IRIS-HEP

G. Watts (UW/Seattle)
Deputy Executive Director of IRIS-HEP
For the IRIS-HEP Team

XVII CPAN Days
COMCHA Session

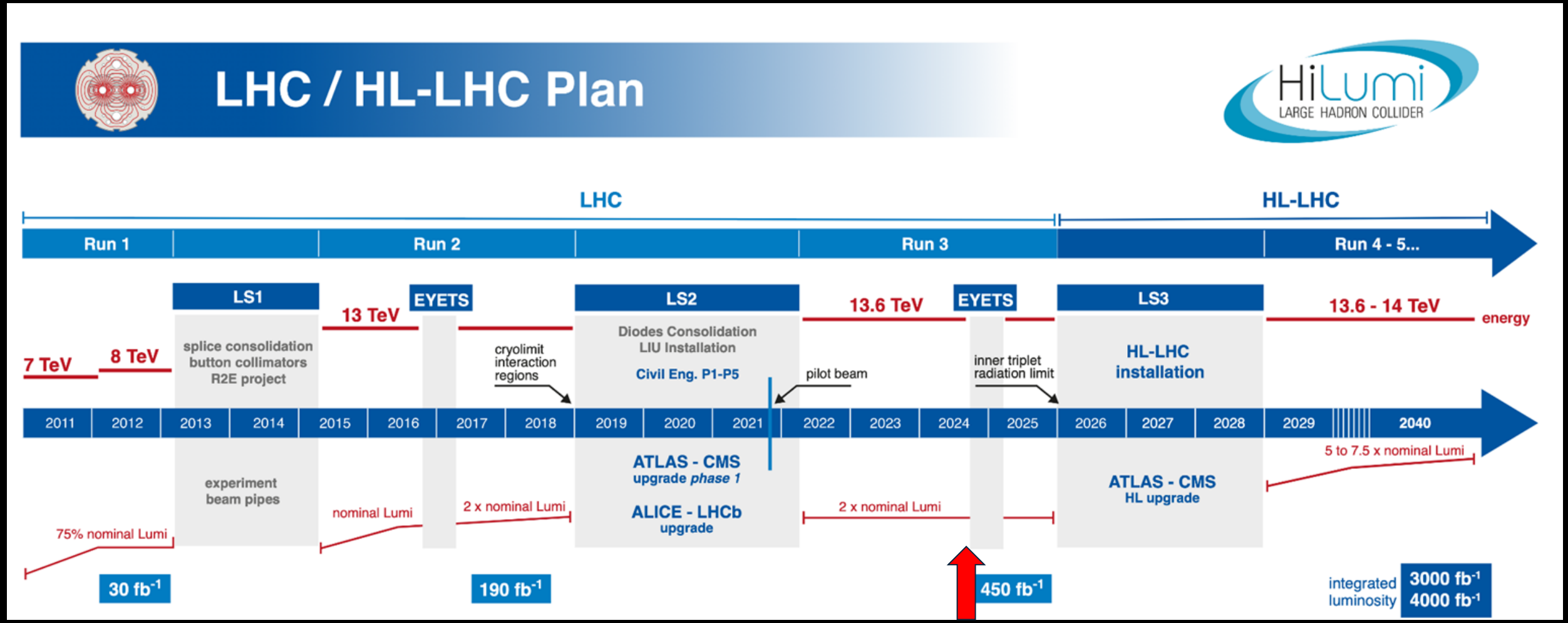
2025-11-19

What is IRIS-HEP?

IRIS-HEP is a software institute funded by the US National Science Foundation to contribute to a ***software upgrade*** for the High Luminosity LHC, in parallel to the large hardware (MREFC) investment.

What is IRIS-HEP?

(January, 2022 version)



COVID ?

CTDR

S2I2-HEP

Institute Conceptualization and
Community White Paper Process

IRIS-HEP Institute

Design | Execution

IRIS-HEP Institute
(Phase 2)

Snowmass

U.S. HEP Community Planning Process

A Short History

NSF funded the S2I2-HEP Conceptualization Project (s2i2-hep.org/) in July 2016

Community charge from the Worldwide LHC Computing Grid in July 2016:

- Anticipate a "software upgrade" in preparation for the HL-LHC
- Identify and prioritize the software research and developments investments
  1. to achieve improvements in software efficiency, scalability and performance and to make use of the advances in CPU, storage and network technologies
  2. to enable new approaches to computing and software that could radically extend the physics reach of the detectors
  3. to ensure the long term sustainability of the software through the lifetime of the HL-LHC

Growing a Global Collaboration

JLab
March, 2018
HSF/OSG/WLCG

UCSD/SDSC
January, 2017
HSF CWP

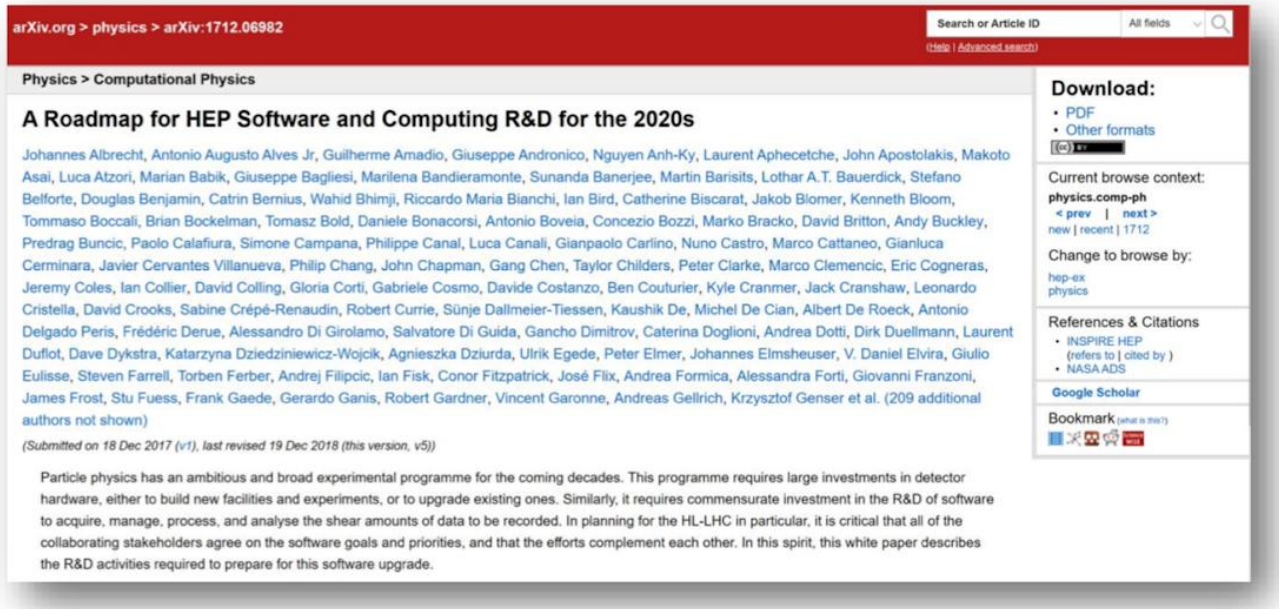
Annecy
June, 2017
HSF CWP

Naples
March, 2017
WLCG/HSF

Community White Paper: **arXiv 1712.06592**

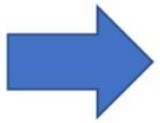
The Community White Paper Process (2016-2017)

1. Grass Roots
2. Strategic Plan



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



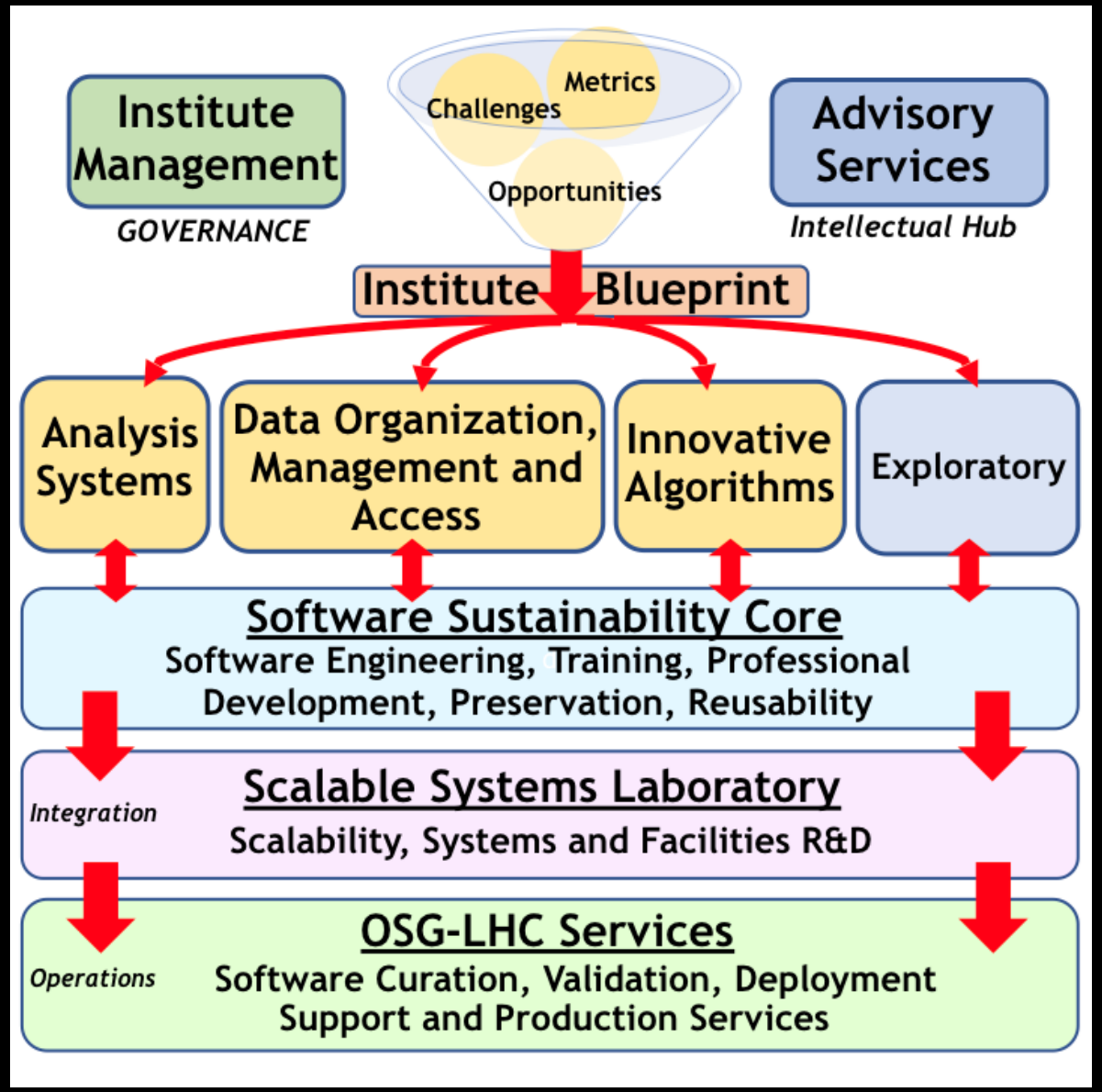
Individual Papers on the arXiv:

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

Main Areas of R&D

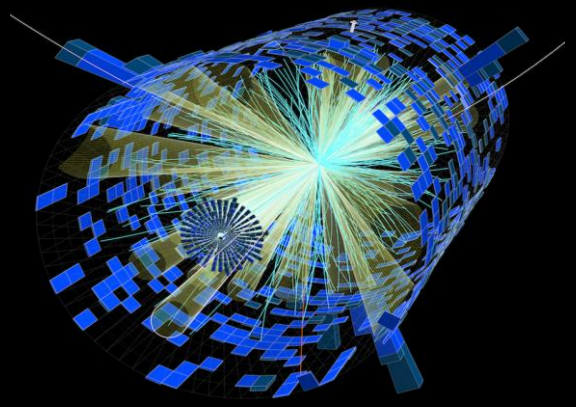
Training, Scaling, Production

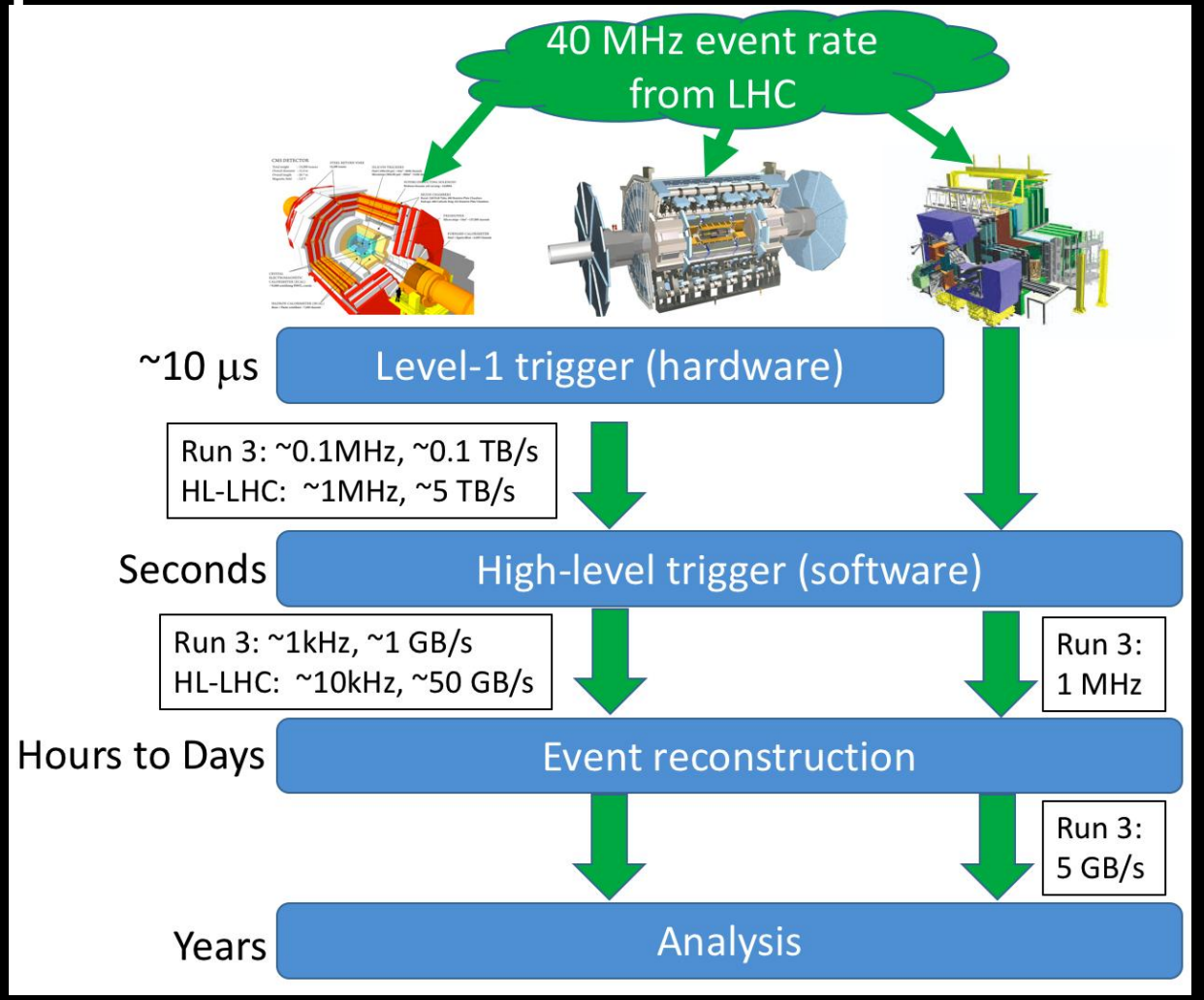


Full Project List

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.





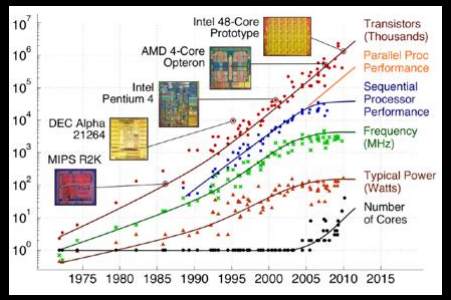
Groups are focused on answering 2 questions

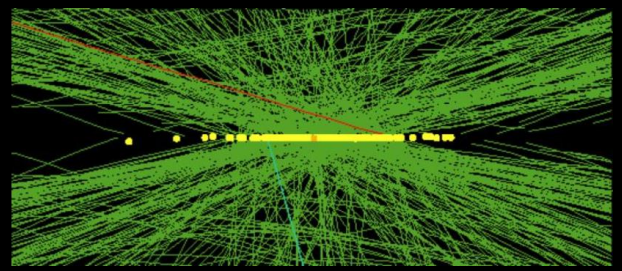
How to redesign tracking algorithms for HL-LHC?

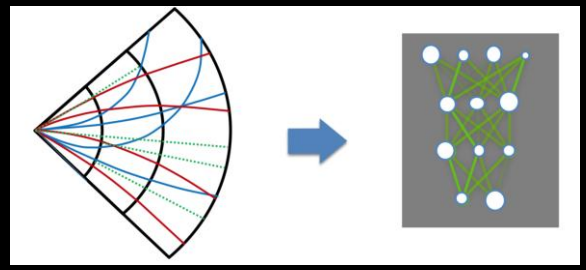
- Determination of charged-particle trajectories (“tracking”) is largest component of event reconstruction
- IRIS-HEP investigations
  - More efficient algorithms
  - More performant algorithms
  - Use of hardware accelerators

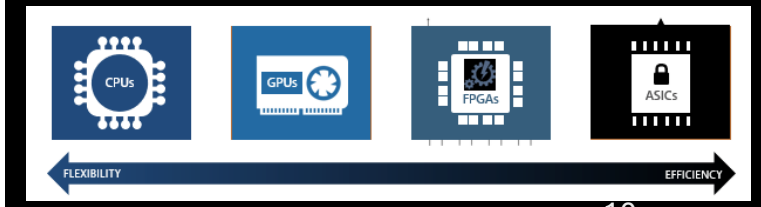
How to make use of major advances in machine learning (ML)?

- Use of ML in HEP may be a major opportunity
  - Capitalize on industry and data science techniques and tools
  - Could reduce CPU needs
  - Could lead to wider use of accelerators
- IRIS-HEP investigations
  - New HEP applications of ML
  - Use of new ML techniques
  - ML on accelerators in realistic HEP apps









Projects

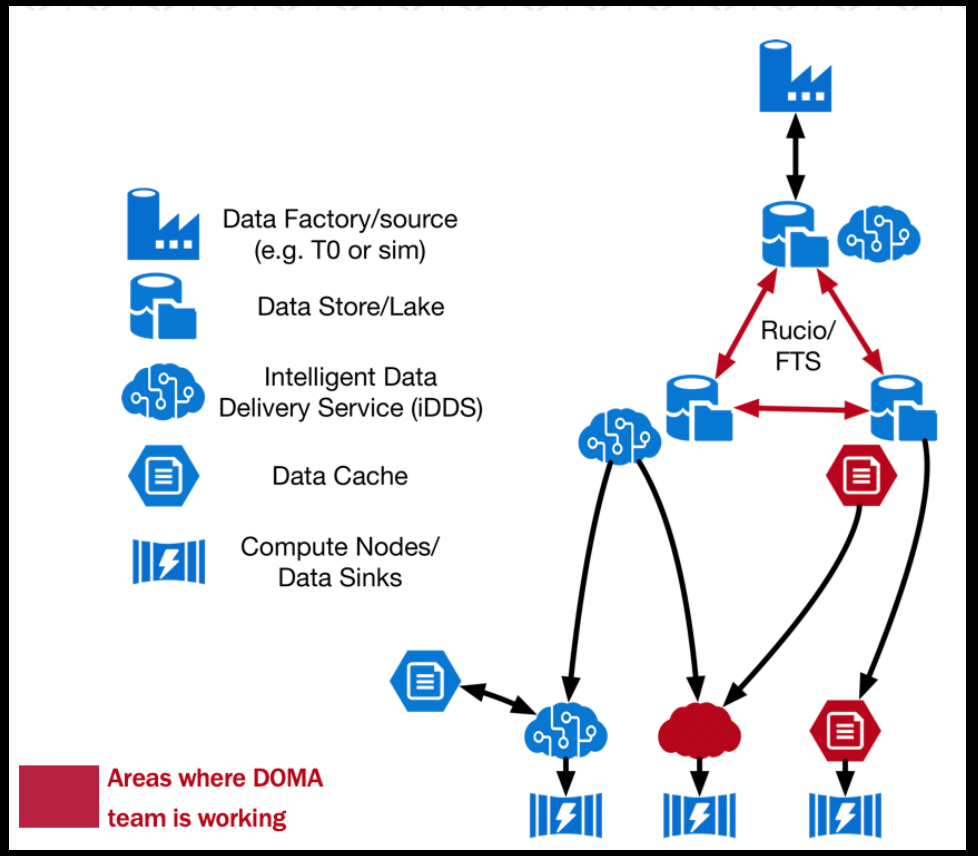
Name	Focus Area(s)	Maturity	Description
Accelerated GNN Tracking	IA	**Exploratory**	accel-gnn-tracking
Accelerators and ML for reconstruction	IA	**Archived**	Accelerated calorimeter reconstruction using Machine Learning as a Service
ACTS	IA	**Development**	Development of experiment-independent, thread-safe track reconstruction.
GPU Trigger Project	IA	**Testing**	Allen: a GPU trigger for LHCb
Line-Segment tracking	IA	**Development**	Segment linking tracking for CMS
Machine Learning for jets	IA	**Development**	Machine learning for jets
mkFit	IA	**Deployed**	Modernizing Kalman filter tracking for CMS
PV-Finder	IA	**Testing**	CNNs to find primary vertices

Full Project List

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



Projects

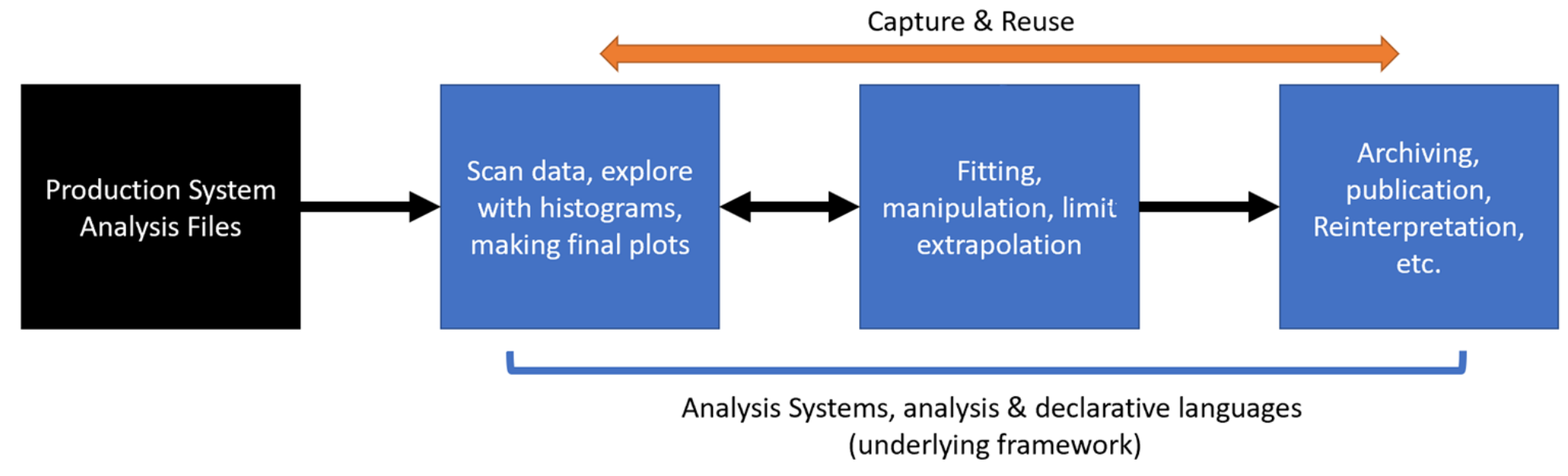
The Analysis Grand Challenge

Name	Focus Area(s)	Maturity	Description
200 Gbps Challenge	AS, DOMA	**Testing**	200gbps
Analysis Grand Challenge	DOMA, AS	**Development**	Analysis Grand Challenge
Caching Data for LHC Analysis	DOMA, OSGLHC	**Deployed**	Cached-based placement of analysis datasets.
Coffea-Casa	DOMA, AS	**Testing**	A Prototype of Analysis Facility
Intelligent Data Delivery Service	DOMA	**Deployed**	Delivering Data. Better.
Modeling Data Workflows	DOMA, OSGLHC	**Archived**	Modeling HL-LHC Data flows
ServiceX	DOMA	**Testing**	Delivering columnar data on demand
SkyhookDM	DOMA	**Development**	Programmable Storage for Databases and Datasets
Third Party Copy	DOMA, OSGLHC	**Deployed**	Envisioning a new way to move LHC data

Full Project List

Analysis Systems

Develop sustainable analysis tools to extend the physics reach of the HL-LHC



- 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.

Analysis Systems projects span all stages of end-user analysis.

Projects – Python Ecosystem

Name	Focus Area(s)	Maturity	Description
ADL Benchmarks	AS	**Deployed**	Functionality benchmarks for analysis description languages
Awkward Array	AS	**Deployed**	Manipulate arrays of complex data structures
Awkward-Dask	AS	**Testing**	Developing a new high-level Dask collection for Awkward Arrays
Columnar PHYSLITE	AS	**Development**	Columnar data analysis workflows with ATLAS PHYSLITE
Functional ADL	AS	**Deployed**	Functional Analysis Description Language
Histogram projects	AS	**Deployed**	Histogramming efforts
ROOT on Conda-Forge	AS	**Deployed**	Use ROOT in Conda through Conda-Forge
Scikit-HEP	AS	**Deployed**	Pythonic analysis tools
uproot	AS	**Deployed**	Read and write ROOT files in Python
Vector	AS	**Deployed**	Manipulate vectors

Projects - Statistics

Name	Focus Area(s)	Maturity	Description
abcd-pyhf	AS	**Development**	Likelihood-based ABCD method with pyhf
AmpGen	AS	**Deployed**	Generation and fitting for multibody hadron decays
cabinetry	AS	**Deployed**	Building and steering template fits
MadMiner	AS	**Deployed**	Likelihood-free Inference
ppx	AS	**Deployed**	Cross-platform Probabilistic Programming eXecution protocol
pyhf	AS	**Deployed**	Differentiable likelihoods

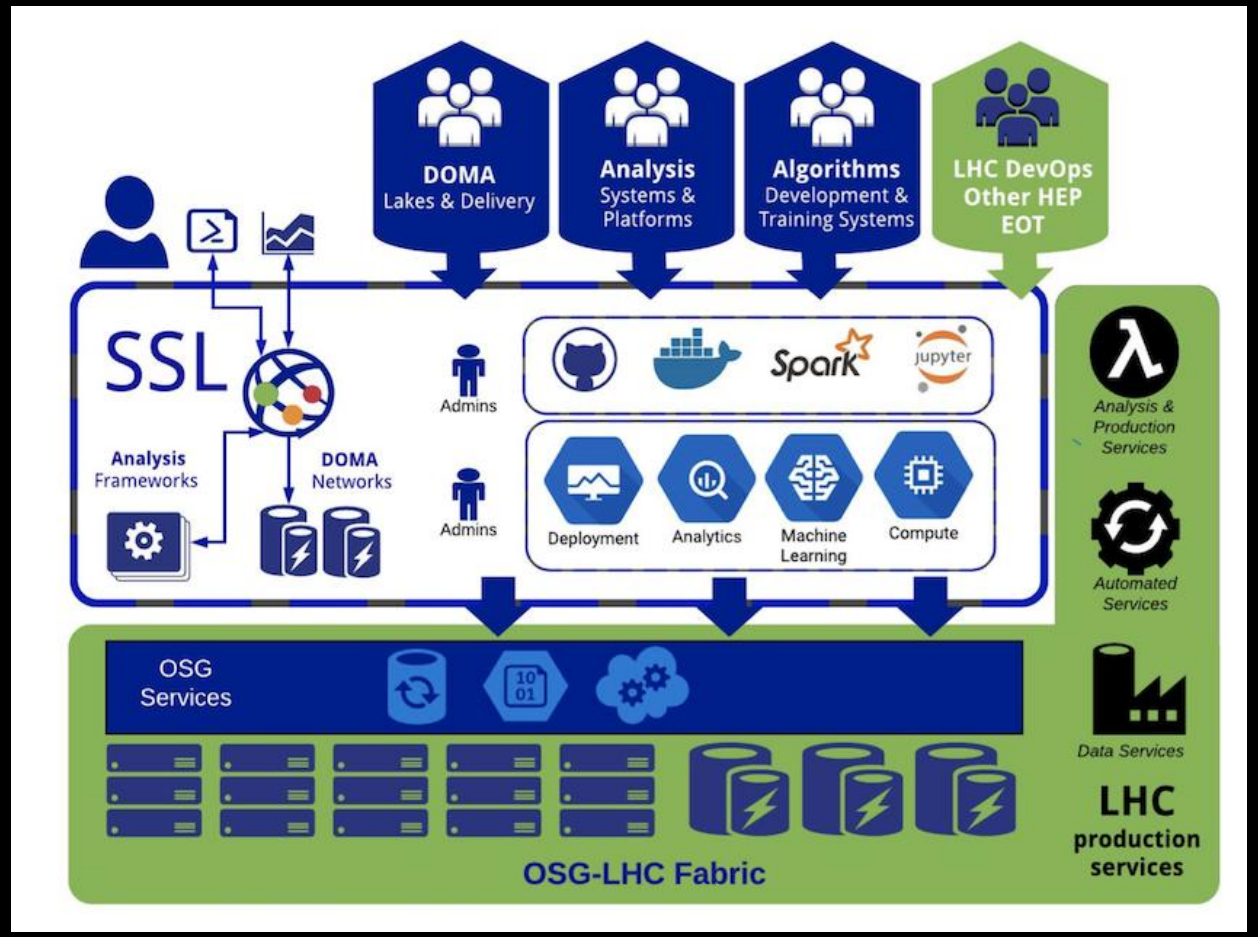
Along with other projects…

Full Information

Scalable Systems Laboratory (SSL)

Goal: Facilities Research

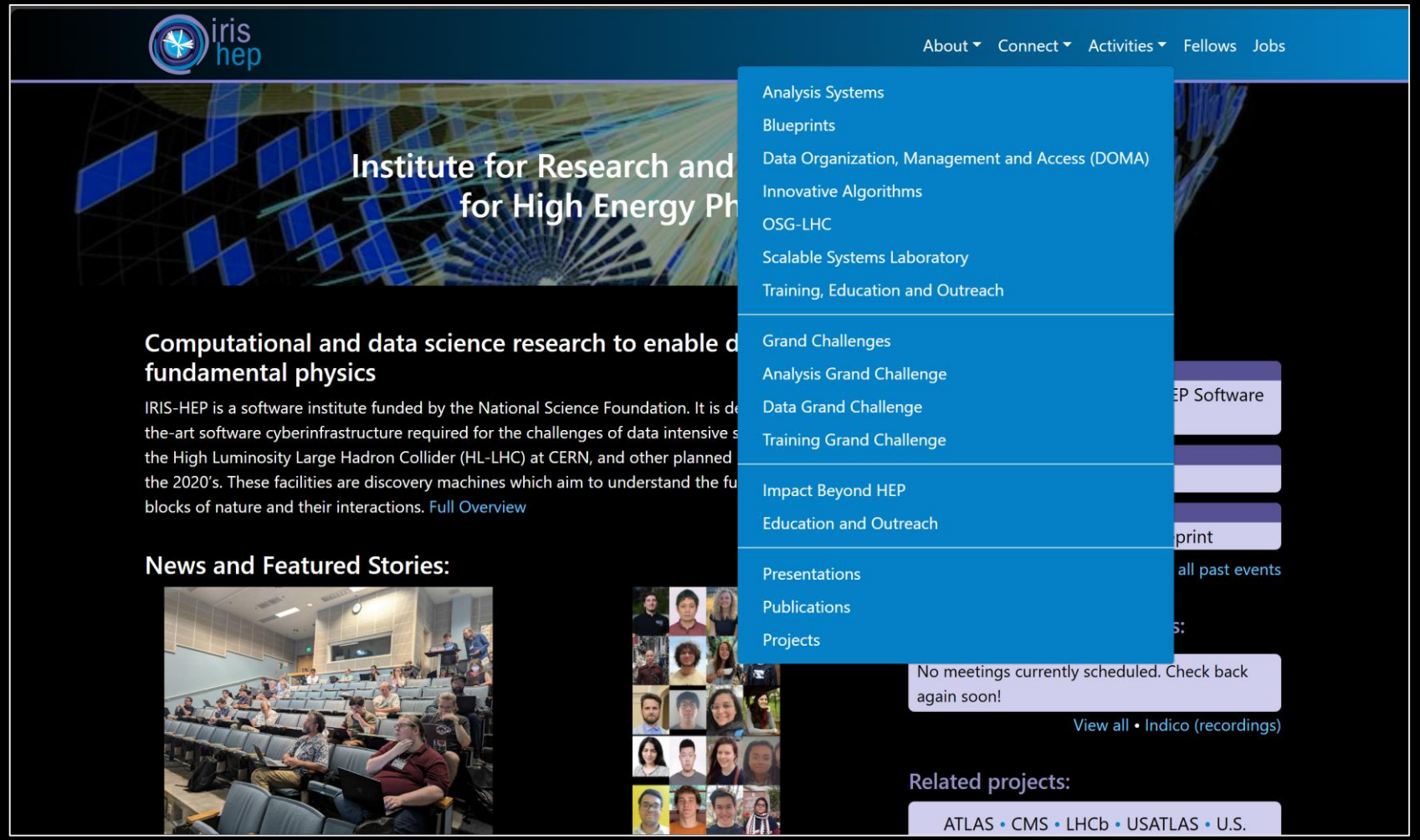
- New paradigms for running facilities (e.g. k8s), distribution, etc.
- Provides access to infrastructure and environments
- Organizes software and resources for scalability testing
- Does foundational systems R&D on accelerated services
- Provides the integration path to the OSG-LHC production infrastructure



Finding Out More

Start with our Web Page:

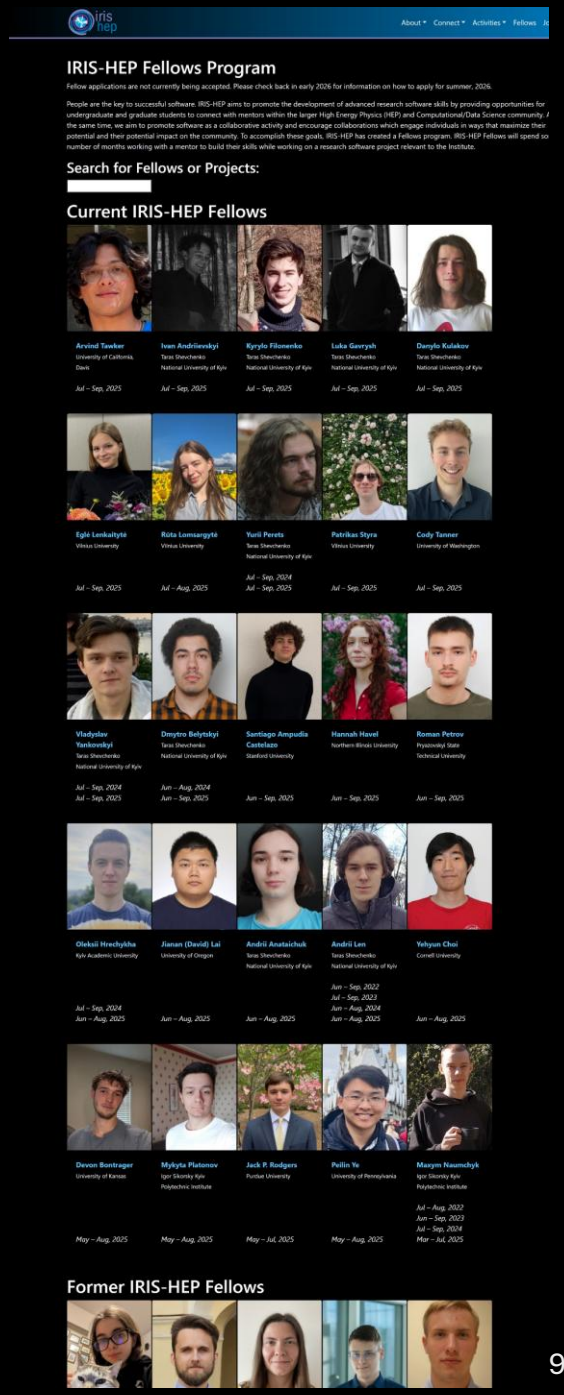
https://iris-hep.org



(Summer) Fellows Programs

Started as a Graduate Program
- COVID turned it into an undergraduate program
- This has been widely successful!

1. Gather projects from potential mentors in ~January
2. Student applications in March
   - Force interaction between potential students and mentors to draw up an application
   - Students must commit to 2 months of paid full-time work
3. Selection in late spring
4. Students work for 2 months over the summer



IRIS-HEP Fellow: Cody Tanner

Example Project

Fellowship dates: Jul – Sep, 2025

Home Institution: University of Washington

Project: Differentiable Modeling of Systematic Uncertainties in ATLAS Object Corrections

Modern ATLAS analyses depend on object corrections that are currently implemented through non-differentiable procedures like histogram lookups and conditional logic, limiting their integration into gradient-based pipelines. This project proposes a neural network model that replicates ATLAS object corrections, including systematic uncertainties, for small-R jets in a differentiable and computationally efficient form. Starting from an existing baseline trained on the JZ2 dataset, the model will be refined through architectural tuning, loss reweighting, and incorporation of per-object uncertainties to approach sub-percent residuals in jet kinematics. A final case study will use the model to reconstruct Z→jj peaks, evaluating the physics impact of improved corrections and uncertainty modeling. This work provides a foundation for embedding fast, uncertainty-aware corrections directly into end-to-end ATLAS workflows.

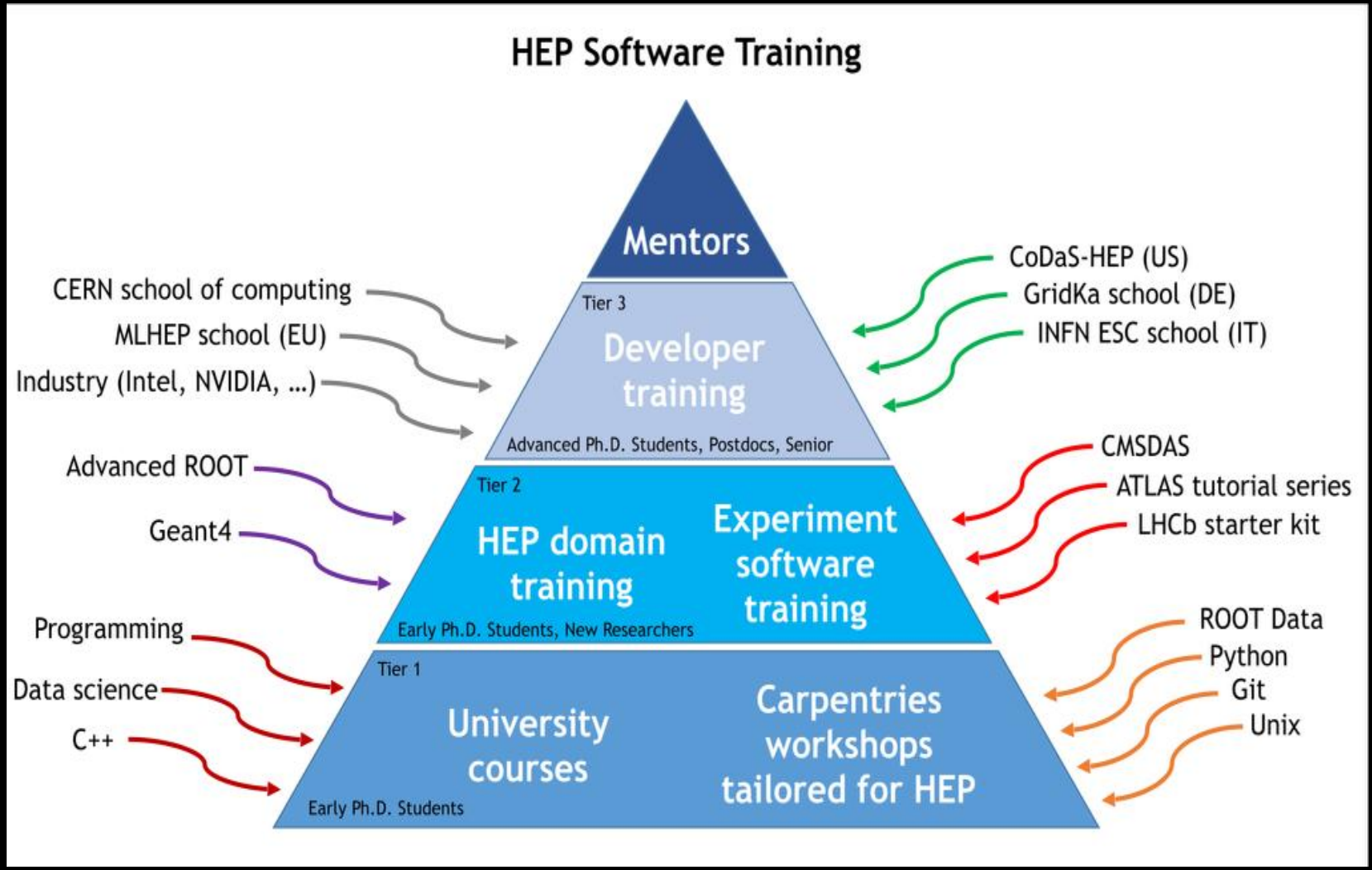
More information: My project proposal

Mentors:

- Gordon Watts (University of Washington)

Training Events

We sponsor, collaborate, run, many training events!



- Python Training
- Software Carptentries Training
- Co-located training at US ATLAS and US CMS summer meetings
- CoDaS HEP
- Coding Camps with Quarknet
- Etc...

Events

December, 2025

- 3 Dec, 2025 - A Coordinated Ecosystem for HEP Software and Computing ()

November, 2025

- 10 Nov - 14 Nov, 2025 - Connecting the Dots 2025 (*University of Tokyo*)

October, 2025

- 8 Oct - 10 Oct, 2025 - IRIS-HEP Institute Retreat (*Geneva, Switzerland*)

September, 2025

- 9 Sep - 9 Sep, 2025 - HSF/IRIS-HEP Python for Analysis Training (*Virtual*)
- 3 Sep - 4 Sep, 2025 - HSF/IRIS-HEP Software Basics Training (Virtual) (*Virtual*)

July, 2025

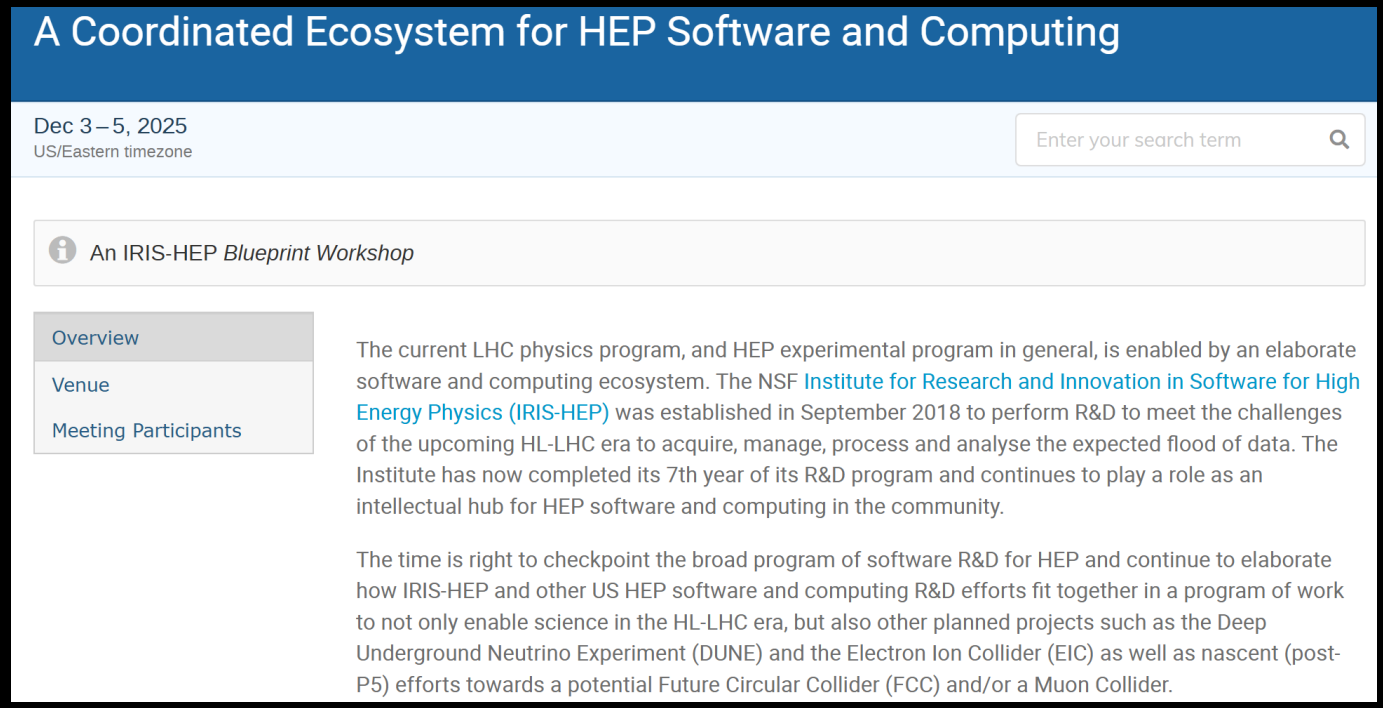
- 28 Jul - 31 Jul, 2025 - JuliaHEP 2025 Workshop (*Princeton University*)
- 24 Jul - 25 Jul, 2025 - US ATLAS / IRIS-HEP Analysis Software Training Event 2025 (*University of Michigan*)
- 21 Jul - 26 Jul, 2025 - Coding Camp 1 - Virtual (*Virtual*)
- 21 Jul - 25 Jul, 2025 - CoDaS-HEP 2025 - Computational and Data Science Training for High Energy Physics (*Princeton University*)
- 14 Jul - 17 Jul, 2025 - PyHEP.dev 2025 - "Python in HEP" Developer's Workshop (*University of Washington*)
- 13 Jul - 18 Jul, 2025 - 2025 Quarknet Data Camp at Fermilab (*Fermilab*)
- 13 Jul - 18 Jul, 2025 - Coding Camp 1 in Spanish (virtual) at UPRM Puerto Rico (*University of Puerto Rico Mayaguez*)
- 7 Jul - 9 Jul, 2025 - Coding Camp at Siena College (NY) (*Siene College (NY)*)

June, 2025

- 30 Jun - 2 Jul, 2025 - Coding Camp at Brookhaven-Stony Brook (*Brookhaven-Stony Brook NY*)
- 23 Jun - 25 Jun, 2025 - Coding Camp (Spanish) at Physics Department, UPRM (Mayaguez, PR) (*University of Puerto Rico Mayaguez*)
- 23 Jun - 26 Jun, 2025 - Coding Camp at University of Alabama (Tuscaloosa) (*University of Alabama (Tuscaloosa)*)
- 18 Jun - 20 Jun, 2025 - HSF/IRIS-HEP Software Basics Training (*In person at CERN Kjell Johnsen Auditorium*)
- 11 Jun - 13 Jun, 2025 - Coding Camp in Spanish at University Gardens High School - Hato Rey Sur - San Juan - Puerto Rico (*University Gardens High School Puerto Rico*)
- 9 Jun - 11 Jun, 2025 - Coding Camp at Rice University Houston (*Rice University (Houston, TX)*)
- 8 Jun - 13 Jun, 2025 - Fifth MODE Workshop on Differentiable Programming for Experiment Design (*Kolymbari, Crete, Greece*)
- 2 Jun - 4 Jun, 2025 - Coding Camp at CROEM High School - Mayaguez - Puerto Rico (*CROEM High School Mayaguez Puerto Rico*)

Intellectual Hub

- We sponsor workshops and conferences
  - PyHEP
  - MODE
  - Connect the Dots
  - Etc.
- We organize Blueprint Meetings



Blueprint Workshops

- Gather experts inside and outside the field
- Help set the direction of R&D in the future
- Build field wide consensus

Conclusions

- We are US based HL-LHC focused distributed Software Institute
  - We do work with others as well (EPIC/DUNE, etc.)
- We collaborate with just about everyone on our projects!
  - Project pages have links of who to contact!
- Subscribe to our announcements list
  - Meetings, blueprints, fellows program, etc.
- Contact me with any questions!