



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 #7





IRIS-HEP Steering Board Meeting #7

G. Watts

For the IRIS-HEP Executive Board

2020-09-08

“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

Paolo Calafiura (LBNL)
US ATLAS Ops Program

Simone Campana (CERN)
WLCG

James Catmore
ATLAS

Oliver Gutsche (FNAL)
US CMS Ops Program

Patrick Koppenburg (NIKHEF)
LHCb

Graeme Stewart (CERN)
HSF

Ken Bloom (U. Nebraska-Lincoln)
The OSG Council



Welcome

steering-board@iris-hep.org

(you)

exec-board@iris-hep.org

(us)



Next Meeting Dates

November 24, 2020
Feb 16, 2021

Early in the new year will fill out 2021 meeting dates after the HSF community calendar is updated



Today

- Year 3 plans
- Y3 Plans from Innovative Algorithms and Analysis Systems
- Discussion

IRIS-HEP Steering Board Meeting #7

Tuesday 8 Sep 2020, 18:00 → 20:00 Europe/Zurich

Vidyó

Description [Live Minutes Can Be Found Here.](#) Please Help Out!

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	Vision and Progress - Year 3 Plans for Innovative Algorithms	🕒 30m	✎
Speakers: David Lange (Princeton University (US)), Heather Gray (UC Berkeley/LBNL)				
18:50	→ 19:20	Vision and Progress - Year 3 Plans - Analysis Systems	🕒 30m	✎
Including a discussion on the Data Analysis Grand Challenge				
Speaker: Kyle Stuart Cranmer (New York University (US))				
19:20	→ 19:40	Feedback and Discussion the Steering Board	🕒 20m	✎
Speakers: Brian Paul Bockelman (University of Nebraska Lincoln (US)), Gordon Watts (University of Washington (US)), Peter Elmer (Princeton University (US))				



A few highlights

Closing out Year 2

OSG-LHC demonstrated a prototype of an entire GSI-free site and is driving the worldwide transition from GSI/X.509 and GridFTP technologies.

This work also touches on DOMA and WLCG, of course (web tokens, etc.)

U.S. LHC site (the U.S. CMS Tier-2 at University of Nebraska) has been able partially migrate its production bulk data transfers from GridFTP to HTTP-TPC; approximately 20% of the CMS production traffic to Nebraska is over HTTP.

For the first time in HEP, the full likelihood for an ATLAS search was uploaded to HEPData. This used the pyhf data format and was featured on the CERN homepage

Efforts in Analysis Systems to both improve tools used by analyzers and increase ways to preserve physics analyses.



A few highlights

Closing out Year 2

The `func_adl` language has been integrated into DOMA's ServiceX project and tested on a 9.5 TB ATLAS xAOD sample. It was able to transform the sample in 2 hours into columns suitable for analysis in the python ecosystem. It was run on the UChicago SSL cluster, River. The bottleneck was the ability for an external system to furnish ServiceX with files to transform at a high enough rate.

The LHCb experiment chose the GPU-based Allen framework, to which IRIS-HEP is contributing, for its Run 3 high-level trigger facility.

Work with US ATLAS, Analysis Systems, DOMA, and SSL.



A few highlights

Closing out Year 2

The `func_adl` language has been integrated into DOMA's ServiceX project and tested on a 9.5 TB ATLAS xAOD sample. It was able to transform the sample in 2 hours into columns suitable for analysis in the python ecosystem. It was run on the UChicago SSL cluster, River. The bottleneck was the ability for an external system to furnish ServiceX with files to transform at a high enough rate.

IRIS-HEP collaborators hosted two international workshops on data reconstruction algorithms in HEP: on machine learning approaches to jets ([ML4Jets](#)) and on pattern recognition problems including tracking ([Connecting the Dots](#)).

Work with US ATLAS, Analysis Systems, DOMA, and SSL.

We have also hosted a number of innovative training and tutorial workshops



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'.

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)



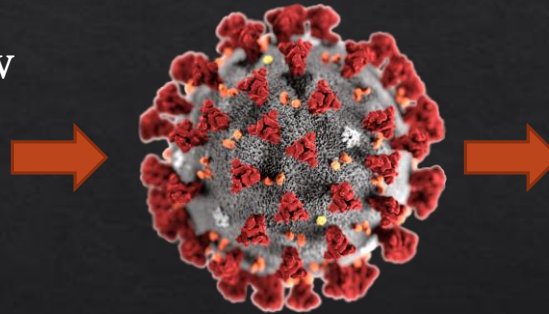
IRIS-HEP Summer

Blueprint Meeting: [Sustainable Software in HEP](#) (~90 registered)

Fellows Program

Normal Fellows Program:

- Goal: graduate students learn new skills from IRIS-HEP R&D
- Room & Board for 2 months
- Travel to an institute university



Class of COVID-19:

- Goal: undergraduate and graduate students learn new skills from IRIS-HEP R&D
- Mostly located locally or remotely
- No travel, but pay salary
- Was able to support 13
- LHCb PV, LHCb GPU trigger, GNN on FPGAs, pyhf testing, ServiceX testing, compression on nanoAOD, SkyHook testing & Spark interface, OSG Monitoring, ACTS track seeding & fitting & integration, CMS HGICAL CAN fast sim, awkward1 on GPUs, GNN on FPGA's, REANA,
- See [fellows page](#), [final presentations](#).



IRIS-HEP Year 3

Some “dates” of interest

Nov 24, 2020	Steering Board Meeting #8
Feb 16, 2020	Steering Board Meeting #9
Late Feb (?)	NSF Yearly Review
End of April, start of May	Yearly Retreat


We would like to do experiment feedback again.

Other meetings being planned



Review and Y3 Planning

From SB#6 Talk

1. The NSF 18 Month Review Results
2. Focus Area Discussions
3. PI Discussions
4. US Ops Programs Discussion
5. Steering Board Meeting
6. Full-Team Retreat
7. PEP Draft 
 - Iterated over by PI's
 - Given to the NSF
 - Iterate until approved
 - Year 3 starts Sept 1, 2020

PEP will be close to final at our next Steering Board Meeting – an obvious topic.



Year 3

Two comments from
the IRIS-HEP NSF
Review we've spent the
summer working on

- 1 Now that we have run for 2 years, improve milestones and metrics
- 2 Grand Challenges to help tie activities at the institute and the US operations programs, and the experiments together.



• Milestones & Metrics

All IRIS-HEP NSF milestones and metrics are public

➔ We will post Y3 shortly

Type	Time	Description	WBS task	Risks	Metric
M	ongoing	G1.1: Host quarterly Steering Board meetings	1.3		
M	ongoing	G1.2: Community Engagement (topical meetings, news items on website, Letters of support)	1.3		M1.1, M1.4, M1.5
D	ongoing	G1.3: Execute on Quarterly/Annual reporting responsibilities	1.2		
D	ongoing	G1.4: Maintain project staffing	1.2		
M	Y3Q1	G1.5: Organize and execute a Blueprint Workshop on Future Analysis Facilities	1.2, 1.3		
M	Y3Q2	G1.6: Successfully pass IRIS-HEP 30 Month Review			
M	Y3Q3	G1.7: Organize and execute a Blueprint Workshop on FPGA Accelerators	1.2, 1.3		
M	Y3Q3	G1.8: Organize and execute the IRIS-HEP Y3 Institute Retreat	1.2, 1.3		
M	Y3Q4	G1.9: Organize and execute an IRIS-HEP/NSF/DOE Blueprint Workshop	1.2, 1.3		
M	Y3Q4	G1.10: Year 3 IRIS-HEP Fellows Cohort	1.3		
M	Y3Q4	G1.11: Host annual Advisory Panel meeting	1.2		
D	Y3Q4	G1.12: Execute Year 4 subawards	1.1		
M	Y4	G1.13: Blueprint workshops (Grand Challenges Planning & Progress)	1.2, 1.3		
M	Y4Q2	G1.14: Successfully pass IRIS-HEP 42 Month Review			
M	Y4Q3	G1.15: Organize and execute the IRIS-HEP Y4 Institute Retreat	1.2, 1.3		
M	Y4Q4	G1.16: Year 4 IRIS-HEP Fellows Cohort	1.3		
M	Y4Q4	G1.17: Host annual Advisory Panel meeting	1.2		
D	Y4Q4	G1.18: Execute Year 5 subawards	1.1		
M	Y5Q3	G1.19: Organize and execute the IRIS-HEP Y5 Institute Retreat	1.2, 1.3		
M	Y5Q4	G1.20: Year 5 IRIS-HEP Fellows Cohort	1.3		
M	Y5Q4	G1.21: Host annual Advisory Panel meeting	1.2		
M	Y5	G1.22: Blueprint workshops (Grand Challenges Progress)	1.2, 1.3		

Management

➔ Blueprint meeting on Analysis Facilities

➔ DOE/NSF/IRIS-HEP Workshop

Year 4 Milestones

- We've written them to indicate where we are doing
- Will refine as part of the Year 4 processes



DOMA

Type	Time	Description	WBS task	Risks	Metrics
M	Y3Q1	G3.1: Second major workflow implemented in iDDS.	W3.1	R3.3	
D	Y3Q2	G3.2: ServiceX can natively send columnar results into SkyHook.	W3.1	R3.1, R3.4	M3.1, M3.2
M	Y3Q2	G3.3: Half of the transfer volume to one LHC site using HTTP-TPC (US internal transfers only).	W3.3	R3.2, R3.3	M3.3
D	Y4Q1	G3.4: iDDS implements a DAOD preprocessing workflow.	W3.1	R3.3	
M	Y4Q1	G3.5: All XCache instances updated to XRootD version 5 and deployed with non-X509 client auth.	W3.5	R3.3	M3.4
M	Y4Q2	G3.6: Demonstrate ability to sustain aggregate 100Gbps data flows from a source storage using HTTP-TPC.	W3.3	R3.2	M3.3
M	Y4Q3	G3.7: Deploy SkyHook at an analysis facility at a U.S. LHC site.	W3.2	R3.4	M3.1
M	Y4Q4	G3.8: ServiceX has a user community of 5 external analysis groups	W3.1	R3.1	M3.2
D	Y4Q4	G3.9: Demonstrate hardware-accelerated ServiceX performance at 30Gbps.	W3.1, W3.4	R3.4	
M	Y5Q1	G3.10: One U.S. LHC site retires GridFTP servers.	W3.3	R3.2	M3.3
M	Y5Q1	G3.11: Demonstrate ability to filter / process data at rate necessary for analysis challenge using SkyHook.	W3.2	R3.4	M3.1
D	Y5Q3	G3.12: Deploy prototype accelerated ServiceX at an analysis facility targeting baseline performance from Year 4.	W3.1	R3.4	M3.2

ServiceX and SkyHook integration

50% of one site transfers via HTTP

iDDS and DAOD preprocessing workflow (Y4)



SSC

Improve the existing workshops with feedback, latest techniques

Type	Time	Description	WBS task	Risks	Metrics
M	Y3Q1	G5.1: Review of basic carpentry non-HEP curriculum, material comprised of software carpentry modules/material taught at the past events	W5.6	R5.1	M5.6
M	Y3Q1	G5.2: Review of basic carpentry HEP curriculum material comprised of basic HEP related modules/material taught at the past events	W5.6	R5.1	M5.6
D	Y3Q1	G5.3: Website with carpentry curriculum (HSF Collab). The website (https://hepsoftwarefoundation.org/training/curriculum.html) is a reference for material for any future related workshop	W5.6		M5.7
M	Y3Q1	G5.4: Data/Software Carpentry outreach workshops for K-12 Physics teachers in Summer at UPRM	W5.1		M5.3 M5.5
M	Y3Q2	G5.5: Carpentries based on above curriculum. This will be a self-guided workshop where a group takes our material and uses/adds/modifies for their own needs	W5.1		M5.7
M	Y3Q2	G5.6: Evaluate outcome of workshops based on the curriculum by having the workshop organiser take our feedback survey	W5.3		M5.7
M	Y3Q1	G5.7: First Sustainability Blueprint in July to assess and guide future software efforts to include a strong element of sustainability to enable their adaptability to new challenges, longevity and efficiency	W5.5		M5.1
D	Y3Q1	G5.8: Community of instructors https://hepsoftwarefoundation.org/training/community.html	W5.2		M5.5
M	Y3Q1	G5.9: Programming Outreach Workshop in December 2020 on using google colab and CMS Data	W5.1		M5.3 M5.5
D	Y3Q2	G5.10: Review of Software Material for Programming Outreach, the material used couple of times would be reviewed based on survey and feedback from teachers who attend	W5.6	R5.1	M5.6
M	Y3Q2	G5.11: Programming Outreach Workshops based on reviewed material in Fall of 2020	W5.1		M5.3 M5.5
M	Y3Q4	G5.12: Second Training Blueprint workshop to assess the success of workshops based off of our basic training material , lessons learnt and focus on challenges, plan and design of HEP domain specific training topics	W5.5		M5.1
M	Y3Q4	G5.13: Brainstorming session with experts, fellows to revisit software and data challenges and correspondingly design future training material and workshops and domain specific topics	W5.3		M5.1
M	Y3Q4	G5.14: Organize and run CoDaS-HEP School (Parallel Programming, ML, Big Data Tools) at Princeton with experts and trained fellows	W5.1	R5.2	M5.2 M 5.4

Using Google Collab and CMS Data – Training Workshop



SSL

Type	Time	Description	WBS task	Risks	Metrics
D	Y3Q1	G6.1: Split River cluster into model and development environments	W6.1		
M, D	Y3Q2	G6.2: Establish a second SSL site	W6.3	R6.1	M6.1
D	Y3Q3	G6.3: Provision user training environments with federated ID, job submission, interactive notebooks	W6.4	R6.2	M6.2
D	Y4Q2	G6.4: Provide federated scheduling of SSL clusters	W6.1, W6.2, W6.3	R6.2	M6.3
D	Y3Q4	G6.5: If cost/benefit relative to accelerated ServiceX on cluster baseline indicates, build prototype system and benchmark its performance	W6.5	R6.3	
D	Y4Q3	G6.6: Coordinate with AS, DOMA, and operations programs to benchmark performance of prototype system components to be used for Analysis Grand Challenge.	W6.1, W6.3	(R2.2)	(M2.2)
D	Y5Q2	G6.7: Coordinate with AS, DOMA, and operations programs to execute the Analysis Grand Challenge	W6.3		(M2.2)

Establish second SSL site

Federation of SSL sites



OSG

Type	Time	Description	WBS task	Risks	Metrics
D	Y3Q1	G7.1: Have at least 1 U.S. site each for ATLAS and CMS in a 100 Gbps perfSONAR mesh	W7.2	R7.5, R7.6	M.7.14
D	Y3Q4	G7.2: Define IRIS-HEP's role in networking for data challenge	W7.2, W7.5	R7.7	—
M	Y3Q2	G7.3: OSG 3.6.0 is released	W7.2, W7.4	R7.4, R7.6, R7.7	M.7.5, M.7.7, M.7.8, M.7.10
D	Y5Q1	G7.4: End support for the OSG 3.5 release series	W7.2, W7.4	R7.4, R7.5, R7.6	M.7.5, M.7.10
D	Y4Q2	G7.5: Have at least 5 instances of refactored XRootD monitoring in production use	W7.2, W7.5	R7.5, R7.6, R7.7	M.7.3
D	Y4Q2	G7.6: Complete the SDLC from engagement to production deployment for one IRIS-HEP software product	W7.2	R7.4, R7.5, R7.6	M.7.5, M.7.8
D	Y5Q2	G7.7: Have one example of an IRIS-HEP-supported service in production using the new container paradigm	W7.2, W7.5	R7.5, R7.6	M.7.4, M.7.6

Data Processing Challenge

OSG 3.5 -> 3.6



Grand Challenges

Review Committee: Global challenges to knit together disparate parts of the institute

- Demonstrate useful work for the LHC Run 4 Era
- Force building of tools for longer-term than just the challenge
- Span more than one area of the institute
- Have stakeholder participation
- Fine to span multiple years (encouraged)

Analysis Grand Challenge

In today's Analysis Systems Talks

Data Processing Grand Challenge

Next Steering Board Meeting



NSF: IAIFI



[About](#) [People](#) [Research](#) [Fellows](#)



The NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)

- Two members of IRIS-HEP are founding members of IAIFI
- As they spin-up we will look for points of collaboration



Questions? Comments?

