

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

PI: Peter Elmer (Princeton), co-PIs: Brian Bockelman (Morgridge Institute), Gordon Watts (U.Washington) with UC-Berkeley, University of Chicago, University of Cincinnati, Cornell University, Indiana University, MIT, U.Michigan-Ann Arbor, 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

http://iris-hep.org

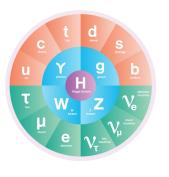
IRIS-HEP was funded as of 1 September, 2018, and is currently ramping up activities





OAC-1836650

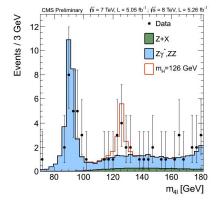
Science Driver: Discoveries beyond the Standard Model of Particle Physics



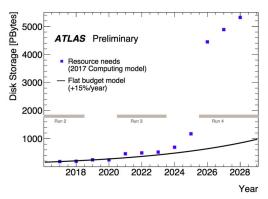
From "Building for Discovery - Strategic Plan for U.S. Particle Physics in the Global Context" - Report of the Particle Physics Project Prioritization Panel (P5):

- 1) Use the Higgs boson as a new tool for discovery
- 2) Pursue the physics associated with neutrino mass
- 3) Identify the new physics of dark matter
- 4) Understand cosmic acceleration: dark matter and inflation
- 5) Explore the unknown: new particles, interactions, and physical principles





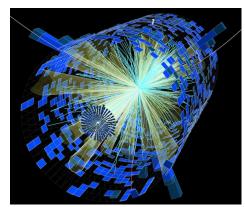
Computational and Data Science Challenges of the High Luminosity Large Hadron Collider (HL-LHC) and other HEP experiments in the 2020s



HIGGS BOSO

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).



arXiv 1712.06982



The Community White Paper Process (2016-2017)

Involved A Diverse

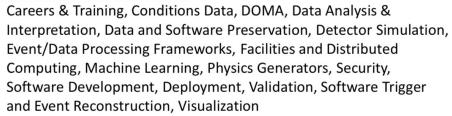
- Computing Management from the Experiments and Labs
- · Individuals interested in the problems
- Members of other compute intensive scientific endeavors
- Members of Industry



rXiv.org > physics > arXiv:1712.06982	Search or Article ID	All fields 🗸 🤇	
	ele I Advanced search)		
Physics > Computational Physics	Dov	vnload:	
A Roadmap for HEP Software and Computing R&D for the 2020s	PD Ott	F her formats	
Johannes Albrecht, Antonio Augusto Alves Jr, Guilherme Amadio, Giuseppe Andronico, Nguyen Anh-Ky, Laurent Aphecetche, John Apostola	kis, Makoto	(cc)) BY	
Asai, Luca Atzori, Marian Babik, Giuseppe Bagliesi, Marilena Bandieramonte, Sunanda Banerjee, Martin Barisis, Lothar A.T. Bauerdick, Stel Beitorte, Douglas Benjamin, Catrin Bernius, Wahid Bhimji, Riccardo Maria Bianchi, Ian Bird, Catherine Biscarat, Jakob Biomer, Kenneth Bio Tommaso Boccali, Brian Bockelman, Tomasz Bold, Daniele Bonacorsi, Antonio Boveia, Concezio Bozzi, Marko Bracko, David Britton, Andy Predrag Buncic, Paolo Calafura, Simone Campana, Philippe Canal, Luca Canali, Gianpaolo Carlino, Nuno Catsro, Marco Cattaneo, Gianlu Cerminara, Javier Cervantes Vilanueva, Phili Chang, John Chanpman, Gang Chen, Taylor Childers, Peter Clarke, Marco Citemencic, Eric C Jeremy Coles, Ian Collier, David Colling, Gioria Corti, Gabriele Cosmo, Davide Costanzo, Ben Couturier, Kyle Cranmer, Jack Cranshaw, Lec	m, physic Buckley, epre a Chan gneras, hep-ex nardo physic		
Cristella, David Crooks, Sabine Crépé-Renaudin, Robert Currie, Sûnje Dallmeier-Tiessen, Kaushik De, Michel De Cian, Albert De Roeck, An Delgado Peris, Frédéric Derue, Alessandro Di Girolamo, Salvatore Di Guida, Gancho Dimitrov, Caterina Doglioni, Andrea Dotti, Dirk Duellma Duflot, Dave Dykstra, Katarzyna Dziedziniewicz-Wojcik, Agnieszka Dziurda, Ulrik Egede, Peter Elmer, Johannes Elmsheuser, V. Daniel Elvir Eulisse, Steven Farrell, Torben Ferber, Andrej Flucci, Ian Fisk, Conor Fitzpatrick, José Fitx, Andrea Formica, Alessandra Forti, Giovanni Fra	nn, Laurent a, Giulio	ences & Citations PIRE HEP ers to cited by) SA ADS	
James Frost, Stu Fuess, Frank Gaede, Gerardo Ganis, Robert Gardner, Vincent Garonne, Andreas Gellrich, Krzysztof Genser et al. (209 ad		le Scholar	
authors not shown) (Submitted on 18 Dec 2017 (v1), last revised 19 Dec 2018 (this version, v5))	Book	mark (what is this?)	
Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in det	ector		

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 while paper describes the R&D activities required to prepare for this software upgrade.

Individual Papers on the arXiv:



Community White Paper & the Strategic Plan

Growing a Global Collaboration



JLab March, 2018 HSF/OSG/WLCG



UCSD/SDSC January, 2017 HSF CWP

Naples March, 2017 WLCG/HSF

Annecy June, 2017 HSF CWP



IRIS-HEP

Sustainable Software R&D objectives

1) Development of <u>innovative algorithms</u> for data reconstruction and triggering;

 Development of highly performant <u>analysis</u> <u>systems</u> that reduce "time-to-insight" and maximize the HL-LHC physics potential; and

3) Development of <u>data organization, management</u> <u>and access systems</u> 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.



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

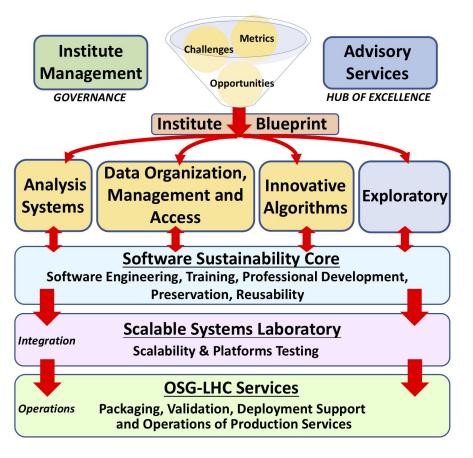
Intellectual Hub for the HEP Community





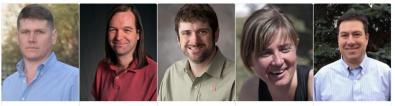
The plan for IRIS-HEP reflects a community vision developed by an international community process organized by the HEP Software Foundation (<u>https://hepsoftwarefoundation.org</u>). The S2I2-HEP conceptualization project (<u>http://s2i2-hep.org</u>) derived a Strategic Plan from the community roadmap 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 Structure and Executive Board



Executive Board

The IRIS-HEP Executive Board manages the day to day activities of the Institute.



Princeton University Peter.Elmer@cern.ch	Gordon Watts University of Washington	Brian Bockelman Morgridge Institute	Heather Gray University of California, Berkeley	David Lange Princeton University David.Lange@cern.ch
Institute PI and Executive Director	Institute co-PI and Deputy Executive Director	Institute co-PI and DOMA R&D Area Lead	Innovative Algorithms Area co-Lead	Innovative Algorithms Area co-Lead



Kyle Cranmer New York University

Area Lead

Sudhir Malik University of Puerto Rico at Mayaguez

Mark Neubauer Rob Gardner University of Illinois at Urbana-Champaign

Frank Wuerthwein University of Chicago University of California San Diego

Analysis Systems

Training, Education and Outreach Coordinator

Blueprint Coordinator

SSL Area Lead

OSG-LHC Area Lead and OSG Executive Director

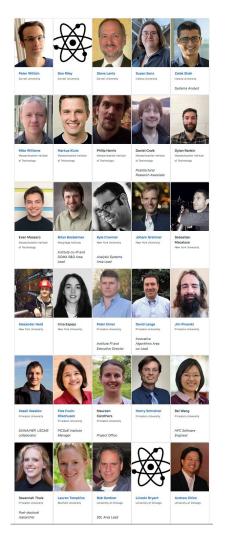


IRIS-HEP Team

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

About 30 FTEs of effort spread over a larger number of people from 18 universities/institutions

		6		2
Emma Torro	Tim Cartwright	Brian Lin	Tim Theisen	Mátyás (Mat)
University of Washington	University of Wisconsin- Mariana	University of Wisconsin- Madison	University of Wisconsin- Madison	Selmeci University of Wisconsin-
mannyan	1000001	1000001		Madaon
	OSG Deputy	OSG Software Team	OSG Release	Software Integration
Post-doc	Executive Director	Manager	Manager	Developer
\bigotimes				
Carl Edquist	Aaron Moate			
University of Wisconsin- Mediaon	University of Waconsin- Madison			
(PROVIDE)	Systems Administrator			



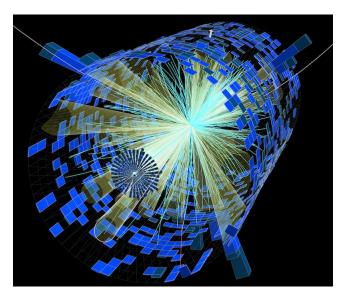
		C	R	
Marc Weinberg University of Chicago	Mike Sokoloff University of Cincinnati	Marian Stahl University of Cincinnati	Mark Neubauer University of Illinois at	Daniel S. Katz University of Illinois at
			Urbana-Dhempaign Blueprint Coordinator	Urbana-Dhampoign
			E	
Ben Galwesky National Center for	Markus Atkinson University of Illinois at	Matthew Feickert University of Illinois at	Shawn McKee University of Michigan-	Ken Bloom University of Nebraska -
Supercomputing Applications	Urbane-Champelign	Urbana-Chempelign Postdoctoral	Ann Arbor OSG Networking	Lincoln
		Postdoctoral researcher	OSG Networking Area Coordinator	
Derek Weitzel University of Nebraska - Lincoln	Oksens Shedura University of Nebraska - Uncoln	Marian Zvada University of Nebraska - Lincoln	Huljun Zhu University of Nebraska - Lincoln	Heather Gray University of California, Berkeley
	DIANA/HEP collaborator	Systems integrator		Innovetive Algorithms Area co-Lead
Ø				
Xiaocong Al UC Berkeley	Carlos Maltzahn University of California, Santa Cruz	Jeff LeFevre University of California, Sente Cruz	Xiaowei(Aaron) Chu University of Californie, Serie Cour	Edgar Fajardo University of California, San Diego
Postdoc and ACTS			Ph.D Student	OSC Software Team Developer
		-	9	
Avi Yagil University of California,	Frank Wuerthwein University of California,	Matevz Tadel University of California,	Slava Krutelyov University of California,	Mario Mascioveochio
Son Diego	Sen Diego OSG-LHC Aree Lead and OSB Executive Director	San Diego	San Diego	University of California, Sen Diego
	E.			R
Igor Sfiligoi University of California,	Diego Davila University of Californie,	Sudhir Malik University of Puerte Rice	Gordon Watts University of	Mason Proffitt University of
San Diego	San Diego	at Mayaguez	Weshington	Weshington
Lead Scientific	Scientific Software	Training, Education	Institute co-Pl and	

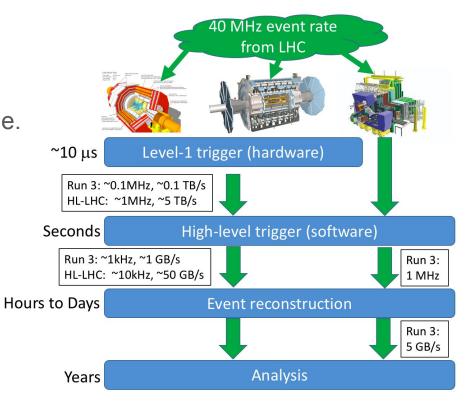


Innovative Algorithms - Trigger/Reconstruction



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





Innovative Algorithms

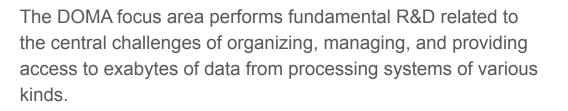


These algorithms face a number of new challenges during HL-LHC:

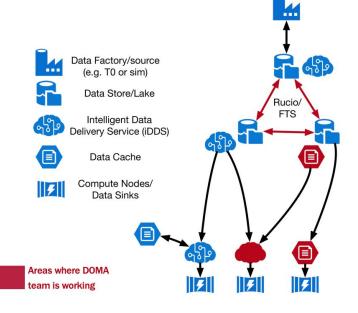
- Upgraded accelerator capabilities, with more collisions per bunch crossing ("pile-up")
- Detector upgrades, including new detector technologies and capabilities Increased event rates to be processed
- Emerging computing architectures

Specific R&D investments include collaboration with ACTS, continuation of the parallel Kalman Filter tracking project, work on HLS4ML, etc.

Data Organization, Management and Access (DOMA)



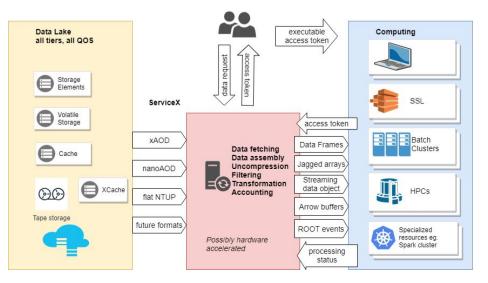
- **Data Organization**: Improve how HEP data is serialized and stored.
- **Data Access**: Develop capabilities to deliver filtered and transformed event streams to users and analysis systems.
- **Data Management**: Improve and deploy distributed storage infrastructure spanning multiple physical sites. Improve inter-site transfer protocols and authorization



DOMA: Intelligent Data Delivery

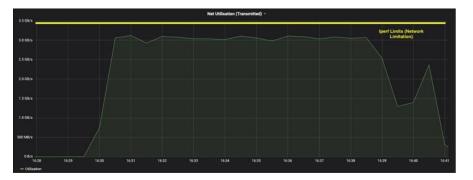
- In the HL-LHC era, we must deliver more events and at lower latencies - if the analysts want to make progress!
 - Low-latency delivery of events requires transformation from long-term archival formats that we want to decrease data size.
 - Data should be transformed and delivered at the storage level, 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 (IDDS) that delivers the desired LHC events without requiring complex transforms or skims from the user.





DOMA: Moving Bulk WLCG Data

- 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.



working group

Homepage of the WLCG





Analysis Systems R&D



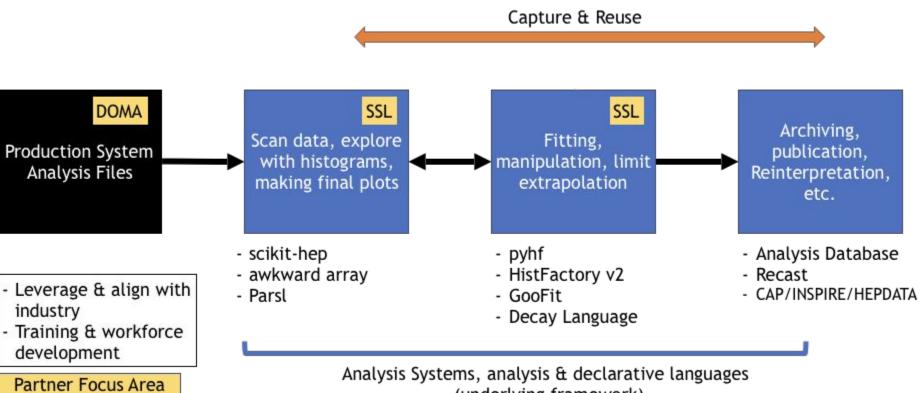
Primary Goals: Develop sustainable analysis tools to extend the physics reach of the HL-LHC experiments by

- creating greater functionality
- reducing time-to-insight
- lowering the barriers for smaller teams, and
- streamlining analysis preservation, reproducibility, and reuse.

Compared to DOMA and Innovative Algorithms (which has more targeted reco/trigger goals), the Analysis Systems group is dealing with more "greenfield" area where there is a very heterogeneous set of use cases and relevant components.

The nature of IRIS-HEP Analysis Systems tasks is more exploratory: "big R" (R&d)

Analysis Systems Data Flow and Projects

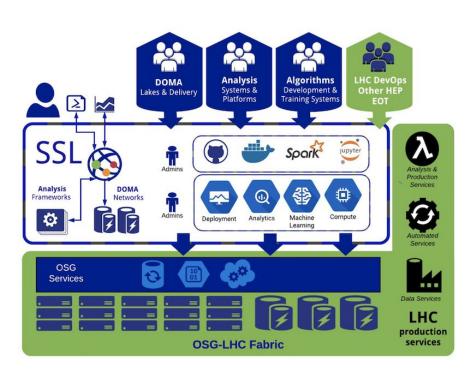


(underlying framework)

Scalable Systems Laboratory (SSL)

Goal: Provide the Institute and the HL-LHC experiments with scalable platforms needed for development in context

- Provides access to infrastructure and environments
- Organizes software and resources for scalability testing
- Does foundational systems R&D on accelerated services
- Provides the integration path to the OSG-LHC production infrastructure





Open Science Grid (OSG) for the LHC

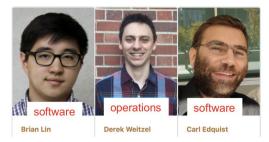




The People in OSG







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

Selmeci

A total of 6 FTE across 11 people. These people have worked together and with the LHC program years.

(Slide from FKW)

Open Science Grid (OSG) and IRIS-HEP





OSG Effort Portfolio

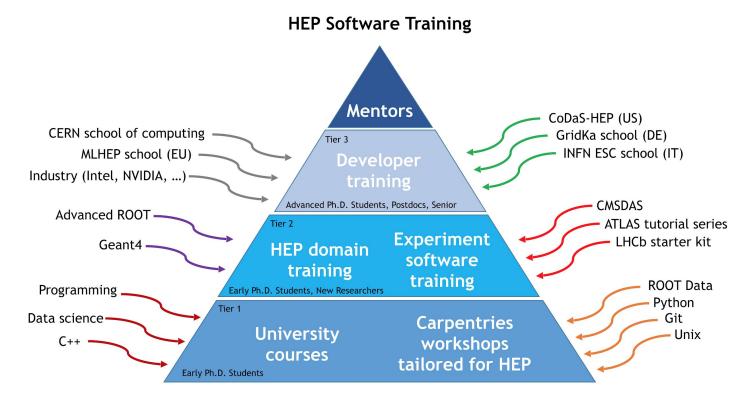


A		В	С	D	E	F	G	н
		Σ	Technology	Operations	Support	Networking	Security	Management
	Σ	17.56	7.46	4.05	2.82	1.24	1.00	0.98
OSG N5Y		6.86	2.48	2.05	1.82	0.00	0.20	0.32
IRIS-HEP		5.94	3.08	1.00	0.00	0.39	0.80	0.67
SAND		0.75	0.00	0.00	0.00	0.75	0.00	0.00
OSG-NP		1.00	0.50	0.50	0.00	0.00	0.00	0.00
TNRP		1.00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

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

(Slide from FKW)

Training and Education - Sustainability/Scalability



This is a general framework for training, but from the NSF we have funds from both IRIS-HEP (OAC-183665) and a separate project FIRST-HEP (OAC-1829707, OAC-1829729, <u>http://first-hep.org</u>) which can work towards implementing this model.

Training, Education and Outreach Events

Upcoming events:

- 19-21 Aug, 2019 ATLAS Software Carpentries Training
 - Lawrence Berkeley National Laboratory
 - Indico page

Past events:

- 22-26 Jul, 2019 Computational and Data Science for High Energy Physics (CoDaS-HEP) 2019 School
 - Princeton University
 - Webpage
- 10 Jun, 2019 FIRST-HEP/ATLAS Software Training
 - Argonne National Laboratory
 - Indico page
- 3-4 Jun, 2019 An introduction to programming for STEM teachers
 - University of Puerto Rico at Mayaguez
 - Indico page

- 24-26 Apr, 2019 Machine Learning Hackathon for UPRM Students
 - University of Puerto Rico at Mayaguez
 - Indico page
- 1-2 Apr, 2019 Software Carpentry Workshop
 - Fermi National Accelerator Laboratory
 - Indico page

In collaboration with FIRST-HEP (http://first-hep.org), the Carpentries (https://carpentries.org) and others





CoDaS-HEP 2017



CoDaS-HEP 2018

Current IRIS-HEP Fellows



Raghav Kansal	Pratyush (Reik)
University of California,	Das
San Diego	Institute of Engineerin
	& Management
	(Kolkata)
IRIS-HEP Fellow	
	IRIS-HEP Fellow
Jun-Aug 2019	Jun-Sep 2019



ML Hackathon UPRM



CoDaS-HEP 2019

http://codas-hep.org

IRIS-HEP Community Activities and Events



Upcoming Events:

IRIS-HEP team members are involved in organizing the following events:

- 29 Jul, 2019 IRIS-HEP Tutorial: Fast columnar data analysis with data science tools (Northeastern University / APS DPF 2019)
- 19 Aug 23 Aug, 2019 ATLAS Software Carpentries Training (LBNL)
- 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)
- 13 Dec 14 Dec, 2019 Machine Learning and the Physical Sciences at NeurIPS 2019 (Vancouver Convention Centre)
- 15 Jan 17 Jan, 2020 ML4Jets2020 (in planning) (New York University)

Recent Events:

- 22 Jul 26 Jul, 2019 CoDaS-HEP 2019 (Princeton University)
- 1 Jul 2 Jul, 2019 TrackML Challenge: Grand Finale (CERN)
- 21 Jun 22 Jun, 2019 Blueprint: Analysis Systems R&D on Scalable Platforms (NYU)
- 19 Jun 20 Jun, 2019 Analysis Systems Topical Workshop (NYU)
- 10 Jun, 2019 FIRST-HEP/ATLAS Software Training (Argonne National Lab)
- 3 Jun 4 Jun, 2019 An introduction to programming for STEM teachers (University of Puerto Rico at Mayaguez)
- 6 May 8 May, 2019 Analysis Description Languages Workshop (Fermilab)

See <u>http://iris-hep.org</u> for the full and latest list

IRIS-HEP Topical Meeting Series

Home » Projects » IRIS-HEP » Topical Meetings

Topical Meetings

Two weekly time slots are available for IRIS-HEP topical meetings:

- Mondays 17:30-18:30GVA (Vidyo and 40-R-B10 at CERN)
- Wednesdays 18:00-19:00GVA (Vidyo only)

There is one event in the future. Hide

April 2019

15 Apr Development of new Histogram tools

March 2019

- 25 Mar Introduction to modern CDN Architectures
- 04 Mar The FAST project

February 2019

- 25 Feb Analysis Description Languages
- 18 Feb Integration of C++ Modules into CMSSW
- 13 Feb HLS4ML: Using ML on FPGAs to enhance reconstruction output
- 04 Feb Training for Software, Computing, Computational and Data Science in HEP

January 2019

28 Jan FuncX: High Performance Function as a Service for Science



https://indico.cern.ch/category/10570/

Meetings are announced on the <u>announcements@iris-hep.org</u> mailing list

Recorded videos are available in Youtube (see links on the individual agenda pages)

Connecting with IRIS-HEP



Website: <u>http://iris-hep.org</u>

Public announcement mailing list: <u>announcements@iris-hep.org</u> [Subscribe]

Topical meetings: <u>https://indico.cern.ch/category/10570/</u>

We will be continuing to organize, co-organize and host various events going forward, see the main project website above.





Monday:

• IRIS-HEP Tutorial: Fast columnar data analysis with data science tools (Nick Smith, Jim Pivarski)

Tuesday:

- COFFEA Columnar Object Framework For Effective Analysis (Nick Smith)
- pyhf: a pure Python statistical fitting library from the high energy physics community with tensors and autograd (Giordon Stark)
- Extending RECAST for truth-level analysis reinterpretations (Alex Schuy)

Wednesday:

• ACTS: a common track reconstruction software (Xiaocong Ai)

Thursday:

• Evolving CMS offline computing towards LHC Run3 and HL-LHC (David Lange)

Summary

(W) iris hep

HEP faces major challenges in the 2020s: Data, Compute, Staffing

The HSF executed an important community process that produced the CWP.

The collaborative spirit continues to grow and bodes well for us (as a community) to meet those challenges.

IRIS-HEP

We are focusing on 3 R&D areas from the CWP: Innovative Algorithms, Analysis Systems, and DOMA.

Plus training, a dedicated integration activity and continuity for the OSG services for the LHC.

We are just beginning our activities and are looking forward to collaborating with many of you in the coming years!