



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 #14



IRIS-HEP Steering Board Meeting #14

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)

Sept 13, 2022 (CERN VM WS Conflict)

Dec 13, 2022

February 14, 2023



Today

Feedback from the experiments

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 #14

Tuesday 7 Jun 2022, 18:00 → 20:20 Europe/Zurich

Zoom

Description Live Minutes Can Be Found Here. Please Help Out!
Zoom link distributed by email

18:00	→ 18:20	Introduction	🕒 20m	✎
Speakers: Brian Paul Bockelman (University of Wisconsin Madison (US)), Gordon Watts (University of Washington (US)), Peter Elmer (Princeton University (US))				
18:20	→ 18:50	LHCb Feedback	🕒 30m	✎
Speaker: Patrick Koppenburg (Nikhef National institute for subatomic physics (NL))				
18:50	→ 19:20	HSF Feedback	🕒 30m	✎
Speaker: Graeme A Stewart (CERN)				
19:20	→ 19:40	Discussion and AOB	🕒 20m	✎
Speakers: Brian Paul Bockelman (University of Wisconsin Madison (US)), Gordon Watts (University of Washington (US)), Peter Elmer (Princeton University (US))				



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)



NSF Review

One Day of Presentations

- Series of [YouTube](#) videos by students, postdocs, etc.
- One presentation for each of the 9 programmatic questions, virtual



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](#)



NSF Review

Recommendations

- 15 total
- Many used the phrase “encourage IRIS-HEP to continue...”

Theme: Connecting our work to the Science it Enables

- How does each project contribute to the science?
- How has community input affected science capabilities?
- Update web pages to add this meta data to each project

Theme: Sustainability & Roadmap

- What does the last 1.5 years of IRIS-HEP look like?
- What does the Software Roadmap to HL-LHC look like?
- How do the two play into each other?
- Links to theory community
- “Gap Analysis” for HL-LHC software ecosystem
- We know a lot about HL-LHC now
 - Input from experiments & US Operations Programs
 - Snowmass Whitepapers
 - Grand Challenges
- And we can collect quite a bit (in next ~1.5 years)
 - Internal Retreats (can we have one in person?)
 - CUA US S&C Meeting
 - Steering Board



NSF Review

Theme: Collaboration

- HSF engagement
- Engagement with the LHC experiments
- Engagement with other experiments and software projects where it makes sense
- How has community input affected science capabilities?
- Update web pages to add this meta data to each project
- Improve Software Citation

Theme: Sustainability

- Plan beyond the end of the IRIS-HEP project
- How will specific projects live beyond IRIS-HEP?

Next Steps

- Writing up response to the questions
- Starting the Year 5 Planning Process



Year 5 Planning Process

Due to the review, we are starting the Y5 Planning Process Late

- Focus Area Chats
 - Collate review responses
 - Plans for last year
- Chats with the PI's
 - Understanding how each PI's team fits into the plan
- EB meeting(s) to discuss plan
- Collect draft budget proposals for Y5

This will be a busy summer



Recent Highlights

- mkFit's validation by CMS now complete.
 - It is the default tracking software package for Run 3!
- SkyHook was released as part of Apache Arrow 7.0.0
- HTTP-TPC transition for the WLCG is now complete
 - almost 90% (by volume) of transfers use the new protocol



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.

The screenshot shows the event page for "Analysis Ecosystems Workshop II". The header is blue with the event title. Below the header, the dates "23-25 May 2022" and location "IJCLab" are listed, along with the timezone "Europe/Zurich timezone". A search bar is present on the right. The main content area features the logos for HSF (HEP Software Foundation) and iris hep. Below the logos, a paragraph of text describes the workshop's history and purpose.

Analysis Ecosystems Workshop II

23-25 May 2022
IJCLab
Europe/Zurich timezone

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*.











Analysis Ecosystem Workshop

Over 70 people attended in person

- Smaller number by Zoom (we had issues at the start)

Topics

- Metadata & Systematics
- Analysis Facilities
- ML and Autodiff Workflows
- Reduced Formats (nanoAOD, PHYSLITE)
- UX and Declarative Languages
- Realtime Trigger Analysis

Pickup badges and lunch tickets	
Auditorium Joliot Curie, IJCLab	08:45 - 09:00
Welcome	Michel Jouvin et al. 
Auditorium Joliot Curie, IJCLab	09:00 - 09:15
Introduction to the Workshop	Dr Graeme A Stewart 
Auditorium Joliot Curie, IJCLab	09:15 - 09:45
New Advances in Analysis Facilities	Robert Gardner 
Auditorium Joliot Curie, IJCLab	09:45 - 10:00
Analysis frameworks and Analysis Facilities: user experience	Nick Smith 
Auditorium Joliot Curie, IJCLab	10:00 - 10:15
Impact on Analysis Facilities in the context of DOMA evolution	Alessandra Forti 
Auditorium Joliot Curie, IJCLab	10:15 - 10:30
Coffee Break	
Espace Joliot Curie, IJCLab	10:30 - 11:00
Data Analysis Metadata Review	Thomas Kuhr 
Auditorium Joliot Curie, IJCLab	11:00 - 11:20
Big Questions in Systematics Processing	Stephan Hageboeck 
Auditorium Joliot Curie, IJCLab	11:20 - 11:40
ML Tooling and Use Landscape in HEP	Sean Joseph Gasiorowski 
Auditorium Joliot Curie, IJCLab	11:40 - 12:00



Other Recent Meetings & Workshops

June, 2022

- 3 Jun - 3 Jun, 2022 - [Mini-workshop on Graph Neural Networks for Tracking](#) (*Princeton University*)

May, 2022

- 31 May - 2 Jun, 2022 - [Connecting the Dots 2022](#) (*Princeton University*)
- 23 May - 25 May, 2022 - [Analysis Ecosystems Workshop II](#) (*IJCLab, Orsay, Paris*)

April, 2022

- 25 Apr - 26 Apr, 2022 - [IRIS-HEP AGC Tools 2022 Workshop](#) (*Virtual*)
- 21 Apr - 22 Apr, 2022 - [Matplotlib Training](#) (*Virtual*) (*Virtual*)

March, 2022

- 28 Mar - 30 Mar, 2022 - [Software Carpentry](#) (*Virtual*) (*Virtual*)
- 25 Mar - 25 Mar, 2022 - [Analysis Facilities Forum Kickoff](#) (*Virtual*)
- 5 Mar - 5 Mar, 2022 - [Data Analysis for Lab Research](#) (*Virtual*)

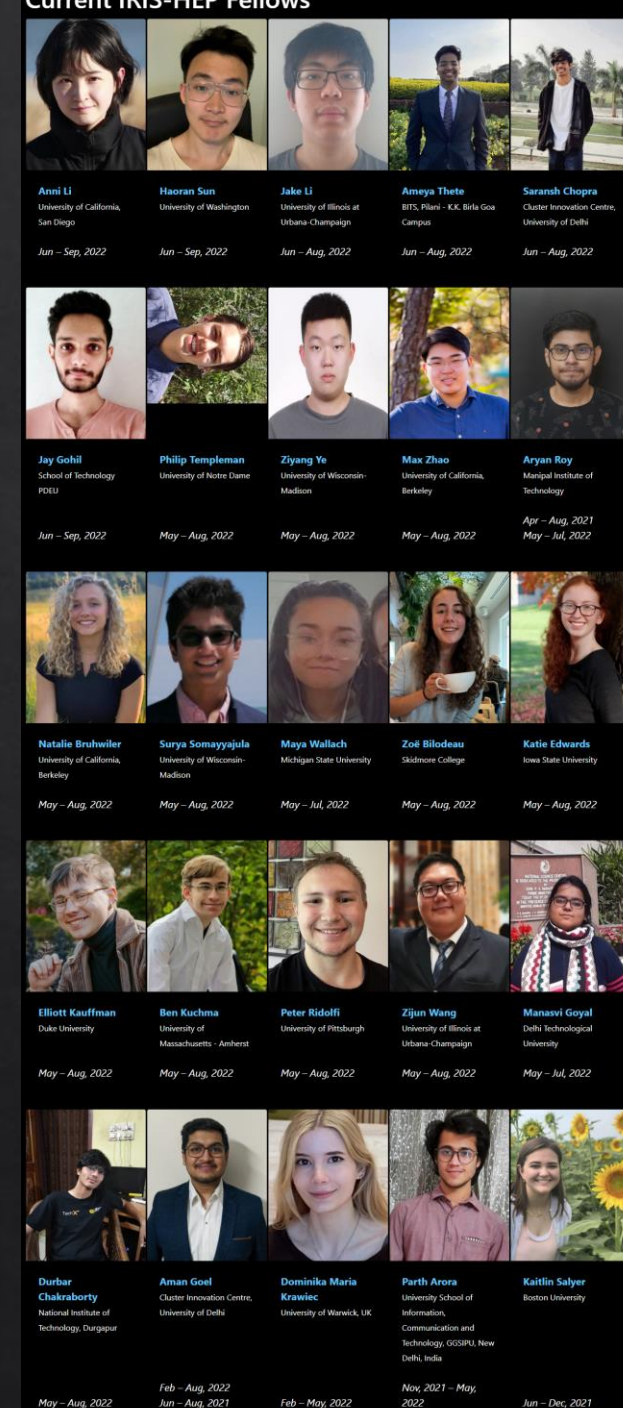


Fellows

New crop of undergraduate IRIS-HEP Fellows!

- Thanks for all the help in recruitment!
- Cross all focus areas of IRIS-HEP!
- Innovative Algorithms
 - PV finder in CMS/ATLAS
 - GNN in DUNE, ML for event-timing in DUNE
 - Generation using GNN, tracking using GNN
 - Muon Collider tracking & electron reco
 - Dead Sensor detection
 - 3D Reco of Atom Cloud for MAGIS-100
- DOMA
 - Network metrics in data federation caches
 - Dask and XROOTD integration

- Analysis Systems
 - Uproot acceleration, awkward & vector acceleration
 - Boost-Histogram improvements
 - Cling (C++) packaging
 - FAIR
 - Metrics for AF's
 - miniAOD transformer for SX, Jupyter plugin for SX
 - pyhf and CMS combine
 - Analysis workflow MINERvA



Questions & Comments?

