



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

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





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!

The screenshot shows the agenda for the IRIS-HEP Steering Board Meeting #13, scheduled for Tuesday, February 15, 2022, from 18:00 to 20:20 in the Europe/Zurich time zone. The meeting is hosted on Zoom. The agenda includes an introduction at 18:00, followed by feedback sessions for ATLAS (18:20) and CMS (18:50), and a discussion and AOB session at 19:20. Each session lists the speaker and their affiliation.

Time	Topic	Speaker(s)	Duration
18:00 → 18:20	Introduction	Brian Paul Bockelman (University of Nebraska Lincoln (US)), Gordon Watts (University of Washington (US)), Peter Elmer (Princeton University (US))	20m
18:20 → 18:50	ATLAS Feedback	Kaushik De (University of Texas at Arlington (US))	30m
18:50 → 19:20	CMS Feedback	Oliver Gutsche (Fermi National Accelerator Lab. (US))	30m
19:20 → 19:40	Discussion and AOB	Brian Paul Bockelman (University of Nebraska Lincoln (US)), Gordon Watts (University of Washington (US)), Peter Elmer (Princeton University (US))	20m



Project Information

The screenshot shows the IRIS-HEP website with a navigation menu. The menu items are: Analysis Systems, Blueprint Activity, Data Organization, Management and Access (DOMA), Innovative Algorithms, Open Science Grid (OSG-LHC), Scalable Systems Laboratory, Training, Education and Outreach, Impact Beyond HEP, Presentations, Publications, and Projects. The 'Data Organization, Management and Access (DOMA)' item is highlighted with a blue background. Below the menu, there is a section titled 'Computational and research to enable fundamental physics' and another titled 'News and Featured Stories:' with a photo of three people.

Data Organization, Management and Access (DOMA)

The HL-LHC era will provide enormous challenges in the area of Data Organization, Management and Access (DOMA). The LHC will provide a significantly increased number of events and increased event complexity, both of which will drive much larger data sizes - with no changes in how the LHC community functions, the total increase in data volume may be a factor of 30.

Given the LHC experiments are, combined, managing nearly an exabyte of data, such a significant increase in volume is unmanageable. New mechanisms and techniques are necessary to more efficiently manage storage resources; the DOMA area in IRIS-HEP is working on the R&D necessary to affect such change.

It is not only data volumes that are potentially disruptive to the HL-LHC physics program; the extraordinarily large number of events (potentially 150 billion simulated and recorded events per year per experiment) presents a challenge in data management for users. Along with the analysis systems team within IRIS, DOMA is working on improved techniques for delivering events to users.

Contact us: doma-team@iris-hep.org

DOMA Projects



Caching Analysis Data

Cached-based placement of analysis datasets.
[More information](#)

Intelligent Data Delivery Service

Delivering Data. Better.
[More information](#)

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

Caching Analysis Data

Significant portions of LHC analysis use the same datasets, running over each dataset several times. Hence, we can utilize cache-based approaches as an opportunity to efficiency of CPU use (via reduced latency) and network (reduce WAN traffic). We are investigating the use of regional caches to store, on-demand, certain datasets. For example, the UCSD CMS Tier-2 and Caltech CMS Tier-2 joined forces to create and maintain a regional cache that benefits all southern California CMS researchers.

These in-production caches have shown to save up to a factor of three of WAN bandwidth compared with traditional data management techniques.

Presentations

- 23 Apr 2020 - "How CMS user jobs use the caches", Edgar Fajardo, XCache DevOps SPECIAL
- 22 Apr 2020 - "XRootD Transfer Accounting Validation Plan", Diego Davila, S&C Blueprint Meeting
- 27 Feb 2020 - "XCache", Edgar Fajardo, IRIS-HEP Poster Session
- 5 Nov 2019 - "Creating a content delivery network for general science on the backbone of the Internet using xcache", Edgar Fajardo, CHEP 2019
- 5 Nov 2019 - "Moving the California distributed CMS xcache from bare metal into containers using Kubernetes", Edgar Fajardo, CHEP 2019
- 12 Sep 2019 - "OSG XCache Discussion", Frank Wuerthwein, IRIS-HEP retreat
- 31 Jul 2019 - "CMS XCache Monitoring Dashboard", Diego Davila, OSG Area Coordination
- 8 Jul 2019 - "XCache Initiatives and Experiences", Frank Wuerthwein, pre-GDB meeting on XCache


(often, but not always)



Slide from SB#12

Analysis Ecosystem

Are we ready for another Ecosystem Workshop?



HEP analysis ecosystem workshop

22–24 May 2017
Amsterdam
Europe/Paris timezone

2017

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



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

IRIS-HEP AGC Workshop

25–26 Apr 2022
Europe/Zurich timezone

Enter your search term 

Overview

Timetable

Participant List


The **IRIS-HEP AGC Tools 2022 Workshop** is dedicated to showcasing tools and workflows related to the so-called “**Analysis Grand Challenge**” (AGC) being organised by **IRIS-HEP** and partners. The AGC focuses on running a physics analysis at scale, including the handling of systematic uncertainties, binned statistical analysis, reinterpretation and end-to-end optimization. The AGC makes use of new and advanced analysis tools developed by the community in the Python ecosystem, and relies on the development of the required cyberinfrastructure to be executed at scale. A specific goal of the AGC is to demonstrate technologies envisioned for use at the HL-LHC.

The agenda is currently work in progress.

If you have any questions, please do not hesitate to get in touch! You can find email addresses of the organizers below.

 **Starts** 25 Apr 2022, 15:30
Ends 26 Apr 2022, 20:00
Europe/Zurich

 **Alexander Held**
Oksana Shadura

 There are no materials yet.



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.



Analysis Ecosystems Workshop II

23-25 May 2022
IJCLab
Europe/Zurich timezone

Enter your search term

Overview

ist

sis-ecosystems...



It is five years since the first [Analysis Ecosystems Workshop](#) organised by the HSF in 2017. Since that time many changes have happened, with the advent of new projects, tools, and data formats, intense activity and progress in established projects and entirely new routes to explore, such as differentiable programming. Still, the challenge of efficient analysis for the HL-LHC era is not yet solved and so the HSF and IRIS-HEP would like to organise the *Second Analysis Ecosystems Workshop*.



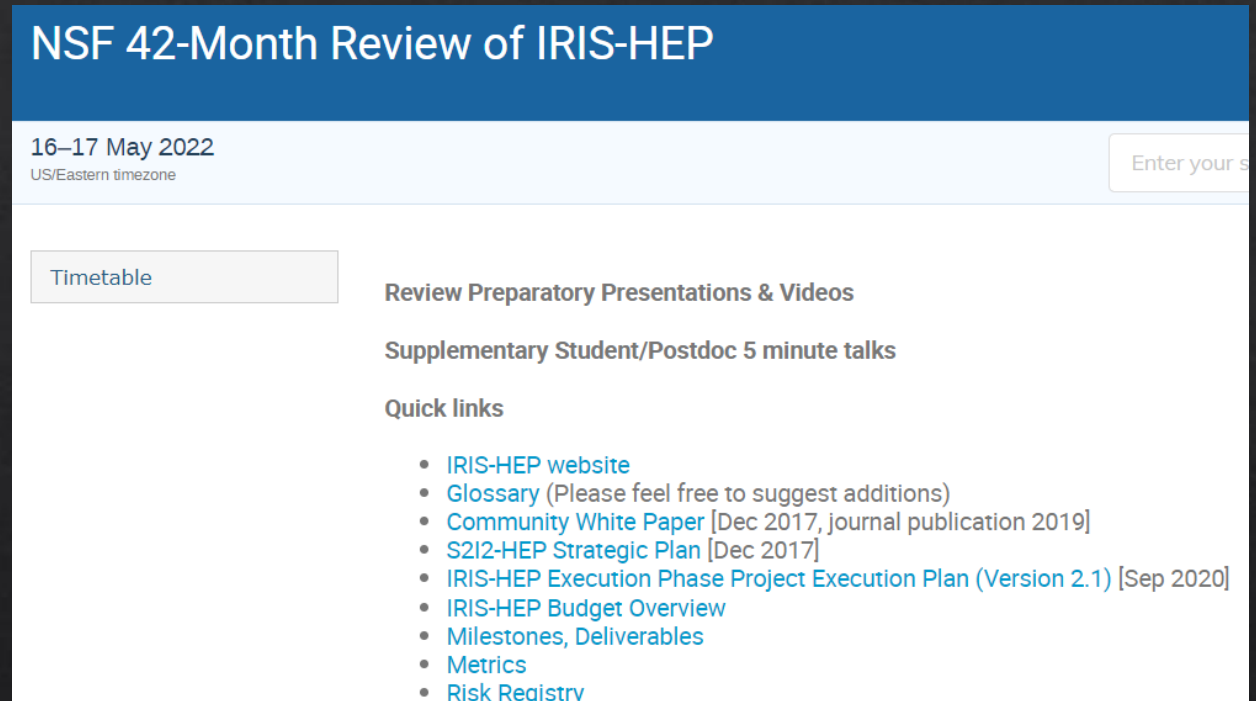
NSF Review

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.)



The screenshot shows the event page for the NSF 42-Month Review of IRIS-HEP. The header is blue with the event title in white. Below the header, the dates '16–17 May 2022' and 'US/Eastern timezone' are displayed. A search bar is visible in the top right corner. The main content area is white and contains a 'Timetable' button on the left. The right side of the page lists 'Review Preparatory Presentations & Videos', 'Supplementary Student/Postdoc 5 minute talks', and a 'Quick links' section with several blue hyperlinks.

NSF 42-Month Review of IRIS-HEP

16–17 May 2022
US/Eastern timezone

Enter your s

Timetable

Review Preparatory Presentations & Videos

Supplementary Student/Postdoc 5 minute talks

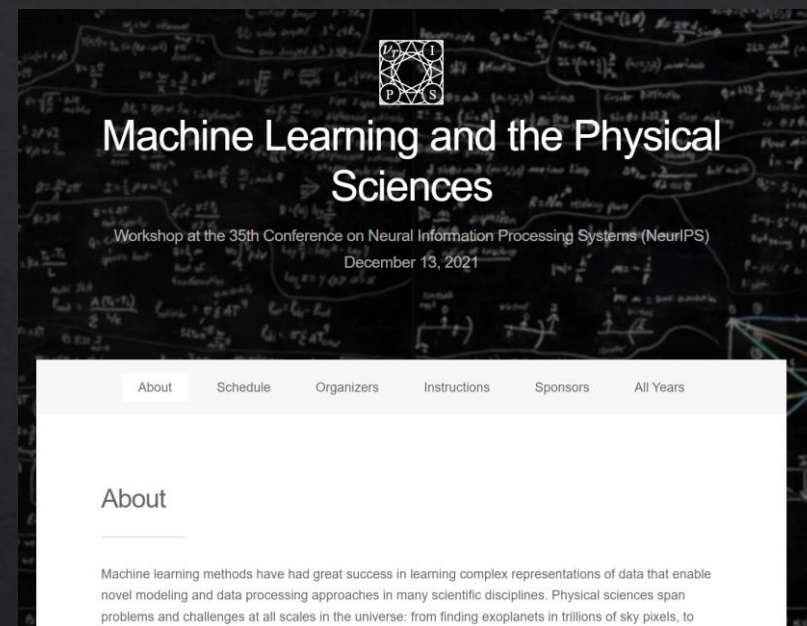
Quick links

- [IRIS-HEP website](#)
- [Glossary](#) (Please feel free to suggest additions)
- [Community White Paper](#) [Dec 2017, journal publication 2019]
- [S2I2-HEP Strategic Plan](#) [Dec 2017]
- [IRIS-HEP Execution Phase Project Execution Plan \(Version 2.1\)](#) [Sep 2020]
- [IRIS-HEP Budget Overview](#)
- [Milestones, Deliverables](#)
- [Metrics](#)
- [Risk Registry](#)








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











Invited Talks

 Megan Ansdell NASA Headquarters	 Bingqing Cheng University of Cambridge	 Surya Ganguli Stanford University	 Max Welling University of Amsterdam	 Laure Zanna New York University
---	---	--	--	--





Panel Discussion

 Jennifer Chayes UC Berkeley	 Marylou Gabrié New York University / Flatiron Institute	 Michela Paganini DeepMind	 Sara Solla Northwestern University	 Moderator: Lenka Zdeborová EPFL
---	--	--	--	---

Organizers

 Atılım Güneş Baydin University of Oxford	 Juan Felipe Carrasquilla Vector Institute / University of Waterloo	 Adji Bousso Dieng Princeton University	 Emine Kucukbenli Harvard University / Boston University
 Gilles Louppe University of Liège	 Benjamin Nachman Lawrence Berkeley	 Brian Nord Fermilab	 Savannah Thais Princeton University / IRIS- HEP

Steering Committee

 Anima Anandkumar Caltech / NVIDIA	 Kyle Cranmer New York University	 Prabhat Ram Microsoft	 Lenka Zdeborová EPFL
---	---	---	--



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!

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



 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 tracc) 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. tracc 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 tracc project. The examples of possible projects are (1) the development of tracc 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 coffee-casa Analysis Facility deployments:** coffee-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?



Data Science Tools for Analysis

Paper used as input to the HL-LHC Computing Review

1	Description and relevance for the HL-LHC	1
2	The HEP analysis software landscape is changing	2
3	The future of HEP analysis tools	9
3.1	File formats	9
3.2	Databases	11
3.3	Distributed computing	12
3.4	Acceleration	13
3.5	Histogramming	14
3.6	Fitting and statistics	16
3.7	Relationship to the community outside of HEP	19
4	Management, risk assessment, and the Grand Challenges	19

arXiv.org > physics > arXiv:2202.02194 Search...
Help | Advance

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?

