



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

(Virtual) 2020 Retreat Introduction

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



OAC-1836650

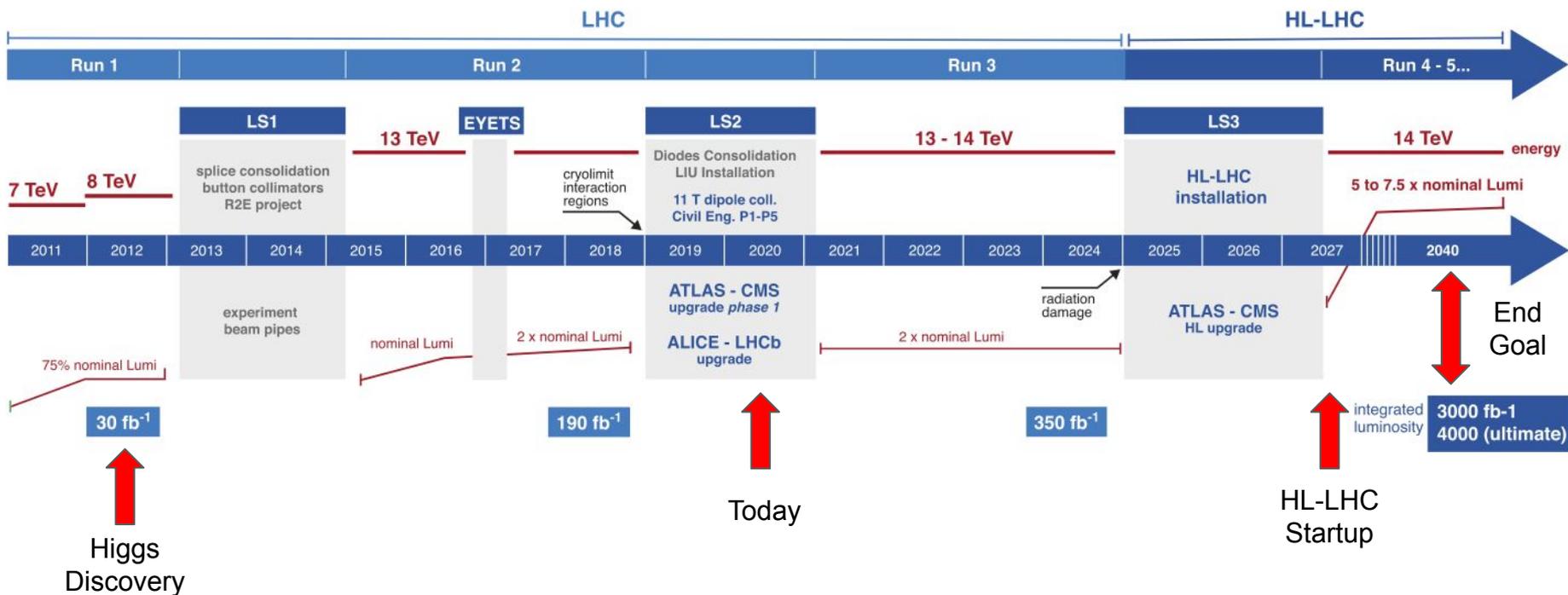
<http://iris-hep.org>



Timeline



LHC / HL-LHC Plan

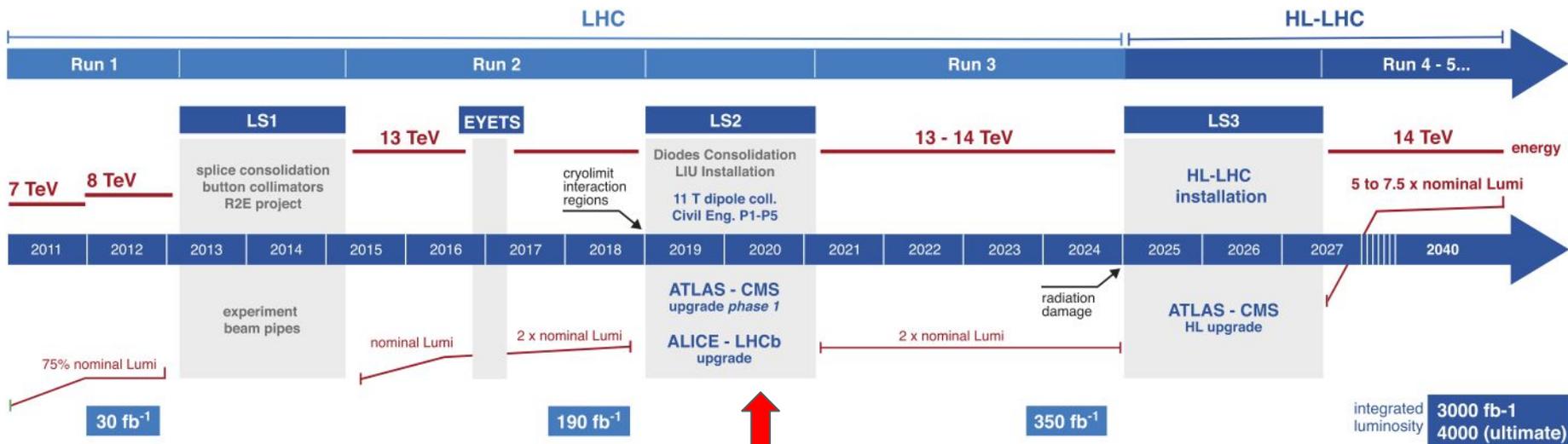


[We will see what happens to Run 3, given the coronavirus delays.]

Timeline

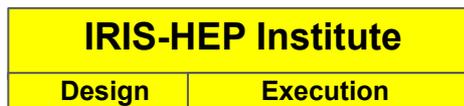


LHC / HL-LHC Plan



Institute Conceptualization and Community White Paper Process

S2I2-HEP



Snowmass

CTDR

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

U.S. HEP Community Planning Process

NSF 18-Month Review of IRIS-HEP



The NSF 18-month Review of IRIS-HEP took place at Princeton on 27-28 February: <https://indico.cern.ch/event/855317/>

This was the first major review of IRIS-HEP since it was funded. This review gated our passage from the so-called “Design Phase” (first 2 years) to the “Execution Phase” (later 3 years). The alternative to “success” was “ramp-down”.

We succeeded and did very well. This is due the excellent work all of you have done in the past 18 months.

The “Execution Phase” implies that we further “up our game”, and we will discuss aspects of how we should do that in this meeting.

[There will be 30 month and 42 month reviews. The latter will be used to decide whether NSF will allow us to submit a renewal proposal for an additional 5 years.]

Report from NSF 18-Month Review of IRIS-HEP



From the “Summary and Future Prospects” in the review report:

“After reviewing the panel charge, reading through the provided materials, listening to presentations and exploring areas of concern, the review panel recommends moving IRIS-HEP to the execution phase of the award. Though the panel identified, and expressed in recommendations, a number of concerns and risks, it concludes that IRIS-HEP is on track to complete its mission as described.

The future for IRIS-HEP is bright. Alignment, definition, and prioritization of projects will be a key to success. It has the potential to transform the tools and techniques for the communities it serves.”

Recommendations from NSF 18-Month Review



Some example recommendations that we should keep in mind during this meeting:

- 1) *“In preparation for the execution phase, IRIS-HEP should establish center-wide grand challenges to assess the progress toward the primary institute high-level goals. Milestones should be objective, technical, and relatable by the broader community and tied back to the grand challenge problems.”*
- 2) *“IRIS-HEP is encouraged to continue to develop its “connectivity diagram”. This is an important way to communicate to future review panels and IRIS-HEP PIs how the project is greater than the sum of its parts. ...”*
- 3) *“We recommend that IRIS-HEP formalize its outreach to external scientific communities. The institute would benefit by creating a formal role for this activity, to enhance participation in planning and training exercises and ensure active involvement from the relevant experimental collaborations.”*
- 4) *“We recommend that IRIS-HEP document their processes for managing the life-cycle of projects from inception, through demonstration of success and subsequent hand-off to operations.”*

[There were also some additional fine-grained recommendations that we will also need to handle, but which don't require broad discussion like these examples.]

IRIS-HEP Vision

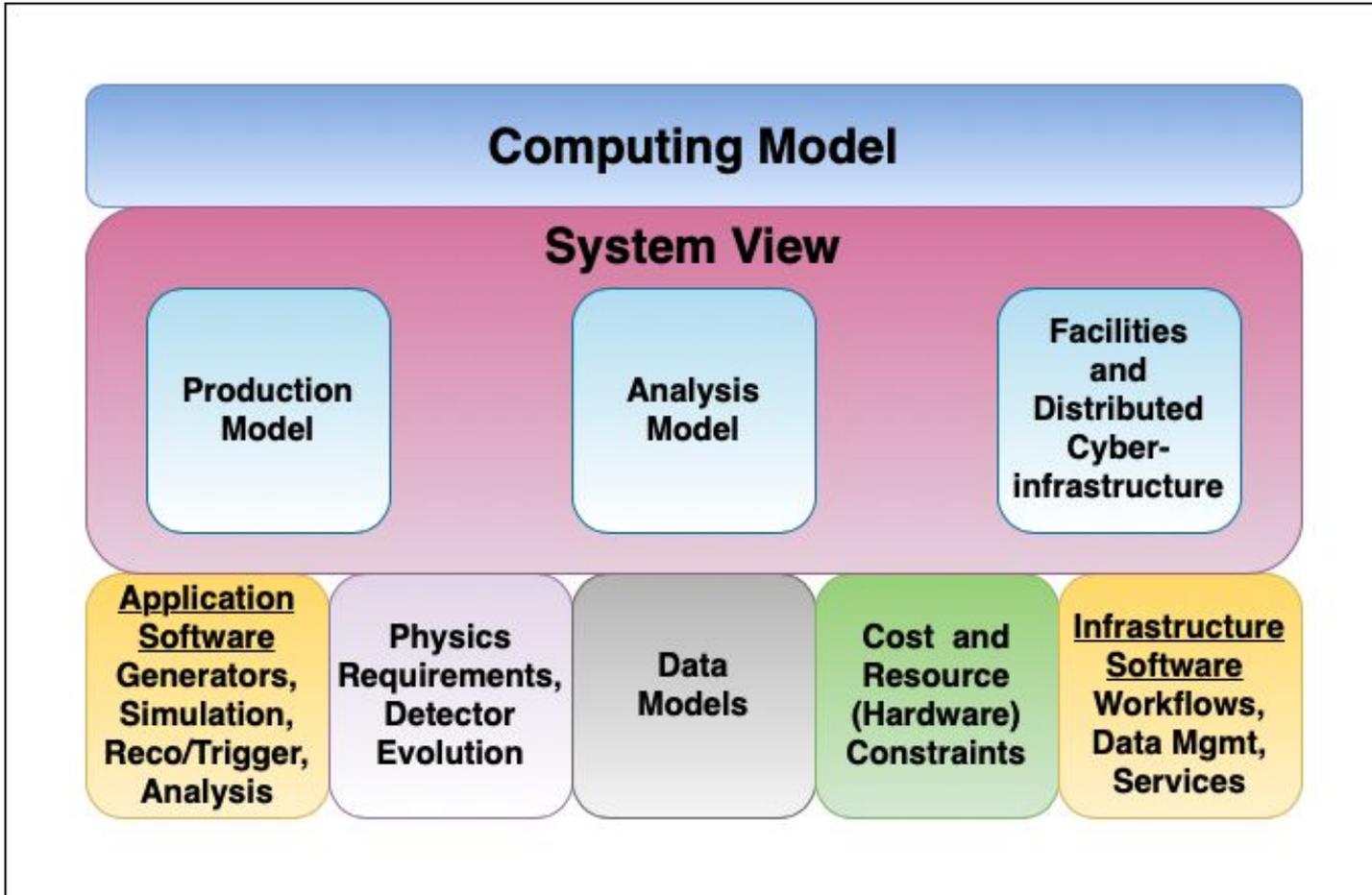


IRIS-HEP addresses key elements of the “Roadmap for HEP Software and Computing R&D for the 2020s” and is implementing the “Strategic Plan for a Scientific Software Innovation Institute (S2I2) for High Energy Physics” submitted to the NSF in December 2017. We aim to:

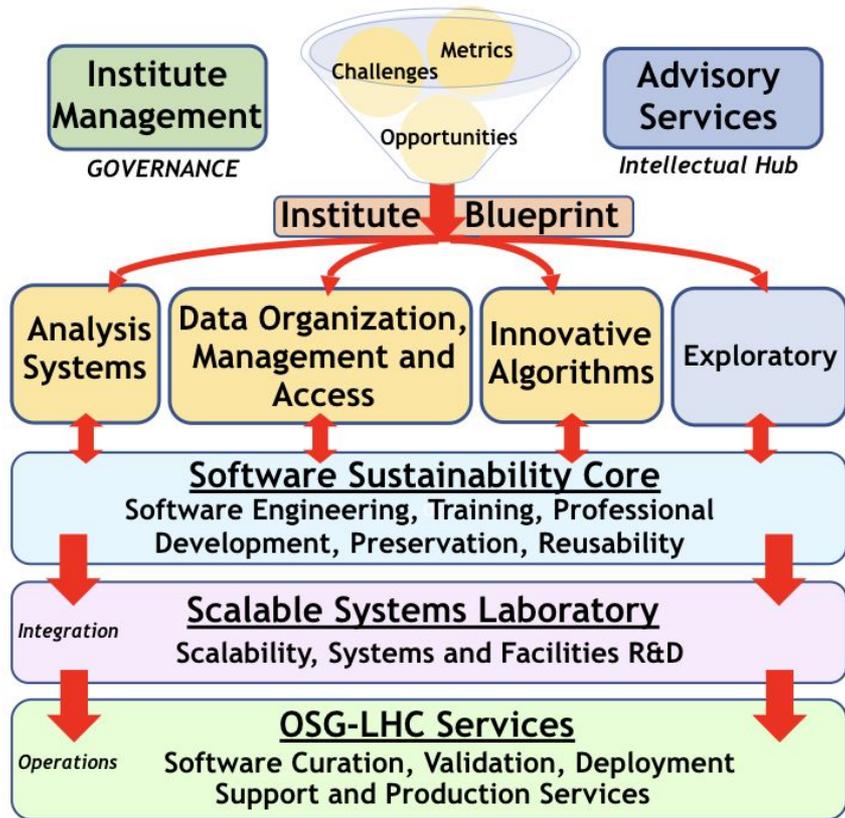
- Enable new approaches to computing and software that maximize, and potentially radically extend, the **physics reach** of the detectors.
- Make improvements in software efficiency, scalability and performance and make use of the advances in CPU, storage and network technologies, that allow the experiments 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.

Software, System View & Computing Models



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 <i>Peter.Elmer@cern.ch</i>	Gordon Watts University of Washington <i>Institute co-PI and Deputy Executive Director</i>	Brian Bockelman Morgridge Institute <i>Institute co-PI and DOMA R&D Area Lead</i>	Robert Tuck Princeton University <i>Associate Project Manager</i>	Floe Fusin-Wischusen Princeton University <i>PICSciE Institute Manager</i>	Rob Gardner University of Chicago <i>SSL Area Lead</i>
					
Heather Gray University of California, Berkeley <i>Innovative Algorithms Area co-Lead</i>	David Lange Princeton University <i>David.Lange@cern.ch</i>	Kyle Cranmer New York University <i>Analysis Systems Area Lead</i>	Sudhir Malik University of Puerto Rico at Mayaguez <i>Training, Education and Outreach Coordinator</i>	Mark Neubauer University of Illinois at Urbana-Champaign <i>Blueprint Coordinator</i>	Frank Wuerthwein University of California, San Diego <i>OSG-LHC Area Lead and OSG Executive Director</i>

The Executive Board meets weekly.

Strategic and Annual Planning



Our model for annual planning involves several steps:

1. Discussions between the Exec. Board area leads and ED/DED of R&D progress in each area (including collaborations developed, community activity, opportunities, etc.)
2. “PI chats” similar to #1, but focused on each university
3. An annual retreat (this meeting) to develop the next year’s plan and (this year) a bigger strategic picture for the next 3 years.
4. Updates to the Statement of Work and budgets for each institution.



The first (Sep. 2019) retreat was mostly focused on team building and detailed “next year” planning, with less emphasis on “Strategic Planning” changes in the middle of the design phase. This retreat (unfortunately virtual) will involve both aspects, as well as preparations to write the “Execution Phase Project Execution Plan”.



Planning at the 2020 IRIS-HEP Retreat

At this Retreat we need to work on planning for two different time scales:

- Year 3 - detailed plans (similar to what we did last year)
- Year 4 & 5 - strategically where we are going and where we will wind up in terms of overall impact of IRIS-HEP at the end of the (first) 5 year project

As a means of structuring the strategic planning, the panel suggested something that we had already in fact already been discussing: the development of “**grand challenges**”. These were used successfully during the pre-LHC ramp-up (and earlier on experiments like BaBar) to either drive large scale integration forward or demonstrate clearly that parts are not working.

We will be discussing various ideas for such “challenges” that we can pursue on a time scale of 2-2.5 years, say, and then we can work backwards from there to define intermediate milestones. These will naturally be goals which we will co-develop with the experiments, the US Operations programs and other partners.



IRIS-HEP Team

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



Steve Lantz
Cornell University

Michael (Tres) Reid
PHD Student

Dan Riley
Cornell University

Peter Wittich
Cornell University

Zelik Shah
Systems Analyst



Bei Wang
Indiana University

Ricky Bala Garg
Stanford University

Lauren Tompkins
Stanford University

Lincoln Bryant
University of Chicago

Andrew Chien
University of Chicago



Brian Sosa
Indiana University

Daniel Crank
Massachusetts Institute of Technology

Philip Harris
Massachusetts Institute of Technology

Markus Kütt
Massachusetts Institute of Technology

Evan Massaro
Massachusetts Institute of Technology



Rob Gardner
University of Chicago

Marc Weinberg
University of Chicago

Gowtham Atturi
University of Cincinnati

Kandrick Li
University of Cincinnati

Milo Skoloff
University of Cincinnati



Dylan Rankin
Massachusetts Institute of Technology

Mika Williams
Massachusetts Institute of Technology

Brian Bockelman
Massachusetts Institute of Technology

Johann Brothner
Massachusetts Institute of Technology

Kyle Cranmer
New York University



Marian Stahl
University of Cincinnati

Markus Atkinson
University of Illinois at Urbana-Champaign

Matthew Feickert
University of Illinois at Urbana-Champaign

Ben Galwey
National Center for Superconducting Applications

David S. Katz
University of Illinois at Urbana-Champaign



Irina Esopo
New York University

Alexander Held
New York University

Sebastian Macke
New York University

Helko Mueller
Research Software Engineer (RSE/LSU)

Sinceri Pérez
Research Software Engineer



Mark Neubauer
University of Michigan - Ann Arbor

Sham McKee
New York University

Ken Bloom
University of Nebraska - Lincoln

Okeana Shadura
University of Nebraska - Lincoln

Derek Weitzel
University of Nebraska - Lincoln



Maureen Carothers
Princeton University

Peter Elmer
Princeton University

Flo Fadin-Wichusen
Princeton University

David Lange
Princeton University

Imara Osborne
Princeton University



Huijun Zhu
University of Nebraska - Lincoln

Marian Zvada
University of Nebraska - Lincoln

Xiaocang Ai
University of California, Berkeley

Nicholas Cooke
University of California, Berkeley

Heather Gray
University of California, Berkeley



Jim Pivarski
Princeton University

Henry Schreiner
Princeton University

Savannah Thais
Post-doctoral researcher

Robert Tuck
Associate Project Manager

Vassil Vassilev
DIANA/HEP USQCD collaborator



Klaener (Aaron) Chu
University of California, Santa Cruz

Ian Jimenez
UC Santa Cruz

Jeff Lawrence
University of California, Santa Cruz

Carlos Maltabes
Santa Cruz

Diego Davila
University of California, San Diego



Edgar Fajardo
University of California, San Diego

Slava Krutelyov
University of California, San Diego

Marlo Masciocchia
University of California, San Diego

Igor Stiflig
University of California, San Diego

Matevs Tadel
University of California, San Diego



Frank Wuerthwein
University of California, San Diego

Ari Yagil
University of California, San Diego

Sushie Malik
University of Pennsylvania at Newark

Mason Proffitt
University of Washington

Emma Torro
University of Washington



Gordon Watts
University of Washington

Tim Cartwright
University of Wisconsin-Madison

Carl Edquist
University of Wisconsin-Madison

Wen Guan
University of Wisconsin-Madison

Brian Lin
University of Wisconsin-Madison



Aaron Meeks
Madison

Malyia (Mat) Sliemers
University of Wisconsin-Madison

Tim Theisen
University of Wisconsin-Madison

Arjun
Systems Administrator

Software Integration Developer

OSG Release Manager

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%

Interaction Between IRIS-HEP and US LHC Ops

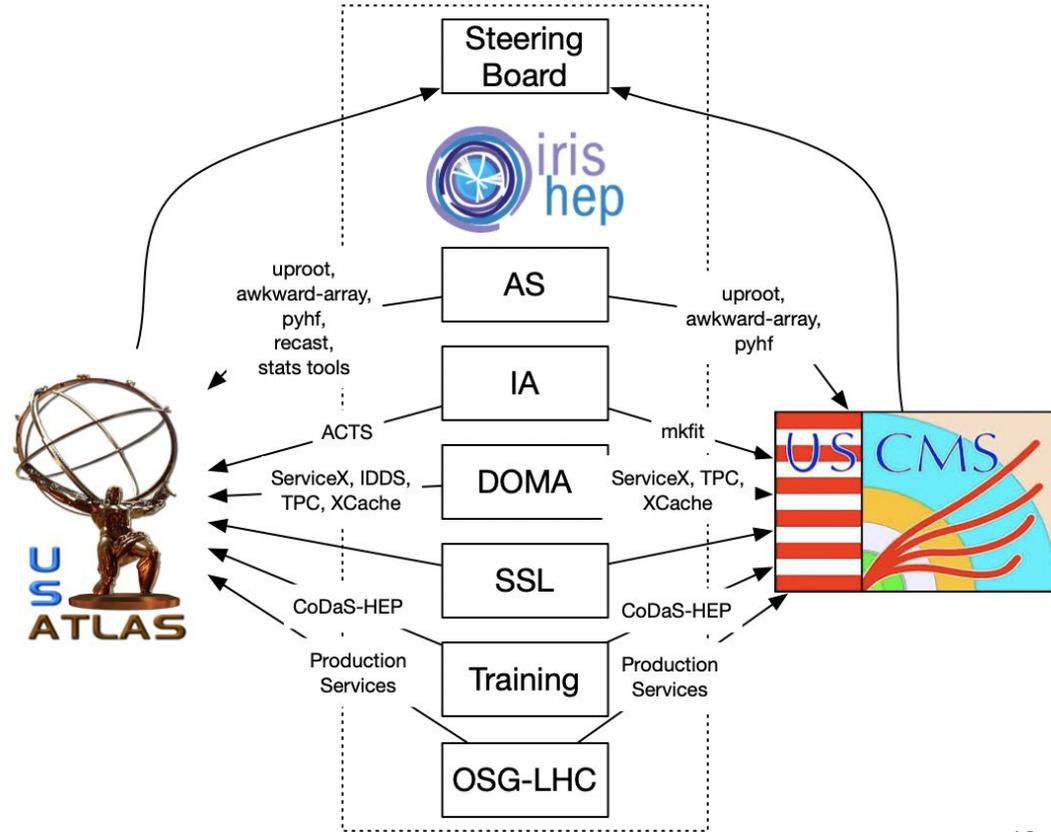
In its first year, IRIS-HEP has established links across its focus areas to both US-LHC operations programs.

- A subset of the projects & contributions where the organizations interact are highlighted in this figure.
- The U.S. LHC Ops programs help guide the R&D activities through the steering board.

IRIS-HEP includes effort to help bring R&D projects through **integration** (SSL) and **production** (OSG-LHC).

- Even then, **requires close collaboration** with the Operations programs to ensure we derive value!

Heather Gray and Brian Bockelman also have roles as co-coordinators of the HL-LHC R&D areas in US-ATLAS and US-CMS



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 (not in 2020, unfortunately) and subsistence/stipend for up to 3 months.

			
Bo Zheng Rice University	David Liu University of Washington	Manjari Trivedi University of Michigan, Ann Arbor	Vladimir Ovechkin University of Washington, Seattle
<i>May - August 2020</i>	<i>May - August 2020</i>	<i>May-August 2020</i>	<i>June-August 2020</i>

(More photos/webpages to come...)

We have a number of Fellows for this summer (and we are still selecting them), as well as students from other sources (university programs, GSoC) this summer. Please make sure to make this an interesting and educational experience for them all.

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.

An in-person event is not possible this year, so we will explore this week what, if anything, can be done as a virtual event.



CoDaS-HEP 2017



CoDaS-HEP 2018

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>)

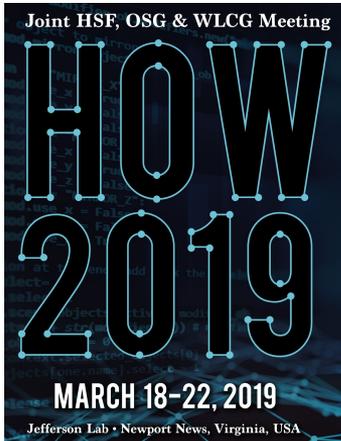


CoDaS-HEP 2019

Intellectual Hub - Building Community & Vision



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



<https://indico.cern.ch/event/759388/>



<https://indico.cern.ch/event/831165/>



1-2 July 2019
CERN
Europe/Paris timezone

Search...

<https://indico.cern.ch/event/813759/>



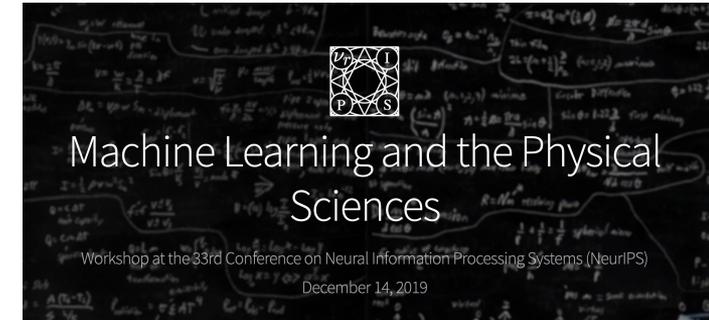
<https://indico.cern.ch/event/769263/>

PyHEP 2019

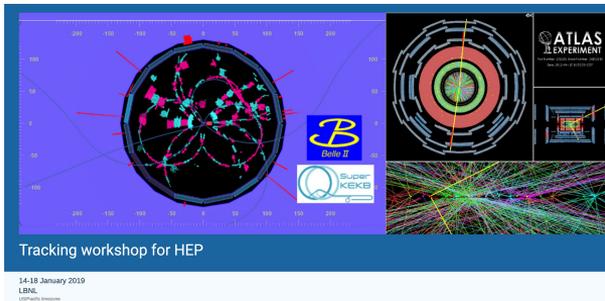
(16-18 October, 2019)

ML4Jets

(15-17 January, 2020)



<https://ml4physicalsciences.github.io/>



<https://indico.physics.lbl.gov/indico/event/712/>



Full list: <https://iris-hep.org/events.html>

<https://indico.cern.ch/event/813325/>

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](#) is now a virtual event: 13-17 July, 2020

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.



A growing community: 38 participants at PyHEP 2018, 55 participants at PyHEP 2019, and we are hoping for at least as many (virtual) participants at PyHEP 2020

Community Activities and Events

July, 2020

- 13 Jul - 17 Jul, 2020 - [PyHEP 2020 Workshop](#) (*Online*)

May, 2020

- 26 May - 29 May, 2020 - [IRIS-HEP Full Team Retreat](#) (*Zoom*)
- 25 May - 29 May, 2020 - [ACTS Tracking for HEP Workshop](#) (*DESY (Hamburg, Germany)*)

April, 2020

- 22 Apr - 24 Apr, 2020 - [Connecting the Dots 2020](#) (*Princeton University*)

February, 2020

- 17 Feb - 19 Feb, 2020 - [Analysis Preservation Bootcamp](#) (*CERN*)

January, 2020

- 15 Jan - 17 Jan, 2020 - [ML4Jets2020](#) (*New York University*)

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)

May 2020

-  11 May [Accelerating Raw Data Analysis with the ACCORDA Software and Hardware Architecture](#)
-  04 May [Rumble: JSONiq \(query language\) on Spark](#)

April 2020

-  29 Apr [Zarr: big array management](#)

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](#)

The last months have been a bit chaotic, to say the least, but we do need to keep planning for topical meetings. Several ideas came out during the PI chats (they should follow up) and we should be thinking about that in this Retreat, too.

In addition, we should plan on some topical meetings in August and September for Fellows, GSoC students, etc. to present their work.

Execution Phase Project Execution Plan (PEP)



Contents

1 The Institute	2		
1.1 Institute Summary	2		
1.2 Services Provided to the Community	2		
1.3 Evaluation Criteria	3		
1.4 Structure of the Institute	3		
2 Management, Staffing and Expenditures Plan	5		
2.1 Project Management Structure and Plans	5		
2.2 Project Governance	6		
2.2.1 Day-to-Day Operations	6		
2.2.2 Advisory Panel and Steering Board	6		
2.3 Staffing Plan	8		
2.4 Staffing and Recruitment / Retention Strategies	9		
2.4.1 Post-doc Mentoring Plan	9		
2.5 Institutional Financial and Business Oversight	10		
2.6 Expenditure Plan	10		
3 Work Breakdown Structure and Project Schedule	11		
3.1 Tasks by WBS area	11		
3.2 Project Schedules	14		
3.2.1 Project Schedule: Management	14		
3.2.2 Project Schedule: Analysis Systems	16		
3.2.3 Project Schedule: DOMA	17		
3.2.4 Project Schedule: Innovative Algorithms	18		
3.2.5 Project Schedule: SSC	20		
3.2.6 Project Schedule: SSL	22		
3.2.7 Project Schedule: OSG-LHC	23		
4 Metrics	25		
4.1 WBS1: Management	25		
4.2 WBS2: Analysis Systems	25		
4.3 WBS3: DOMA	25		
4.4 WBS4: IA	25		
4.5 WBS5: SSC	25		
4.6 WBS6: SSL	26		
4.7 WBS7: OSG-LHC Services	26		
5 Sustainability	26		
5.1 Sustainability in the HEP community	26		
5.2 Sustainability of the OSG	27		
5.3 Sustainability as part of a national software cyberinfrastructure	28		
6 Software and Data	28		
6.1 Analysis Systems Software and Data	28		
6.2 Data Organization, Management, and Access	29		
6.3 Innovative Algorithms Software and Data	31		
6.4 OSG-LHC Software and Data	31		
6.5 IRIS-HEP Data Management Plan	33		
6.5.1 Research Data Artifacts	33		
7 Software Development Lifecycle (SDLC) Methodology	34		
7.1 HEP Collaborations and the Software Development Lifecycle	34		
7.2 IRIS-HEP Software Development Lifecycle	35		
7.2.1 The OSG-LHC Software Development Lifecycle	36		
8 Community Engagement	38		
8.1 The Institute as Intellectual Hub	38		
8.2 OSG-LHC Community Engagement	42		
8.3 Synergistic Activities	42		
9 Security	45		
9.0.1 IRIS-HEP Physical Security Plans	45		
9.0.2 IRIS-HEP Cyber-Security Plans	45		
10 Risk Plan	46		
10.1 Risks for WBS1: Management	46		
10.2 Risks for WBS2: Analysis Systems	47		
10.3 Risks for WBS3: DOMA	48		
10.4 Risks for WBS4: Innovative Algorithms	49		
10.5 Risks for WBS5: Software Sustainability Core (SSC)	51		
10.6 Risks for WBS6: Scalable Systems Laboratory (SSL)	51		
10.7 Risks for WBS7: OSG-LHC	52		
11 Diversity Plan	54		
A Appendix - WBS Accounting and Reporting	54		

We need to summarize aspects of the current planning process in an “Execution Phase Project Execution Plan (PEP)” in the weeks following this meeting. We can consider evolution of the structure (“Design Phase” structure show here) if needed.

Summary

- IRIS-HEP was funded on September 1st, 2018
 - We have passed our first major review! (18 months)
 - We are approaching the end of the “design phase”, and need to plan for the “execution phase”
- Next Steps:
 - We start the “Execution Phase” on 1 September 2020
 - We need to work on our plans for integrating projects in prototype stage into coherent and scalable software for the community, planning on two time scales: for Year 3 (“next year”) and for the next 3 years (incl. Year 4 & 5).
 - This meeting is about building those plans, including a new focus on “Grand Challenges” which can be used to tie together various elements of our work and potentially integrate with the experiments, the Ops programs, etc.
 - This is also a time to discuss any new *opportunities* where we might make an impact and/or any ways in which we organize to be more impactful (internally or through new collaborations)
 - The coronavirus will likely limit travel in the next year or so, at least for large meetings, so we will have to continue to be creative
 - The “Snowmass Process-2021” also provides an opportunity for us to update the Community White Paper/Roadmap.