

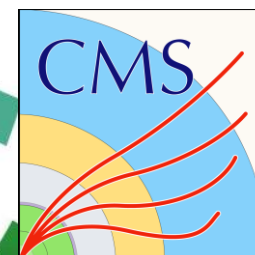


OAC-
1836650

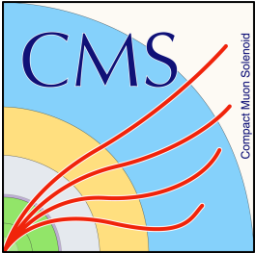
IRIS-HEP US CMS

PET MEETING

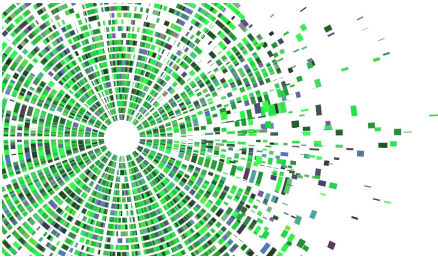
Gordon Watts



Two things...



First time I've ever put that on one of my talk's title pages!



Microsoft's PowerPoint AI suggested this background after typing "US CMS IRIS-HEP"

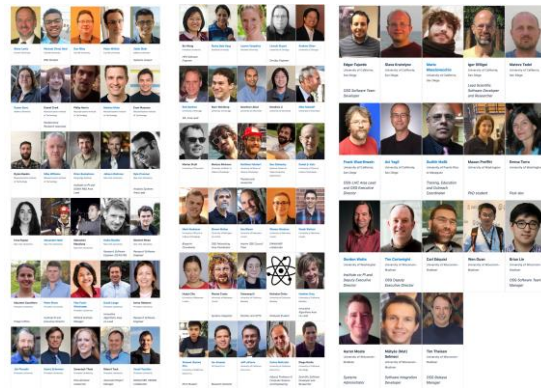
IRIS-HEP

EVERYONE PROBABLY KNOWS THIS

IRIS-HEP

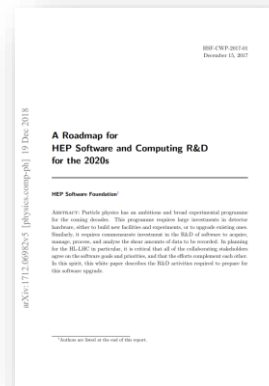
18 Universities across the USA

~28 FTE's
spread
across ~60
people



**Computational and data science
research to enable discoveries
in fundamental physics**

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



Designed to address key elements of
the [Roadmap for HEP Software and
Computing R&D for the 2020's](#).



IRIS-HEP

Institute Aims, in brief:

- Enable new approaches to computing and software that maximize, and potentially radically extend, the **physics reach** of the detectors.
- Make improvements in software efficiency, scalability and performance and make use of the advances in CPU, storage and network technologies, that allow the experiments to **maximize their physics reach within their computing budgets**.
- Significantly improve the long-term **sustainability** of the software through the lifetime of the HL-LHC.
- Play the role of **intellectual hub** for the larger software R&D effort required to ensure the success of the HL-LHC scientific program
- Lead research into deployment of the resulting systems on **distributed high throughput computing**.
- Work to **improve scientific software** more broadly and create **opportunities for a more diverse participation** in scientific software and computing.

Part of the measuring stick we use to decide what projects and challenges to take on



What are we up to?



iris hep

About ▾ Connect ▾ **Activities ▾** Jobs

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

COVID-19 pandemic: IRIS-HEP is responding to the COVID-19 pandemic.
Many IRIS-HEP researchers are also involved with **Science-Responds.org**, a larger community effort!

Computational and data science research to enable discoveries in fundamental physics

IRIS-HEP is a software institute funded by the National Science Foundation. It aims to develop the state-of-the-art software cyberinfrastructure required for the challenges of data intensive scientific research at the High Luminosity Large Hadron Collider (HL-LHC) at CERN, and other planned HEP experiments of the 2020's. These facilities are discovery machines which aim to understand the fundamental building blocks of nature and their interactions. [Full Overview](#)

News and Featured Stories:

IRIS-HEP Poster Session 2020 at Princeton University
IRIS-HEP held a Poster Session at Princeton

Analysis Preservation Bootcamp
The first Analysis Preservation Bootcamp was held at CERN from February 17-19, 2020. Thirty

Upcoming Events:

May 25-29, 2020 **DESY (Hamburg, Germany)**
ACTS Tracking for HEP Workshop

Jul 11-13, 2020 **Univ. of Texas at Austin (Austin, TX)**
PyHEP 2020 Workshop

[View all past events](#)

Upcoming Topical Meetings:

Apr 29, 2020
Zarr: big array management

May 4, 2020
Rumble: JSONiq (query language) on Spark

May 11, 2020
Accelerating Raw Data Analytics

[View all](#) • [Indico \(recordings\)](#) • [Vidyo room](#)

Related projects:









The ATLAS experiment • The Compact Muon Solenoid (CMS) experiment • The Large Hadron Collider beauty (LHCb)

<http://iris-hep.org>

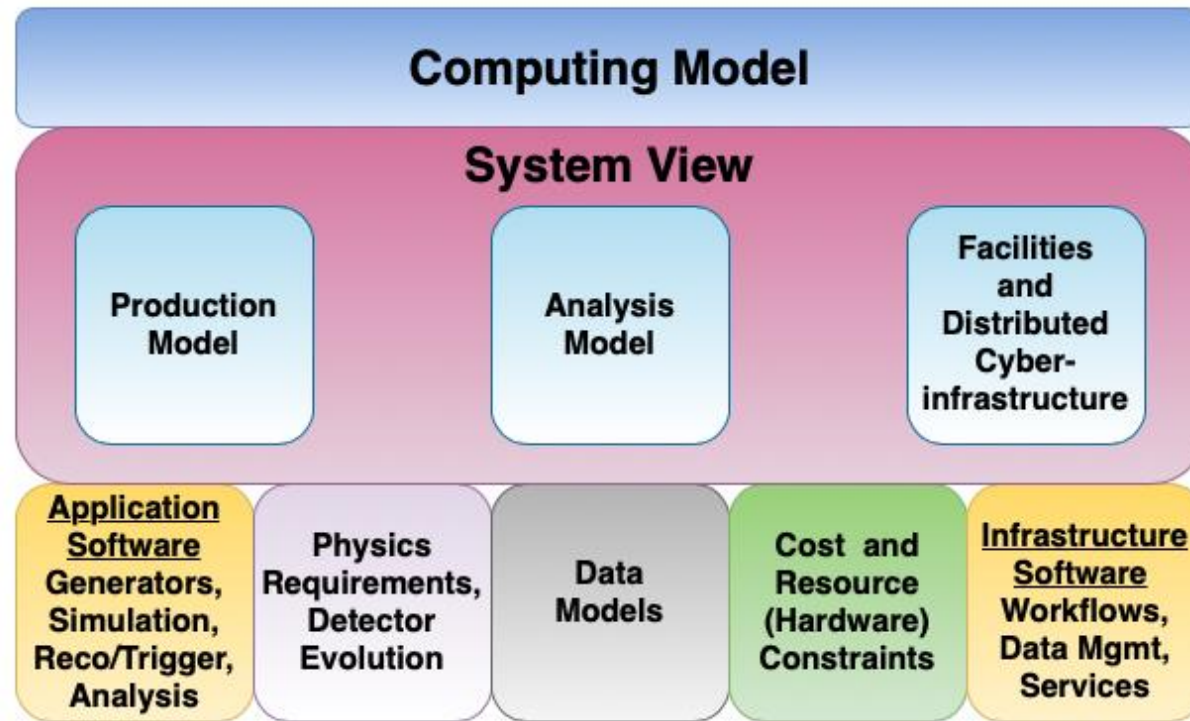
Details and contacts of all many of the projects we are involved in.

Contact us: as-team@iris-hep.org

AS Projects

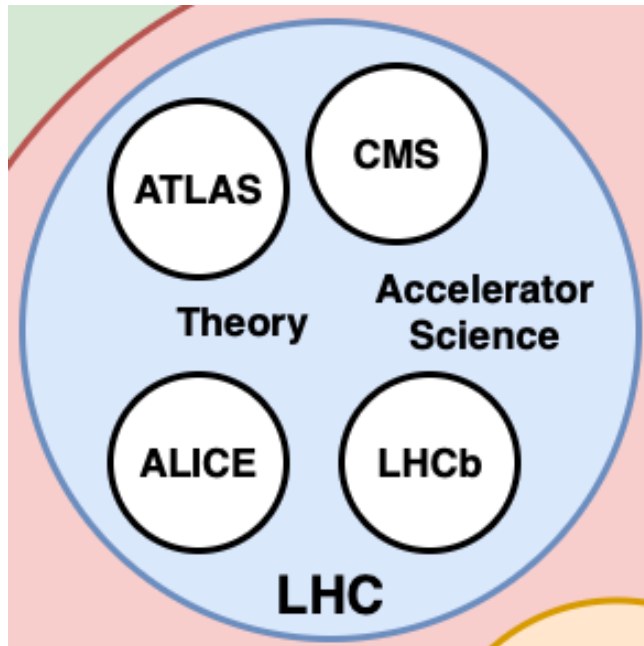
 ADL Benchmarks Functionality benchmarks for analysis description languages More information	 AmpGen Generation and fitting for multibody hadron decays More information	 awesome-hep A curated list of awesome high energy and particle physics software More information	 Awkward Array Manipulate arrays of complex data structures More information
 Decay Language Describe and convert particle decays More information	 exploratory-ml Analysis Reinterpretation More information	 Functional ADL Functional Analysis Description Language More information	 Boost Histogram Histogramming efforts More information

IRIS-HEP R&D Overview



As an organization we operate at all levels

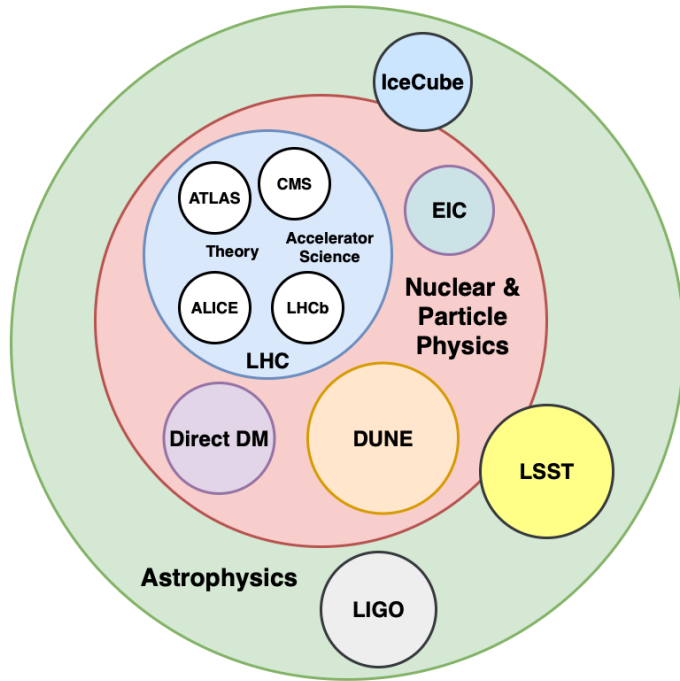
Multi Experiment



We strive for solutions useful to the LHC community*

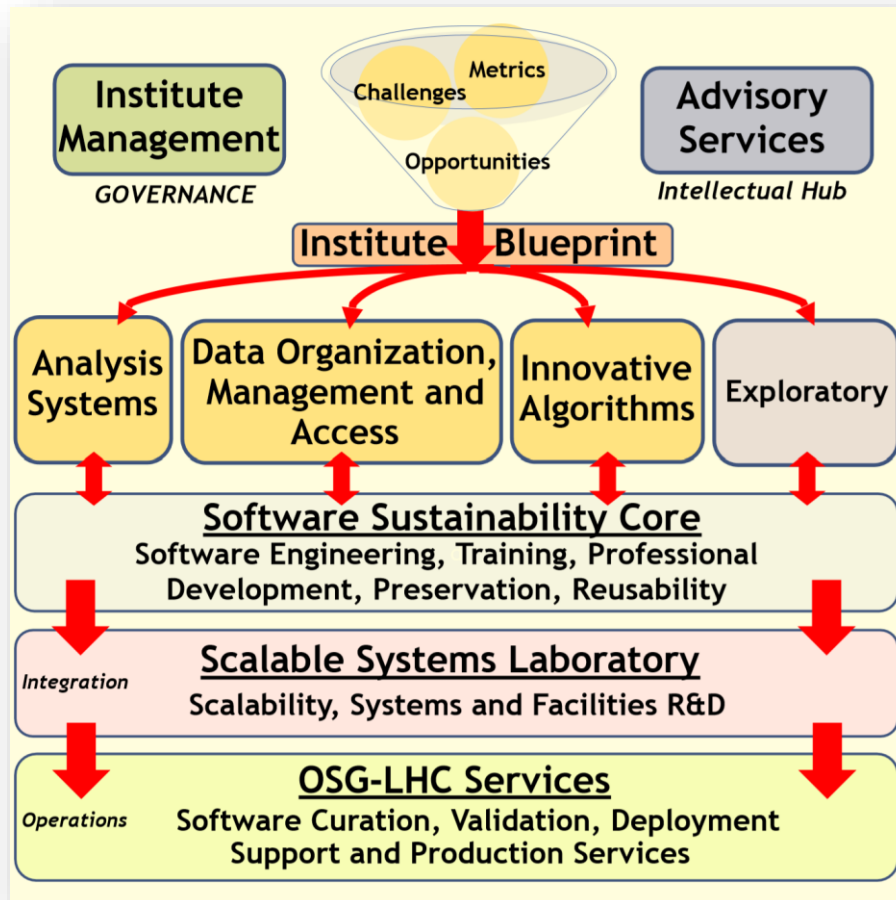
*as best we can

Multi Experiment















With an eye towards an even broader community

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	Robert Tuck Princeton University Associate Project Manager	Floë Fusin-Wischusen Princeton University PICSciE Institute Manager	Rob Gardner University of Chicago SSL 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	Frank Wuerthwein University of California, San Diego OSG-LHC Area Lead and OSG Executive Director












And
Mike Hildreth
And
Kaushik De

Steering Board

Represents the major stakeholders and partners for the IRIS-HEP project. Will meet quarterly with the IRIS-HEP Executive Board to learn the status of the project and **provide feedback on the large-scale priorities** and current strategy of the Institute.

The steering board meets quarterly with the executive board:

- <https://indico.cern.ch/category/10989/>
- <https://iris-hep.org/about/steering-board>

				
Peter Elmer Princeton University <i>Institute PI and Executive Director</i>	Gordon Watts University of Washington <i>Institute co-PI and Deputy Executive Director</i>	Tommaso Boccali INFN-Pisa <i>CMS Experiment</i>	Paolo Calafiura LBNL <i>US ATLAS Ops Program</i>	Simone Campana CERN <i>Worldwide LHC Computing Grid (WLCG)</i>
				
David Costanzo Sheffield <i>ATLAS Experiment</i>	Oliver Gutsche FNAL  <i>US CMS Ops Program</i>	Gerhard Raven VU/NIKHEF <i>LHCb Experiment</i>	Graeme Stewart CERN <i>HEP Software Foundation (HSF)</i>	Ken Bloom University of Nebraska - Lincoln <i>Interim OSG Council Chair</i>

Advisory Board

Provides annual non-stakeholder feedback on the goals and evolving project plans, and evaluates how well the Institute, as a project, is achieving its overall mission as defined with NSF. The Advisory Panel consists of 7 fixed members with an option of inviting ad-hoc additional members as needed for particular topics.

The first in-person meeting with the Advisory Panel took place on 9 September, 2019:

- <https://indico.cern.ch/event/840467/>

Extremely useful feedback for us as a project as we began to get critical mass in terms of staffing.



The next Advisory Panel meeting will take place in summer, 2020, after the IRIS-HEP Retreat and Strategic Planning exercise.

Maybe (COVID-19)

Project Round Up

PROJECTS WE ARE INVOLVED IN, LEADING, ETC.

Innovative Algorithms

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

[ACTS](#) – Experiment independent, parallel, track reconstruction

[ML For Jet Physics](#) – Using ML for jet taggers, boosted objects, etc.

[exploratory-ml](#) – Using ML to transform analysis workflows

[mkFit](#) – Fully Vectorized and Parallel Kalman Filter for use in collider experiments.

[FastPID](#) – Using generative models to simulate PID detectors in LHCb

[ML on FPGA](#) – Fast inference for use in low latency environments like a L1 trigger.

[GPU Trigger](#) – Allen GPU trigger framework for LHCb



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.

[XCache](#) – Regional caches to store, on-demand, datasets (CMS/ATLAS T2 Facilities)

[iDDS](#) – Moving from byte/file level delivery to event delivery

[Modeling Data Workflows](#) – Using Run 2 to predict Run 3, 4, and beyond

[ServiceX](#) – Delivering columnar data on demand.

[SkyhookDM](#) – Ceph for selection, projection, aggregation, indexing of data

[TPC](#) – Third Party Copy – using modern protocols to move data between datacenters



Analysis Systems

Develop sustainable analysis tools to extend the physics reach of the HL-LHC experiments: create greater functionality to enable new techniques; reducing time-to-insight and physics; lowering the barriers for smaller teams; and streamlining analysis preservation, reproducibility, and reuse.

[ADL Benchmarks](#) – Comparison data query tasks in different Analysis Description languages

[DecayLanguage](#) – Convert particle decay descriptions between digital representations

[AMPGEN](#) – Fitting, multibody decays using isobar model (LHCb)

[FunctionalADL](#) – A Functional Analysis Description Language with SQL roots

[Awesome-hep](#) – Awesome list of high energy and particle physics software.

[Histogram Projects](#) – boost-histogram, Aghast, etc.

[Awkward Array](#) – Hierarchical numpy

[MadMiner](#) – Likelihood free inference from Monte Carlo and simulation



Analysis Systems

Develop sustainable analysis tools to extend the physics reach of the HL-LHC experiments: create greater functionality to enable new techniques; reducing time-to-insight and physics; lowering the barriers for smaller teams; and streamlining analysis preservation, reproducibility, and reuse.

[Particle](#) – Accessing PDG info in python

[PPX](#) – Cross-platform Probabilistic Programming eXecution protocol for MC to Inference Engine connections

[pyhf](#) – Python implementation of HistFactory (including toys!) based on TensorFlow and PyTorch as backends

[RECAST](#) – Framework for live-archiving of existing analyses

[ROOT-conda](#) – installing root from conda forge

[Scikit-HEP](#) – Python package which brings in common HEP tools

[uproot](#) – Loads TTree's from ROOT files into awkward arrays (writes too!)



SSL

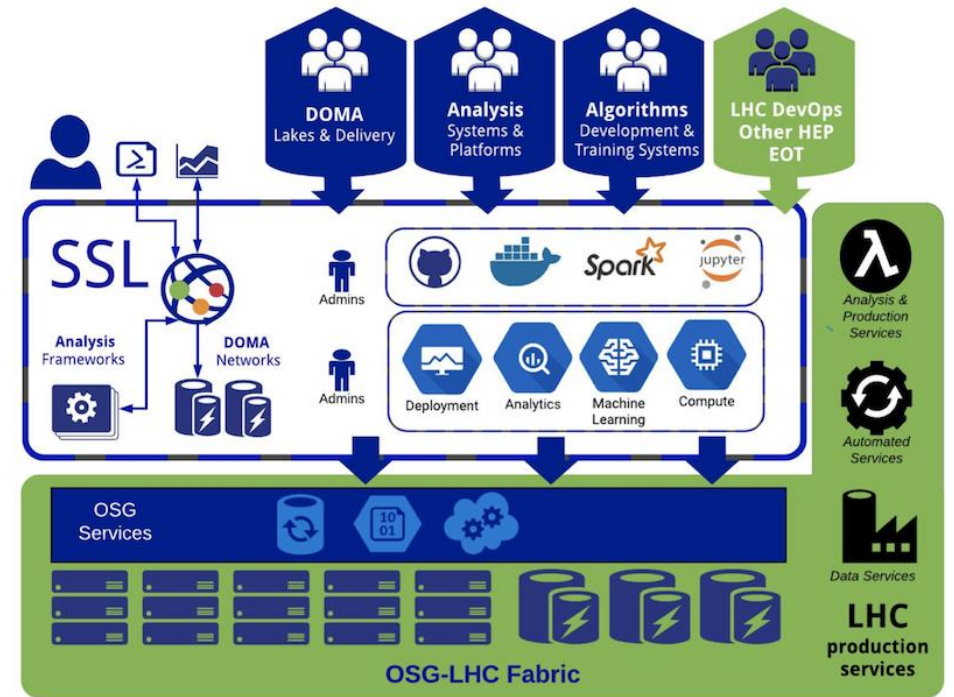
Provide the Institute and the HL-LHC experiments with scalable platforms needed for development in context, perform facilities and systems R&D.

It has grown to more than that:

- Leading the charge to re-imagine the compute facility
- Developing a common substrate for HEP
- Flexible computing, optimal use of resources, minimal sysadmin effort



River (UChicago): REANA, ServiceX, ATLAS analytics, COVID calculations, training platform, the SLATE project. Backfilled by OSG.



OSG

- The OSG is a consortium dedicated to the advancement of all of open science via the practice of Distributed High Throughput Computing, and the advancement of its state of the art.
- IRIS-HEP is the funding mechanism to support LHC needs from the consortium.
 - Effort from IRIS-HEP is roughly $\frac{1}{3}$ of the total effort in OSG today.
- At a high level policy, only shared interests between US ATLAS and US CMS ops programs are within scope of the OSG effort in IRIS-HEP
 - There are activities in the OSG consortium more broadly that serve multiple domains, DOE-NP, and cosmic and intensity frontier experiments in DOE-HEP, plus one of the experiments but not the other. Such activities are not within scope of IRIS-HEP.
- A lot of “operations” R&D goes on inside OSG
 - Globus transition
 - Packages are constantly being retired
 - New workflows – e.g. containers, ML, GPU’s
 - Works tightly with DOMA and SSL as a result

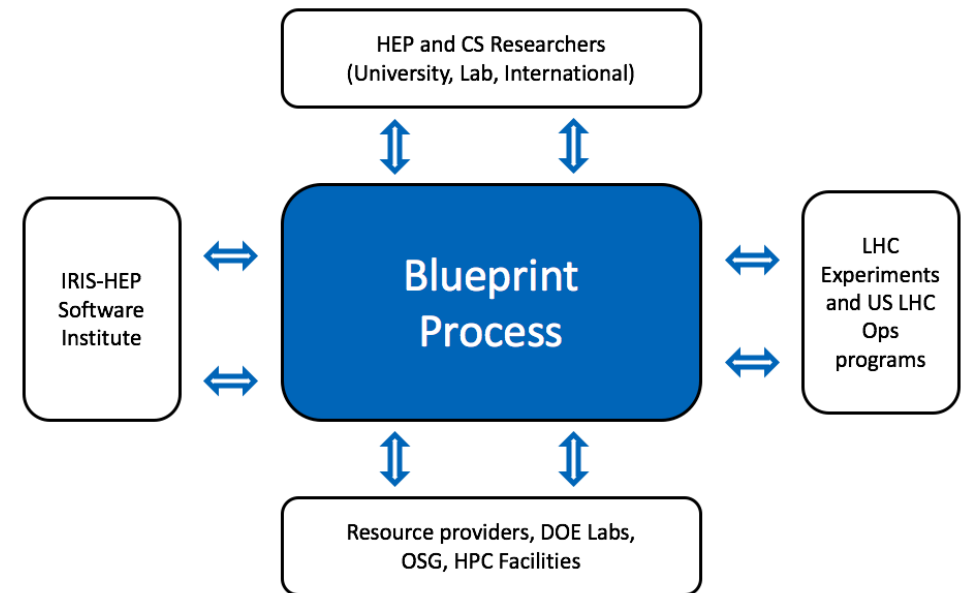


Blueprint Workshops

Small workshops 3-4 times a year with experts from outside and inside the field to facilitate effective collaborations by building and maintaining a common vision.

See [indico](#) for workshops material

- 21 Jun - 22 Jun, 2019 - Blueprint: Analysis Systems R&D on Scalable Platforms (NYU)
- 10 Sep - 11 Sep, 2019 - Blueprint: Accelerated Machine Learning and Inference (Fermilab)
- 23 Oct - 25 Oct, 2019 - Blueprint: A Coordinated Ecosystem for HL-LHC Computing R&D (Catholic University of America, Washington DC)
- 20 Feb, 2020 - Training Blueprint (Vidyo)



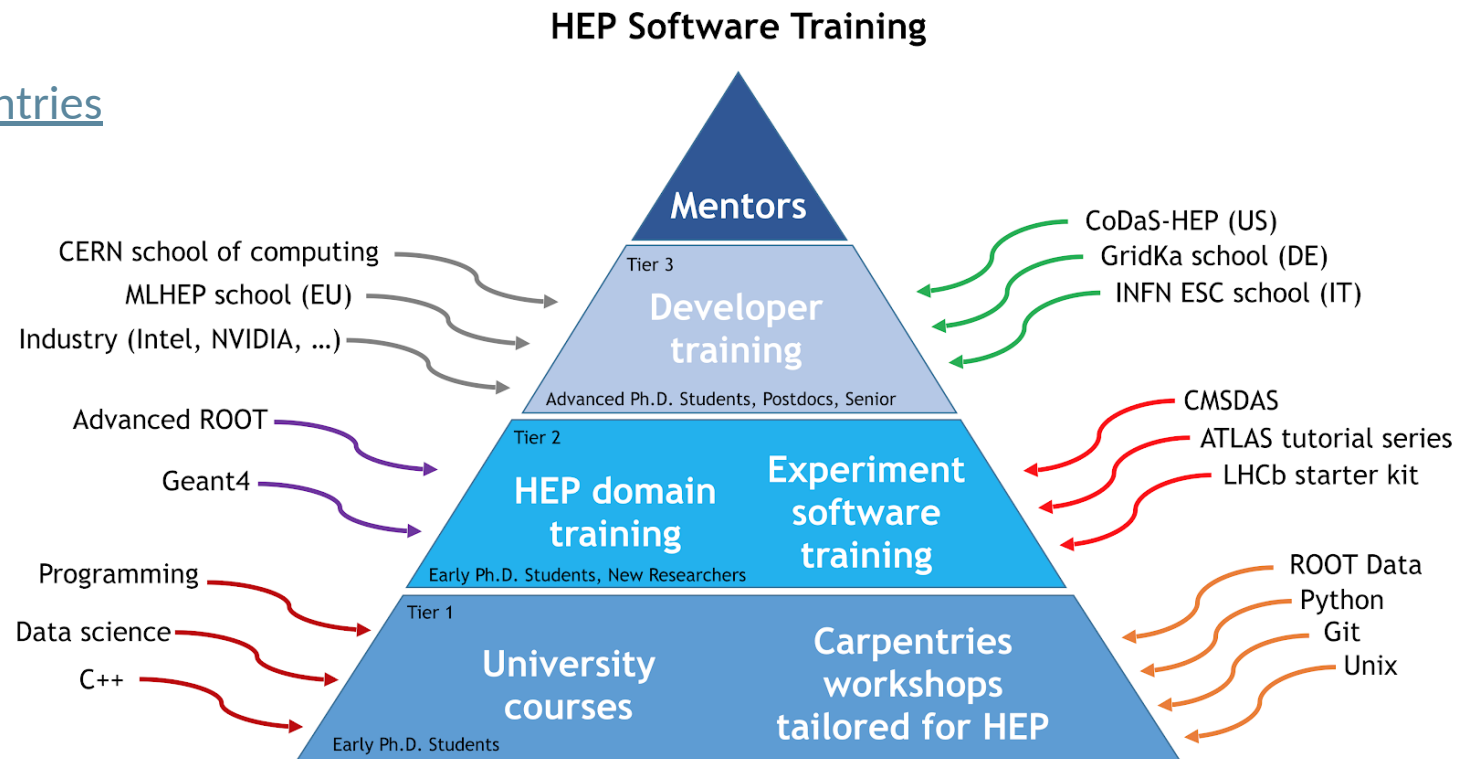
Training

How to build a sustainable and scalable training framework that grows skills through multiple stages of people's careers?

Collaborate with [FIRST-HEP](#) and [Carpentries](#)

CoDaS summer school, FIRST-HEP/ATLAS software training, ML hackathon at University of Puerto Rico, Software Carpentry Workshop, etc.

A number of our members took these courses and have become teachers. SSL often provides infrastructure for the training.



Towards Year 3

AND BEYOND...

High Level Overview

1. The NSF 18 Month Review Results
2. Focus Area Discussions
3. PI Discussions
4. US Ops Programs Discussion
5. Steering Board Meeting
6. Full-Team Retreat
7. PEP Draft

This is a conversation leading to
Year 3 and out plans




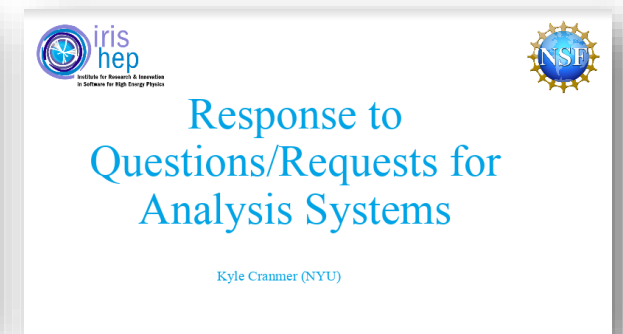
High Level Overview

- | | | |
|------------------------------------|---|---|
| 1. The NSF 18 Month Review Results | → | • Successful (thank you for your help!) |
| 2. Focus Area Discussions | | • Better Metrics and Milestones |
| 3. PI Discussions | | • Grand Challenges |
| 4. US Ops Programs Discussion | | • Evidence of impact outside of HEP |
| 5. Steering Board Meeting | | |
| 6. Full-Team Retreat | | |
| 7. PEP Draft | | |




High Level Overview

1. The NSF 18 Month Review Results
 2. Focus Area Discussions 
 3. PI Discussions
 4. US Ops Programs Discussion
 5. Steering Board Meeting
 6. Full-Team Retreat
 7. PEP Draft
- Focus Area Leads view of how things are progressing
 - In progress
 - Questions like – how are people performing, future grand challenge ideas, etc.



High Level Overview

1. The NSF 18 Month Review Results
 2. Focus Area Discussions
 3. PI Discussions 
 4. US Ops Programs Discussion
 5. Steering Board Meeting
 6. Full-Team Retreat
 7. PEP Draft
- PI's discuss their team's progress in the various focus areas, their vision for the future, how it interfaces with the focus areas
 - Will start in a week or so
 - Questions look very similar to the Focus area questions



High Level Overview

1. The NSF 18 Month Review Results
2. Focus Area Discussions
3. PI Discussions
4. US Ops Programs Discussion →
 - This meeting and the same one with US ATLAS
 - Discussions & interaction at the full-team retreat
 - Any other interactions you think would be good
5. Steering Board Meeting
6. Full-Team Retreat
7. PEP Draft




High Level Overview

1. The NSF 18 Month Review Results
 2. Focus Area Discussions
 3. PI Discussions
 4. US Ops Programs Discussion
 5. Steering Board Meeting 
 6. Full-Team Retreat
 7. PEP Draft
- Occurs on May 26 - [indico](#)
 - Y3 plans will be the main theme



High Level Overview

1. The NSF 18 Month Review Results
 2. Focus Area Discussions
 3. PI Discussions
 4. US Ops Programs Discussion
 5. Steering Board Meeting
 6. Full-Team Retreat 
 7. PEP Draft
- Occurs on May 26-28 - [indico](#)
 - Virtual Meeting – in planning stages
 - US OPS welcome
 - Focus area talks with PI input to present a first draft of Y3 vision
 - With direction for out years
 - Presentations by full team
 - Result: Content of draft PEP



High Level Overview

1. The NSF 18 Month Review Results
 2. Focus Area Discussions
 3. PI Discussions
 4. US Ops Programs Discussion
 5. Steering Board Meeting
 6. Full-Team Retreat
 7. PEP Draft
- 
- Iterated over by PI's
 - Given to the NSF
 - Iterate until approved
 - Year 3 starts Sept 1, 2020



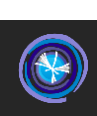
Brainstorming...

I'VE PROVIDED SOME STRUCTURE, BUT IT ISN'T MEANT TO BE
LIMITING...

Challenges

- 10 PB Data reconstruction in one day
 - Connection to a “data lake”? Rucio as a source?
 - Does data lake need a blueprint to get the various experiments talking? Is that something that naturally happens in the WLCG?
- Vertical slice of analysis chain – from ServiceX to fitted limit plots
 - Could we make this auto-diff’able?
- Distributed SSL facility, 3 centers, common k8 substrate

“Grand” just means it involves many areas of IRIS-HEP



Workshops, Training, Blueprint

Blueprints to focus joint work with ATLAS – especially if it might involve outside experts

Common, topical workshops

Joint training efforts

Connections to Snowmass?



More ideas?

Places we could collaborate better:

- ServiceX (Kevin P, Nick Smith), SkyhookDM
- Statistical Tools (pyhf, RECAST/REANA, etc.?)

