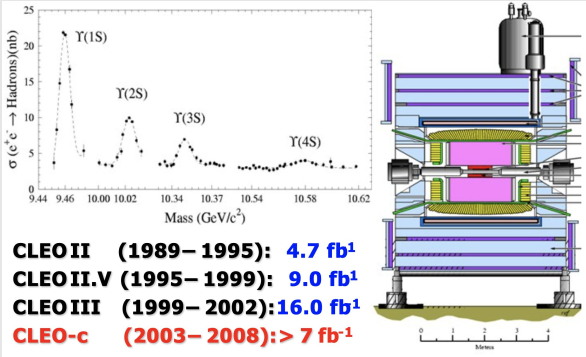
Introduction to the HSF-India Project

David Lange

Princeton University

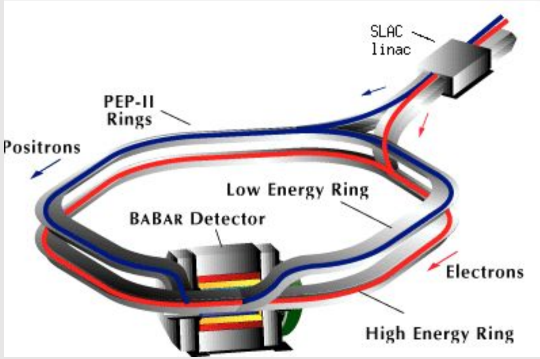
Introductions - David Lange

CLEO-II(.5) @ Cornell



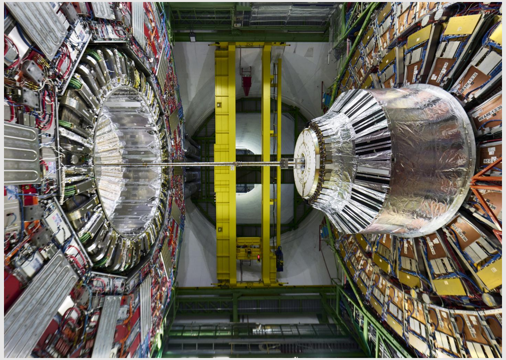
- Silicon detector calibration + operations
- Form-factor analysis in semileptonic decays
- Event generators

BABAR @ SLAC



- CP violation analysis
- RPC detector operations + software
- Event generators ("EvtGen")
- Event reconstruction software

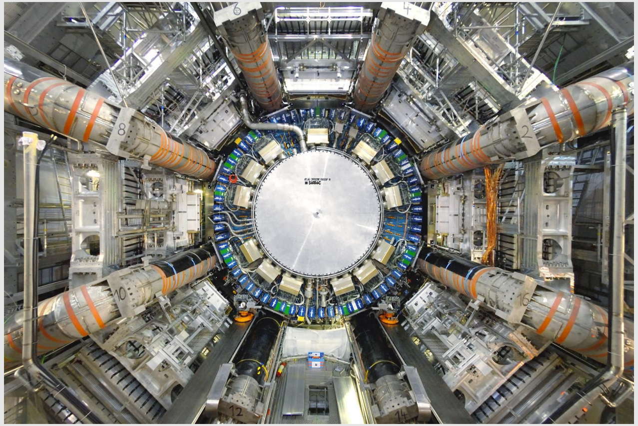
CMS @ CERN



- Event reconstruction software
- Simulation techniques
- Computing resource projections
- IRIS-HEP software institute

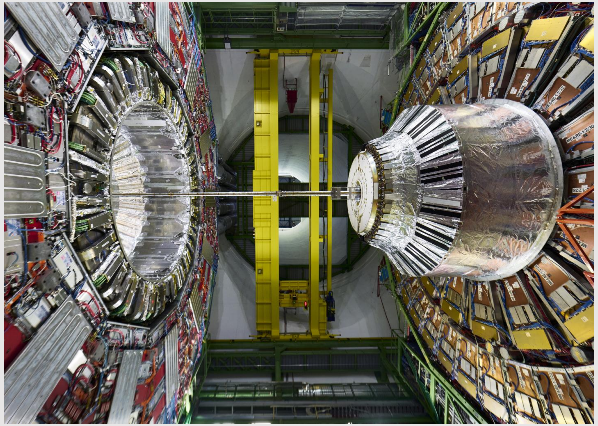
Introductions - Verena Martinez Outschoorn

ATLAS @ CERN



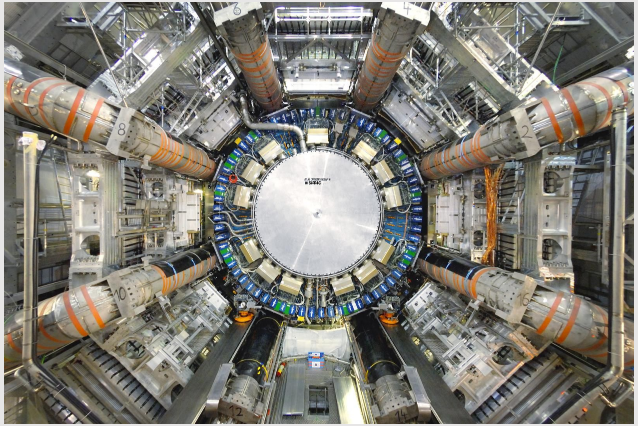
- Muon spectrometer construction and electronics
- Muon reconstruction software
- W boson cross section and charge asymmetry

CMS @ CERN



- Beam spot reconstruction
- Pixel detector design and construction
- Searches for stop pair production

ATLAS @ CERN



- Muon trigger algorithm development and electronics
- Muon software
- Analysis model and software
- Searches for exotic higgs boson decays

Introductions - Rafael Coelho Lopes de Sá

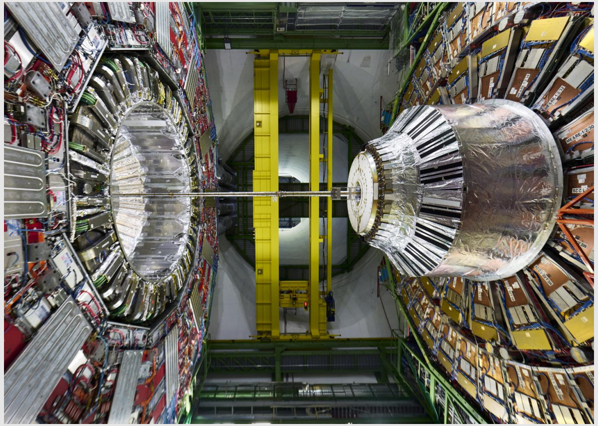
D0 @ FNAL

W mass measurement

EWK precision measurements

LAr Calorimeter Operations

CMS @ CERN



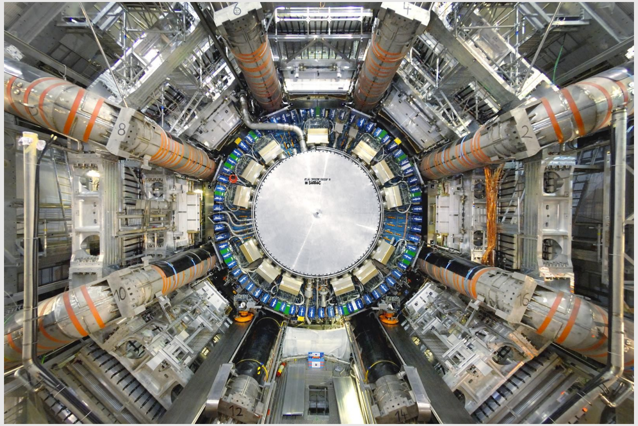
Evidence for electroweak production of WW pairs

Precision diboson cross-section measurements

Electron-photon calibration

Phase-2 Outer Tracker Mechanics

ATLAS @ CERN



Searches for exotic Higgs decays and new light scalar states

New *b*-tagging algorithms

Evidence for off-shell Higgs production

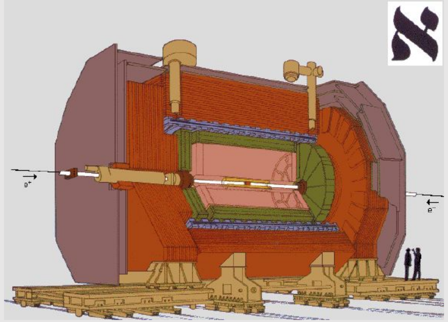
New analysis methods using machine learning

ITk Pixel Mechanics and Cooling

Computing operations

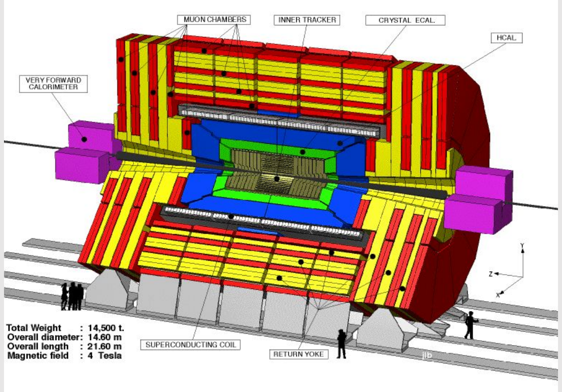
Introductions - Peter Elmer

Aleph@CERN



BaBar@SLAC

CMS@CERN



IRIS-HEP

iris hep
Institute for Research & Innovation in Software for High Energy Physics

Silicon Vertex/Tracker Detectors

Software and Computing

I am an experimental particle physicist (Princeton Physics since 2001, but based in Geneva, Switzerland) focused on computational and data science problems in my field, along with the software/computing systems to solve them.

Organizing the HEP community to address these challenges

The HEP Software Foundation facilitates cooperation and common efforts in High Energy Physics software and computing internationally.

- The HSF (http://hepsoftwarefoundation.org) was created in early 2015 as a means for organizing our community to address the software challenges of future projects such as the HL-HLC. The HSF has the following objectives:
  - Catalyze new common projects
  - Promote commonality and collaboration in new developments to make the most of limited resources
  - Provide a framework for attracting effort and support to S&C projects
  - Provide a structure to set priorities and goals for work in common projects

Community White Paper

January 2017
UCSD

June 2017
Annecy

A Roadmap for HEP Software and Computing R&D for the 2020s

[*Computing and Software for Big Science*](#) **volume 3**, Article 7 (2019)

"The result: a Programme of Work for the field as a whole, a multifaceted approach to addressing growing computing needs on the basis of existing or emerging hardware."

Eckhard Elsen (CERN Director of Research and Computing), editorial published with CWP/Roadmap

Many workshops, involving a diverse group

- International participants
- Computing Management from the Experiments and Labs
- Individuals interested in the problems
- Members of other compute intensive scientific endeavors
- Members of Industry
- http://s2i2-hep.org/
- https://hepsoftwarefoundation.org/

Individual Papers on the arXiv:

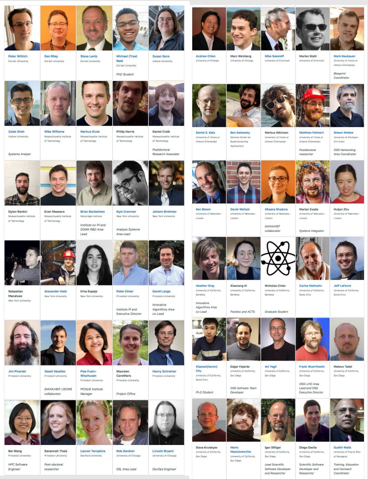
Careers & Training, Conditions Data, DOMA, Data Analysis & Interpretation, Data and Software Preservation, Detector Simulation, Event/Data Processing Frameworks, Facilities and Distributed Computing, Machine Learning, Physics Generators, Security, Software Development, Deployment, Validation, Software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

Community White Paper & the Strategic Plan → IRIS-HEP

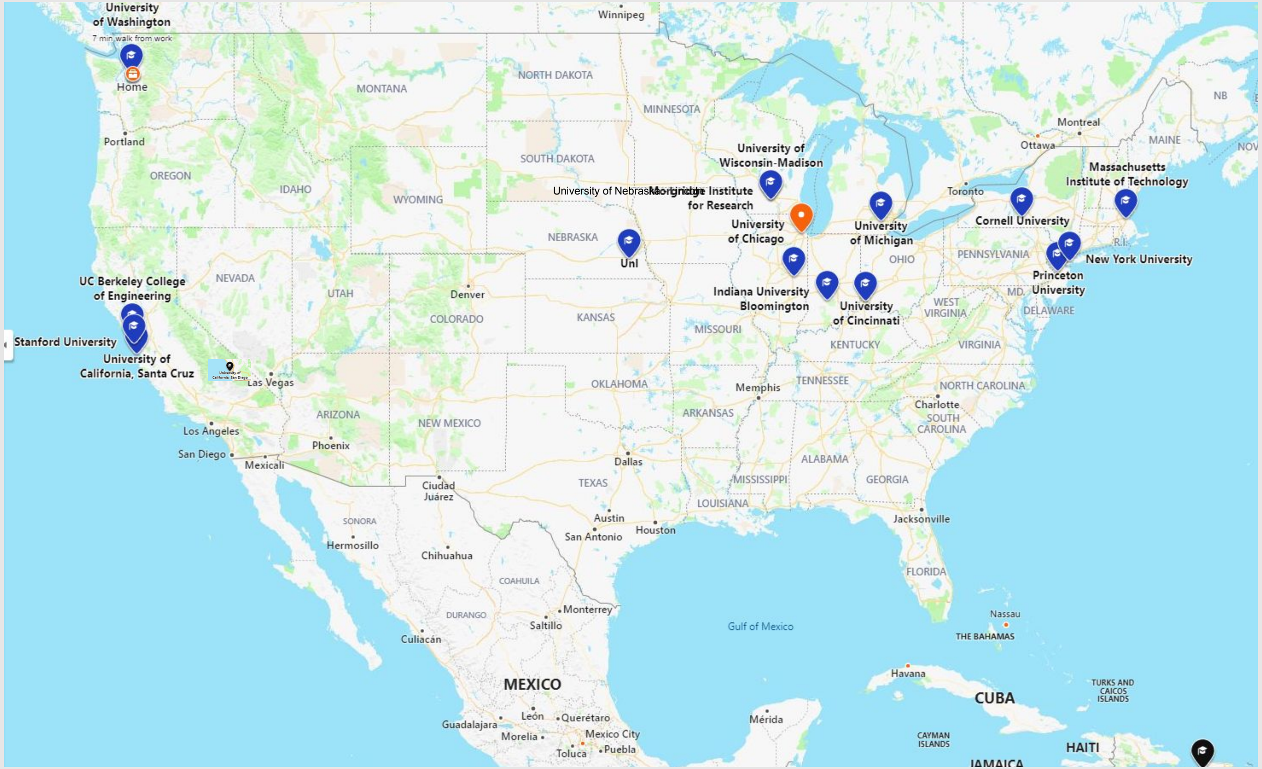
arXiv 1712.06982 arXiv 1712.06592

IRIS-HEP is one of the projects that grew out of the CWP. It is a Virtual/Distributed Institute

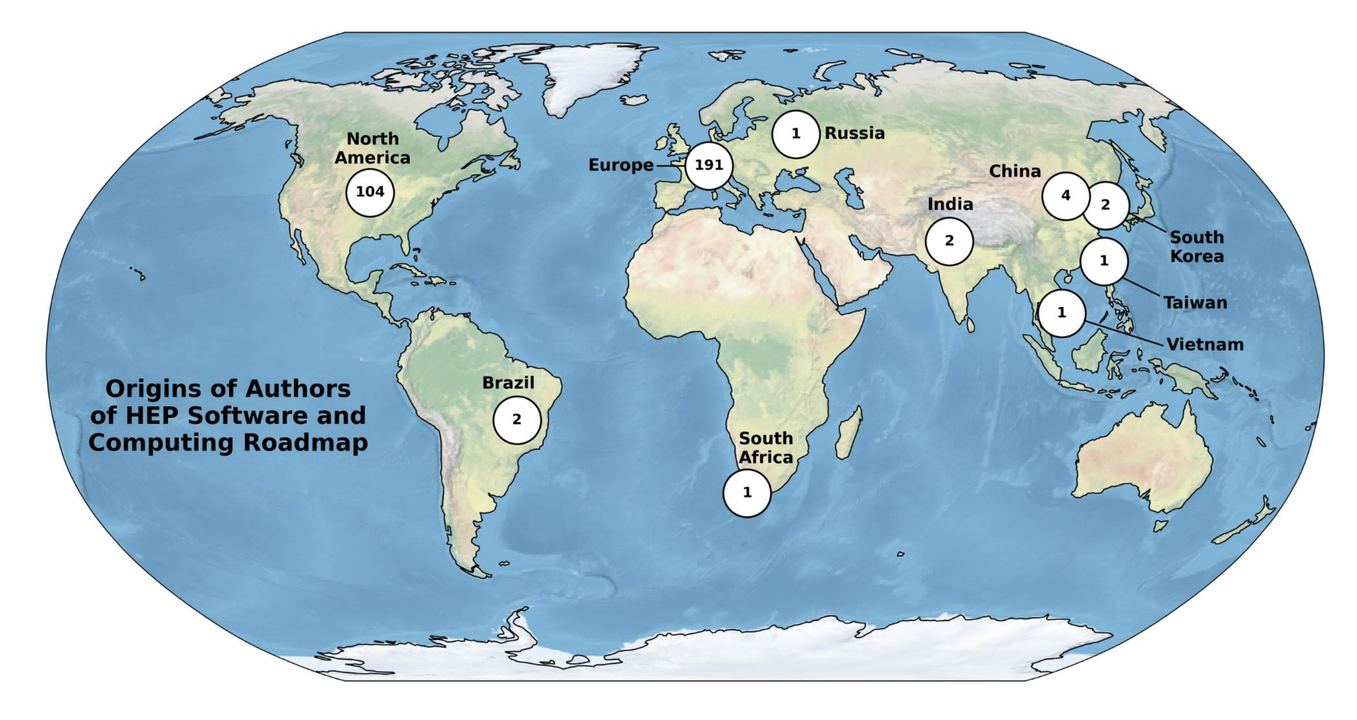
~30 FTE's distributed around the USA.



(many more but wouldn't fit here!)

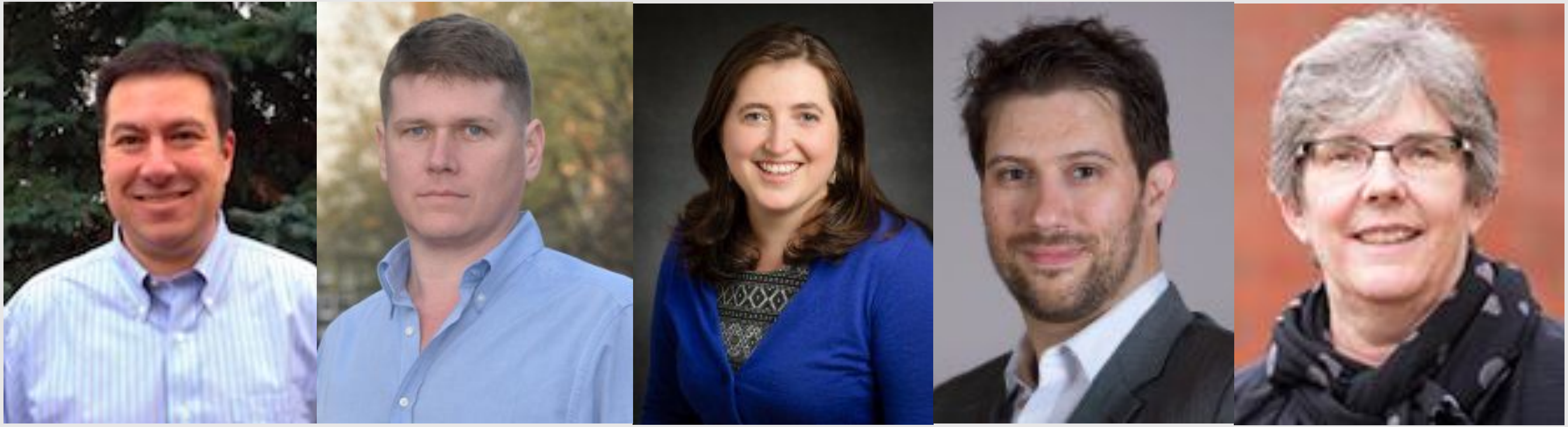


Observation: Nearly all authors of the HSF Community Roadmap where from institutions in Europe and the US



- Researchers participating in HEP experiments are more globally distributed than those that have joined HSF and other collaborative software projects

Team – We wrote a proposal to NSF aiming to broaden this community to include Asia – Starting with India



Princeton University: Peter Elmer, David Lange (PI)
University of Massachusetts, Amherst: Rafael Coelho Lopes de Sa, Verena Martinez Outschoorn
Oregon State University: Heidi Schellman
Primary contacts in India: Kajari Mazumdar, Brij Kishor Jashal

Project aims

HSF-India is a 5 year project that aims to build **international research software collaborations** between US, European, and India based researchers to reach the science goals of experimental particle, nuclear and astroparticle research.

Given the growing complexity of our scientific data and collaborations, these collaborations are increasingly important to raise the collective productivity of our research community.

Intended as a long-term investment in international team science.

What does our project fund?

This funding is different from a typical ‘’research’’ project. Much of our funding is to ***facilitate*** research collaborations, rather than directly fund a specific research activity - eg, around half of the funding is for paying for participant travel, fellowship (training) stipends, etc

One unfortunate restriction is that we can not fund travel that does not involve the US in some way (eg, origin, destination). We expect that CERN to be included as part of “US” given the large US-based researcher presence there.

Aiming for broad experimental and data/computer science engagement

Foster collaborations spanning physics, data science, and computer science to **exploit national capabilities and synergies** across groups

Sustain these **software clusters**, eg, cross-network collaborations, on specific software development projects across transitions where individual students move on and new students arrive.

Important outcomes include

- Cross-network research collaborations
- Foster STEM career outcomes (positive outcomes are not limited to academic progression)
- Students that progress to become mentors, instructors, etc..

Broad experimental engagement: Our proposal called out projects including: HL-LHC, DUNE, EIC, Xenon, DarkSide, INO, and future collider concepts (FCC-ee, ILC, Muon collider, etc). Other experiments with synergies are not excluded apriori

Planned Project Components

1. Topical meetings
2. Training in research software skills
3. Research project database
4. Bidirectional research exchanges (~10/year support for travel/subsistence)
5. Summer or semester student programs

As you might expect, programs will start small and progress as we build collaborations and understand what works and what does not work

Research Themes

We established three broad research themes as a basis for building collaborations.

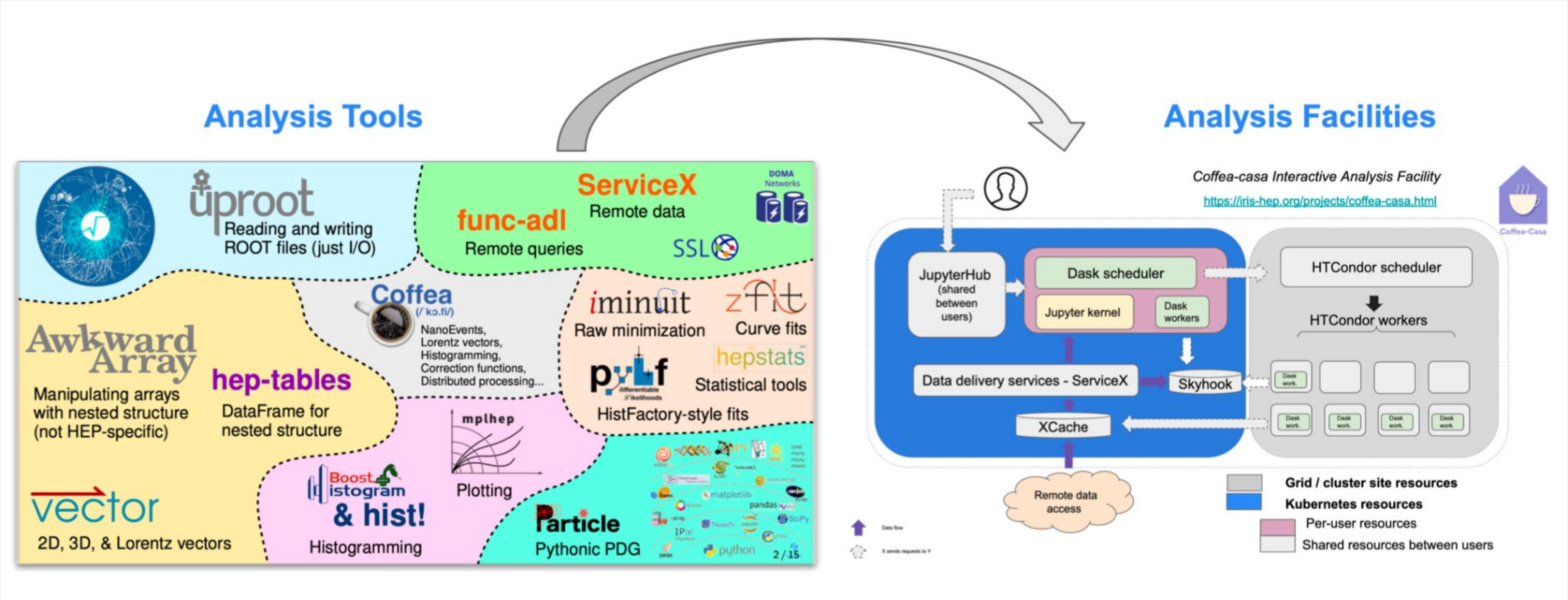
1. Analysis Systems
2. Simulation tools
3. Open Science

It is not explicit, but advanced ML techniques span across these areas..

Clearly these are each quite broad. We want to avoid being “too broad” to the extent that it would impact the ability to seed and then grow research collaborations.

Research Theme - Analysis Systems

To enable our data analysts to realize the maximum scientific potential of the data in the least time

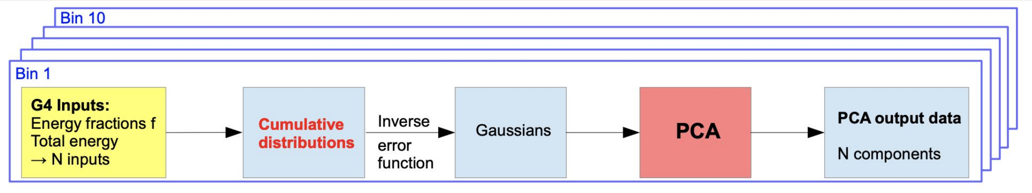


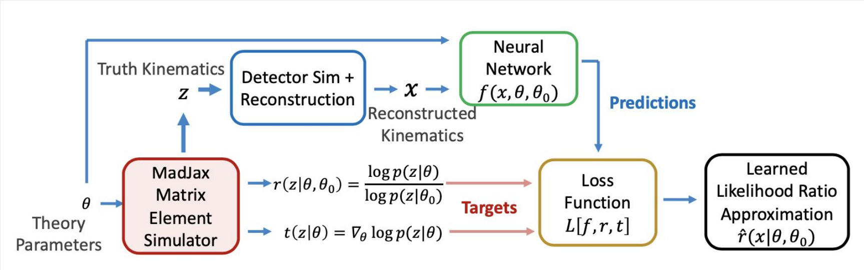
Research Theme - Simulation Tools

Monte Carlo simulation (event generators, interactions with detector material,etc) is a critical tool for the increasingly complex questions being asked by current and planned experiments.

Code and architecture modernization, physics process improvements, education, etc.

Fast simulation or fast generation approaches based on modern machine learning techniques





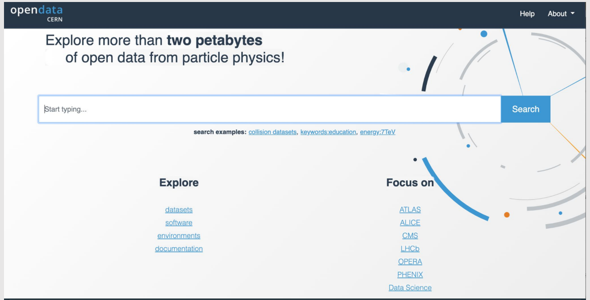
Novel simulation approaches (i.e., simulation based Inference, differentiable programming-based, etc)

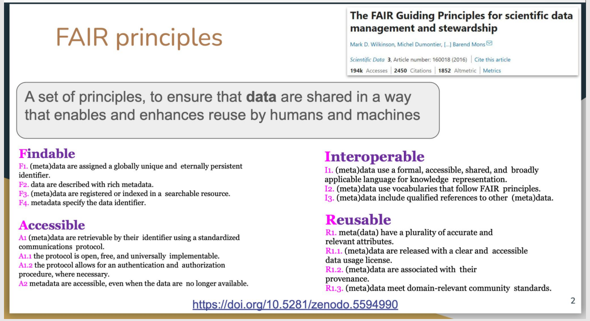
Research theme: Open data and open software

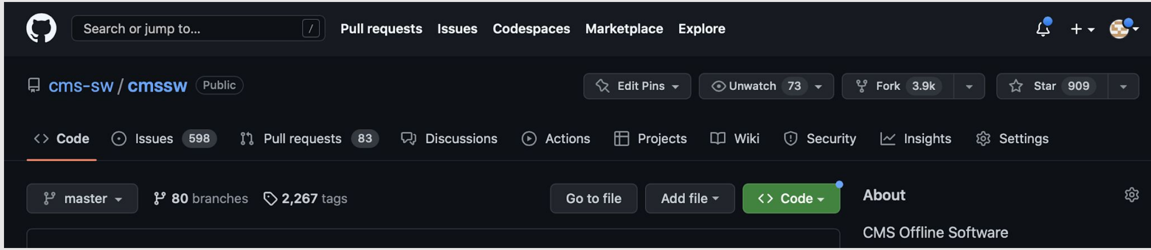
- Reproducibility
- Reusability
- Reinterpretability

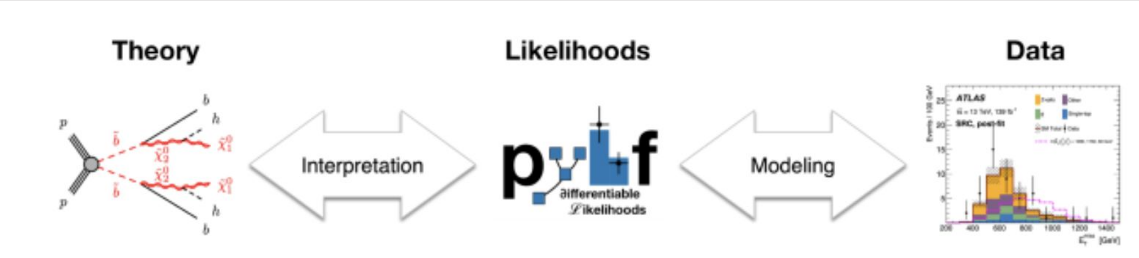
Intentional approach to production, analysis and publication processes fosters research across numerous scientific communities based on HEP our results

- Our community is understanding how best to do this, and what tools are needed to make this "easy" for HEP researchers









Likelihoods are an essential link between theory and ATLAS data (Image: K. Cranmer/ATLAS)

Training as the basis to build Research Software capability

Engage with existing HSF (+friends) training programs to build sustainable collaborations in research software.

Basics

The UNIX Shell
A guide through the basics of the file systems and the shell.
Start learning now!
Contribute!

Version controlling with git
Track code changes, undo mistakes, collaborate. This module is a must.
Start learning now!
Contribute!

Programming with python
Get started with an incredibly popular programming language.
Start learning now!
Contribute!

SSH
Introduction to the Secure Shell (SSH)
Status: Early development
Start learning now!
Contribute!

Machine learning
Get behind the buzzword and teach machines to work for you intelligently!
Start learning now!
Watch the videos!
Contribute!

Matplotlib for HEP
Make science prettier with beautiful plots!
Status: Beta testing
Start learning now!
Contribute!

ROOT
The most famous data analysis framework used in HEP.
Start learning now!
Contribute!

Software Development and Deployment

Version controlling with git
Track code changes, undo mistakes, collaborate. This module is a must.
Start learning now!
Contribute!

Advanced git
Learn to work with branches and more with this interactive webpage.
Start learning now!
Contribute!

CI/CD (gitlab)
Continuous integration and deployment with gitlab.
Start learning now!
Watch the videos!
Contribute!

CI/CD (github)
Continuous integration and deployment with github actions.
Start learning now!
Watch the videos!
Contribute!

Docker
Introduction to the docker container image system.
Start learning now!
Watch the videos!
Contribute!

Singularity
Introduction to containerization with Singularity/Apptainer.
Status: Early development
Start learning now!
Contribute!

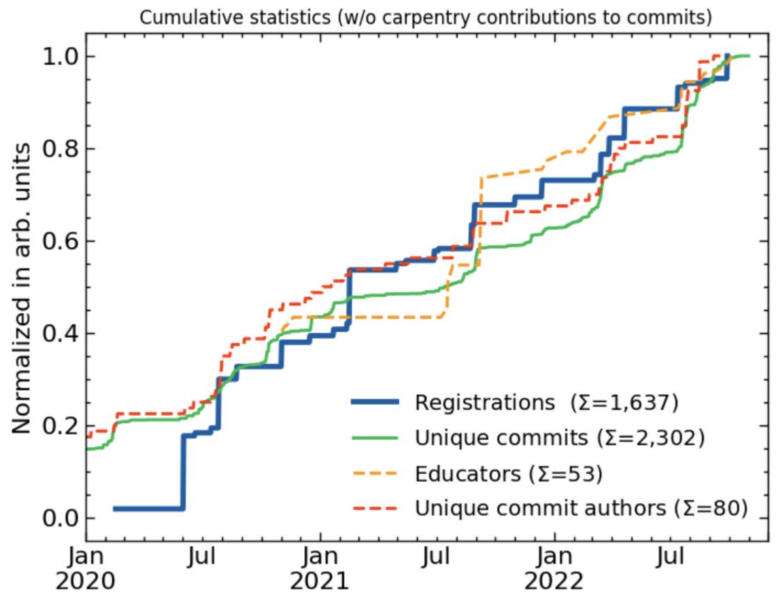
Unit testing
Unit testing in python.
Status: Beta testing
Start learning now!
Contribute!

Level up your python
Advanced bits of python (testing, debugging, logging, and more)
Start learning now!
Contribute!

C++ corner

HEP C++ Course
A full introduction to C++ based on a series of slides and exercises.
Start learning now!
Watch the videos!

Build systems: cmake
Building code is hard. CMake makes it easier.
Start learning now!



The number of commits in our repository and the number of registered learners closely follows the number of our educators. You, too, can make a difference!

Research opportunities - Summer program and Fellows

Many research software groups have benefited tremendously from Google Summer of Code (GSoC) and other short-term (eg, summer) research-at-a-distance programs.

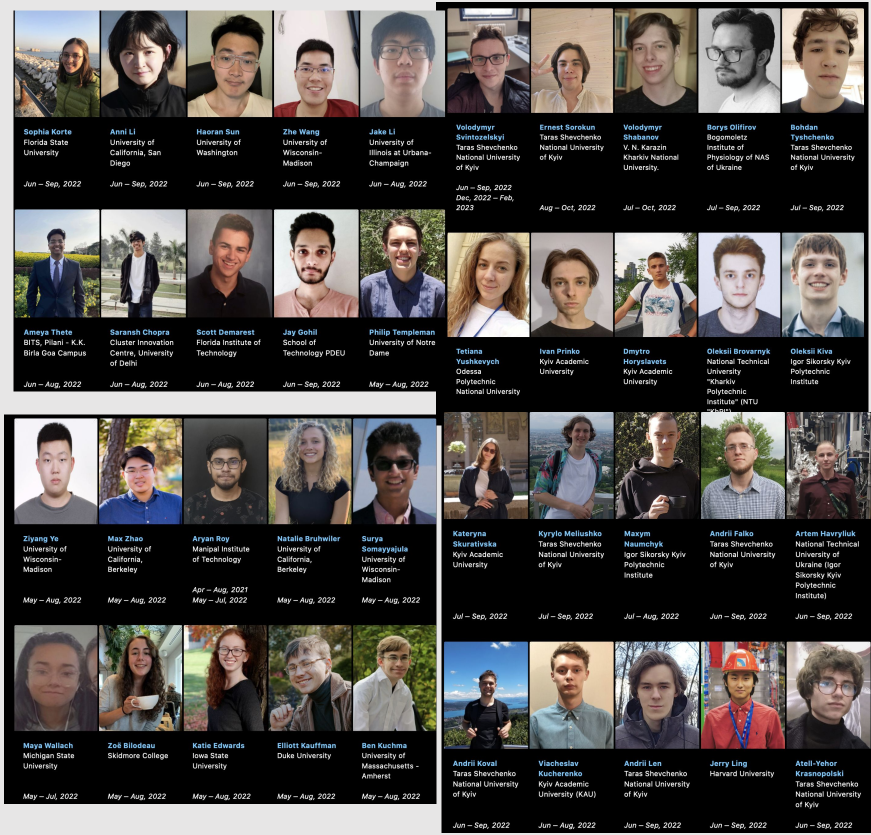
Important difference w.r.t programs like GSoC : We aim to connect research groups not simply have one-off projects. To facilitate this, HSF-India fellow projects will follow a co-mentorship model (eg, US-based mentor + India-based mentor)

We recognize that boot-strapping these fellowships will not be simple at first. If HSF-India is successful, this will become easier as the project matures.

IRIS-HEP Fellows program

https://iris-hep.org/fellows.html

- **Key Insight:** we need to provide incentivized and explicit paths forward for enthusiastic students from the more advanced training schools in HEP (ESC/Bertinoro, CoDaS-HEP, MLHEP, etc.)
  or for people who become engaged with our software projects in other ways.
- **Project focused:** bring students into contact with “mentors” to work on a specific, pre-defined project, allowing them to grow their software skills and experience working in large projects.



Research opportunities – Researcher Exchange

Months long in person mentorships or co-development opportunities. Focused on advanced graduate students, postdocs, early-stage faculty

Bidirectional (Omnidirectional?) program including travel support and researcher subsistence to support for local costs

As with fellows, the idea is to use researcher exchanges as a mechanism to build collaborations between groups, not just individuals.

This is a 5 year project that is just starting

Our goal for this visit is to discuss with interested research groups about how our program could be structured to support your research goals

- General attributes of fellow and research exchange programs that would be of use to your (and our) community
- Specific research ideas, existing projects with common interest to build upon, etc.
- Establish communication channels
- Identify opportunities for training events (eg, ways to co-locate with existing events)