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


http://iris-hep.org
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

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 HL-LHC Challenge

Multiple proton-proton interactions per beam bunch crossing ("pile-up") as seen in a simulation of the ATLAS Inner Tracker.

Higher probability of an interesting interaction, but with consequences: detectors/electronics need to handle the higher rate, higher radiation dose and significantly more complex events.
Timeline

**LHC / HL-LHC Plan**

**Run 1**
- **LS1**: 7 TeV
  - Splice consolidation button collimators R2E project
  - Experiment beam pipes
  - 75% nominal Lumi
  - 30 fb⁻¹

**Run 2**
- **EYETS**: 8 TeV
  - Diodes Consolidation LiU Installation
  - 11 T dipole coll.
  - Civil Eng. P1-P5
  - Cryogenic interaction regions
  - 2 x nominal Lumi
  - 190 fb⁻¹

**Run 3**
- **LS2**: 13 TeV
  - ATLAS - CMS upgrade phase 1
  - ALICE - LHCb upgrade
  - 2 x nominal Lumi
  - 350 fb⁻¹

**HL-LHC**
- **LS3**: 14 TeV
  - HL-LHC installation
  - ATLAS - CMS HL upgrade
  - Integrated luminosity: 3000 fb⁻¹
  - 4000 (ultimate)
  - 5 to 7.5 x nominal Lumi

**Institute Conceptualization and Community White Paper Process**
- **S2I2-HEP**
- **CTDR**
- **IRIS-HEP Institute**
  - Design
  - Execution

**CERN HL-LHC Planning - Computing Technical Design Reports (CTDR) for ATLAS/CMS**

**U.S. HEP Community Planning Process**
- **Snowmass**
In addition to LHC accelerator upgrades, significant hardware (detector) investments are planned for HL-LHC to handle the high-rate environment and expected radiation dose.

This includes the ~$150M NSF MREFC to upgrade key elements of ATLAS and CMS (see CMS example in figure), with a 1 April, 2020 start date.

A key goal of the Institute is to maximize the physics reach and impact of this detector investment.
IRIS-HEP Vision

IRIS-HEP addresses key elements of the “Roadmap for HEP Software and Computing R&D for the 2020s” and is implementing the “Strategic Plan for a Scientific Software Innovation Institute (S2I2) for High Energy Physics” submitted to the NSF in December 2017. We aim to:

● 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.

The Institute also aims to play the role of intellectual hub for the larger software R&D effort required to ensure the success of the HL-LHC scientific program, and lead research into deployment of the resulting systems on distributed high throughput computing. The Institute also works to improve scientific software more broadly and create opportunities for a more diverse participation in scientific software and computing.
How did we get to IRIS-HEP?
Developing a Global R&D Roadmap

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
The HSF (http://hepsoftwarefoundation.org) was created in early 2015 as a means for organizing our community to address the software challenges of future projects such as the HL-HLC. The HSF has the following objectives:

- Catalyze new common projects
- Promote commonality and collaboration in new developments to make the most of limited resources
- Provide a framework for attracting effort and support to S&C common projects (new resources!)
- Provide a structure to set priorities and goals for the work

The HSF is an unfunded, volunteer organization, with a “bottoms-up” structure.
Growing a Global Collaboration

UCSD/SDSC
January, 2017
HSF CWP

Annecy
June, 2017
HSF CWP

JLab
March, 2018
HSF/OSG/WLCG

Naples
March, 2017
WLCG/HSF
All CWP and S2I2 Workshops

- 26-27 Apr, 2018 - Reconstruction, Trigger, and Machine Learning for the HL-LHC
  - Massachusetts Institute of Technology, Boston
- 26-29 Mar, 2018 - Joint WLCG/HSF Workshop 2018
  - Centro Congressi Federico II, Naples, Italy
- 14 Dec, 2017 - Mini-workshop on Building Collaborations for ML in HEP
  - Massachusetts Institute of Technology, Boston
- 28-29 Nov, 2017 - S2I2/DOE-lab mini-workshop on HL-LHC Software and Computing R&D
  - Catholic University of America, Washington DC
- 16-17 Nov, 2017 - Data Organisation, Management and Access (DOMA) in Astronomy, Genomics and High Energy Physics
  - Flatiron Institute (Simons Foundation), New York City
- 23-26 Aug, 2017 - S2I2-HEP Workshop
  - University of Washington, Seattle
- 26-30 Jun, 2017 - HEP Software Foundation Workshop
  - LAPP (Annecy)
- 5-6 Jun, 2017 - CWP Event Processing Frameworks Workshop
  - FNAL
- 22-24 May, 2017 - HEP Analysis Ecosystem Retreat
  - Amsterdam

- 8-12 May, 2017 - DS@HEP 2017 (Data Science in High Energy Physics)
  - FNAL
- 1-3 May, 2017 - 2nd S2I2 HEP/CS Workshop
  - Princeton University
- 28-30 Mar, 2017 - CWP Visualization Workshop
  - CERN (and Vidyo)
- 23 Mar, 2017 - Community White Paper Follow-up at FNAL
  - FNAL
- 20-22 Mar, 2017 - IML Topical Machine Learning Workshop
  - CERN, The workshop includes a CWP session on Machine Learning
- 9 Mar, 2017 - Software Triggers and Event Reconstruction WG meeting
  - LAL/Orsay, session at Connecting The Dots workshop
- 8 Mar, 2017 - S2I2-HEP/OSG/USCMS/USAtlas Panel at OSG All Hands Meeting
  - SDSC/UCSD
- 23-26 Jan, 2017 - HEP Software Foundation Workshop
  - University of California at San Diego / San Diego Supercomputer Center
- 7-9 Dec, 2016 - S2I2 HEP/CS Workshop
  - University of Illinois at Urbana-Champaign
A Roadmap for HEP Software and Computing R&D for the 2020s

HEP Software Foundation

ABSTRACT: Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders agree on the software goals and priorities, and that the efforts complement each other. In this spirit, this white paper describes the R&D activities required to prepare for this software upgrade.

Many of our collaborators and colleagues at U.S. CMS universities and laboratories have contributed to the CWP process and papers
Impact - Physics: Will efforts in this area enable new approaches to computing and software that maximize, and potentially radically extend, the physics reach of the detectors?

Impact - Cost/Resources: Will efforts in this area lead to 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?

Impact - Sustainability: Will efforts in this area significantly improve the long term sustainability of the software through the lifetime of the HL-LHC?

Interest/Expertise: Does the U.S. university community have strong interest and expertise in the area?

Leadership: Are the proposed focus areas complementary to efforts funded by the US-LHC Operations programs, the DOE, and international partners?

Value: Is there potential to provide value to more than one HL-LHC experiment and to the wider HEP community?

Research/Innovation: Are there opportunities for combining research and innovation as part of partnerships between the HEP and Computer Science/Software Engineering/Data Science communities?

US-ATLAS and US-CMS Ops were integral partners in developing this strategic plan.
Many workshops, involving a diverse group:
- International participants
- Computing Management from the Experiments and Labs
- Individuals interested in the problems
- Members of other compute intensive scientific endeavors
- Members of Industry
- http://s2i2-hep.org/
- https://hepsoftwarefoundation.org/

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
- arXiv 1712.06982
- arXiv 1712.06592

“The result: a Programme of Work for the field as a whole, a multifaceted approach to addressing growing computing needs on the basis of existing or emerging hardware.”

Eckhard Elsen (CERN Director of Research and Computing), editorial published with CWP/Roadmap
The Software Institute
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.

The plan for IRIS-HEP reflects a community vision developed by an international community process organized by the HEP Software Foundation ([https://hepsoftwarefoundation.org](https://hepsoftwarefoundation.org)). The S2I2-HEP conceptualization project ([http://s2i2-hep.org](http://s2i2-hep.org)) derived a Strategic Plan from the community roadmap which would leverage the strengths of the U.S. university community. IRIS-HEP aims to 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.

---

IRIS-HEP funded as a 5 year project from 1 Sep, 2018
IRIS-HEP Structure & Goals

S.1. Perform R&D in several critical areas towards meeting the challenges of the HL-LHC. The initial focus areas will be “Analysis Systems”, “Innovative Algorithms” and “Data Organization, Management and Access”.

S.2. Provide an integration path to deliver the output from R&D activities into the distributed and scientific production infrastructures.

S.3. Deliver a set of operational services needed for DHTC for the LHC experiments on the Open Science Grid.

S.4. Serve as an intellectual hub for the wider HEP and scientific community, providing expertise and training, and fostering wider collaborations around its areas of activity.
The Executive Board meets weekly.
IRIS-HEP Team

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

About 28 FTEs of funded effort spread over a larger number of people from 18 universities/institutions

Gender Diversity

Exec Board: 10%
Subaward PIs: 16.7%
Full Team: 17%

For comparison:
CoDaS-HEP 2019: 25.9%
US-CMS Physicists 2017: 16%
US-CMS Grad Students 2017: 17%
A Whirlwind Tour of IRIS-HEP
Innovative Algorithms
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
IRIS-HEP Innovative Algorithms Highlights

Parallel tracking contributions to MkFit

- Develop track finding/fitting implementations that work efficiently on many-core architectures (vectorized and parallelized algorithms):
- 4x faster track building w/ similar physics performance in realistic benchmark comparisons

<table>
<thead>
<tr>
<th>GPU</th>
<th>Development of the Kalman Filter</th>
<th>Porting ACTS seeding code to run on GPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tracking contributions to ACTS</td>
<td></td>
</tr>
</tbody>
</table>

Tracking contributions to ACTS

- Developing connections with other experiments (e.g. Belle-2, JLAB) who may be interested in using ACTS

ML on FPGAs contributions to HLS4ML/FastML

- Identifying specific use cases and operational scenarios for use of FPGA-based algorithms in experiment software trigger, event reconstruction or analysis algorithms

<table>
<thead>
<tr>
<th>ML4Jets establishing and curating common metrics and data sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim to connect with diverse segments of machine learning community. Strong connections with theoretical community interested in jet physics</td>
</tr>
<tr>
<td>Tree Neural network approach demonstrated on reference dataset</td>
</tr>
</tbody>
</table>

ML on FPGAs contributions to HLS4ML/FastML

- Developing connections with other experiments (e.g. Belle-2, JLAB) who may be interested in using ACTS

<table>
<thead>
<tr>
<th>GPU</th>
<th>Development of the Kalman Filter</th>
<th>Porting ACTS seeding code to run on GPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tracking contributions to ACTS</td>
<td></td>
</tr>
</tbody>
</table>

ML4Jets establishing and curating common metrics and data sets

- Aim to connect with diverse segments of machine learning community. Strong connections with theoretical community interested in jet physics
- Tree Neural network approach demonstrated on reference dataset


Data Organization, Management, and Access
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.
DOMA: Intelligent Data Delivery

- In the HL-LHC era, we must deliver more events and at lower latencies if the analysts want to make progress!
  - Low-latency delivery of events requires transformation from long-term archival formats that we want to decrease data size.
  - Data should be transformed and delivered at the storage level, not at the workstation.
  - Users should be enabled to work on a multitude of data formats (esp. non-ROOT) without having to write them to disk.
- We are currently prototyping an Intelligent Data Delivery service to:
  - Extract events from a data lake for fine-grained processing
  - Deliver events to analysis facilities at a high data rate.

**IMPACT / Status:**
- Working to integrate intelligent data delivery with ATLAS’s PanDA for fine-grained processing.
- Can transform and deliver ATLAS xAOD events for analysis.
- Working with Coffea team to deliver CMS NanoAOD events to Jupyter notebooks.
DOMA (Data Organization, Management, Access)

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.

ServiceX / Intelligent Data Delivery

Low-latency delivery of numpy-friendly data transformed from experiment custom formats enabling the use of community supported data science tools.

(joint effort with Analysis Systems)
There is a strong movement in the community to move from niche protocols for bulk data movement to more standardized ones such as HTTP.

- Bockelman co-leads the working group within the WLCG for “third party copy’ (TPC).

During IRIS-HEP, HTTP-TPC has gone from small test transfers to scale tests on servers to scale tests in the WLCG DOMA community.

- Demonstrated HTTP’s ability to achieve speeds similar to GridFTP on dedicated server hardware.

**IMPACT / Status:**

- Worked to make HTTP-TPC available in the storage systems used by U.S. LHC sites.
- With WLCG, worked to finalize a common, interoperable authorization scheme based on OAuth2 and JWT.
Analysis Systems
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.
Analysis Systems - Data Query

- **ServiceX (IDDS)** - Collaboration with DOMA to serve analysis data in a “columnar form”.
- **Awkward Array** - HEP-style column-wise data in the python ecosystem for manipulating the data
- **Coffea** - column-oriented framework for analysis (developed initially at FNAL in the US CMS context)
  - Builds on top of other backends allowing execution on Spark- or HTCondor-based resources.

Full chain to make a Z mass peak in electron data!
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.

Statistical Modeling Language and Tool Limit Extraction

Rewritten from C++ in Python to use TensorFlow or PyTorch as back end.

GPU acceleration comes for “free”

Just released and being incorporated into analyses now

Built into SciKit-HEP, a suite of packages that are being adopted by the community
Analysis Systems - Statistical Models

Implementation of widely used statistical tool in modern frameworks

NumPy, TensorFlow, PyTorch

Installation:

\$> \text{pip install pyhf}

By leveraging these tools, we inherit benefits

Auto-Differentiation:

Tensor libraries from ML community provide exact gradients $\frac{\partial L}{\partial \mu}, \frac{\partial L}{\partial \theta_i}$ for use in minimization.

Optimizers

pyhf likelihood are simple tensor-value python functions. Can use multiple minimization algorithms, such as scipy.minimize or MINUIT

Hardware Acceleration

For ML-library tensor backends, Computational graph can be transparently placed on hardware accelerators: GPUs and TPUs for order of magnitude speed-up in computation.

Reducing time to insight!

ROOT: 10+ hours
pyhf: < 30 minutes
Finalist for best paper award at Super Computing 2019
- Largest scale Bayesian inference ever using in a universal probabilistic programming language
  - Applied to complex LHC Physics use case: Sherpa code base of ~1M lines of code in C++
- 230x speedup for synchronous data parallel training of a 3DCNN-LSTM neural network
  - 1,024 nodes (32,768 CPU cores)
  - 128k minibatch size, largest for this NN architecture
  - One of the largest-scale use of PyTorch built-in MPI
- Novel protocol (PPX) to execute & control existing, large-scale, scientific simulator code bases

https://arxiv.org/abs/1907.03382
Scalable Systems Laboratory
Scalable Systems Laboratory (SSL)

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

- 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
SSL base platform deployed

- Repurposed CS R&D cluster
  - 3k cores, 2x40g to campus 100g
  - SciDMZ; Kubernetes for flexibility for services and compute

- Deployment of AS & DOMA services (REANA & ServiceX) & ATLAS analytics via SLATE & Helm

- Backfilled by OSG when not in use by IRIS-HEP

Federated ID access (institutional, CERN account), edge services hosting, Unix account provisioning, OSG-LHC software environment
SSL Cyberinfrastructure for training

OSG-LHC
Open Science Grid (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 ⅓ 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.
OSG scope within IRIS-HEP

The People in OSG

Operations = UNL
Security = Indiana University
Software = U. Wisconsin – Madison
Networking = U. Michigan

A total of 6 FTE across 11 people.
These people have worked together
and with the LHC program years.
OSG Highlight: Transitioning away from Globus

- Globally, the LHC today depends on GSI for authentication and gridFTP for bulk data transfer.
  - Both are no longer supported by their original developers.
- OSG inherited the source code, and is maintaining it within the context of “Grid Community Toolkit” that was created for this purpose in 2018.
- We developed a roadmap for replacement of both GSI and gridFTP that has been socialized globally, and across science domains.
  - August 22nd 2019: Roadmap and schedule presented to LHC ops program via OSG council
  - September 12th 2019: Roadmap and schedule presented to WLCG via GDB
  - January 2020: First demo of a US-LHC site free of GSI and gridFTP (prototype & proof-of-concept)
  - January 2021: OSG production software without GSI and gridFTP is released
  - January 2022: End of support of GSI and gridFTP in OSG releases.
OSG-Highlight: Central Service Operations Paradigm

- Traditionally, OSG provided software, testing, deployment & operations documentation/training/support.
  - Tier-1/2/3 sites instantiated the services for LHC community based on this software.
- **New paradigm more cleanly separates hardware, infrastructure services, and science support:**
  - RAW hardware capacity is provided by cloud, HPC, Universities & National labs (T1/2/3).
  - Centralized service organization(s) deploy & operate services required to turn RAW capacity into useable capacity for the US LHC community.
    - Some services run by ops program, others by entities like OSG that cater to multiple domains in NSF, and multiple frontiers in DOE-HEP (and DOE-NP).
  - Physics support & training
    - Domain specific projects (e.g. IRIS-HEP & LHC ops program) provide support & training in the tools and software necessary to do the science.
- IRIS-HEP started exploring this new paradigm (collaboration between SSL & OSG):
  - Support containers in addition to RPMs to instantiate services at T1/2/3
  - Developed container security policy document
  - Working under leadership of Rob Gardner (SSL/SLATE) & Romain Wartel (CERN) & Jim Basney (TrustedCI) within “WLCG SLATE Security Working Group” to create a new security model that supports this new paradigm.
Blueprint and Intellectual Hub
Blueprint Activity - Maintaining a Common Vision

- Small "blueprint" workshops 3-4 times per year with key personnel and experts
- Facilitate effective collaborations by building and maintaining a common vision
- Answer specific questions within the scope of the Institute’s activities or within the wider scope of HEP software & computing.

- 10 Sep - 11 Sep, 2019 - Blueprint: Accelerated Machine Learning and Inference (Fermilab)
- 20 Feb, 2020 - Training Blueprint (Vidyo)
How does the ensemble of US Software R&D efforts fit together to implement the HL-LHC Software/Computing roadmap described in the Community White Paper and meet the challenges of the HL-LHC? Which areas are not covered by US R&D efforts?

How do the US Software R&D efforts collaborate with each other and with international efforts? How do these efforts align with and leverage national exascale, national NSF OAC priorities and trends in the broader community?

How should the US R&D efforts be structured and organize in order to impact planned updates (all in ~2021/2022) to the HSF Community White Paper, the software/computing part of the US Snowmass process and HL-LHC experiment-specific software/computing TDRs?

~40 participants from US National Labs, the two US-LHC Operations Programs, the IRIS-HEP team, and NSF & DOE program directors/managers. The meeting was organized with the involvement and blessing of the DOE program managers.

This built on a similar workshop at CUA which we organized in 2017 during the conceptualization phase for IRIS-HEP.

Checkpoint the relationship of the IRIS-HEP efforts and collaborations with the two US-LHC Operations programs.

Explore how IRIS-HEP interacts with other NSF efforts and planned and ongoing DOE efforts (SciDAC, LDRD, ECP, ESNET, HEP-CCE, facilities, etc.)

Discussion-oriented: a report will be forthcoming.
Intellectual Hub - Building Community & Vision

Sponsorship and/or (co-)organization of relevant community workshops and activities:

- Joint HSF, OSG & WLCG Meeting
  - [Link](https://indico.cern.ch/event/759388/)

- **Connecting The Dots 2020**
  - [Link](https://indico.cern.ch/event/831165/)

- **ML4Jets**
  - [Link](https://indico.cern.ch/event/769263/)

- **PyHEP 2019**
  - (16-18 October, 2019)
  - [Link](https://indico.cern.ch/event/813759/)

- **ML4Jets**
  - (15-17 January, 2020)
  - [Link](https://ml4physicalsciences.github.io/)

- **Machine Learning and the Physical Sciences**
  - [Link](https://ml4physicalsciences.github.io/)

- **Latin American Workshop on Software and Computing challenges in High-Energy Particle Physics (LAWSCHEP 2019)**
  - [Link](https://indico.cern.ch/event/813325/)

Full list: [https://iris-hep.org/events.html](https://iris-hep.org/events.html)
PyHEP Workshop Series

PyHEP is a series of workshops started in 2018 to discuss and promote the usage of Python in the HEP community at large. It has been supported by DIANA/HEP, and now IRIS-HEP, in collaboration with HSF.

PyHEP 2020 will soon be announced: 11-13 July, 2020 in Austin, TX, partially overlapping with the SciPy 2020 conference, also in Austin, TX

This is not just a “programming language” issue, it is a key place where HEP can explore how to interact with, learn from, contribute to, and perhaps lead areas in the larger scientific, data science and ML communities. (Including use of open data, experimentalist - theorist interactions, etc.)

A consistent message from our students and postdocs who transition to industry and other fields is that we teach them great skills, but they are limited initially by only knowing HEP-only tools.
IRIS-HEP is establishing itself as an intellectual hub for HEP software & computing.

We begin from a strong position built during the CWP & institute conceptualization period. Through strong partnerships with other projects, our stakeholders and activities like the blueprint workshops we continue to build this intellectual hub.

https://iris-hep.org/collaborations
Training
Major Activities

Training for HEP community
- Software Carpentry Workshops
  - Github/Unix/Python/Plotting
  - Universities, National Labs
  - 180 people trained
  - Female participants highly encouraged

Outreach
- Programming for STEM teachers
- Underrepresented Communities
- Scientific Software Club at UPRM

Careers and Jobs
- IRIS-HEP Fellows (HEP/non-HEP)
  - Graduate
  - Undergraduates

Collaboration
- FIRST-HEP (http://first-hep.org)
- HEP Software Foundation (https://hepsoftwarefoundation.org/)
How to build a sustainable and scalable training framework that grows skills through multiple stages of people’s careers?

This is a general framework for training, but from the NSF we have funds from both IRIS-HEP (OAC-1836650) and a separate project FIRST-HEP (OAC-1829707, OAC-1829729, http://first-hep.org) which are working towards implementing this model.
ML Hackathon UPRM 2019

http://codas-hep.org
The school was created as part of an NSF PIF grant (PHY-1521042), and is now being sustained and evolved through both IRIS-HEP (OAC-1836650) and a separate NSF CyberTraining project FIRST-HEP (OAC-1829707, OAC-1829729, http://first-hep.org)

CoDaS-HEP School

http://codas-hep.org

The CoDaS-HEP school aims to provide a broad introduction to these critical skills as well as an overview of applications High Energy Physics. Specific topics to be covered at the school include:

- Parallel Programming
- Big Data Tools and Techniques
- Machine Learning - Technology and Methods
- Practical skills: performance evaluation, use of git for version control

The program includes both lectures and practical hands-on exercises.
Mentoring - IRIS-HEP Fellows Program

**Key Insight:** we need to provide incentivized and explicit paths forward for enthusiastic students from the more advanced training schools (ESC/Bertinoro, CoDaS-HEP, MLHEP, etc.) or for people who become engaged with our software projects in other ways.

**Project focused:** bring students into contact with “mentors” to work on a specific, pre-defined project, allowing them to grow their software skills and project experience. The fellow supports, when possible, travel and subsistence for a 3 month extended stays in the mentor’s institution.

We began this activity as part of DIANA/HEP (an NSF SI2-SSI project) and are continuing it in IRIS-HEP.

We have had 3 fellows thus far during the ramp-up of IRIS-HEP, and have just announced a call for Fellow proposals for summer, 2020 and beyond.
## IRIS-HEP Topical Meeting Series

Meetings are announced on the [announcements@iris-hep.org](mailto:announcements@iris-hep.org) mailing list. Recorded videos are available in YouTube (see links on the individual agenda pages).

### February 2020
- 19 Feb  ACTS tracking
- 12 Feb  HL-LHC tracking
- 03 Feb  Modeling computing resource needs/costs

### January 2020
- 29 Jan  Primary vertex finding at LHCb
- 27 Jan  Allen project

### November 2019
- 27 Nov  Unfolding
- 18 Nov  Graph Neural Nets
- 11 Nov  Accelerating Raw Data Analytics (to be rescheduled)

### October 2019
- 09 Oct  Software for Dark Matter Experiments

### September 2019
- 25 Sep  ML4Jets benchmarks
- 16 Sep  Summer student project presentations

### August 2019
- 21 Aug  Summer student project presentations
- 19 Aug  Summer student project presentations

### July 2019
- 31 Jul  A Common Tracking Software (ACTS) Project
Broader Impacts

Outreach in General
- U. Puerto Rico: carpentries material for coding, ML hackathon (first ML activity among Phys/CS/Eng students)
- Connect to School Teachers, and research undergraduates in STEM, includes Hispanic and other underrepresented populations.
- Exploring potential QuarkNet scale-out

Intellectual Leadership
- Focus Areas are intellectual centers for their work (topical meetings, blueprint meetings, slack, our website, etc.)
- Management roles in community organizations such as the WLCG.
- Training as a way of leading people to a common set of successful solutions to typical physics problems
- Impact people at national labs (FNAL, BNL, etc.) as well as universities.

Outreach in General
- Outreach to ensure software/practices adoption
  - Blueprint meeting early in 2020 to discuss this
  - Targeted to ramp up after the blueprint meeting

Training
- ~300 people trained
- Physicists
- Future and current developers

Diversity, broadening participation: recruiting
- Fellows program promotes development of advanced research software skills, and provide a connection with a mentor/role model
- Institute has a Code of Conduct
- Institute promotes software development as an inclusive and collaborative activity
And Back to the Top
Summary

- **IRIS-HEP** was funded on September 1st, 2018
  - We are approaching the end of the design phase
  - Projects in all phases (design, prototype, and production) exist.
  - We are fully staffed, ~30 FTE’s
  - Full description of projects available on our website, [http://iris-hep.org](http://iris-hep.org)

- **Community Impact**
  - Software is being adopted by others, in some cases dramatically.
  - Facilities work in SSL and OSG is leading the international field

- **Community Outreach**
  - We’ve reached almost 1000 people with our workshops, and another 300 with our training efforts
  - We continue to organize Blueprint workshops to build community consensus.

- **Next**
  - Start “Execution Phase” September 2020
  - Work on integrating projects in prototype stage into coherent and scalable software for the community
  - The “Snowmass Process-2021” provides an opportunity for us to update the Community White Paper/Roadmap.