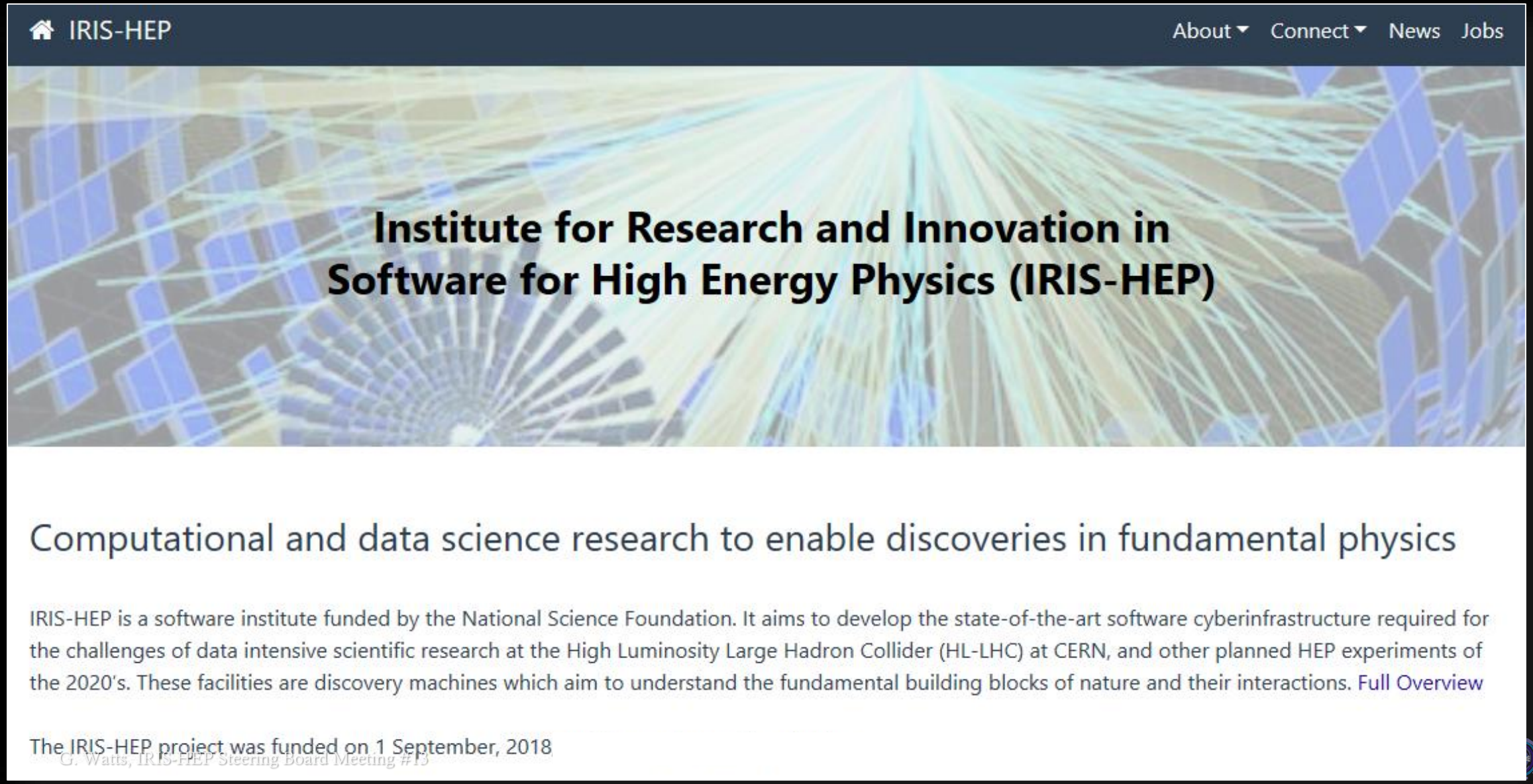

IRIS-HEP
About
Connect
News
Jobs
Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)
Computational and data science research to enable discoveries in fundamental physics
IRIS-HEP is a software institute funded by the National Science Foundation. It aims to develop the state-of-the-art software cyberinfrastructure required for the challenges of data intensive scientific research at the High Luminosity Large Hadron Collider (HL-LHC) at CERN, and other planned HEP experiments of the 2020's. These facilities are discovery machines which aim to understand the fundamental building blocks of nature and their interactions. Full Overview
The IRIS-HEP project was funded on 1 September, 2018

Institute for Research & Innovation
in Software for High Energy Physics

IRIS-HEP Steering Board Meeting #13

G. Watts

For the IRIS-HEP Executive Board

2022-02-14

"The IRIS-HEP Steering Board represents the Institute's stakeholders to provide, to the Executive Board, the stakeholder's input on the priorities, execution, and strategy of the Institute."

Thank You

Danilo Piparo (CERN)
CMS

Paolo Calafiura (LBNL)
US ATLAS Ops Program

Simone Campana (CERN)
WLCG

Alessandro Di Girolamo (CERN)
ATLAS

Oliver Gutsche (FNAL)
US CMS Ops Program

Patrick Koppenburg (NIKHEF)
LHCb

Graeme Stewart (CERN)
HSF

Ken Herner (FNAL)
The OSG Council

Welcome

steering-board@iris-hep.org

(you)

exec-board@iris-hep.org

(us)

Next Meeting Dates
(proposed)

June 7, 2022
Sept 13, 2022 (CERN VM WS Conflict)
Dec 13, 2022
February 14, 2023

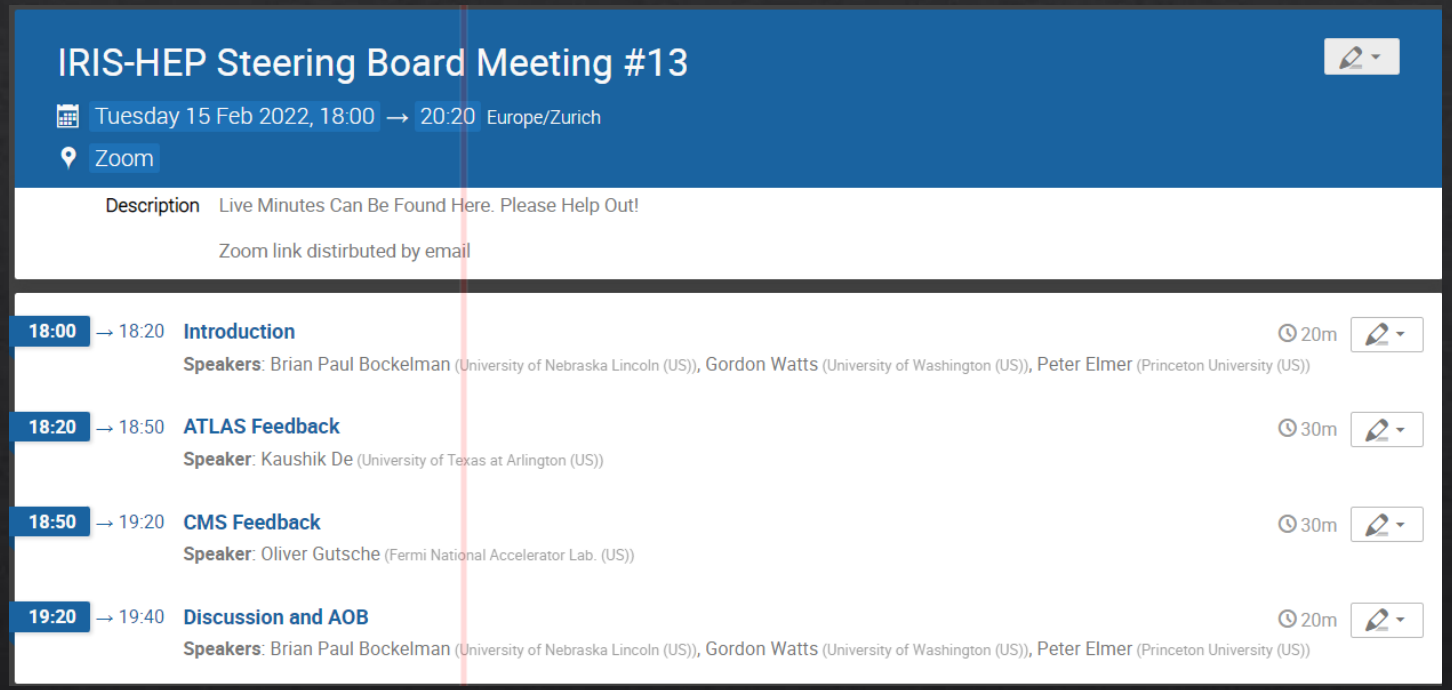
Will circulate these by email after the meeting for comments!

Today

Feedback from the experiments

Could we do LHCb and HSF next meeting?

Please use the google doc circulated in the email to add comments or make any notes you want us to track!



IRIS-HEP Steering Board Meeting #13

Tuesday 15 Feb 2022, 18:00 → 20:20 Europe/Zurich

Zoom

Description Live Minutes Can Be Found Here. Please Help Out!

Zoom link distirbuted by email

18:00 → 18:20 **Introduction** 20m

Speakers: Brian Paul Bockelman (University of Nebraska Lincoln (US)), Gordon Watts (University of Washington (US)), Peter Elmer (Princeton University (US))

18:20 → 18:50 **ATLAS Feedback** 30m

Speaker: Kaushik De (University of Texas at Arlington (US))

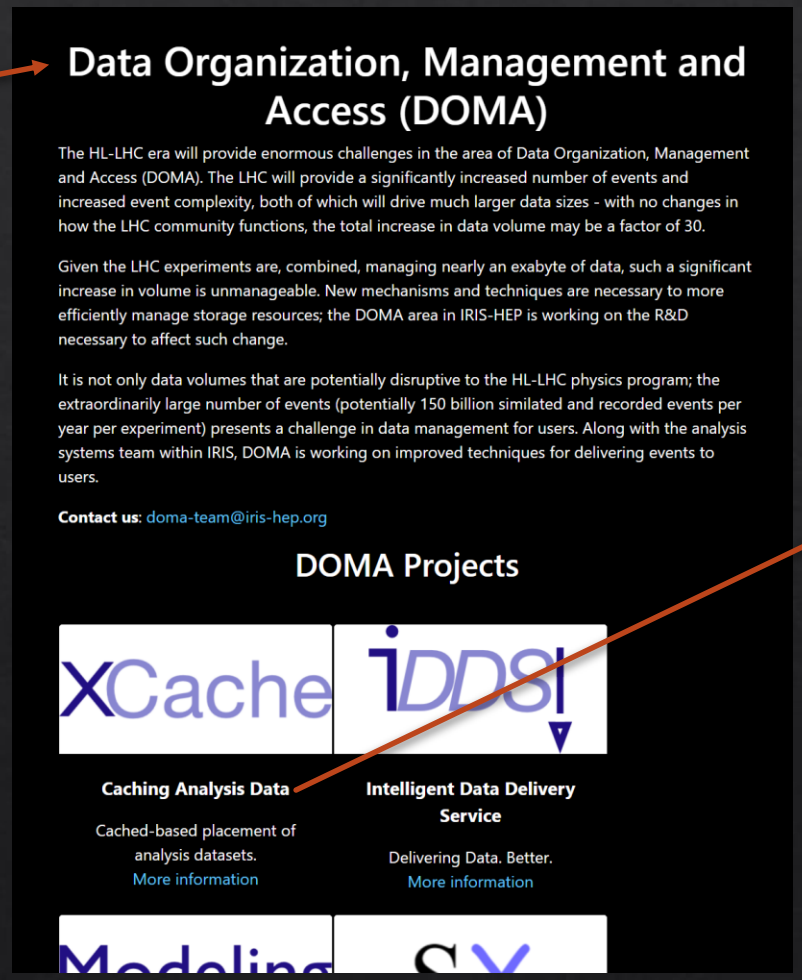
18:50 → 19:20 **CMS Feedback** 30m

Speaker: Oliver Gutsche (Fermi National Accelerator Lab. (US))

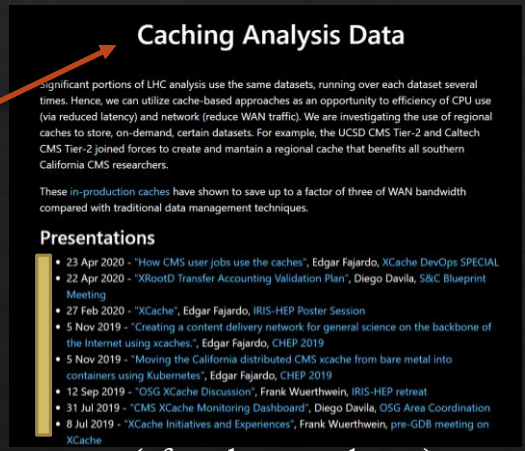
19:20 → 19:40 **Discussion and AOB** 20m

Speakers: Brian Paul Bockelman (University of Nebraska Lincoln (US)), Gordon Watts (University of Washington (US)), Peter Elmer (Princeton University (US))

Project Information



Per-project information is available on all IRIS-HEP projects.



(often, but not always)

Slide from SB#12

Analysis Ecosystem

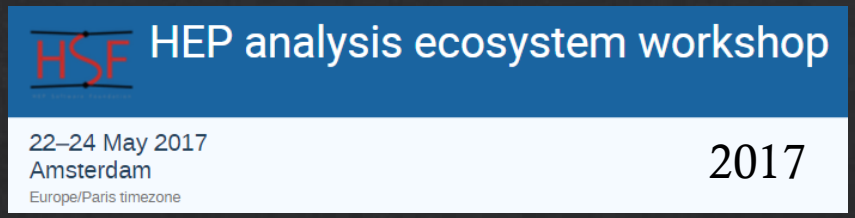
Are we ready for another Ecosystem Workshop?

Some of the topics from last time:

- Where we are today: tools and data formats
- Covered Energy Frontier and Intensity Frontier
- Out look for the next 5-7 years for hardware and software
- Survey of modern tools
- Missing pieces

A report was generated along with a summary slide deck

Just starting to gauge community interest



- It has been almost 5 years
- Python/ROOT vision for HL-LHC much more in focus
- Grand Challenges
- New challenges (e.g. analysis facilities, scaling, differentiability)
- New languages (e.g. Julia, Go)
- Snowmass

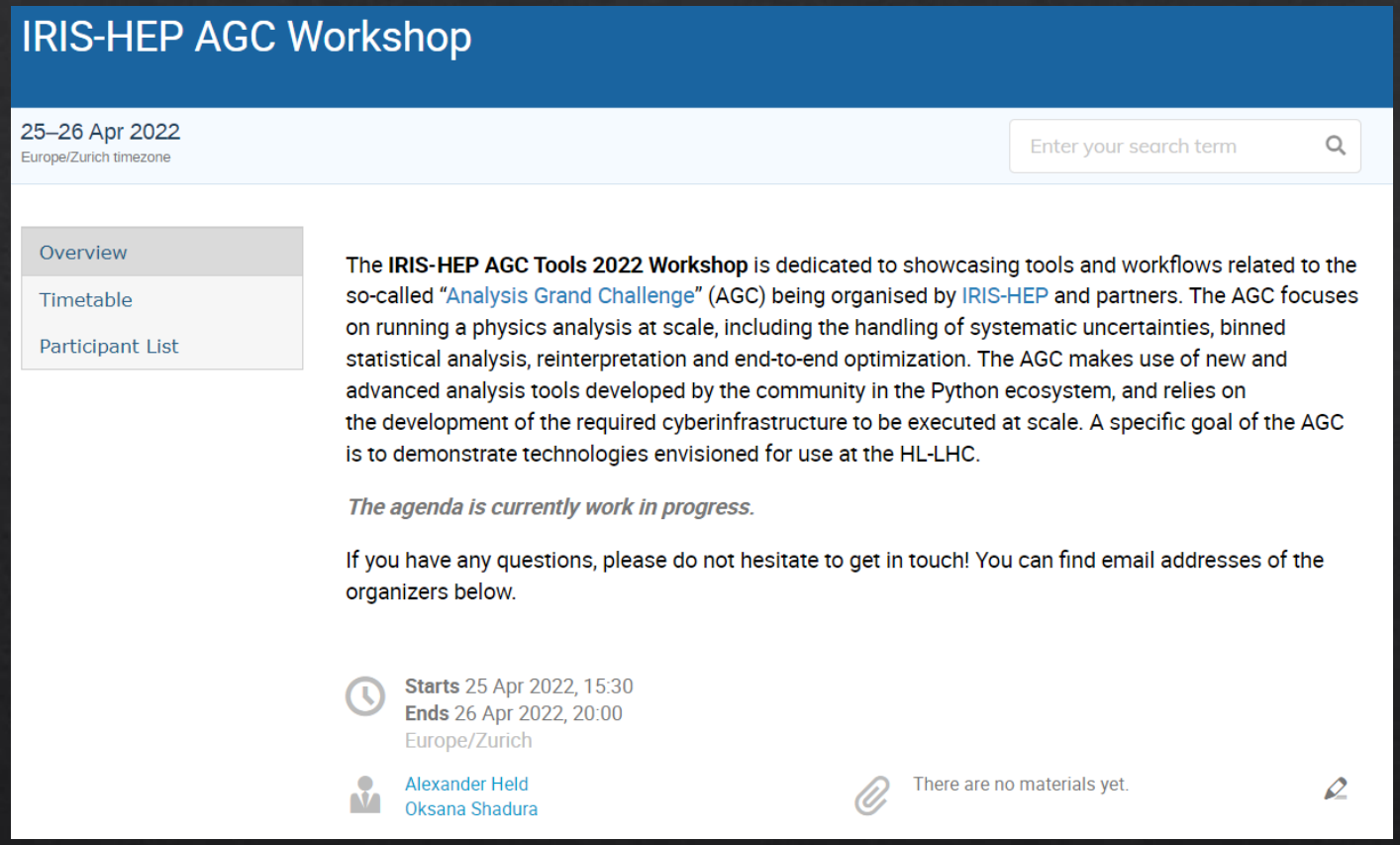
Week of canceled CHEP? (2nd Week of May)
Really wanted it hybrid!!!

https://indico.cern.ch/event/1126109/

Analysis Grand Challenge Workshop #2

April 25-26
Virtual Only

Similar to previous workshop
Updated demos
Continued focus on getting more people introduced to new tools



https://indico.cern.ch/e/aew2

Analysis Ecosystem Workshop

May 23-25, IJCLab, Paris
Hybrid

Topics for the workshop will include, amongst others:

- Analysis Facilities
- ML tools and differentiable computing workflows
- "Real-time" trigger-level analysis
- Analysis User Experience and Declarative Languages
- Analysis on reduced formats or specialist inputs
- Bookkeeping and systematics handling

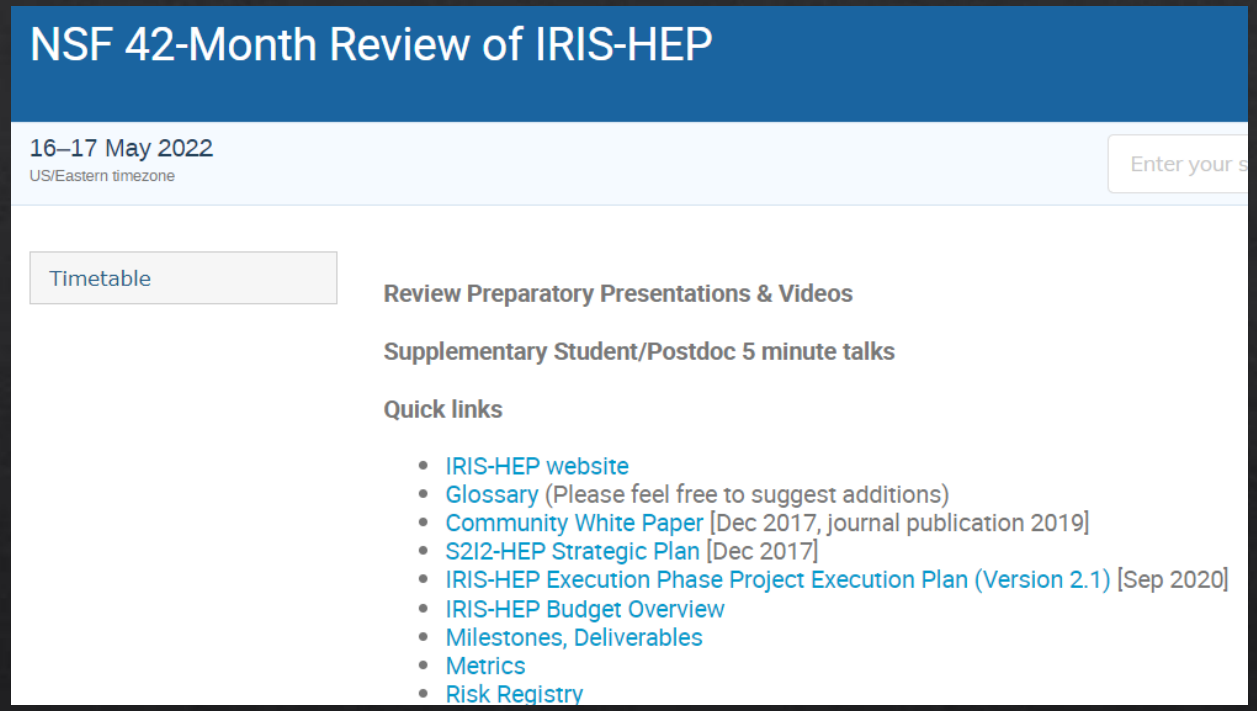
As a workshop, there will be limited presentations, lots of time for discussion and a written outcome that summarises the workshop's conclusions and points the way forward.

https://indico.cern.ch/event/1114340/

NSF Review

Our next "big" review is set
- Will cover all aspects of the program
- Likely to be online (sadly)

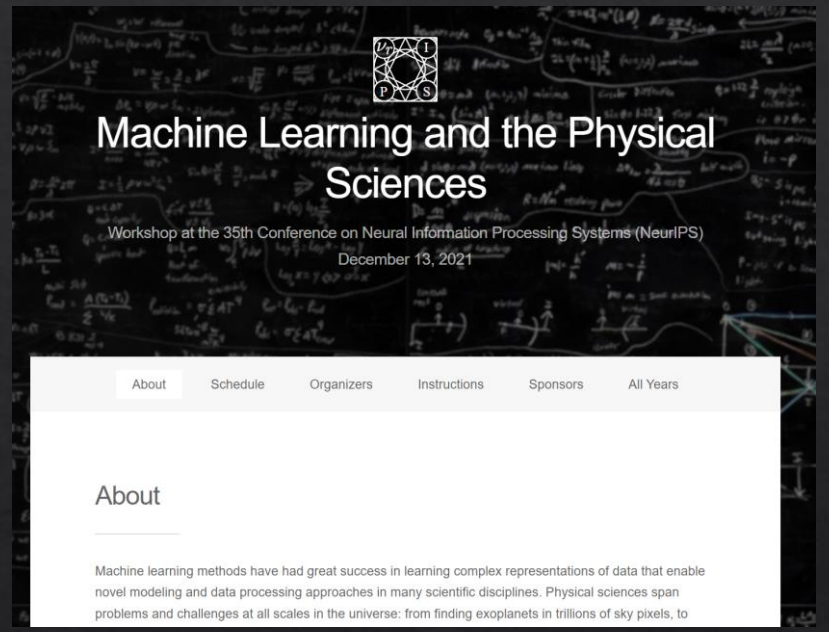
We are starting a full project "scrubbing"
- Website Content is being refreshed
- Lifecycle indications for all projects will be added
- Updates to presentation lists, paper lists, and fellow's pages (and etc.)

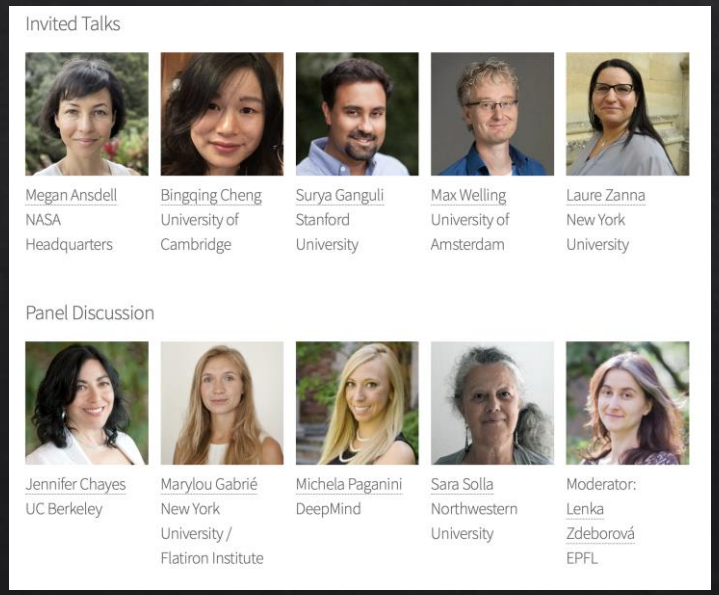


Happened on December 13th, 2021 → Link

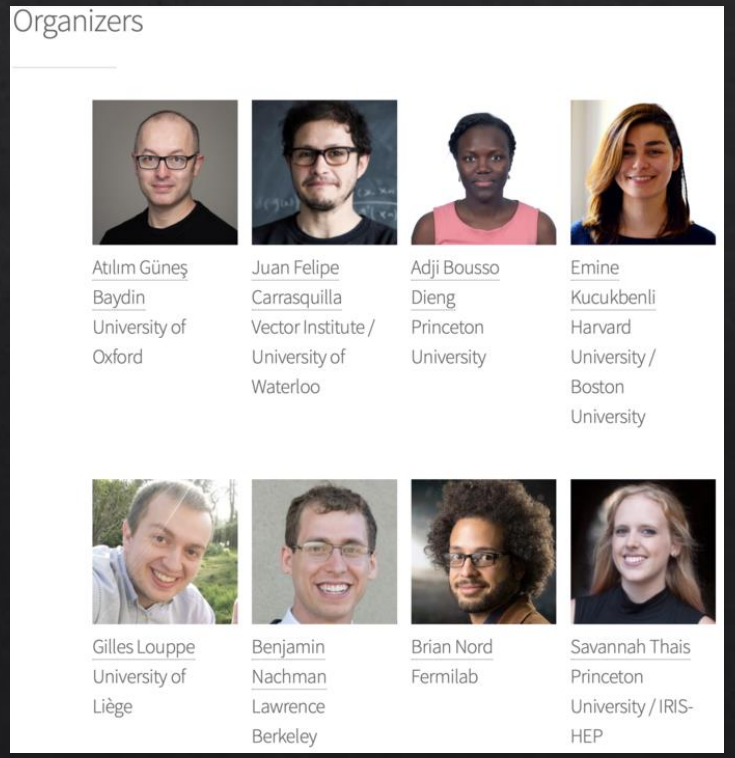
NeurIPS: ML and the Physical Sciences

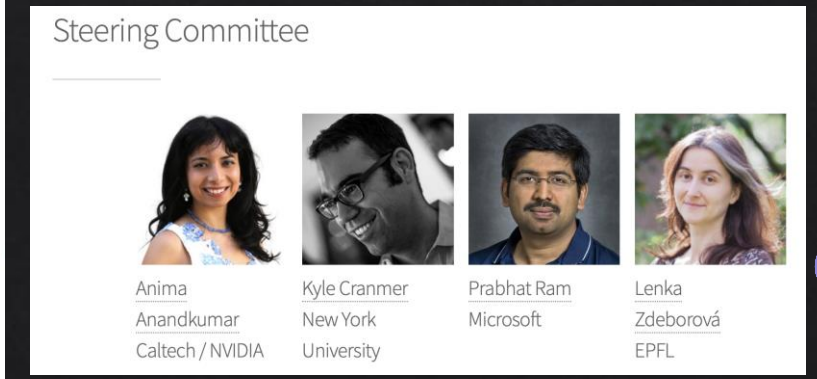
- 152 papers accepted for presentation and discussion
- Over 200 submitted
- Sponsorship from Vector Institute, APS's Data Science group, IRIS-HEP
- Last year they had in excess of 400 attendees





G. Watts, IRIS-HEP Steering Board Meeting #13

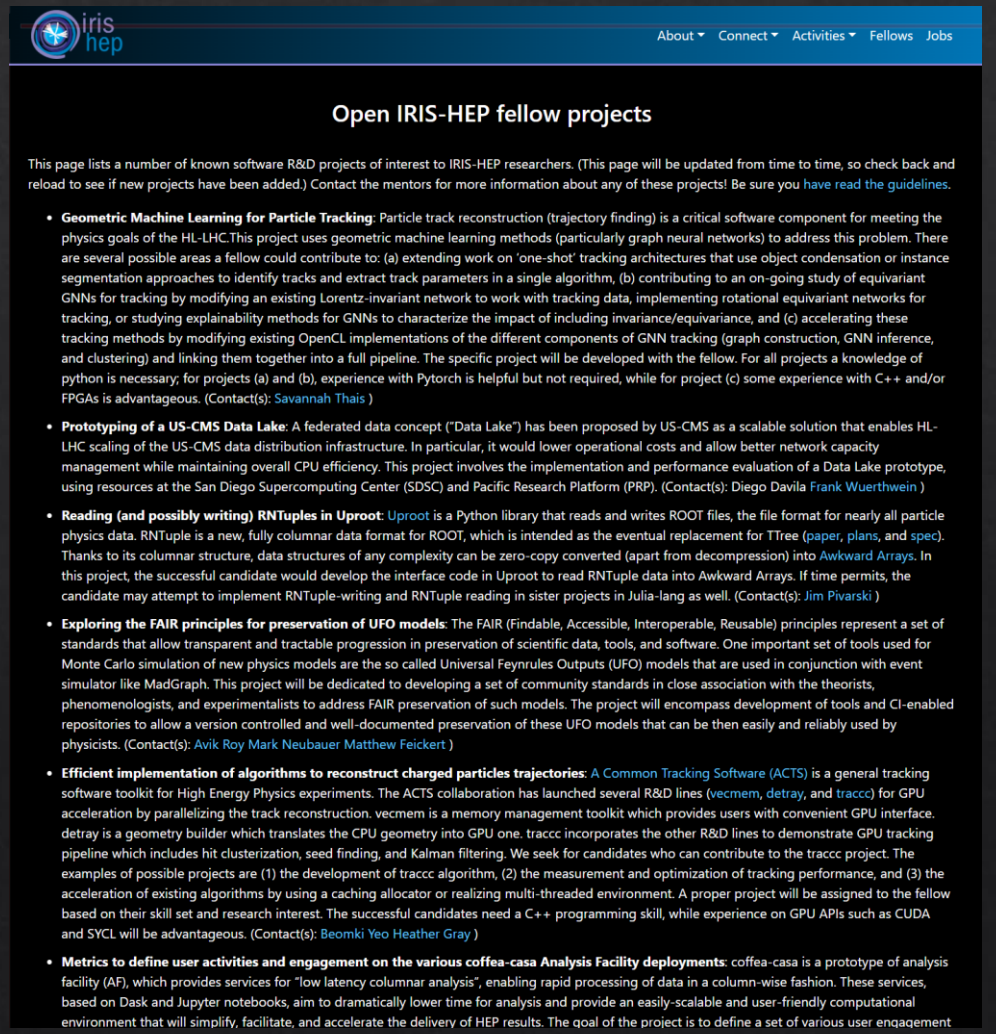




Questions? Comments?

We are about to put out our call for IRIS-HEP fellows

- Lots of projects posted
- From all focus areas
- Know a good student? Want to distribute it to students associated with your work or experiment?
- Official announcement will be sent to the iris-hep announcement list. Please do recirculate to anyone or group that might be interested!



iris hep — About ▾ Connect ▾ Activities ▾ Fellows Jobs

Open IRIS-HEP fellow projects

This page lists a number of known software R&D projects of interest to IRIS-HEP researchers. (This page will be updated from time to time, so check back and reload to see if new projects have been added.) Contact the mentors for more information about any of these projects! Be sure you have read the guidelines.

- **Geometric Machine Learning for Particle Tracking**: Particle track reconstruction (trajectory finding) is a critical software component for meeting the physics goals of the HL-LHC.This project uses geometric machine learning methods (particularly graph neural networks) to address this problem. There are several possible areas a fellow could contribute to: (a) extending work on 'one-shot' tracking architectures that use object condensation or instance segmentation approaches to identify tracks and extract track parameters in a single algorithm, (b) contributing to an on-going study of equivariant GNNs for tracking by modifying an existing Lorentz-invariant network to work with tracking data, implementing rotational equivariant networks for tracking, or studying explainability methods for GNNs to characterize the impact of including invariance/equivariance, and (c) accelerating these tracking methods by modifying existing OpenCL implementations of the different components of GNN tracking (graph construction, GNN inference, and clustering) and linking them together into a full pipeline. The specific project will be developed with the fellow. For all projects a knowledge of python is necessary; for projects (a) and (b), experience with Pytorch is helpful but not required, while for project (c) some experience with C++ and/or FPGAs is advantageous. (Contact(s): Savannah Thais)
- **Prototyping of a US-CMS Data Lake**: A federated data concept ("Data Lake") has been proposed by US-CMS as a scalable solution that enables HL-LHC scaling of the US-CMS data distribution infrastructure. In particular, it would lower operational costs and allow better network capacity management while maintaining overall CPU efficiency. This project involves the implementation and performance evaluation of a Data Lake prototype, using resources at the San Diego Supercomputing Center (SDSC) and Pacific Research Platform (PRP). (Contact(s): Diego Davila Frank Wuerthwein)
- **Reading (and possibly writing) RNTuples in Uproot**: Uproot is a Python library that reads and writes ROOT files, the file format for nearly all particle physics data. RNTuple is a new, fully columnar data format for ROOT, which is intended as the eventual replacement for TTree (paper, plans, and spec). Thanks to its columnar structure, data structures of any complexity can be zero-copy converted (apart from decompression) into Awkward Arrays. In this project, the successful candidate would develop the interface code in Uproot to read RNTuple data into Awkward Arrays. If time permits, the candidate may attempt to implement RNTuple-writing and RNTuple reading in sister projects in Julia-lang as well. (Contact(s): Jim Pivarski)
- **Exploring the FAIR principles for preservation of UFO models**: The FAIR (Findable, Accessible, Interoperable, Reusable) principles represent a set of standards that allow transparent and tractable progression in preservation of scientific data, tools, and software. One important set of tools used for Monte Carlo simulation of new physics models are the so called Universal Feynrules Outputs (UFO) models that are used in conjunction with event simulator like MadGraph. This project will be dedicated to developing a set of community standards in close association with the theorists, phenomenologists, and experimentalists to address FAIR preservation of such models. The project will encompass development of tools and CI-enabled repositories to allow a version controlled and well-documented preservation of these UFO models that can be then easily and reliably used by physicists. (Contact(s): Avik Roy Mark Neubauer Matthew Feickert)
- **Efficient implementation of algorithms to reconstruct charged particles trajectories**: A Common Tracking Software (ACTS) is a general tracking software toolkit for High Energy Physics experiments. The ACTS collaboration has launched several R&D lines (vecmem, detray, and traccc) for GPU acceleration by parallelizing the track reconstruction. vecmem is a memory management toolkit which provides users with convenient GPU interface. detray is a geometry builder which translates the CPU geometry into GPU one. traccc incorporates the other R&D lines to demonstrate GPU tracking pipeline which includes hit clusterization, seed finding, and Kalman filtering. We seek for candidates who can contribute to the traccc project. The examples of possible projects are (1) the development of traccc algorithm, (2) the measurement and optimization of tracking performance, and (3) the acceleration of existing algorithms by using a caching allocator or realizing multi-threaded environment. A proper project will be assigned to the fellow based on their skill set and research interest. The successful candidates need a C++ programming skill, while experience on GPU APIs such as CUDA and SYCL will be advantageous. (Contact(s): Beomki Yeo Heather Gray)
- **Metrics to define user activities and engagement on the various coffea-casa Analysis Facility deployments**: coffea-casa is a prototype of analysis facility (AF), which provides services for "low latency columnar analysis", enabling rapid processing of data in a column-wise fashion. These services, based on Dask and Jupyter notebooks, aim to dramatically lower time for analysis and provide an easily-scalable and user-friendly computational environment that will simplify, facilitate, and accelerate the delivery of HEP results. The goal of the project is to define a set of various user engagement

People

Kyle Cranmer is moving to the American Family Insurance Data Science Institute at the University of Wisconsin-Madison

Analysis Systems leadership:

- We have a plan in place and are making sure it is ready to go.

Frank Wurthwein has been named the director of the San Diego Supercomputer Center

IRIS-HEP OSG-LHC leadership roll is unchanged.

Connections To Other Similar Organizations

- HSF
- PyHEP
- HEP-CCE (USA)
- FIRST-HEP (USA)
- UK SWIFT (UK)
- PUNCH4NFDI (Ger)
- FAIR4HEP (USA)

Others that we should be connecting to?

https://arxiv.org/abs/2202.02194

Data Science Tools for Analysis

Paper used as input to the HL-LHC Computing Review

Physics > Data Analysis, Statistics and Probability

[Submitted on 4 Feb 2022]

HL-LHC Computing Review Stage 2, Common Software Projects: Data Science Tools for Analysis

Jim Pivarski, Eduardo Rodrigues, Kevin Pedro, Oksana Shadura, Benjamin Krikler, Graeme A. Stewart

This paper was prepared by the HEP Software Foundation (HSF) PyHEP Working Group as input to the second phase of the LHCC review of High-Luminosity LHC (HL-LHC) computing, which took place in November, 2021. It describes the adoption of Python and data science tools in HEP, discusses the likelihood of future scenarios, and recommendations for action by the HEP community.

Comments: 25 pages, 7 figures; presented at this https URL (LHCC Review of HL-LHC Computing)
Subjects: **Data Analysis, Statistics and Probability (physics.data-an)**; High Energy Physics - Experiment (hep-ex)
Report number: FERMILAB-CONF-22-061-SCD
Cite as: arXiv:2202.02194 **[physics.data-an]**
(or arXiv:2202.02194v1 **[physics.data-an]** for this version)

Submission history

From: Jim Pivarski [view email]
[v1] Fri, 4 Feb 2022 15:39:45 UTC (1,692 KB)

Questions & Comments?