



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, and is ramping up its activities.

G. Watts, IRIS-HEP Steering Board Meeting #9



IRIS-HEP Steering Board Meeting #9

G. Watts

For the IRIS-HEP Executive Board

2021-02-16

“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

June 1, 2021

September 7, 2021

November 30, 2021

February 15, 2022



Today

- Feedback from HSF
- Feedback from LHCb
- Y3 Plans from SSL
- Discussion and AOB

IRIS-HEP Steering Board Meeting #9

Tuesday 16 Feb 2021, 18:00 → 20:20 Europe/Zurich

Zoom

Description Live Minutes Can Be Found [Here](#). Please feel free to help out!
Zoom link distributed by email about a week before the meeting.

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	→ 19:00	Experimental Feedback	⌚ 40m	📎
		HSF Feedback	⌚ 20m	📎
Speaker: Graeme A Stewart (CERN)				
		LHCb Feedback	⌚ 20m	📎
Speaker: Patrick Koppenburg (Nikhef National Institute for subatomic physics (NL))				
19:00	→ 19:30	SSL - Year 3 Plans	⌚ 30m	📎
Speaker: Robert William Gardner Jr (University of Chicago (US))				
19:30	→ 19:50	Feedback from the Steering Board	⌚ 20m	📎
Speakers: Brian Paul Bockelman (University of Nebraska Lincoln (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'.

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)



IRIS-HEP Year 3

Some “dates” of interest

March/April/May	NSF Yearly Review
End of April/May/June	Yearly Retreat
June 1 st	Steering Board Meeting #10
Late Spring/Summer	Year 4 Planning Process
Late Spring	Call for IRIS-HEP Fellows

} We would like to do experiment feedback again (LHCb & HSF?)



COVID: Restructured Fellows Program II

- Room & Board for graduate student to visit IRIS-HEP institution
- Work with an expert
- Carry knowledge back to home institution
- Tuition not covered
- A small fraction of undergraduates



- Supply salary
- Mostly to support upper division undergraduates
- Modeled more along the lines of Google Summer of Code
- Made special effort to reach out beyond the normal recruitment lists

Thanks for your help in the last call: we had 25 people apply, and 16 have started!



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

- Understanding Systematic Errors
 - Taxonomy
 - How can we *not* produce 800 trees for all variations?
- Progress towards defining an analysis
 - Where can we get sufficient size data for a real analysis (Open Data) (~100's of TB)
 - Milestones for this quarter
- Analysis Facilities Blueprint Meeting
 - Report is in preparation
- 1 once/month meeting between US Ops Programs and IRIS-HEP

Data Grand Challenge

- Work continues in the context of the WLCG
- Updates at the next SB meeting



Action Items From SB#8

- Measuring Opportunistic Resource usage in the OSG (DONE)
 - Two meetings connecting ATLAS and CMS have occurred
 - Ongoing: OSG is publishing more information. CMS has progress. ATLAS knows about the information.
- Mini-Workshop discussing end-to-end reconstruction (DONE)
 - Urge forming discussions at various conferences (vCHEP). IRIS-HEP will re-evaluate role in a year
- IRIS-HEP work with pyarrow (SkyHookDM)
 - Overlap with HEP-CCE/IOS group. Schedule topical meetings (IN PROGRESS)
 - Make sure HEP-CCE gets an invite to IRIS-HEP All-Hands meeting (IN PROGRESS)
 - Overlap with Innovative Algorithms (tracking/ ACTS as demo for GPU porting) – connect the groups (DONE)
- US Operations Postdocs – Schedule topical meetings to build community, publicize information, form for discussion
 - See [here](#) for already scheduled talks(look for HL-LHC R&D topics tag) (DONE)



Questions? Comments?

