Poster #40

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

OAC-1836650

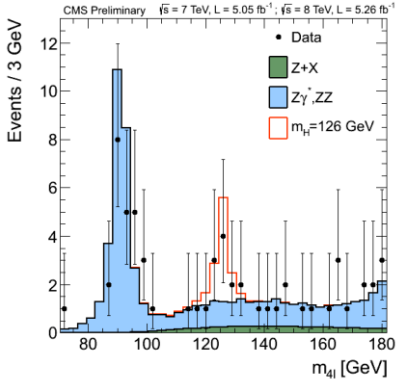
CSSI Meeting, Feb 14, 2020

Science Driver: Discoveries beyond the Standard Model of Particle Physics

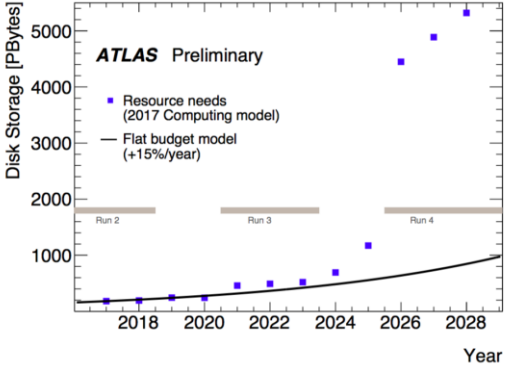
From "Building for Discovery - Strategic Plan for U.S. Particle Physics in the Global Context" - Report of the Particle Physics Project Prioritization Panel (P5):

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

Seattle Snowmass 2021

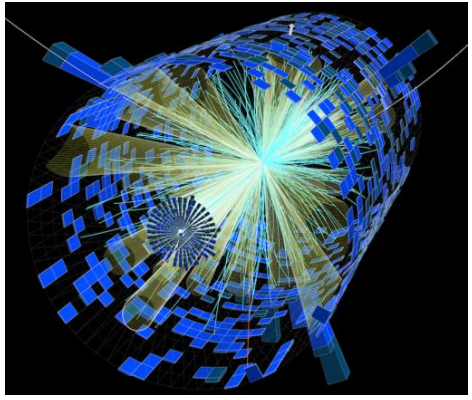


Computational and Data Science Challenges of the High Luminosity Large Hadron Collider (HL-LHC) and other HEP experiments in the 2020s



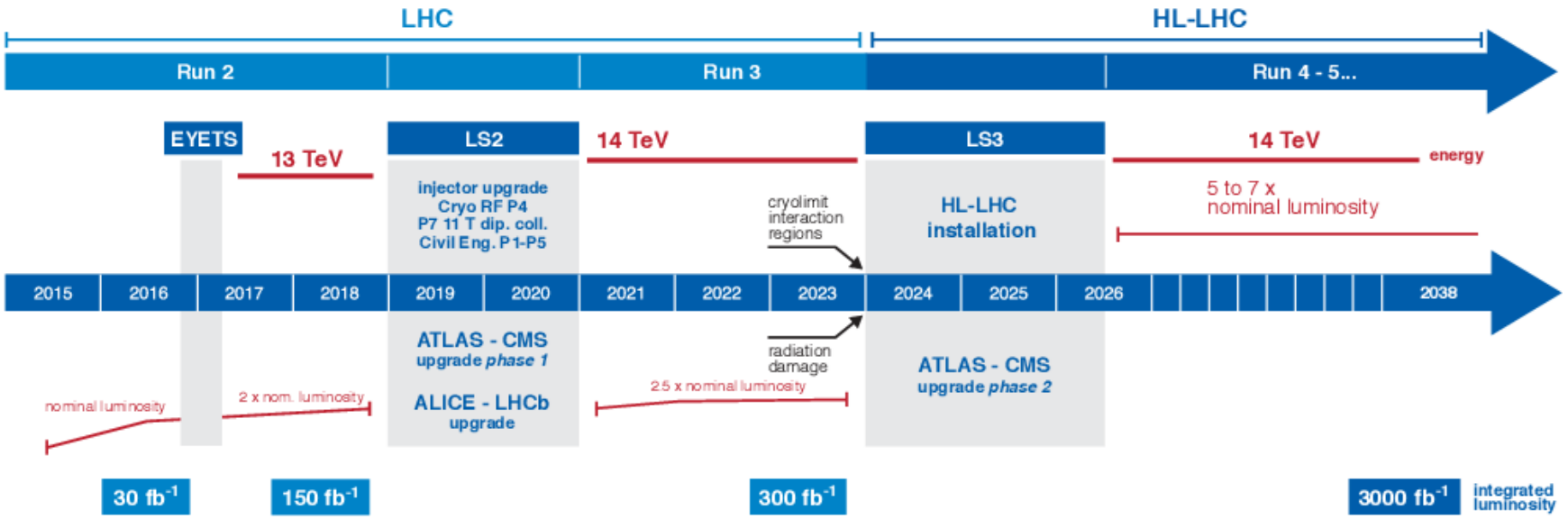
The HL-LHC will produce exabytes of science data per year, with increased complexity: an average of 200 overlapping proton-proton collisions per event.

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).



Timeline

LHC / HL-LHC Plan



CTDR

S2I2-HEP

Institute Conceptualization and Community White Paper Process

IRIS-HEP Institute

Design | **Execution**

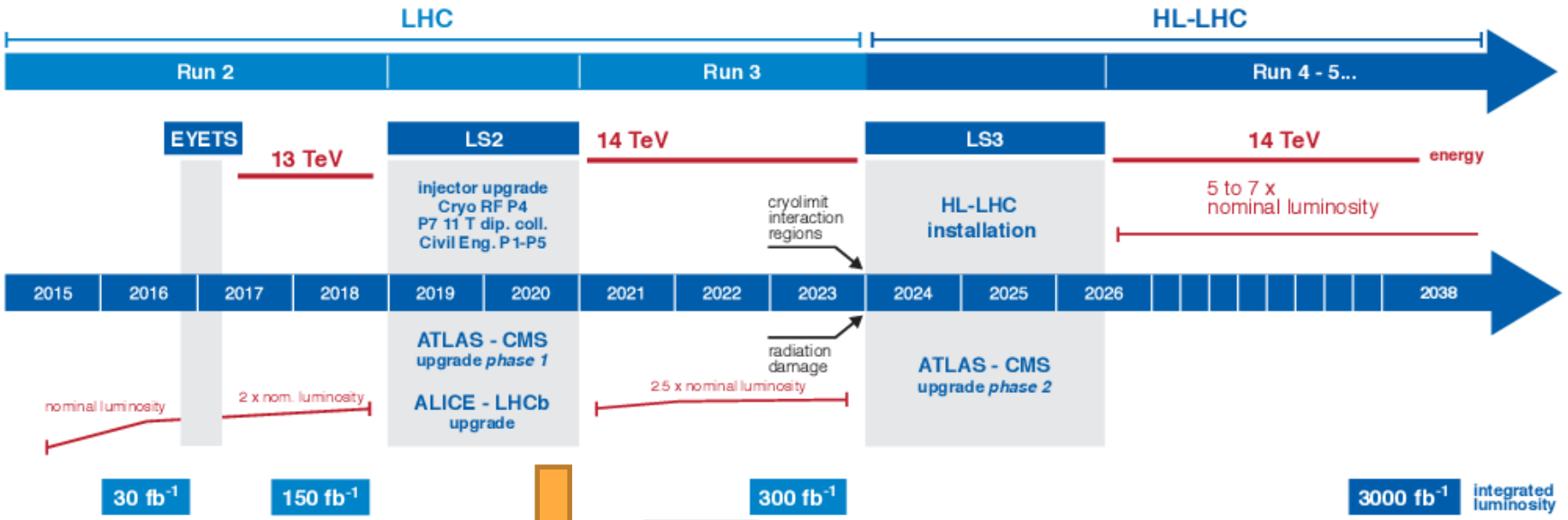
CERN HL-LHC Planning - Computing Technical Design Reports (CTDR) - ATLAS/CMS

Snowmass

U.S. HEP Community Planning Process

Timeline

LHC / HL-LHC Plan



CTDR

S2I2-HEP

IRIS-HEP Institute

Design | **Execution**

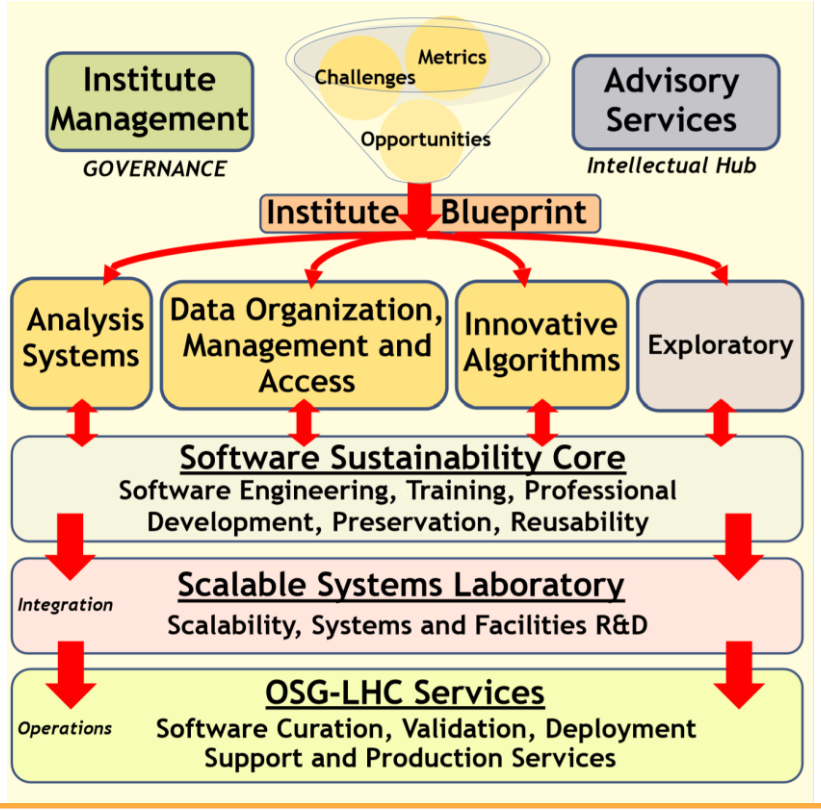
Snowmass

Institute Conceptualization and Community White Paper Process

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

U.S. HEP Community Planning Process

Structure And Focus Areas



Research & Software Development

Sustainability

Scale Up

Operations

Application Software

Infrastructure Software

Our Audience: LHC Physicists and LHC Facility Operations Groups

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.

All software is open source

Experiment's Production System

Data Query, histogramming, plotting, statistical models, fitting, archiving, reproducibility, publication

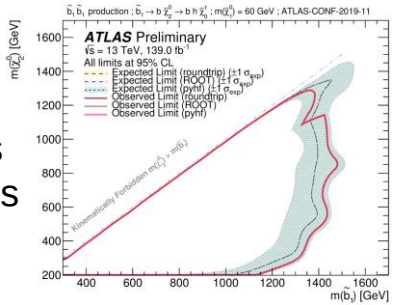
pyhf
∂ifferentiable
ℒikelihoods

Statistical Modeling Language and Tool
→ Limit Extraction

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

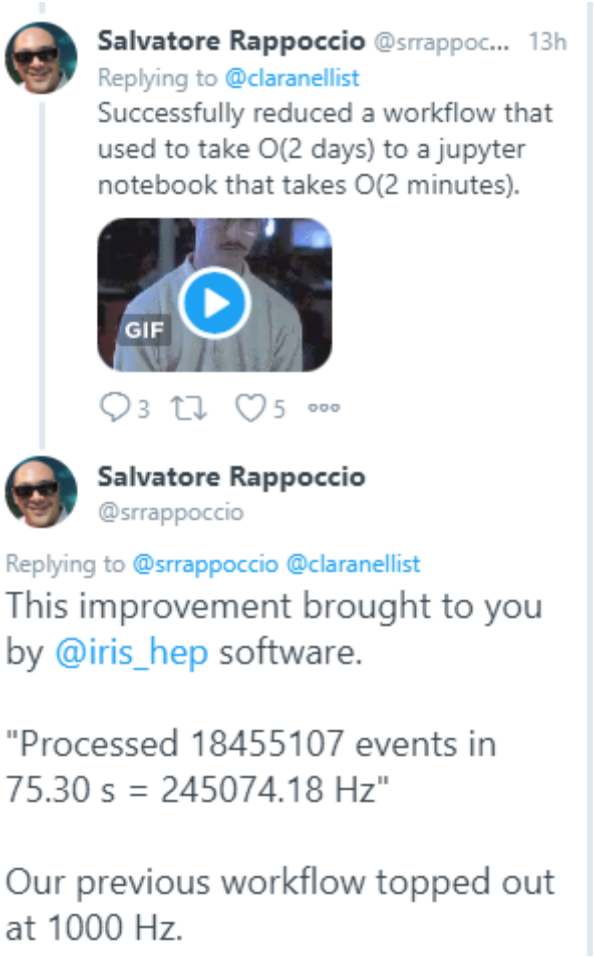
GPU acceleration comes for "free"

C++: 10+ hours
pyhf: 30 minutes



Just released and being incorporated into Analyses Now

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



uproot
awkward array
coffee

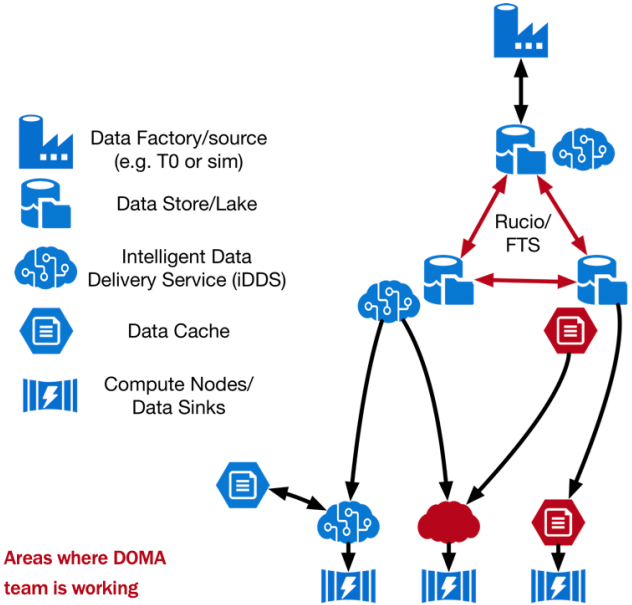
DIANAHEP
And IRIS-HEP

Prototype Phase – Used in analysis by early adopters

DOMA (Data Organization, Management, Access)

Fundamental R&D related to the central challenges of organizing, managing, and providing access to exabytes of data from process systems of various kinds.

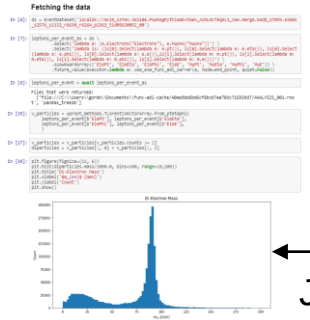
- Data Organization: Improve how HEP data is serialized and stored.
- Data Access: Develop capabilities to deliver filtered and transformed eve streams to users and analysis systems.
- Data Management: Improve and deploy distributed storage infrastructure spanning multiple physical sites. Improve inter-site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

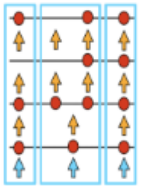


Jupyter Notebook

Innovative Algorithms – Trigger & Reconstruction

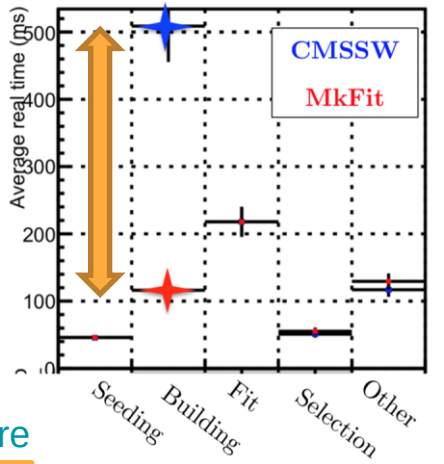
Algorithms for real-time processing of detector data in the software trigger and offline reconstruction are critical components of HEP's computing challenge.

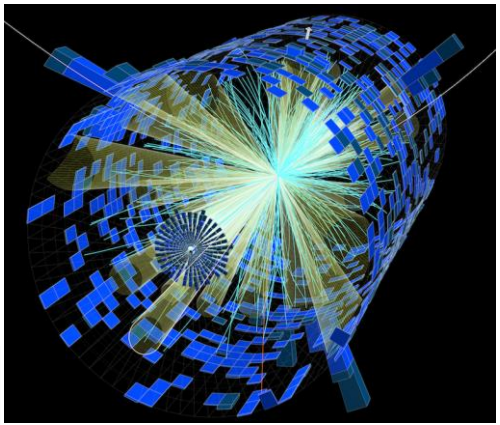
- How to redesign tracking algorithms for HL-LHC?
- How to make use of major advances in machine learning (ML)?



mkFit – Parallel Track Fitting

- 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





Pileup in the HL-LHC will increase combinatorics dramatically

Now being integrated into CMS production software

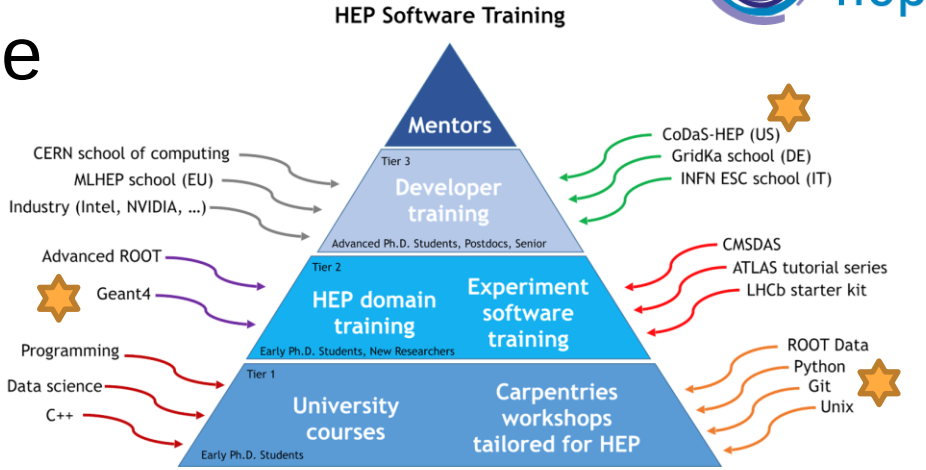
Will supply tracking enhancements for ~3500 physicists

~300 have attended various small trainings we've run or sponsored

Software Sustainability Core

Training

Sample Topics: Git, OpenMP, SciPy, ML, Random Numbers, Columnar Data Analysis, Vectorization, etc.

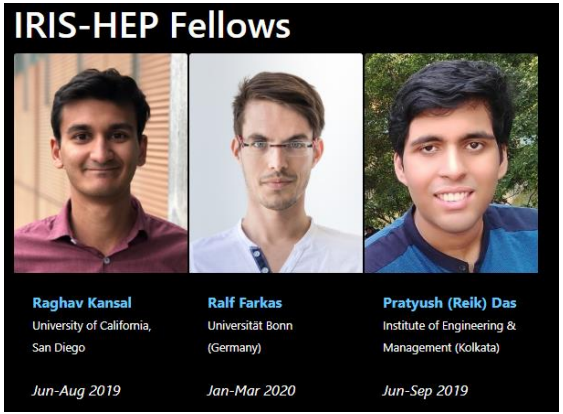


Direct Value to IRIS-HEP

We've had previous students become teachers, and previous students are now team-members in IRIS-HEP.
Not just value to the community!

Fellows Program

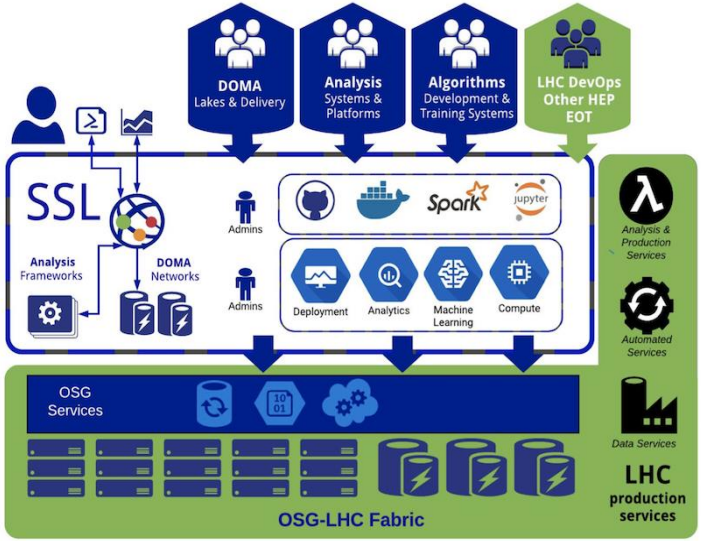
Provides opportunities for undergraduate and graduate students to connect with mentors within the larger HEP and Computational/Data Science community.



Scalable Systems Laboratory

Goal: Provide the Institute and the HL-LHC experiments with scalable platforms needed for development in context. Facilities R&D

River – a repurposed UChicago CS research cluster now being used to test/run IRIS-HEP projects.



CoDaS-HEP school environment, ServiceX test bed.

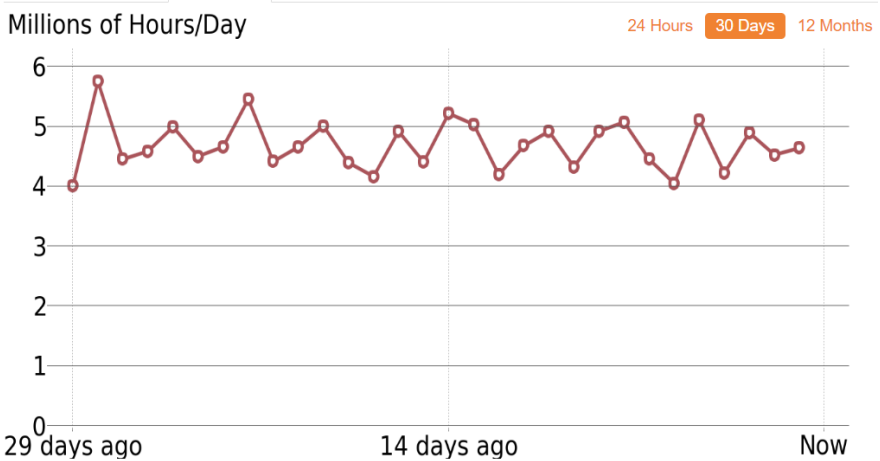
Kubernetes based cluster, can run the OSG-LHC environment, school environments, etc. Experimenting with “no-ops” management.

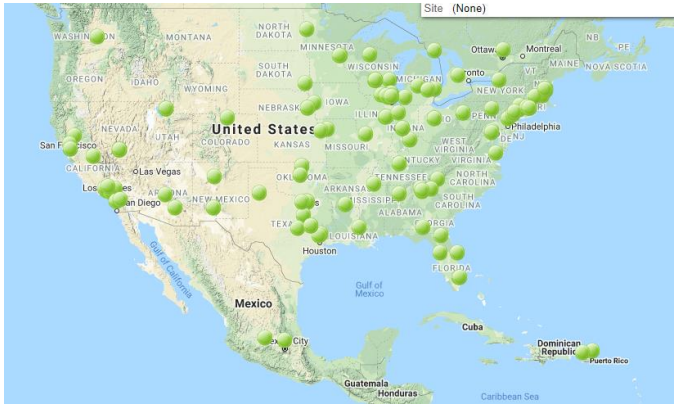
Collaborating with a CyberTraining project (OAC-1829707, 1829729) as well as a growing number of international collaborators.

Open Science Grid - LHC

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 supports LHC operations and development of the consortium.

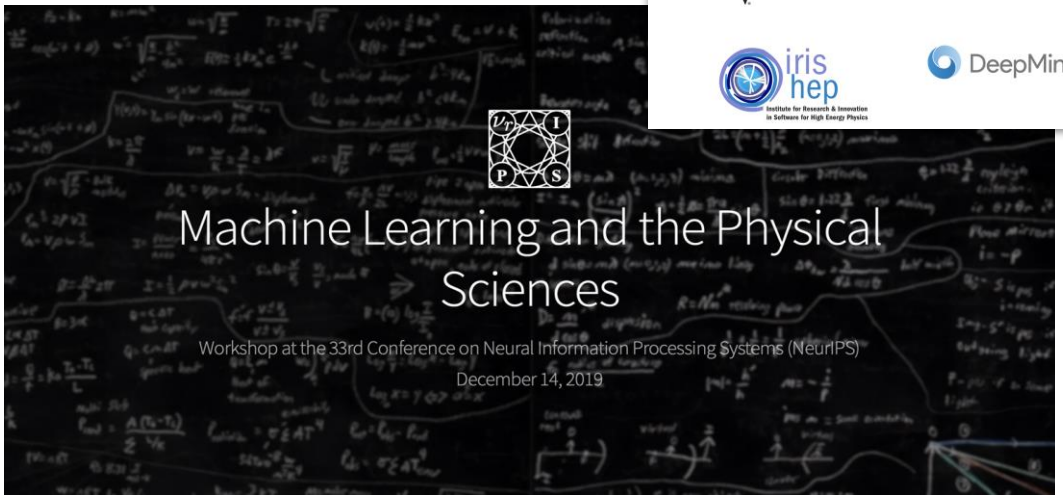




- Work to separate local site hardware and software support by moving services into containers.
- Transitioning security service to use tokens

Particle physicists all over the world depend on these services and scheduling of processing hours (~10,000)

Some (biased) Impact Highlights

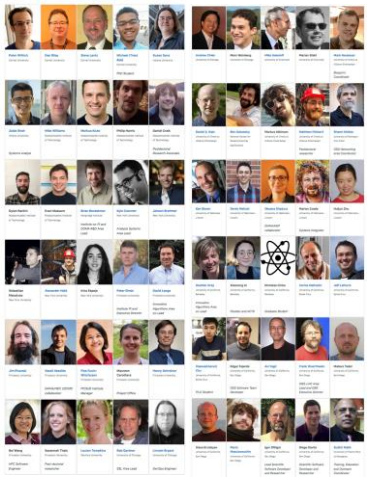


@NeurIPS

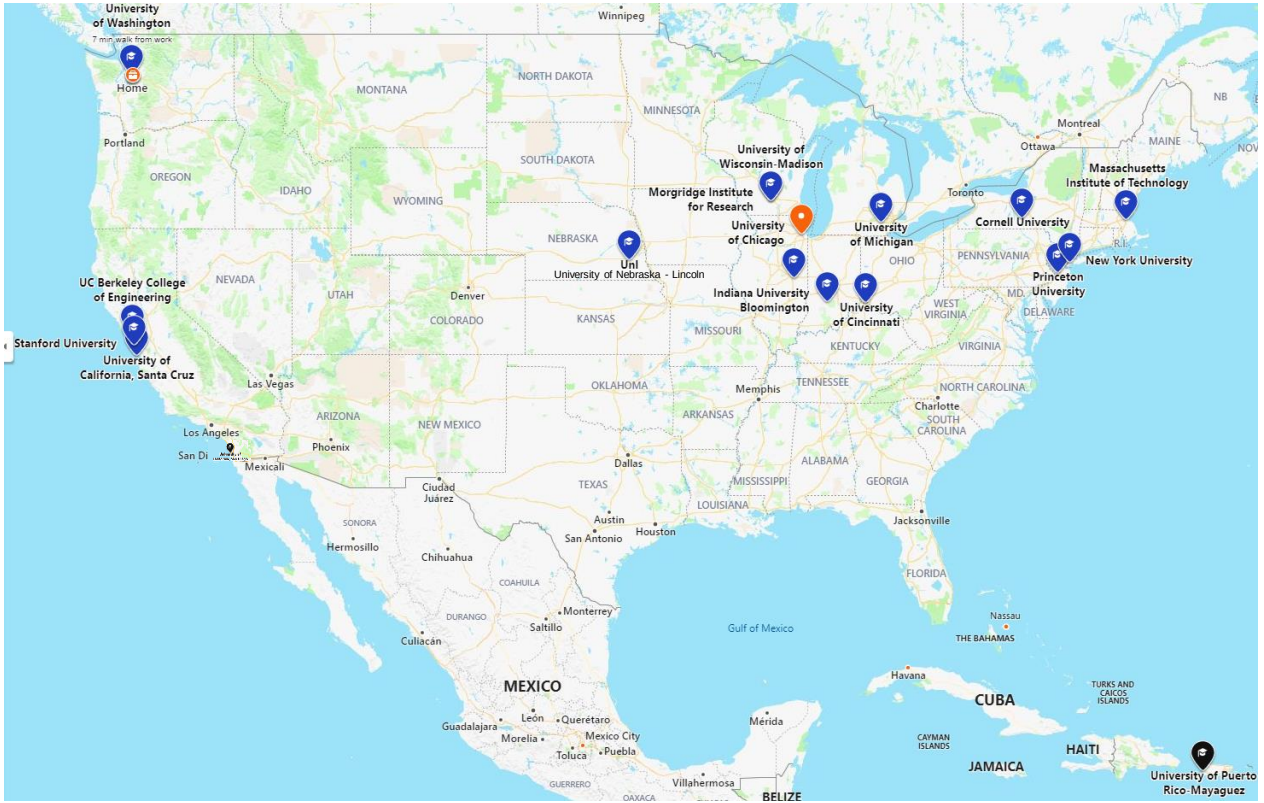
Co-Sponsored: interest in ML in physics and the sciences is very high in the global community.

Virtual Institute

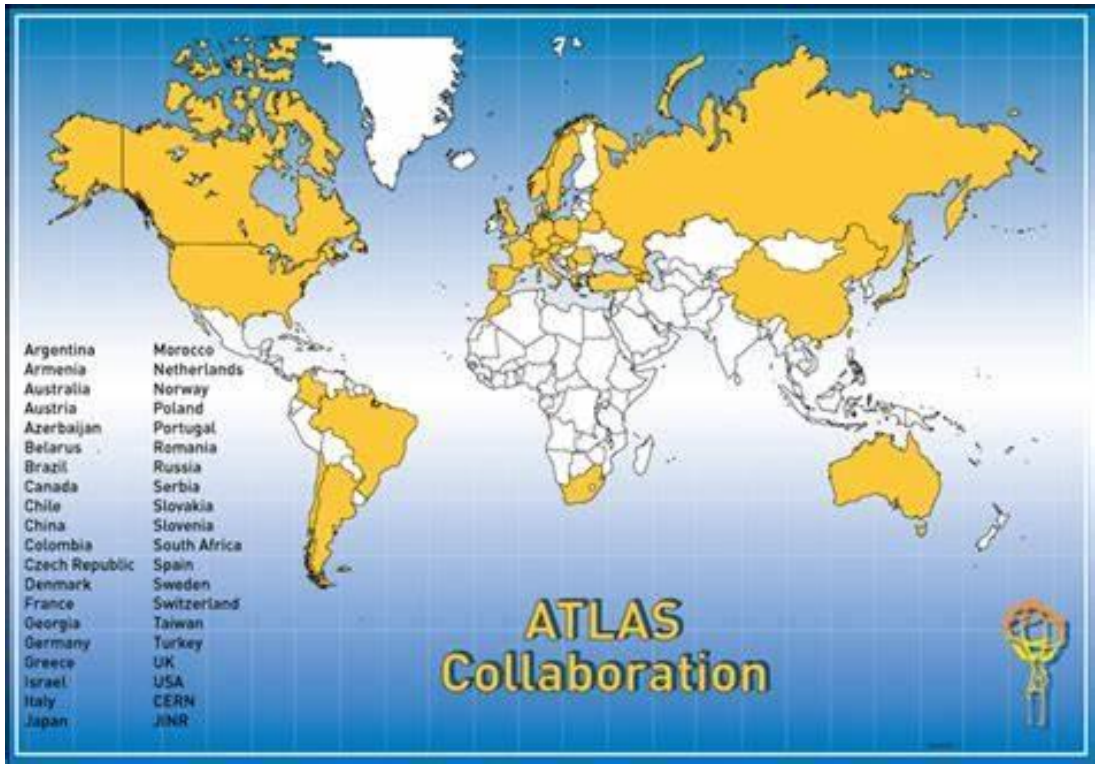
~30 FTE's distributed around the USA.



(many more but wouldn't fit here!)



For a Global Field



Global community is ~O(30K)

Community Building

IRIS-HEP came out of the S2I2-HEP: Conceptualization Process

This was a community building exercise:

- 17 workshops from 2016-2017
- More than 20 papers of ideas submitted to the physics archive
- Roadmap published in “Computing and Software for Big Science”

Computing and Software for Big Science (2019) 3:7
https://doi.org/10.1007/s41781-018-0018-8

ORIGINAL ARTICLE

A Roadmap for HEP Software and Computing R&D for the 2020s

The HEP Software Foundation[1] · Johannes Albrecht[2,3] · Antonio Augusto Alves Jr[4] · Guilherme Amadio[5] · Giuseppe Andronico[6,7] · Nguyen Anh-Ky[8,9] · Laurent Aphecetche[10] · John Apostolakis[11] · Makoto Asai[12] · Luca Atzori[11] · Marian Babik[11] · Giuseppe Bagliesi[13,14] · Marilena Bandieramonte[11] · Sunanda Banerjee[15] · Martin Barisits[11] · Lothar A. T. Bauerdick[16] · Stefano Belforte[17] · Douglas Benjamin[18] · Catrin Bernius[19] · Wahid Bhimji[20] · Riccardo Maria Bianchi[21] · Ian Bird[11] · Catherine Biscarat[22] · Jakob Blomer[11] · Kenneth Bloom[23] · Tommaso Boccali[24] · Brian Bockelman[23] · Tomasz Bold[25] · Daniele Bonacorsi[26] · Antonio Boveia[27] · Concezio Bozzi[28] · Marko Bracko[29,30] · David Britton[31] · Andy Buckley[32] · Predrag Buncic[11] · Paolo Calafiura[20] · Simone Campana[11] · Philippe Canal[16] · Luca Canali[11] · Gianpaolo Carlino[33] · Nuno Castro[34,35] · Marco Cattaneo[11] · Gianluca Cerminara[11] · Javier Cervantes Villanueva[11] · Philip Chang[36] · John Chapman[37] · Gang Chen[38] · Taylor Childers[1] · Peter Clarke[39] · Marco Clemencic[11] · Eric Cogneras[40] · Jeremy Coles[41] · Ian Collier[42] · David Colling[39] · Gloria Corti[11] · Gabriele Cosmo[11] · Davide Costanzo[43] · Ben Couturier[11] · Kyle Cranmer[44] · Jack Cranshaw[1] · Leonardo Cristella[45] · David Crooks[46] · Sabine Crépé-Renaudin[10] · Robert Currie[47] · Sünje Dallmeier-Tiessen[11] · Kaushik De[23,*] · Michel De Cian[11] · Albert De Roeck[11] · Antonio Delgado Peris[11] · Frédéric Derue[10] · Alessandro Di Girolamo[11] · Salvatore Di Guida[48] · Gancho Dimitrov[49] · Caterina Doglioni[31] · Andrea Dotti[50] · Dirk Duellmann[11] · Laurent Duflot[10] · Dave Dykstra[16] · Katarzyna Dziedziniewicz-Wojcik[11] · Agnieszka Dziurda[11] · Ulrik Egede[51] · Peter Elmer[19] · Johannes Elmsheuser[52] · V. Daniel Elvira[16] · Giulio Eulisse[11] · Steven Farrell[53] · Torben Ferber[54] · Andrej Filipčič[55] · Ian Fisk[56] · Conor Fitzpatrick[57] · José Flix[58] · Andrea Formica[59] · Alessandra Forti[60] · Giovanni Franzoni[61] · James Frost[62] · Stu Fuess[16] · Frank Gaede[63] · Gerardo Ganis[11] · Robert Gardner[64] · Vincent Garonne[65] · Andreas Gellrich[63] · Krzysztof Genser[16] · Simon George[66] · Frank Geurts[67] · Andrei Gheata[11] · Mihaela Gheata[11] · Francesco Giacomini[68] · Stefano Giagu[69,70] · Manuel Giffels[71] · Douglas Gingrich[72] · Maria Girone[11] · Vladimir V. Gligorov[73] · Ivan Glushkov[74] · Wesley Gohn[75] · Jose Benito Gonzalez Lopez[11] · Isidro González Caballero[76] · Juan R. González Fernández[65] · Giacomo Govi[16] · Claudio Grandi[68] · Hadrien Grasland[40] · Heather Gray[36] · Lucia Grillo[31] · Wen Guan[77] · Oliver Gutsche[16] · Vardan Gyurjyan[16] · Andrew Hanushevsky[78] · Farah Hariri[79] · Thomas Hartmann[80] · John Harvey[11] · Thomas Hauth[63] · Benedikt Hegner[11] · Beate Heinemann[52] · Lukas Heinrich[52] · Andreas Heiss[60] · José M. Hernández[2] · Michael Hildreth[81] · Mark Hodgkinson[74] · Stefan Hoeche[82] · Burt Holzman[16] · Peter Hristov[11] · Xingtao Huang[111] · Vladimir N. Ivanchenko[83,11] · Todor Ivanov[112] · Jan Ivarsen[84] · Brij Jashal[85] · Bodhitha Jayatilaka[16] · Roger Jones[86] · Michel Jouvin[16] · Soon Yung Jun[16] · Michael Kagan[78] · Charles William Kalderon[31] · Meghan Kane[87] · Edward Karavakis[11] · Daniel S. Katz[88] · Dorian Kcira[89] · Oliver Keeble[11] · Borut Paul Kersevan[90] · Michael Kirby[16] · Alexei Klimentov[91] · Markus Klute[92] · Ilya Komarov[93] · Dmitri Konstantinov[94] · Patrick Koppenburg[95] · Jim Kowalkowski[16] · Luke Kreczko[96] · Thomas Kuhr[97] · Robert Kutschke[16] · Valentin Kuznetsov[98] · Walter Lampl[51] · Eric Lancon[91] · David Lange[100] · Mario Lassnig[11] · Paul Laycock[11] · Charles Leggett[66] · James Letts[101] · Birgit Lewendel[11] · Teng Li[102] · Guilherme Lima[16] · Jacob Linacre[74] · Tomas Linden[40] · Miron Livny[103] · Giuseppe Lo Presti[11] · Sebastian Lopienski[11] · Peter Love[104] · Adam Lyon[16] · Nicolò Magini[105] · Zachary L. Marshall[52] · Edoardo Martelli[11] · Stewart Martin-Haugh[106] · Pere Mato[11] · Kajari Mazumdar[48] · Thomas McCauley[36] · Josh McFayden[11] · Shawn McKee[64] · Andrew McNab[74] · Rashid Mehdiyev[74] · Helge Meinhard[11] · Dario Menasce[26] · Patricia Mendez Lorenzo[11] · Alaettin Serhan Mete[74] · Michele Michelotto[70] · Jovan Mitrevski[44] · Lorenzo Moneta[11] · Ben Morgan[74] · Richard Mount[91] · Edward Moyse[36] · Sean Murray[16,52] · Armin Nairz[11] · Mark S. Neubauer[64] · Andrew Norman[107] · Sérgio Novaes[108] · Mihaly Novak[11] · Arantza Oyanguren[109] · Nurcan Ozturk[110] · Andres Pacheco Pages[26,30] · Michela Paganini[120] · Jerome Pansanel[27] · Vincent R. Pascuzzi[52] · Glenn Patrick[91] · Alex Pearce[11] · Ben Pearson[74] · Kevin Pedro[16] · Gabriel Perdue[96] · Antonio Perez-Calero Yzquierdo[109] · Luca Perrozzi[71] · Troels Petersen[95] · Marko Petric[11] · Andreas Petzold[42] · Jónatan Piedra[113] · Leo Piilonen[114] · Danilo Piparo[11] · Jim Pivarski[100] · Witold Pokorski[11] · Francesco Polci[115] · Karolos Potamianos[31] · Fernanda Psihas[56] · Albert Puig Navarro[116] · Günter Quast[63] · Gerhard Raven[95,117] · Jürgen Reuter[63] · Alberto Ribon[11] · Lorenzo Rinaldi[69] · Martin Ritter[89] ·

Extended author information available on the last page of the article

Springer

“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 Roadmap

Part of IRIS-HEP’s mandate is to continue this process

- Blueprint meetings to build field-wide consensus on specific problems.
- The Fellows Program
- Topical Meetings: seminars on topics of interest.
- Sponsorship of conferences and workshops like PyHEP 2020, and LAWSCHEP 2019.

~900 have attended various small workshops we’ve run or sponsored

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
- 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.