

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 #9





IRIS-HEP Steering Board Meeting #10

G. Watts

For the IRIS-HEP Executive Board

2021-06-08

“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

Oliver Gutsche (FNAL)
US CMS Ops Program

Paolo Calafiura (LBNL)
US ATLAS Ops Program

Kaushik De
(UTA)

Patrick Koppenburg (NIKHEF)
LHCb

Simone Campana (CERN)
WLCG

Graeme Stewart (CERN)
HSF

Alessandro Di Girolamo (CERN)
ATLAS

Zachary Marshall
(LBNL)

Ken Herner (FNAL)
The OSG Council



Welcome

steering-board@iris-hep.org

(you)

exec-board@iris-hep.org

(us)



Next Meeting Dates

September 7, 2021
November 30, 2021
February 15, 2022



Today

- DOMA
- IA
- Training Challenge Discussion

The screenshot shows a Zoom meeting agenda for "IRIS-HEP Steering Board Meeting #10". The meeting is scheduled for Tuesday, June 8, 2021, from 18:00 to 20:20 in the Europe/Zurich time zone. The description states: "Live Minutes Can Be Found [Here](#). Please Help Out! Zoom link distributed by email." The agenda items are as follows:

Time	Topic	Speakers	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 → 18:40	DOMA - Status and Plans	Brian Paul Bockelman (University of Wisconsin Madison (US))	20m
18:40 → 19:00	Innovative Algorithms - Status and Plans	David Lange (Princeton University (US))	20m
19:00 → 19:20	Discussion and Feedback from the Steering Board	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 in blue. 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

Review Was Held May 27-28

- Half was pre-prepared videos for the review committee
- Half was in-person talks to answer directly the review questions
 - Which come directly from our PEP
- Thank you everyone for support
 - We included reports of this meeting in the review when they asked how we collect feedback



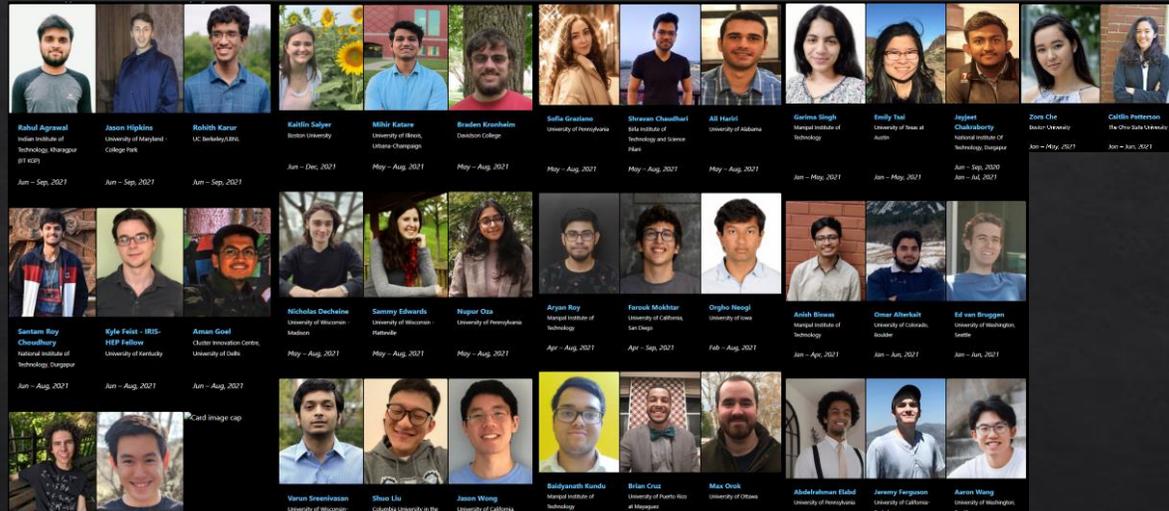
Some of the prepared videos are publically viewable
(like this one from Analysis Systems)

One side effect: Major update and scrubbing of our website



Fellows

Recently finished the call for summer fellows: we have quite a crop!
Several you helped us get the message out – thank you!



COVID-19 Version of Fellows:

- Graduate to Undergraduate
- In person to remote
- Room & board to stipend

Allowed a marked increase in number of people we were able to reach.

- When and how do we resume this program?

Next steps: looking for a more diverse set of mentors for the next round
(fall/winter)



Fellow Projects

Area	Project
AS	HISTOGRAMS USING NUMBA
AS	Integrating REANA Backend into ROB for evaluating workflows in the cloud
AS	Deep Learning Implementations for Sustainable Matrix Element Method Calculations
AS	Implementing ServiceX data as a source for ROOT's RDataFrame
AS	Modernizing FastJet Interfaces with pybind11 and interfacing with Awkward Arrays
AS	Automating Awkward Array Testing
AS	Advancing the Hist library
DOMA	Continuous Testing of Facility's Functionality Including Data Delivery Services Available in Coffea-Casa Analysis Facility
DOMA	Coffea-Casa Analysis Facility IRIS-HEP Fellow Proposal
DOMA	SkyhookDM - Ability to Push back query execution to Client in case of overload OSDs
DOMA	Facilitate continuous benchmarking/regression testing for the critical components of SkyhookDM
IA	Implementation of ML algorithms for ambiguity resolution in ACTS track reconstruction
IA	Implement hashing-based particle track reconstruction in ACTS
IA	Anomaly Detection with Spiking Neural Networks
IA	Developing Rotationally Symmetric Graph Neural Networks for Particle Track Reconstruction
IA	Graph Generative Models for Fast Detector Simulations in Particle Physics
IA	Graph Methods for Particle tracking
IA	CMS Jet Reconstruction with Quantile Neural Networks
IA	Developing Symmetric Graph Neural Networks for Charged Particle Tracking
IA	Analyzing Neutrino Interactions
IA	Muon Collider Tracking Software
IA	Particle Graph Autoencoders for Real-Time Jet Anomaly Detection
IA	Accelerating End-to-End Deep Learning Reconstruction using Graph Neural Networks
IA	Exploring the Role of Conformal Space Mapping in Deep Learning Models for Instance Segmentation of Particle Tracks



Training Challenge

Partly as a response to comments here (from HSF), partly as us taking a hard look at how the training is working...

Q: What is best way to integrate with existing experiment training/induction activities (and scale)?

We are now working to define, with the larger community, a series of specific goals for the period 2021-2023 in four categories and to work with the community to achieve them.

- **Scalability** - We aim for sufficient scalability in the training activities such that all students and postdocs can receive training in both the introductory material and the more advanced material. In the steady state we expect a required scale approximately equal to the number of incoming students each year.
- **Sustainability** - We aim to develop community processes by which both the instructors involved in training activities, and the training materials themselves, are continually renewed and meet the other two goals.
- **Training Scope** - We aim for a curriculum (introductory, intermediate, advanced) that broadly meets the needs of the community and evolves over time as needed.
- **Diversity and Inclusion** - The participation in the training should be representative of our community and (as we engage earlier in the pipeline) should work to represent the society at large



Year 4 Planning Process

Nearing the end of the Year 4 planning process

- Chats with the PI's have occurred
- Retreat had the Focus Area First Draft Plans
 - Lots of Year 4 plans
 - Even some discussion of Year 5
- Focus Area chats have occurred
- Now collecting draft budget proposals for Y4

Early summer we should be done!

Changes:

- ~\$150K for analysis facility seeds @ Nebraska and @Chicago
- Appointing two conveners for the Analysis Grand Challenge and one for the Data Grand Challenge.
- Building a Training Grand Challenge, and will find someone new to support this



Grand Challenges

Global challenges to knit together disparate parts of the institute

Analysis Grand Challenge

The analyzer wants to optimize an analysis end-to-end for a targeted signal hypothesis (including systematics) on an HL-LHC sized dataset so that they can obtain sensitive observed results for that signal while still being able to later reinterpret the analysis for various signal hypotheses.

Minor updates in next slides...

Data Processing Grand Challenge

Process a year's worth of LHC data from both experiments

Minor updates in next slides...



Grand Challenges

Analysis Grand Challenge

- We've updated the web pages describing the status of the grand challenges.
- Analysis Facility Discussions Continue
 - As does the report writing
- See Oksana's CHEP Plenary talk
 - [*Coffea-casa: An Analysis Facility Prototype*](#)
 - *DASK, ServiceX, Coffea, Awkward, all working together*
 - *Co-located at a T2, so DASK overflows into T2 queues smoothly (see demo in talk)*

Data Grand Challenge

- Work continues in the context of the WLCG.
- Testing of links between various endpoints (cloud and grid) is continuing
 - Helping to identify places that will need the most improvement for the data movement aspects of the challenge

Progress is a bit light due to the review...



Action Items From SB#9

- HSF related feedback
 - Community needs a way to strengthen C++ training
 - Can we link to Carpentries better than we currently are?
 - Specifically: best pedagogical practices? How do we evaluate our effectiveness?
 - Analysis Systems
 - Better document sustainability post-IRIS-HEP
 - Improve cooperation on declarative analysis
 - Common set of analysis benchmarks
 - More info on how we track and maintain connected with ROOT developments
 - Grand Challenge
 - How do you get to multi-user from single-analysis?
 - Establish discussions with HSF as they build their own plan-of-work
 - UW SWIFT project (Davide?) should describe their work
 - EP tasks and efforts - expose IRIS-HEP (contact: Graeme and Jacob)
- LHCb related Feedback
 - Follow up on JAX work for doing particle ID (uproot -> iminuit -> JAX -> matplotlib)
 - Are there analysis tool integrations with IRIS-HEP possible?
 - Analysis Preservation
 - REANA?
 - LHCb will add a "free" discussion inside LHCb and invite some folks from IRIS-HEP

We haven't started to seriously work on these yet due to the review



Questions? Comments?

