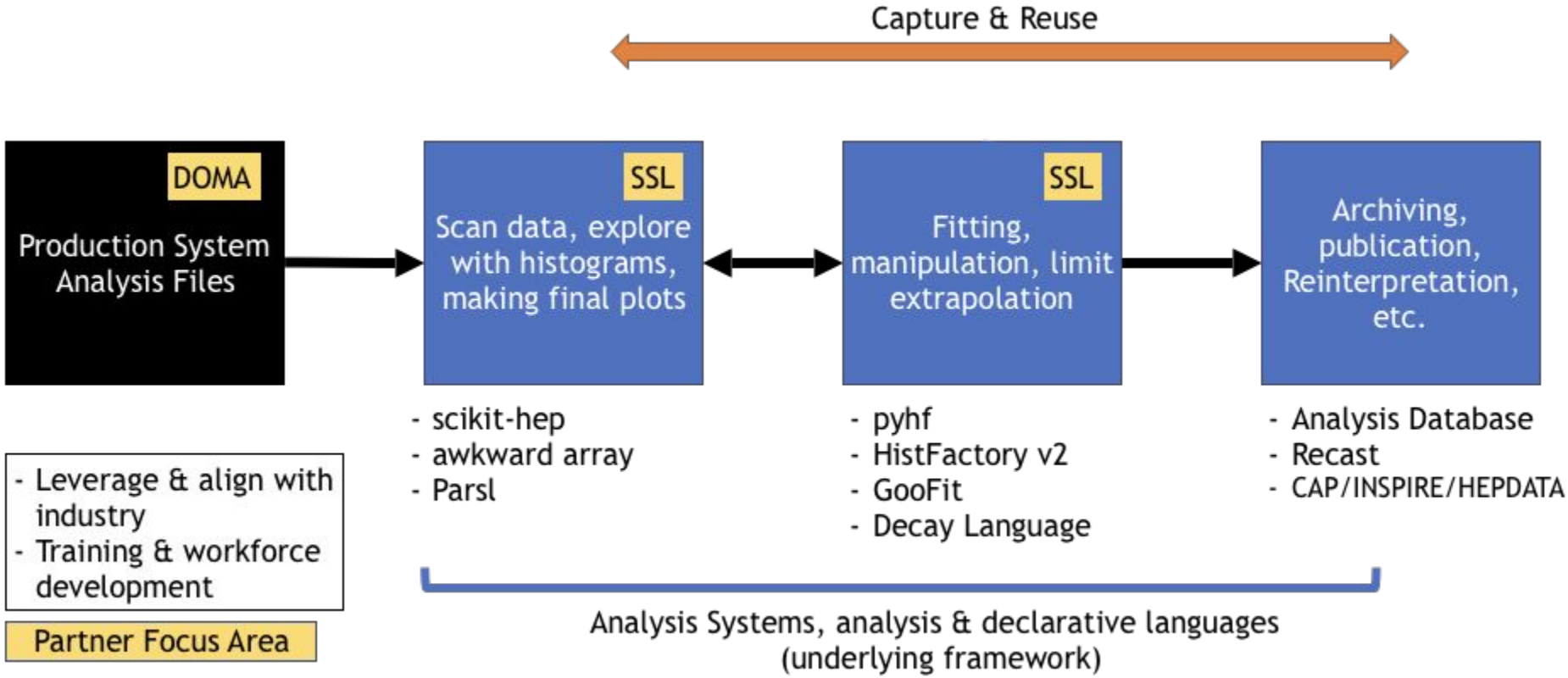
Analysis Systems R&D Goals

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)

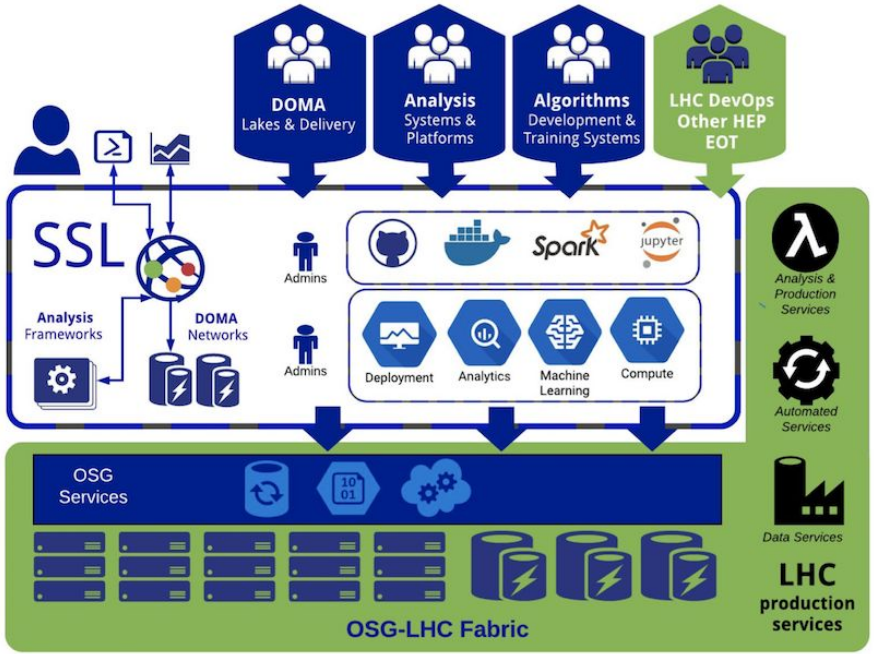
Analysis Systems Data Flow and Projects



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) and IRIS-HEP

OSG Effort Portfolio

A	B	C	D	E	F	G	H
	Σ	Technology	Operations	Support	Networking	Security	Management
Σ	17.56	7.46	4.05	2.82	1.24	1.00	0.98
OSG N5Y	6.86	2.48	2.05	1.82	0.00	0.20	0.32
IRIS-HEP	5.94	3.08	1.00	0.00	0.39	0.80	0.67
SAND	0.75	0.00	0.00	0.00	0.75	0.00	0.00
OSG-NP	1.00	0.50	0.50	0.00	0.00	0.00	0.00
TNRP	1.00	0.70	0.20	0.00	0.10	0.00	0.00
CESER	2.00	0.70	0.30	1.00	0.00	0.00	0.00

IRIS-HEP is roughly 1/3 of total effort in OSG

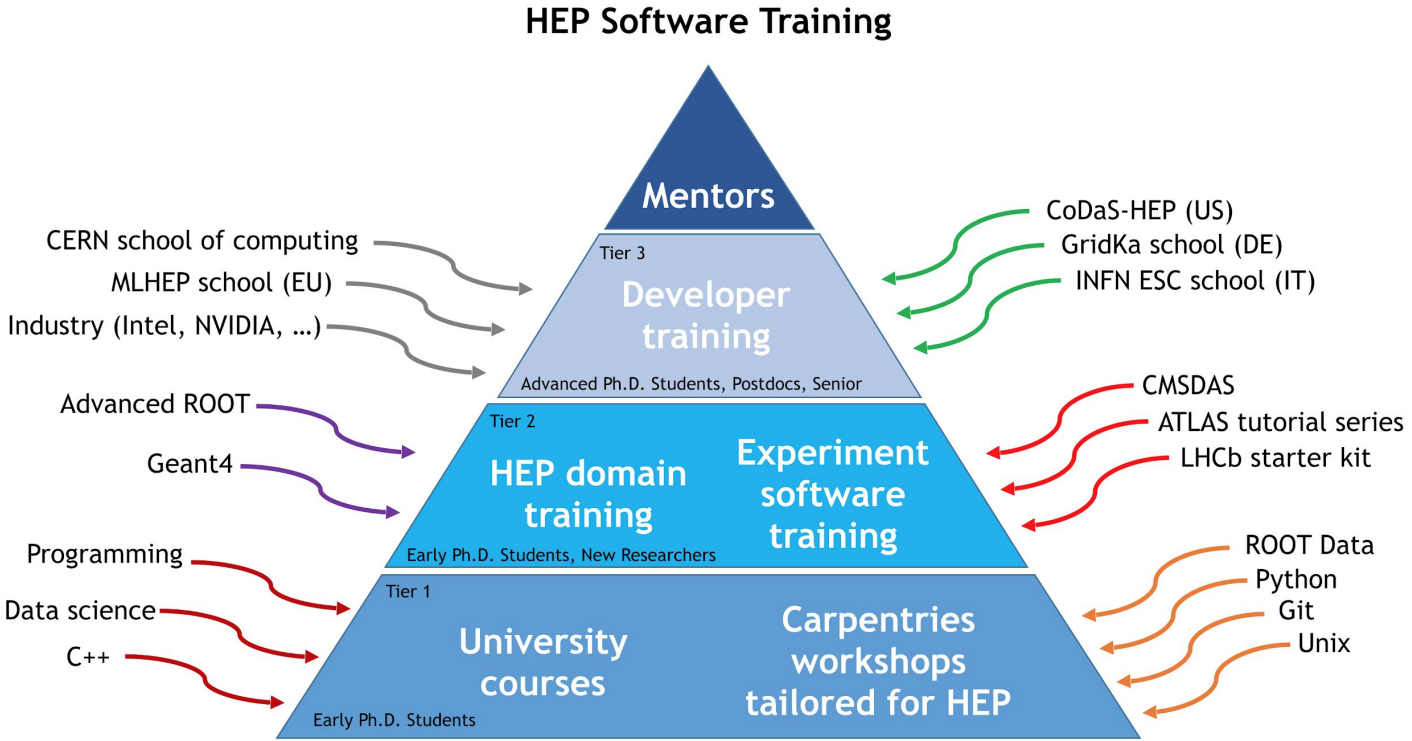
(Slide from FKW)

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



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.

IRIS-HEP and OSG

IRIS-HEP will provide to OSG:

- Continuity - The OSG is a successful working team. As it reorganizes around a different funding model, IRIS-HEP will provide effort to support the common services needed for the LHC, and as part of the larger team.

IRIS-HEP expects from OSG:

- (Hopefully mutually beneficial) engagement with our R&D activities which are exploring how we will accomplish our research program in the 2020s
- An environment in which we can engage with other scientific programs in the US on how cyberinfrastructure will evolve