



# Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)

PI: Peter Elmer (Princeton), co-PIs: Brian Bockelman  
(Morgridge Institute), Gordon Watts (U.Washington) with  
UC-Berkeley, University of Chicago, University of Cincinnati,  
Cornell University, Indiana University, MIT, U.Michigan-Ann  
Arbor, U.Nebraska-Lincoln, New York University, Stanford  
University, UC-Santa Cruz, UC-San Diego, U.Illinois at  
Urbana-Champaign, U.Puerto Rico-Mayaguez and  
U.Wisconsin-Madison



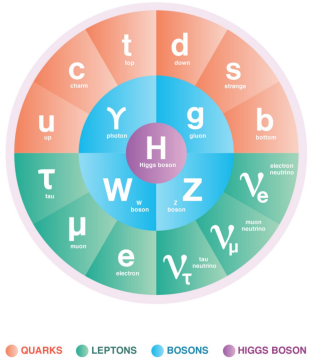
OAC-1836650

<http://iris-hep.org>

*IRIS-HEP was funded as of 1 September,  
2018, and is currently ramping up activities*

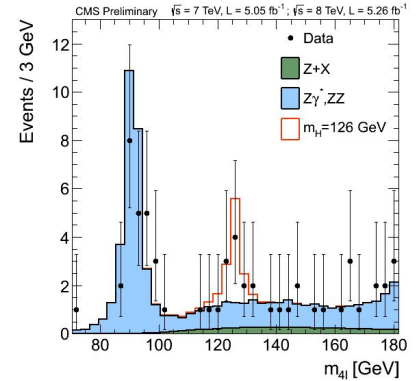


# Science Driver: Discoveries beyond the Standard Model of Particle Physics

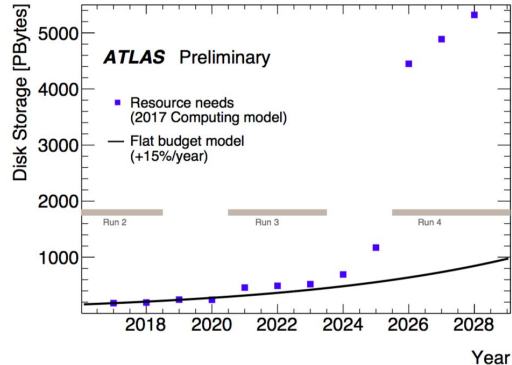


From “Building for Discovery - Strategic Plan for U.S. Particle Physics in the Global Context” - Report of the Particle Physics Project Prioritization Panel (P5):

- 1) Use the Higgs boson as a new tool for discovery
- 2) Pursue the physics associated with neutrino mass
- 3) Identify the new physics of dark matter
- 4) Understand cosmic acceleration: dark matter and inflation
- 5) Explore the unknown: new particles, interactions, and physical principles

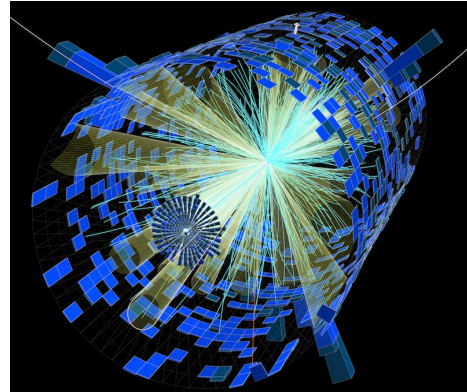


## Computational and Data Science Challenges of the High Luminosity Large Hadron Collider (HL-LHC) and other HEP experiments in the 2020s

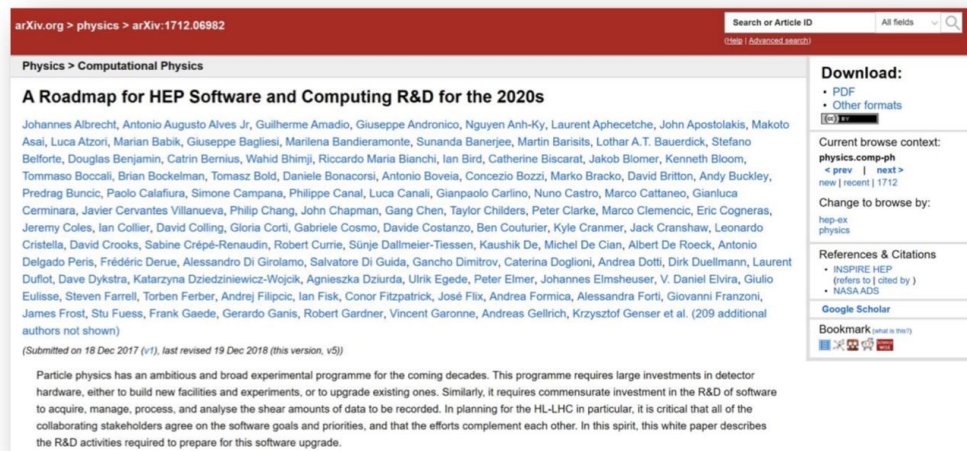


The HL-LHC will produce exabytes of science data per year, with increased complexity: an average of 200 overlapping proton-proton collisions per event.

During the HL-LHC era, the ATLAS and CMS experiments will record ~10 times as much data from ~100 times as many collisions as were used to discover the Higgs boson (and at twice the energy).

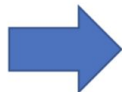


# The Community White Paper Process (2016-2017)



Involved A Diverse

- Computing Management from the Experiments and Labs
- Individuals interested in the problems
- Members of other compute intensive scientific endeavors
- Members of Industry



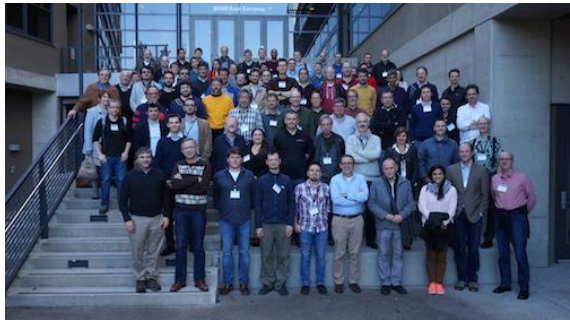
## Individual Papers on the arXiv:

Careers & Training, Conditions Data, DOMA, Data Analysis & Interpretation, Data and Software Preservation, Detector Simulation, Event/Data Processing Frameworks, Facilities and Distributed Computing, Machine Learning, Physics Generators, Security, Software Development, Deployment, Validation, Software Trigger and Event Reconstruction, Visualization

## Community White Paper & the Strategic Plan



# Growing a Global Collaboration



JLab  
March, 2018  
HSF/OSG/WLCG



UCSD/SDSC  
January, 2017  
HSF CWP



Annecy  
June, 2017  
HSF CWP

Naples  
March, 2017  
WLCG/HSF



# IRIS-HEP

## Sustainable Software R&D objectives

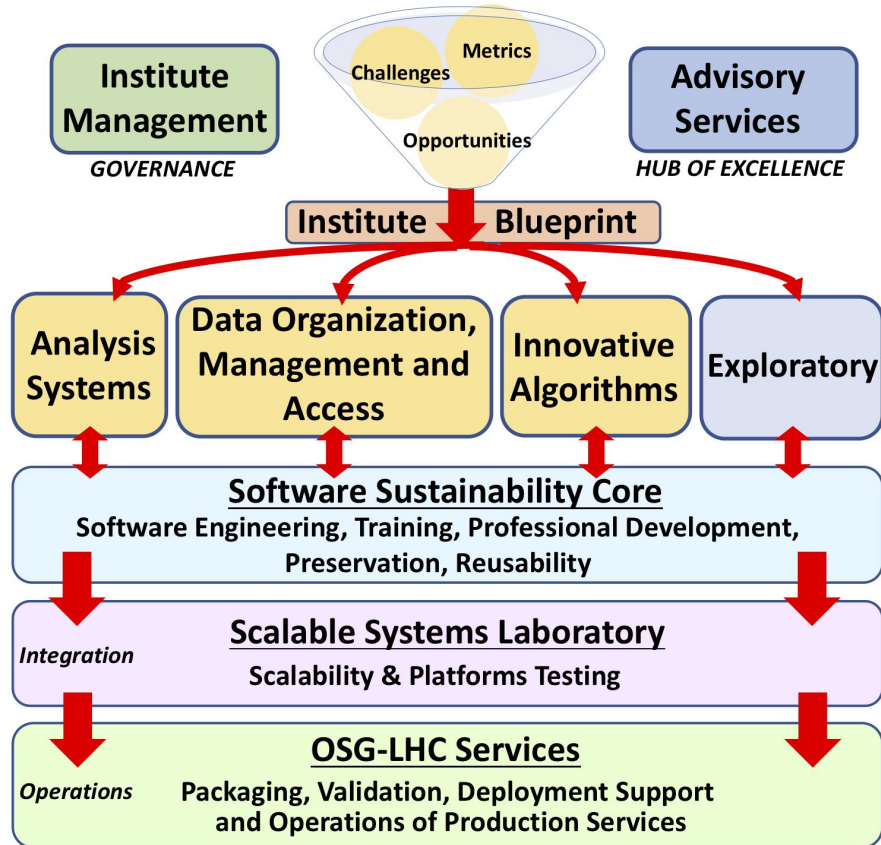
- 1) Development of innovative algorithms for data reconstruction and triggering;
- 2) Development of highly performant analysis systems that reduce “time-to-insight” and maximize the HL-LHC physics potential; and
- 3) Development of data organization, management and access systems for the community’s upcoming Exabyte era.
- 4) Integration of software and scalability for use by **the LHC community on the Open Science Grid**, the Distributed High Throughput Computing infrastructure in the U.S.

Intellectual Hub for the HEP Community













The plan for IRIS-HEP reflects a community vision developed by an international community process organized by the HEP Software Foundation (<https://hepsoftwarefoundation.org>). The S2I2-HEP conceptualization project (<http://s2i2-hep.org>) derived a Strategic Plan from the community roadmap which would leverage the strengths of the U.S. university community. IRIS-HEP aims to function as an intellectual hub for the national and international HEP community, through training, community workshops and the development of wider collaborations with the larger computer and data science communities.

# IRIS-HEP Structure and Executive Board



## Executive Board

The IRIS-HEP Executive Board manages the day to day activities of the Institute.

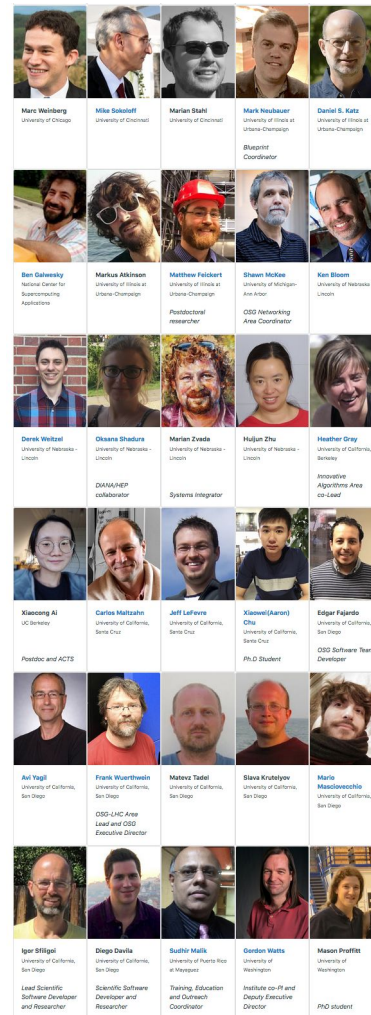
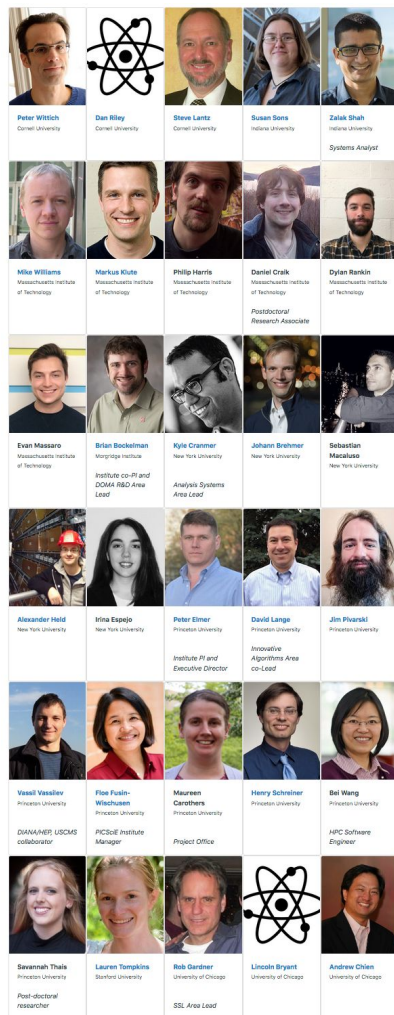
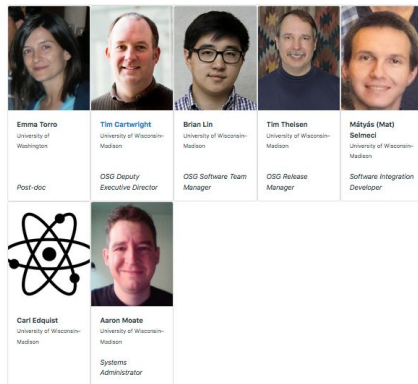
|   |  |  |  |   |
|---|--|--|--|---|
|  |                                   |                     |                         |                                        |
| <b>Peter Elmer</b><br>Princeton University<br><br><i>Peter.Elmer@cern.ch</i>        | <b>Gordon Watts</b><br>University of Washington<br><br><i>Institute co-PI and Deputy Executive Director</i>          | <b>Brian Bockelman</b><br>Morgridge Institute<br><br><i>Institute co-PI and DOMA R&amp;D Area Lead</i> | <b>Heather Gray</b><br>University of California, Berkeley<br><br><i>Innovative Algorithms Area co-Lead</i> | <b>David Lange</b><br>Princeton University<br><br><i>Innovative Algorithms Area co-Lead</i>                               |
|  |                                   |                     |                         |                                        |
| <b>Kyle Cranmer</b><br>New York University<br><br><i>Analysis Systems Area Lead</i> | <b>Sudhir Malik</b><br>University of Illinois at Mayaguez<br><br><i>Training, Education and Outreach Coordinator</i> | <b>Mark Neubauer</b><br>University of Illinois at Urbana-Champaign<br><br><i>Blueprint Coordinator</i> | <b>Rob Gardner</b><br>University of Chicago<br><br><i>SSL Area Lead</i>                                    | <b>Frank Wuerthwein</b><br>University of California, San Diego<br><br><i>OSG-LHC Area Lead and OSG Executive Director</i> |



# IRIS-HEP Team

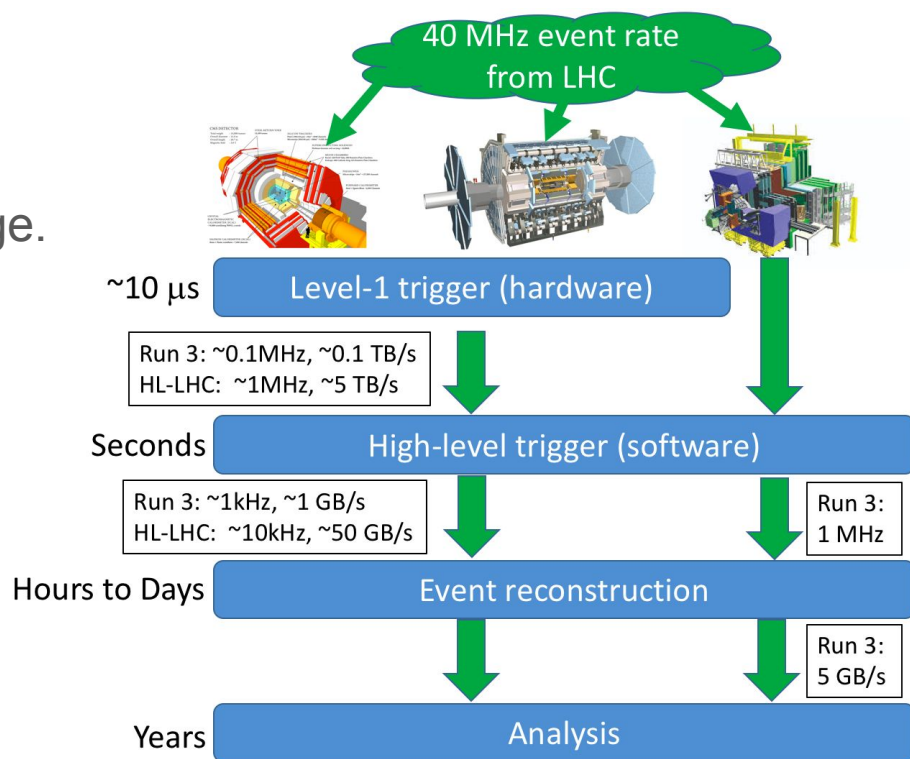
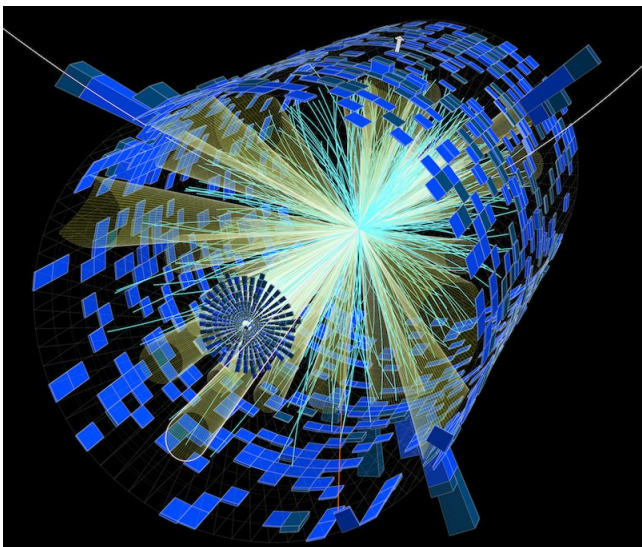
<http://iris-hep.org/about/team>

About 30 FTEs of effort  
spread over a larger number  
of people from 18  
universities/institutions



# Innovative Algorithms - Trigger/Reconstruction

Algorithms for real-time processing of detector data in the software trigger and offline reconstruction are critical components of HEP's computing challenge.





# Innovative Algorithms

These algorithms face a number of new challenges during HL-LHC:

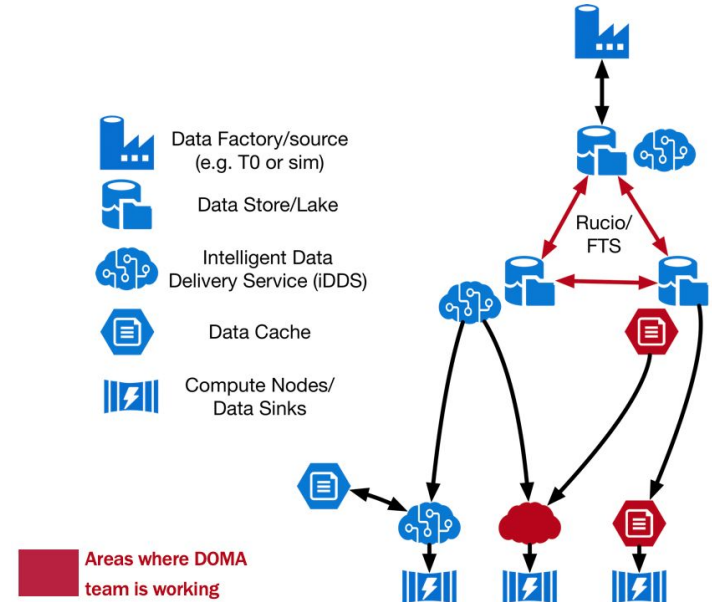
- Upgraded accelerator capabilities, with more collisions per bunch crossing (“pile-up”)
- Detector upgrades, including new detector technologies and capabilities  
Increased event rates to be processed
- Emerging computing architectures

Specific R&D investments include collaboration with ACTS, continuation of the parallel Kalman Filter tracking project, work on HLS4ML, etc.

# Data Organization, Management and Access (DOMA)

The DOMA focus area performs fundamental R&D related to the central challenges of organizing, managing, and providing access to exabytes of data from processing systems of various kinds.

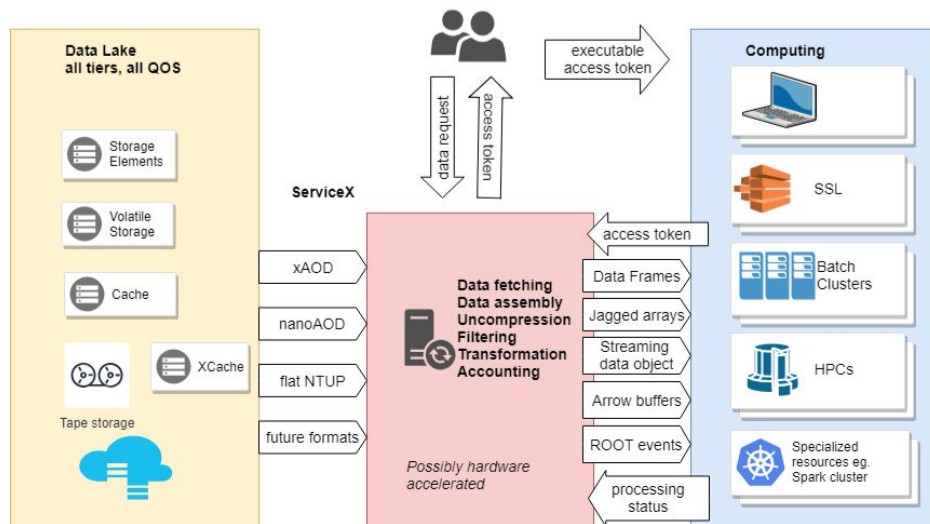
- **Data Organization:** Improve how HEP data is serialized and stored.
- **Data Access:** Develop capabilities to deliver filtered and transformed event streams to users and analysis systems.
- **Data Management:** Improve and deploy distributed storage infrastructure spanning multiple physical sites. Improve inter-site transfer protocols and authorization



# DOMA: Intelligent Data Delivery



- In the HL-LHC era, we must deliver more events - and at lower latencies - if the analysts want to make progress!
  - Low-latency delivery of events requires transformation from long-term archival formats that we want to decrease data size.
  - Data should be transformed and delivered at the storage level, not at the workstation.
  - Users should be enabled to work on a multitude of data formats (esp. non-ROOT) without having to write them to disk.
- We are currently prototyping an Intelligent Data Delivery Service (IDDS) that delivers the desired LHC events without requiring complex transforms or skims from the user.



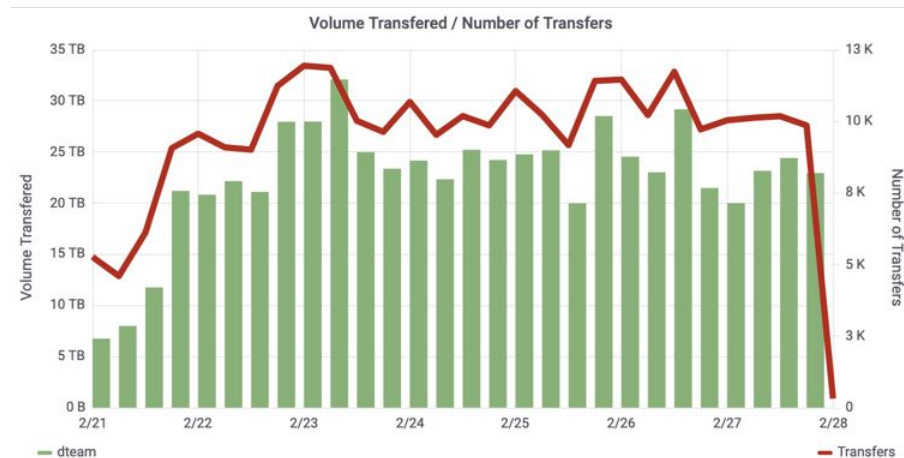


# DOMA: Moving Bulk WLCG Data

[Homepage](#) of the WLCG working group



- There is a strong movement in the community to move from niche protocols for bulk data movement to more standardized ones such as HTTP.
  - Bockelman co-leads the working group within the WLCG for “third party copy” (TPC).
- During IRIS-HEP, HTTP-TPC has gone from small test transfers to scale tests on servers to scale tests in the WLCG DOMA community.
  - Demonstrated HTTP’s ability to achieve speeds similar to GridFTP on dedicated server hardware.



# Analysis Systems R&D

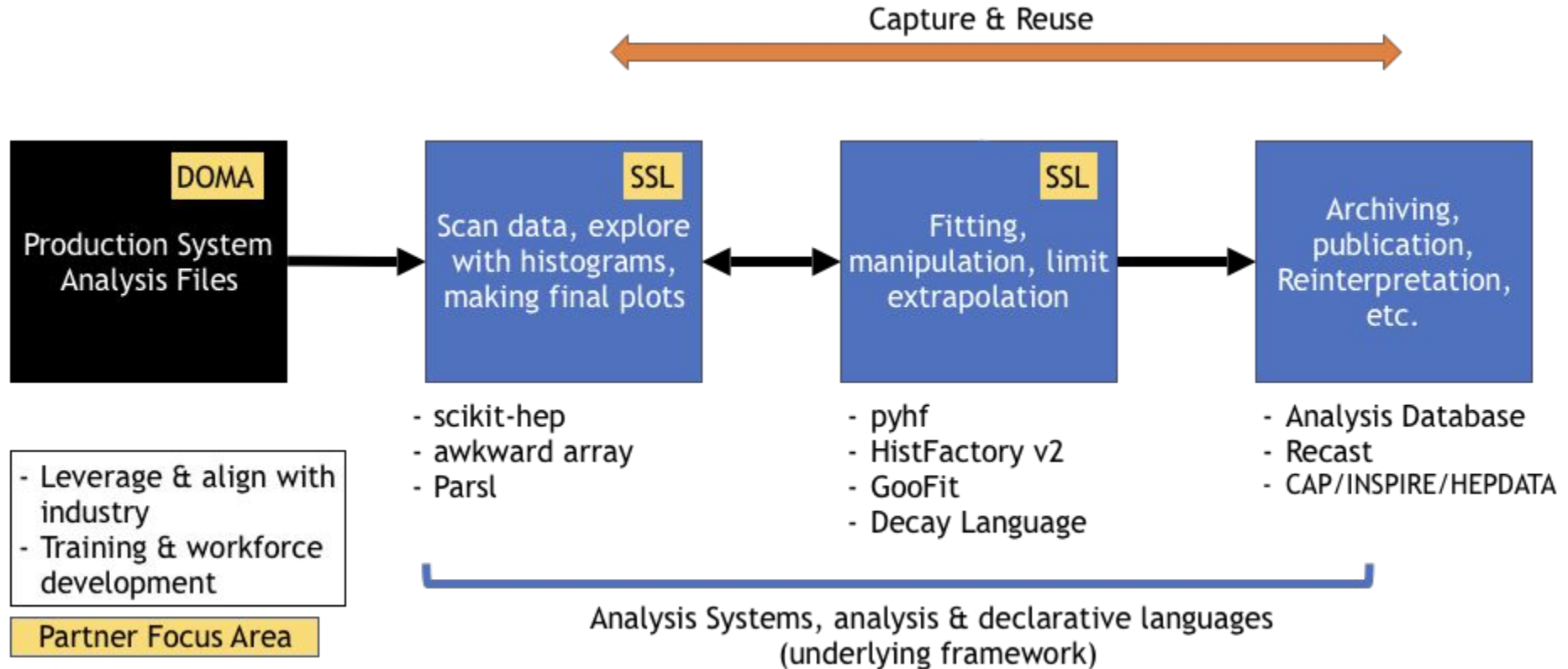
**Primary Goals:** Develop sustainable analysis tools to extend the physics reach of the HL-LHC experiments by

- creating greater functionality
- reducing time-to-insight
- lowering the barriers for smaller teams, and
- streamlining analysis preservation, reproducibility, and reuse.

Compared to DOMA and Innovative Algorithms (which has more targeted reco/trigger goals), the Analysis Systems group is dealing with more “greenfield” area where there is a very heterogeneous set of use cases and relevant components.

The nature of IRIS-HEP Analysis Systems tasks is more exploratory: “big R” (R&d)

# Analysis Systems Data Flow and Projects



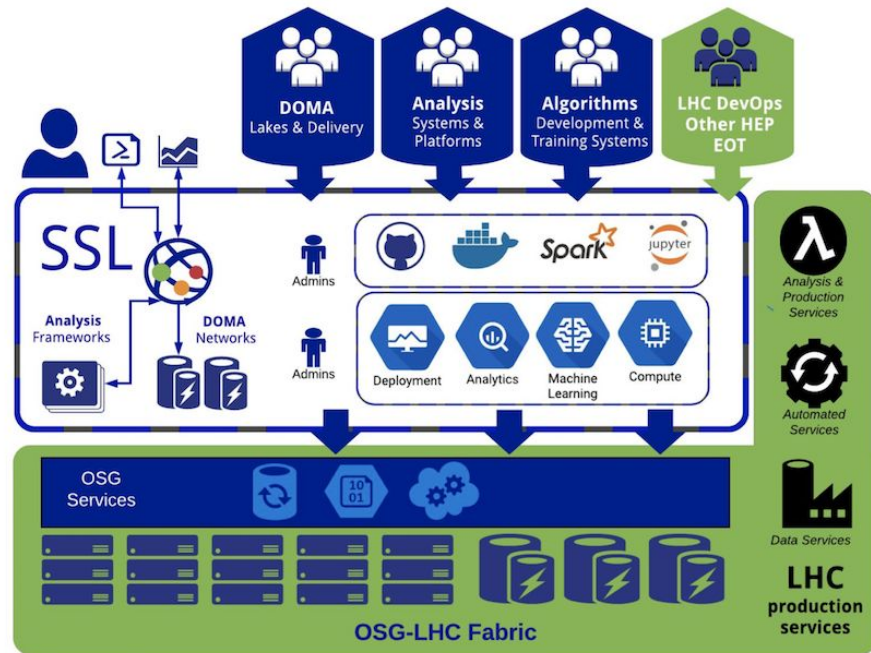


# Scalable Systems Laboratory (SSL)



Goal: Provide the Institute and the HL-LHC experiments with scalable platforms needed for development in context

- Provides access to infrastructure and environments
- Organizes software and resources for scalability testing
- Does foundational systems R&D on accelerated services
- Provides the integration path to the OSG-LHC production infrastructure



# Open Science Grid (OSG) for the LHC



## The People in OSG



|  |  |   |  |  |   |  |  |
|--|--|---|--|--|---|--|--|
| <br>Deputy Director<br>Tim Cartwright | <br>Security<br>Zalak Shah      | <br>Security<br>Susan Sons   | <br>operations<br>John Thiltges | <br>operations<br>Marian Zvada | <br>software<br>Mátyás (Mat) Selmeči | <br>releases<br>Tim Theisen | <br>networking<br>Shawn McKee |
| <br>software<br>Brian Lin             | <br>operations<br>Derek Weitzel | <br>software<br>Carl Edquist |  |  |   |  |  |

**Operations = UNL**  
**Security = Indiana University**  
**Software = U. Wisconsin – Madison**  
**Networking = U. Michigan**

A total of 6 FTE across 11 people.  
These people have worked together  
and with the LHC program years.

(Slide from FKW)

# Open Science Grid (OSG) and IRIS-HEP



## OSG Effort Portfolio



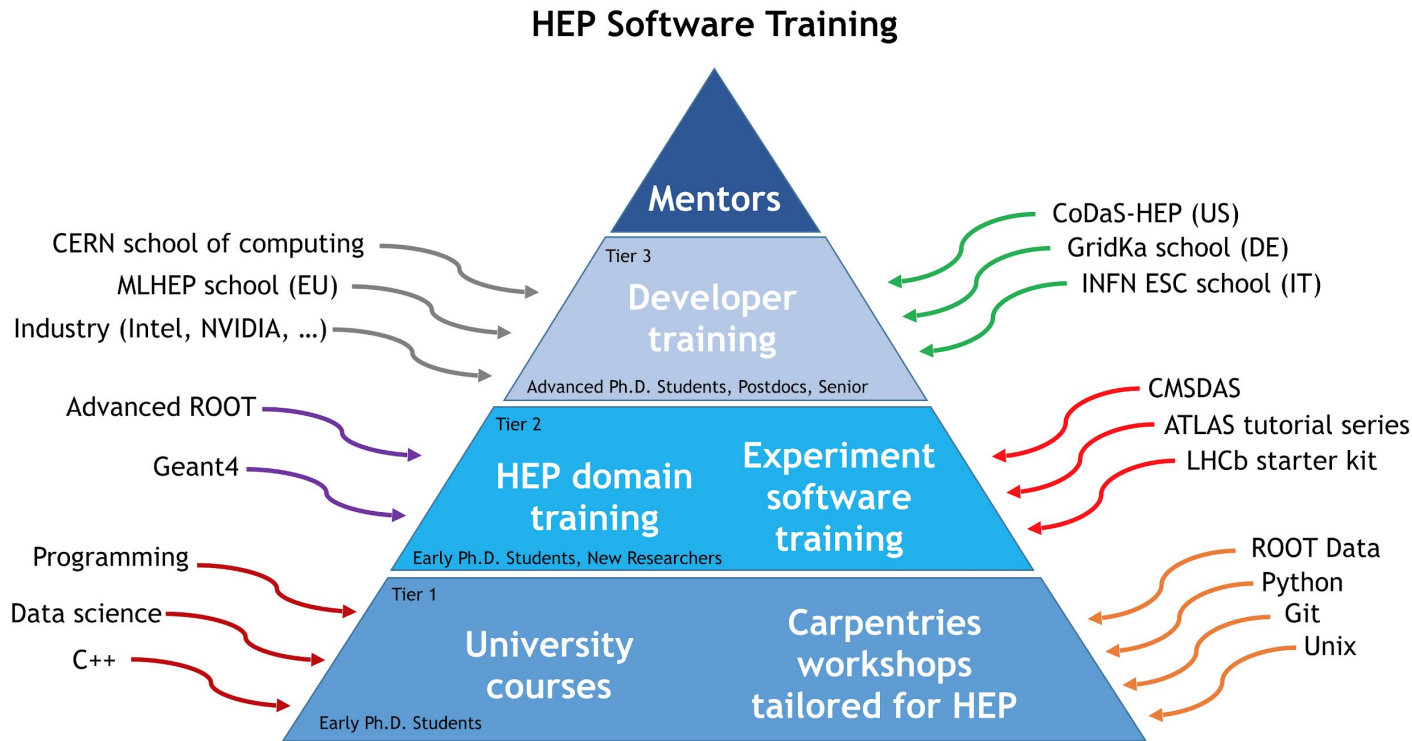
| A        | B        | C          | D          | E       | F          | G        | H          |
|----------|----------|------------|------------|---------|------------|----------|------------|
|          | $\Sigma$ | Technology | Operations | Support | Networking | Security | Management |
| $\Sigma$ | 17.56    | 7.46       | 4.05       | 2.82    | 1.24       | 1.00     | 0.98       |
| OSG N5Y  | 6.86     | 2.48       | 2.05       | 1.82    | 0.00       | 0.20     | 0.32       |
| IRIS-HEP | 5.94     | 3.08       | 1.00       | 0.00    | 0.39       | 0.80     | 0.67       |
| SAND     | 0.75     | 0.00       | 0.00       | 0.00    | 0.75       | 0.00     | 0.00       |
| OSG-NP   | 1.00     | 0.50       | 0.50       | 0.00    | 0.00       | 0.00     | 0.00       |
| TNRP     | 1.00     | 0.70       | 0.20       | 0.00    | 0.10       | 0.00     | 0.00       |
| CESER    | 2.00     | 0.70       | 0.30       | 1.00    | 0.00       | 0.00     | 0.00       |
|          |          |            |            |         |            |          |            |

**IRIS-HEP is roughly 1/3 of total effort in OSG**

(Slide from FKW)



# Training and Education - Sustainability/Scalability



This is a general framework for training, but from the NSF we have funds from both IRIS-HEP (OAC-183665) and a separate project FIRST-HEP (OAC-1829707, OAC-1829729, <http://first-hep.org>) which can work towards implementing this model.

# Training, Education and Outreach Events

## Upcoming events:

- 19-21 Aug, 2019 - ATLAS Software Carpentries Training
  - Lawrence Berkeley National Laboratory
  - [Indico page](#)

## Past events:

- 22-26 Jul, 2019 - Computational and Data Science for High Energy Physics (CoDaS-HEP) 2019 School
  - Princeton University
  - [Webpage](#)
- 10 Jun, 2019 - FIRST-HEP/ATLAS Software Training
  - Argonne National Laboratory
  - [Indico page](#)
- 3-4 Jun, 2019 - An introduction to programming for STEM teachers
  - University of Puerto Rico at Mayaguez
  - [Indico page](#)
- 24-26 Apr, 2019 - Machine Learning Hackathon for UPRM Students
  - University of Puerto Rico at Mayaguez
  - [Indico page](#)
- 1-2 Apr, 2019 - Software Carpentry Workshop
  - Fermi National Accelerator Laboratory
  - [Indico page](#)

In collaboration with FIRST-HEP (<http://first-hep.org>), the Carpentries (<https://carpentries.org>) and others



CoDaS-HEP 2017



CoDaS-HEP 2018

## Current IRIS-HEP Fellows



**Raghav Kansal**  
University of California,  
San Diego

*IRIS-HEP Fellow*  
*Jun-Aug 2019*



**Pratyush (Reik)  
Das**  
Institute of Engineering  
& Management  
(Kolkata)

*IRIS-HEP Fellow*  
*Jun-Sep 2019*



ML Hackathon UPRM



CoDaS-HEP 2019

<http://codas-hep.org>

# IRIS-HEP Community Activities and Events



## Upcoming Events:

IRIS-HEP team members are involved in organizing the following events:

- 29 Jul, 2019 - [IRIS-HEP Tutorial: Fast columnar data analysis with data science tools](#) (*Northeastern University / APS DPF 2019*)
- 19 Aug - 23 Aug, 2019 - [ATLAS Software Carpentries Training](#) (*LBNL*)
- 10 Sep - 11 Sep, 2019 - [Blueprint: Accelerated Machine Learning and Inference](#) (*Fermilab*)
- 23 Oct - 25 Oct, 2019 - [Blueprint: A Coordinated Ecosystem for HL-LHC Computing R&D](#) (*Catholic University of America, Washington DC*)
- 13 Dec - 14 Dec, 2019 - [Machine Learning and the Physical Sciences at NeurIPS 2019](#) (*Vancouver Convention Centre*)
- 15 Jan - 17 Jan, 2020 - [ML4Jets2020](#) (in planning) (*New York University*)

## Recent Events:

- 22 Jul - 26 Jul, 2019 - [CoDaS-HEP 2019](#) (*Princeton University*)
- 1 Jul - 2 Jul, 2019 - [TrackML Challenge: Grand Finale](#) (*CERN*)
- 21 Jun - 22 Jun, 2019 - [Blueprint: Analysis Systems R&D on Scalable Platforms](#) (*NYU*)
- 19 Jun - 20 Jun, 2019 - [Analysis Systems Topical Workshop](#) (*NYU*)
- 10 Jun, 2019 - [FIRST-HEP/ATLAS Software Training](#) (*Argonne National Lab*)
- 3 Jun - 4 Jun, 2019 - [An introduction to programming for STEM teachers](#) (*University of Puerto Rico at Mayaguez*)
- 6 May - 8 May, 2019 - [Analysis Description Languages Workshop](#) (*Fermilab*)

See <http://iris-hep.org> for the full and latest list



# IRIS-HEP Topical Meeting Series



[Home](#) » [Projects](#) » [IRIS-HEP](#) » [Topical Meetings](#)

## Topical Meetings

Two weekly time slots are available for IRIS-HEP topical meetings:

- Mondays - 17:30-18:30GVA (Vidyo and 40-R-B10 at CERN)
- Wednesdays - 18:00-19:00GVA (Vidyo only)

There is one event in the future. [Hide](#)

April 2019

 15 Apr [Development of new Histogram tools](#)

March 2019

 25 Mar [Introduction to modern CDN Architectures](#)

 04 Mar [The FAST project](#)

February 2019

 25 Feb [Analysis Description Languages](#)

 18 Feb [Integration of C++ Modules into CMSSW](#)

 13 Feb [HLS4ML: Using ML on FPGAs to enhance reconstruction output](#)

 04 Feb [Training for Software, Computing, Computational and Data Science in HEP](#)

January 2019

 28 Jan [FuncX: High Performance Function as a Service for Science](#)

<https://indico.cern.ch/category/10570/>

Meetings are announced on the [announcements@iris-hep.org](mailto:announcements@iris-hep.org) mailing list

Recorded videos are available in Youtube (see links on the individual agenda pages)

# Connecting with IRIS-HEP



Website: <http://iris-hep.org>

Public announcement mailing list: [announcements@iris-hep.org](mailto:announcements@iris-hep.org) [[Subscribe](#)]

Topical meetings: <https://indico.cern.ch/category/10570/>

We will be continuing to organize, co-organize and host various events going forward, see the main project website above.

# IRIS-HEP-related talks/events APS/DPF 2019



## Monday:

- IRIS-HEP Tutorial: Fast columnar data analysis with data science tools (Nick Smith, Jim Pivarski)

## Tuesday:

- COFFEA - Columnar Object Framework For Effective Analysis (Nick Smith)
- pyhf: a pure Python statistical fitting library from the high energy physics community with tensors and autograd (Giordon Stark)
- Extending RECAST for truth-level analysis reinterpretations (Alex Schuy)

## Wednesday:

- ACTS: a common track reconstruction software (Xiaocong Ai)

## Thursday:

- Evolving CMS offline computing towards LHC Run3 and HL-LHC (David Lange)

# Summary

HEP faces major challenges in the 2020s: Data, Compute, Staffing

The HSF executed an important community process that produced the CWP.

The collaborative spirit continues to grow and bodes well for us (as a community) to meet those challenges.

## IRIS-HEP

We are focusing on 3 R&D areas from the CWP: Innovative Algorithms, Analysis Systems, and DOMA.

Plus training, a dedicated integration activity and continuity for the OSG services for the LHC.

We are just beginning our activities and are looking forward to collaborating with many of you in the coming years!