Building Research Software Collaborations

Peter Elmer (Peter.Elmer@cern.ch)

PRINCETON UNIVERSITY

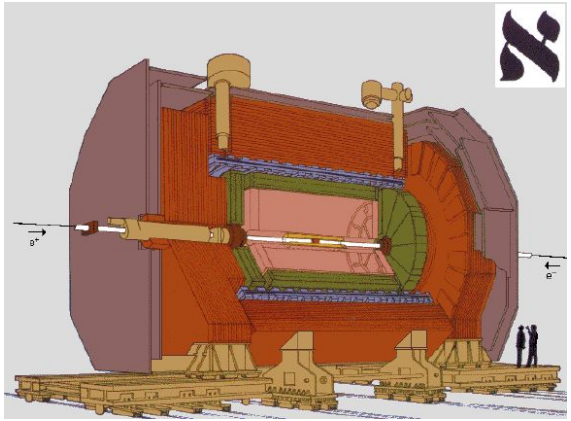
HSF-India Workshop at U.Delhi - 23 May, 2024

Introductions…. Who am I?

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.

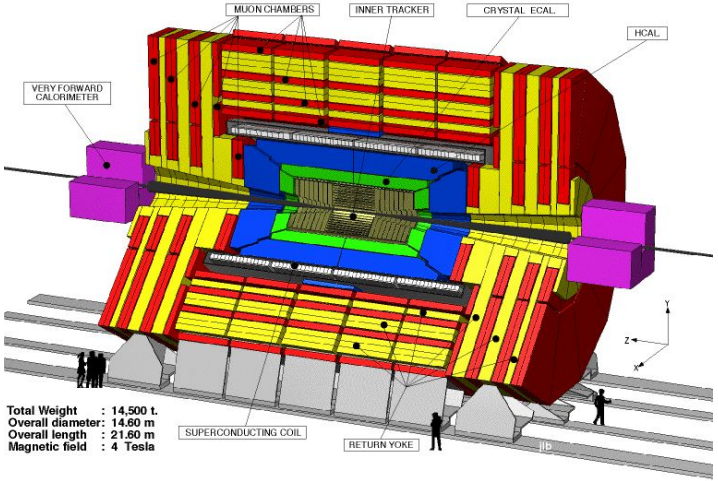
Researchers in experimental particle physics tend to introduce themselves to each other with reference to the series of experiments with which they have collaborated. So here is my own version of that:

Aleph@CERN

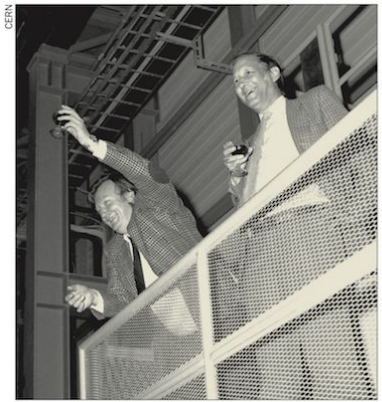


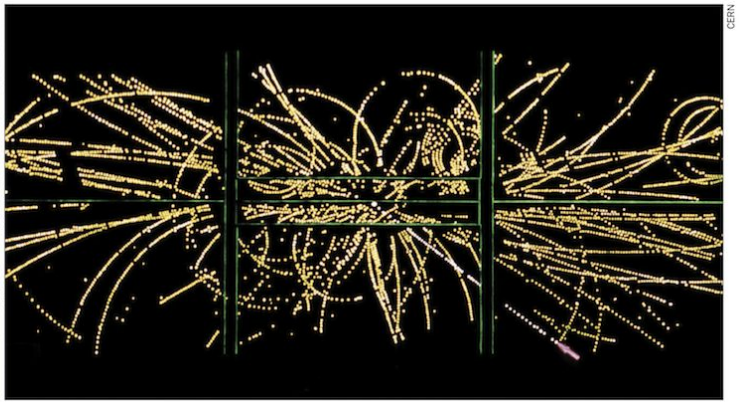
BaBar@SLAC

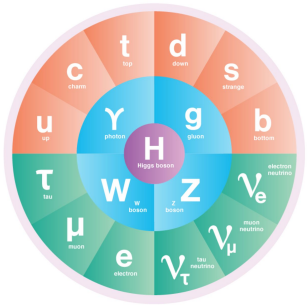
CMS@CERN

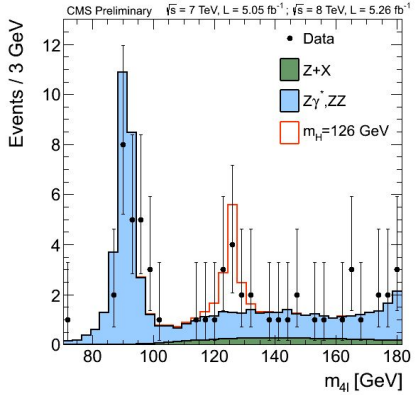


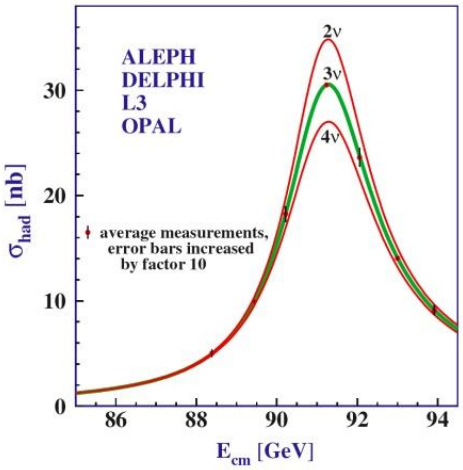
Experimental Development of the Standard Model

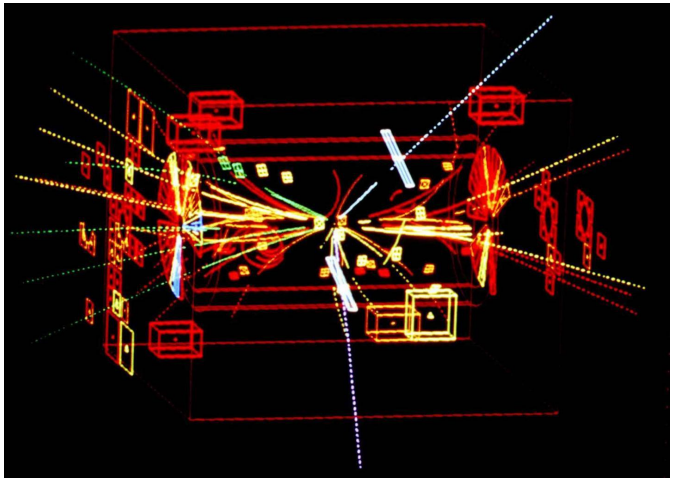


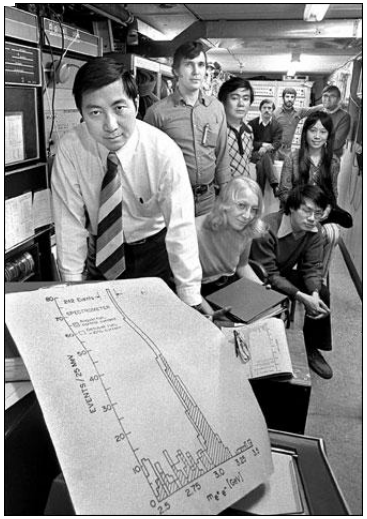




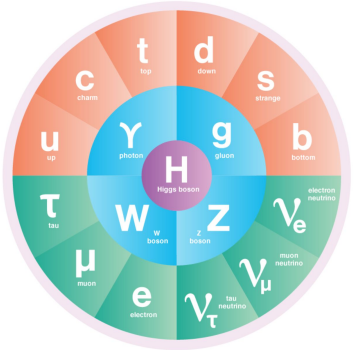




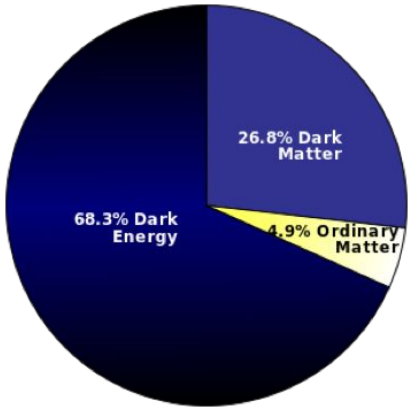


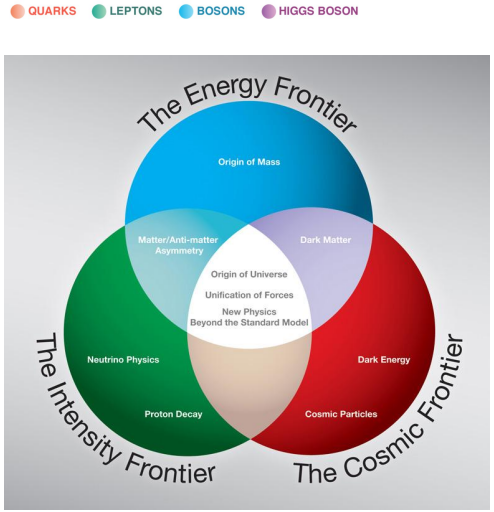


Science Drivers - Beyond the Standard Model of Particle Physics



While the Standard Model of Particle Physics describes, often with incredible precision, the vast majority of experimental observations, it is known to be incomplete. It does not (for example) include gravity, and it does not explain neutrino masses, the matter-antimatter asymmetry or dark matter/energy.



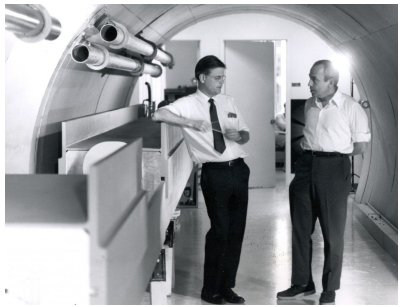


From “Building for Discovery - Strategic Plan for U.S. Particle Physics in the Global Context” - Report of the Particle Physics Project Prioritization Panel (P5):

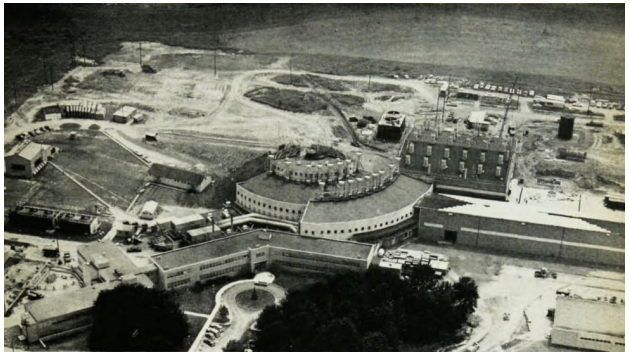
1) Use the Higgs boson as a new tool for discovery
2) Pursue the physics associated with neutrino mass
3) Identify the new physics of dark matter
4) Understand cosmic acceleration: dark energy and inflation
5) Explore the unknown: new particles, interactions, and physical principles

Big Science and Accelerators - larger and larger facilities

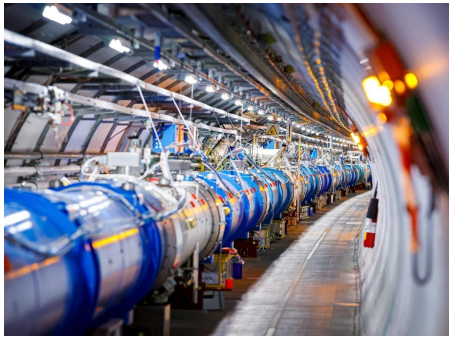
E.O. Lawrence and the cyclotron
at U.C. Berkeley



The birth of the “National Accelerator
Laboratory” in the US (now Fermilab)

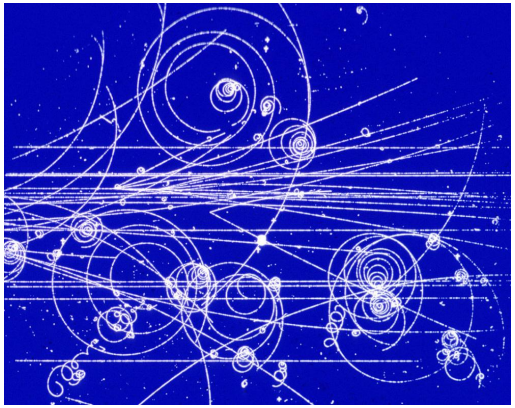


The Princeton-Pennsylvania Accelerator,
Milton White 1964 Physics Today 17(8): 27



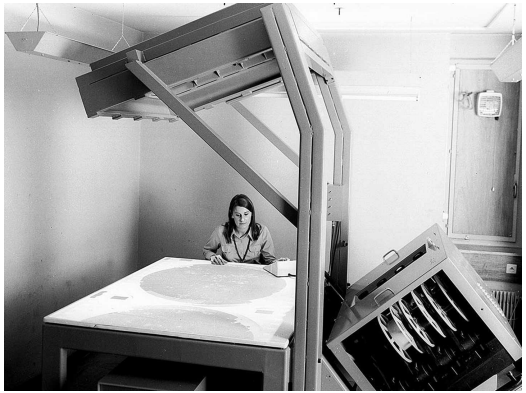
The Large Hadron Collider and
CERN first as a European laboratory
and then as a “world laboratory”

Instrumentation - Detectors (+ Electronics)

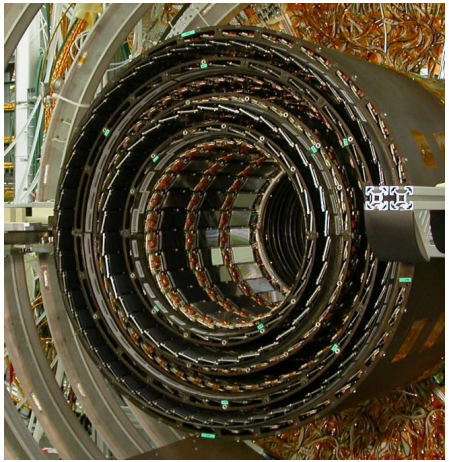


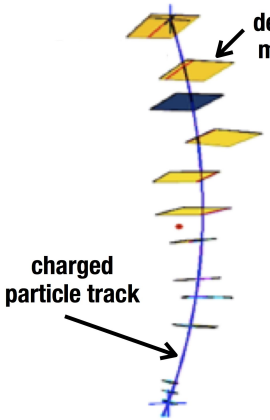
Bubble chamber photography

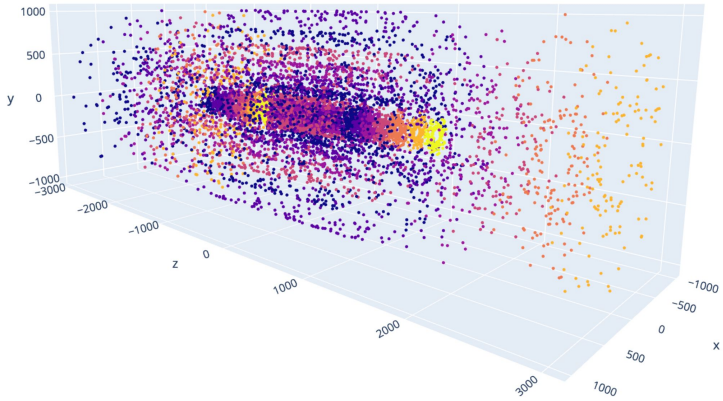
By-hand “scanning” of the photos



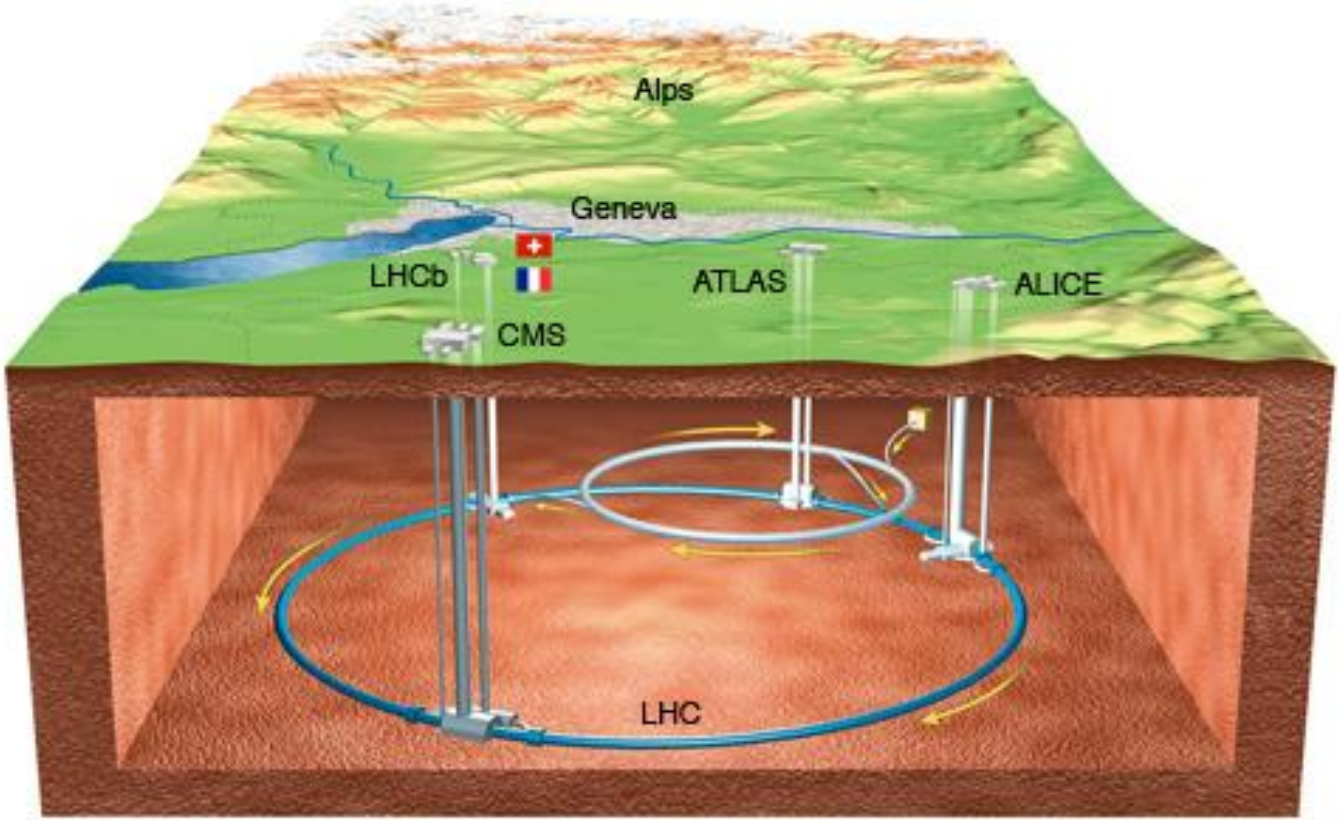
Modern detectors with (digital) electronic readout

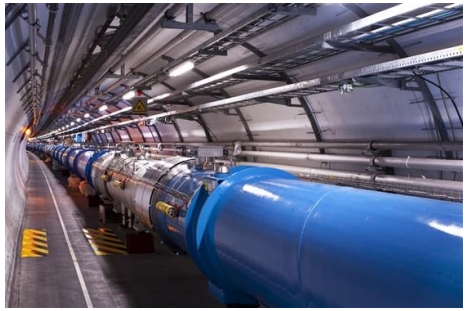






The Large Hadron Collider at CERN





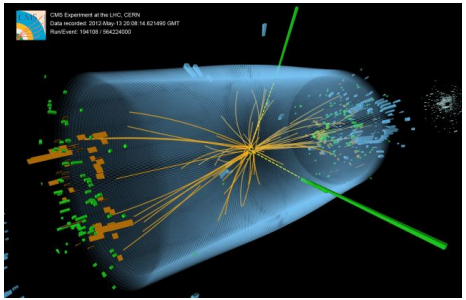
27 km circumference tunnel, with larger caverns at points around the ring for detectors

100-150m underground

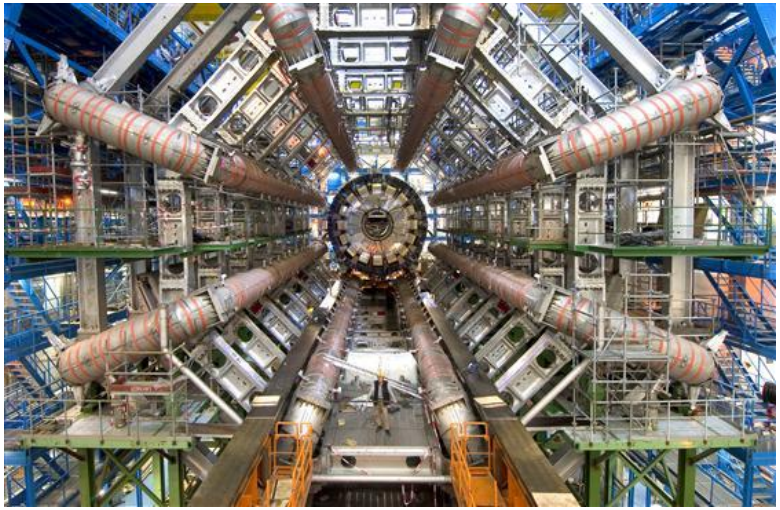
Built in the 1980s for a previous collider (LEP) and expanded with new caverns and access in the early 2000s

Highest energy collider (design 14 TeV, operating at 13TeV) currently available

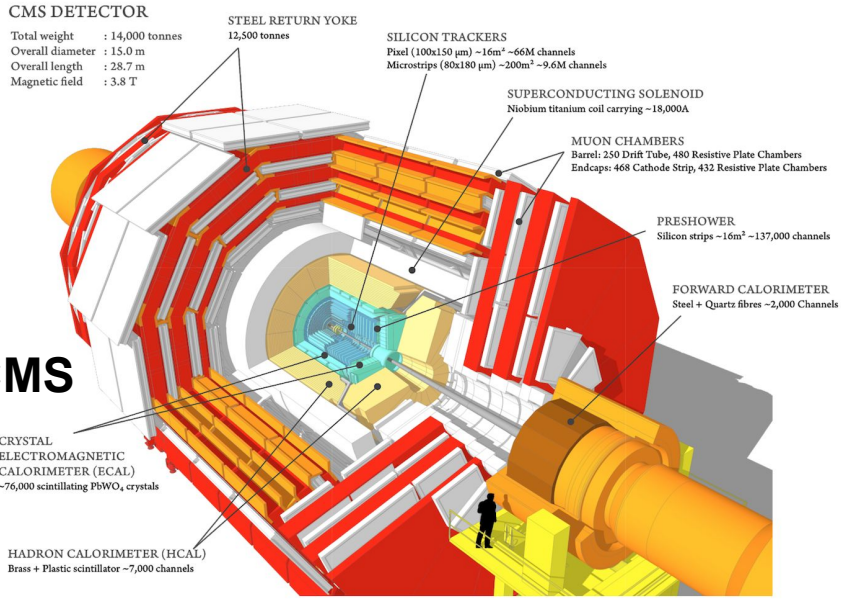
Large Hadron Collider Experiments Are Massive Data Generators



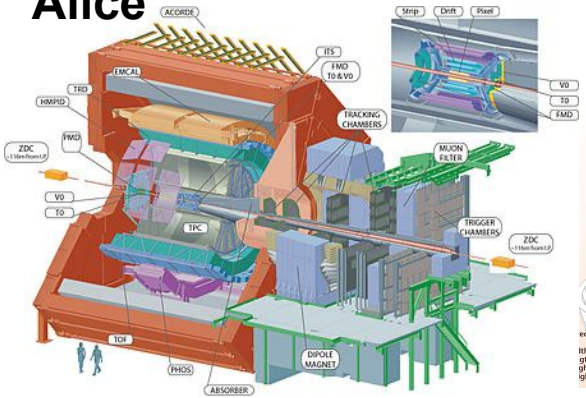
Atlas

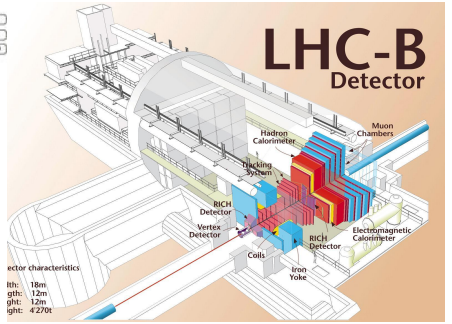


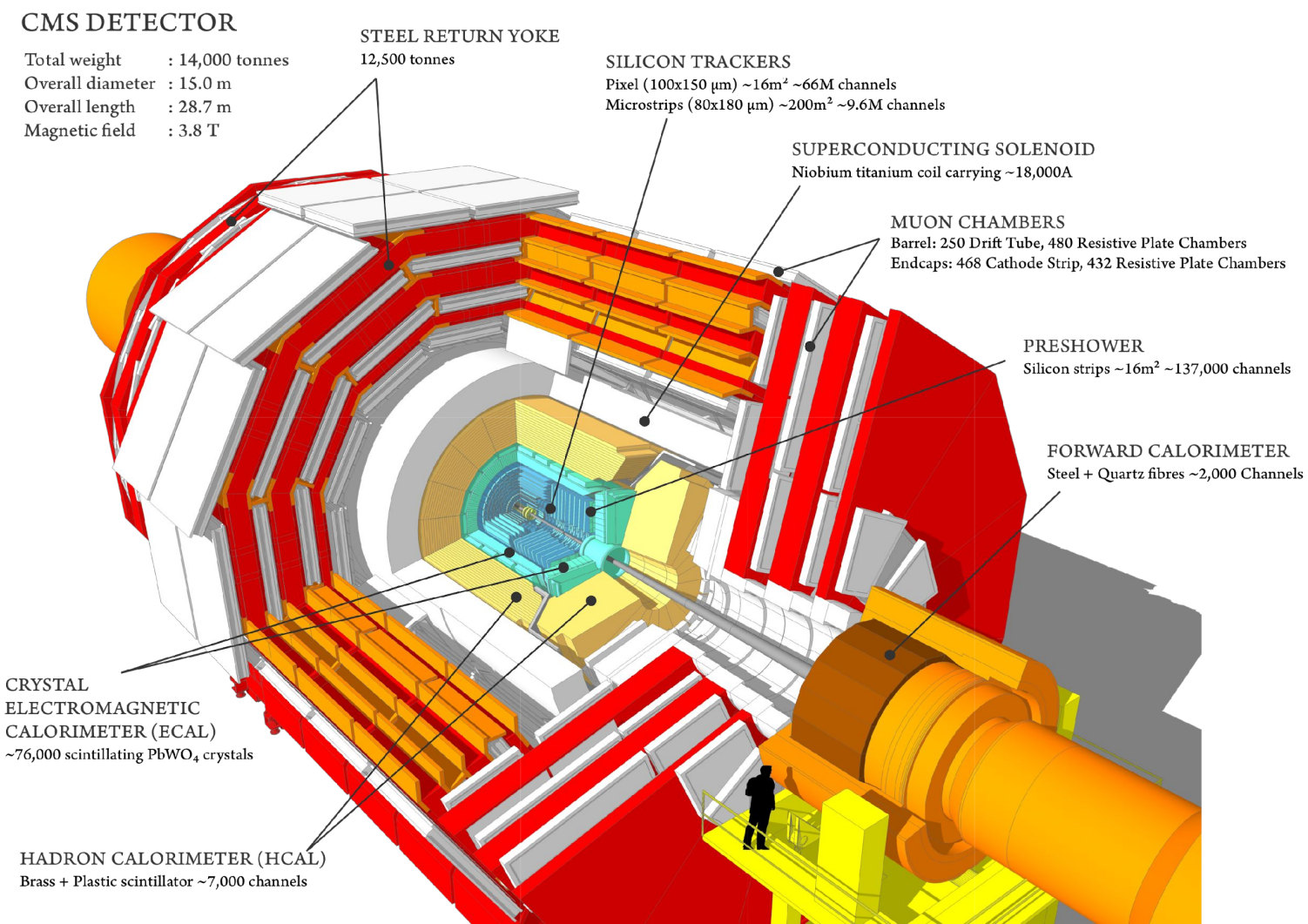
CMS

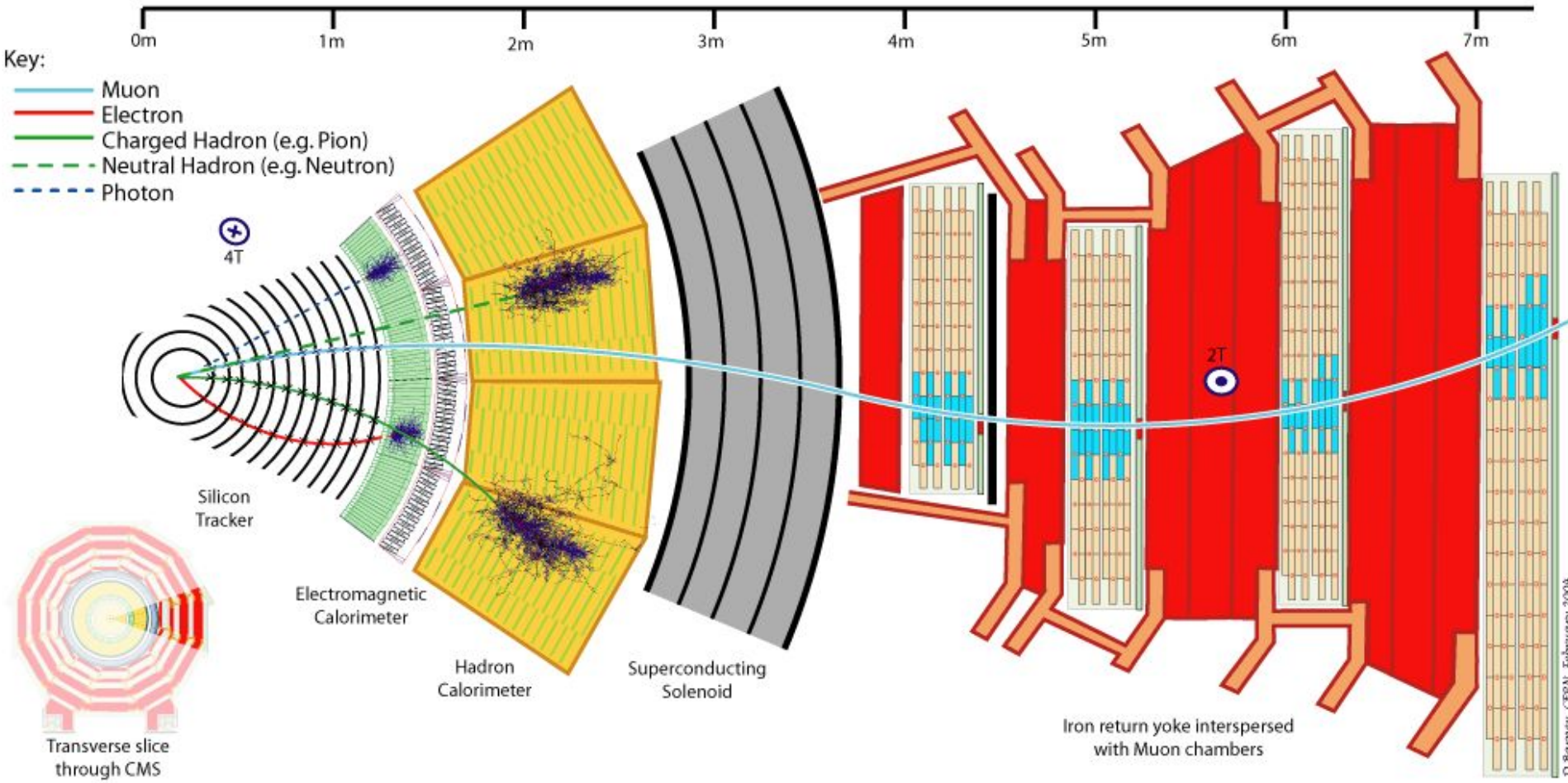


Alice





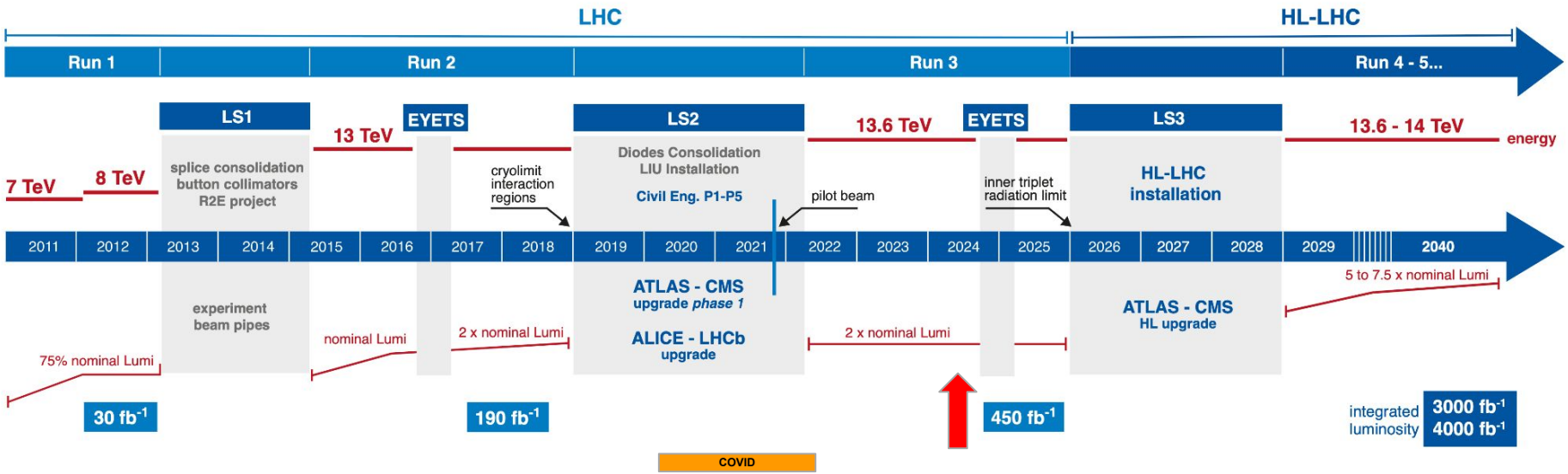

CMS DETECTOR
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels
SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers
PRESHOWER
Silicon strips ~16m² ~137,000 channels
FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels
CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76,000 scintillating $PbWO_4$ crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels


0m
1m
2m
3m
4m
5m
6m
7m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
Transverse slice through CMS
D.Barney, CERN, February 2004

CMS Collaboration - 2013

(January, 2022 version)

LHC / HL-LHC Plan



S2I2-HEP

IRIS-HEP Institute	IRIS-HEP Institute
Design	Execution

Institute Conceptualization and Community White Paper Process

Snowmass

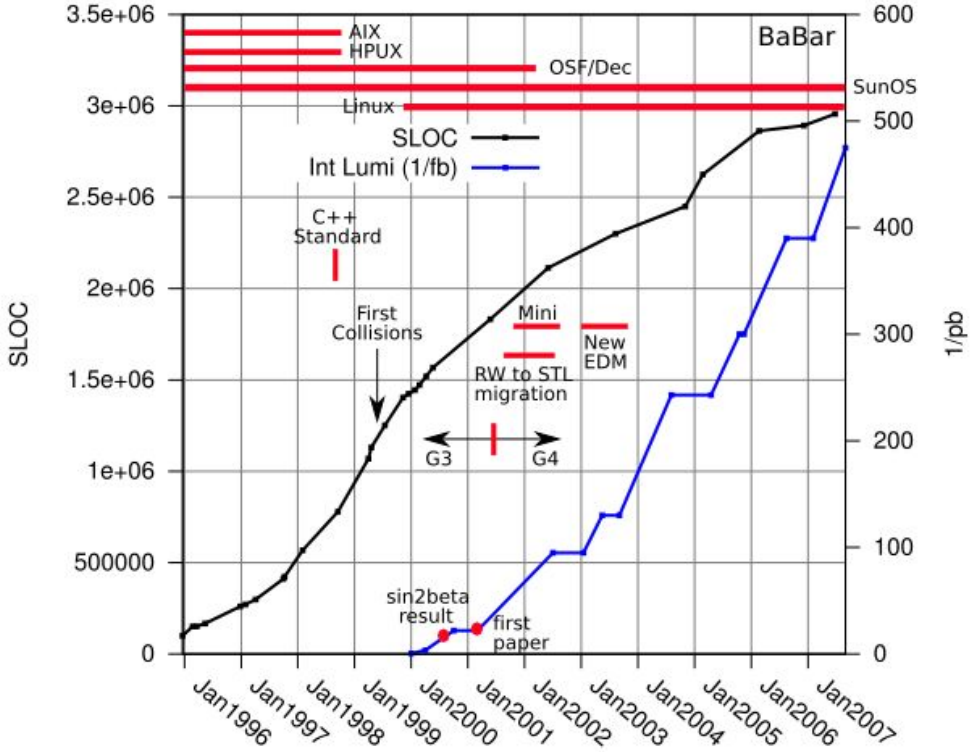
U.S. HEP Community Planning Process

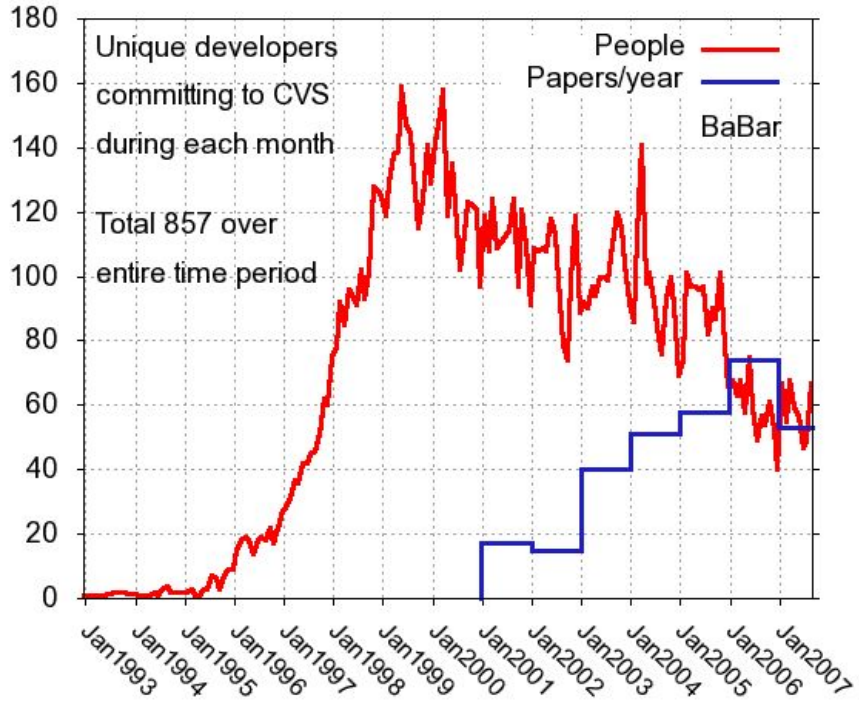

And where does computing come into this?
Cyberinfrastructure?

Computing hardware is a *consumable.*

Research Software is the *true* cyberinfrastructure.

Software of the BaBar experiment at SLAC





But what does all of this software do?

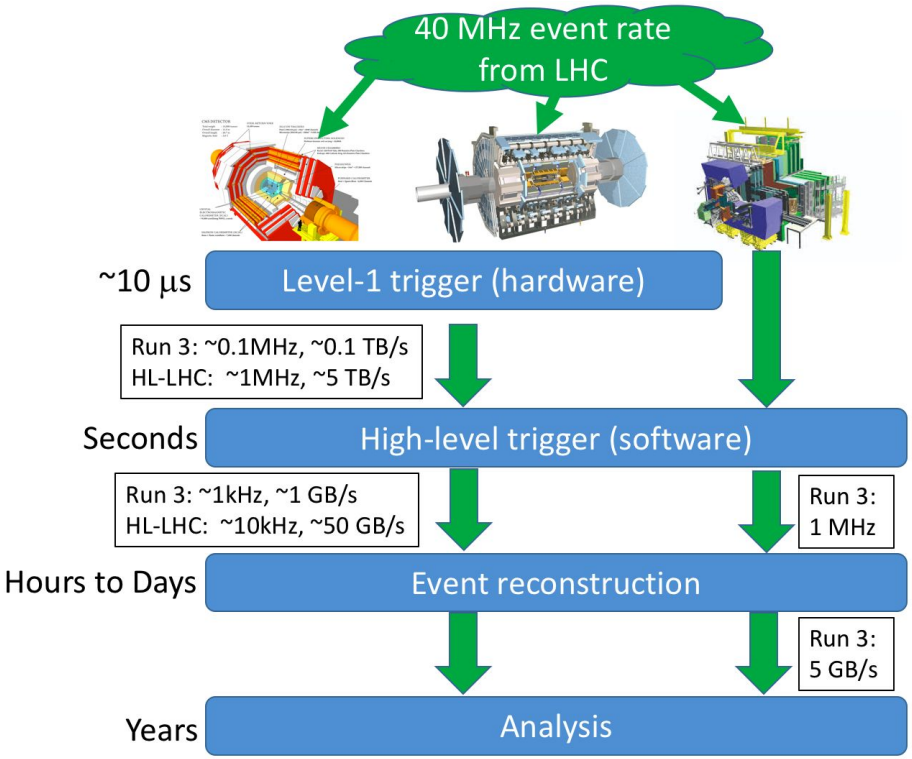
“Trigger” - a real-time/pseudo-real-time filter

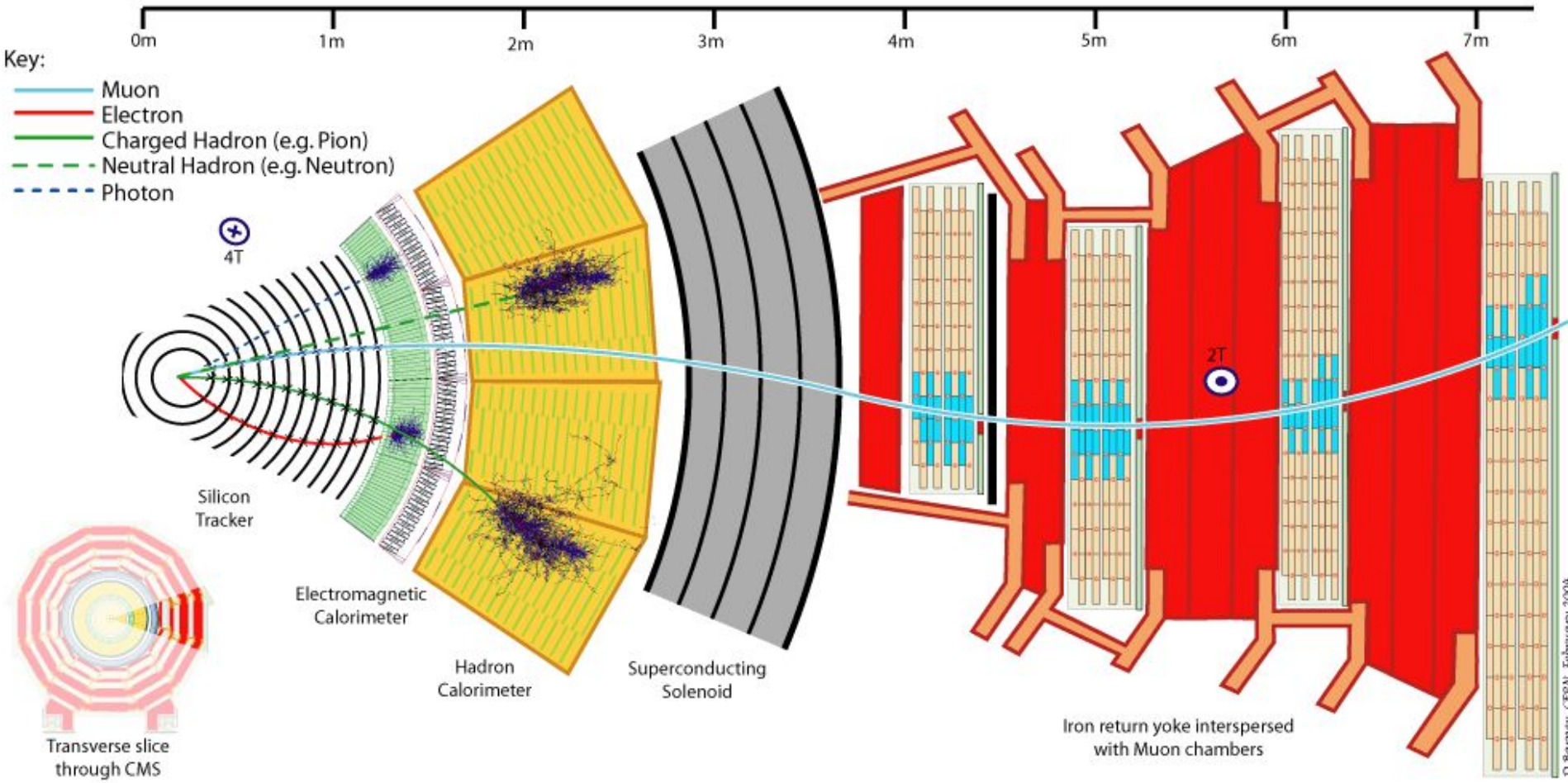
The LHC collides proton bunches at 40MHz, i.e. a bunch crossing every 25ns

Common codebases are often used for data reduction (high level trigger) and detailed offline processing (reconstruction) of detector data and simulation

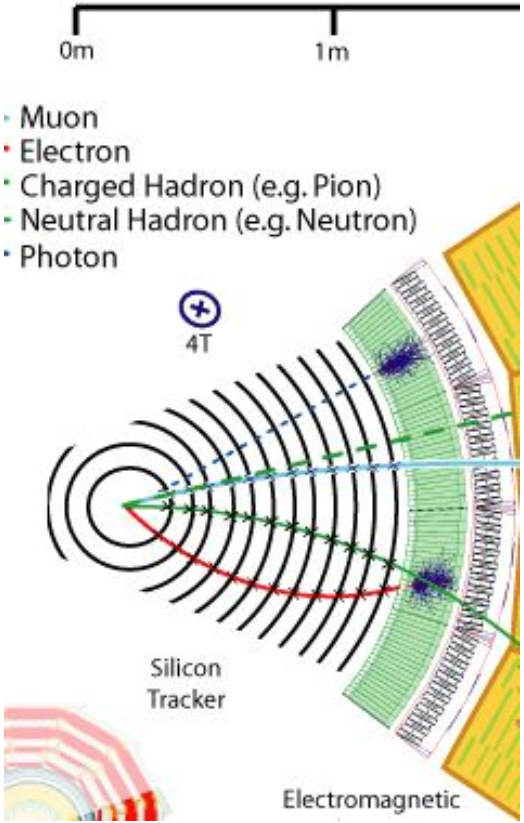
Each consists of numerous algorithms, primarily developed by HEP research community researchers with varying technical skill sets.

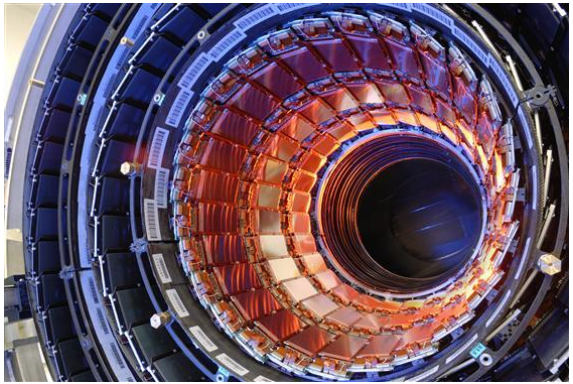
Software implemented in FPGAs for Level-1 Trigger and CPUs/GPUs for the High Level Trigger.

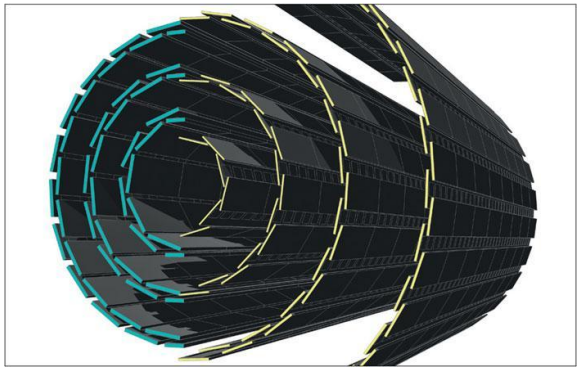


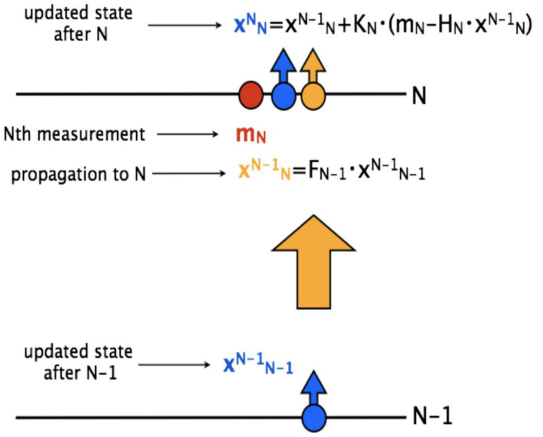

0m
1m
2m
3m
4m
5m
6m
7m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
Transverse slice through CMS
D.Barney, CERN, February 2004

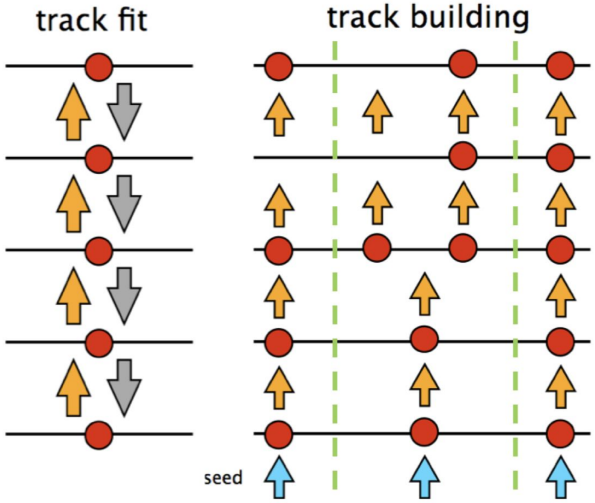
“Reconstruction”



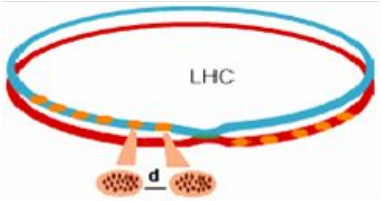


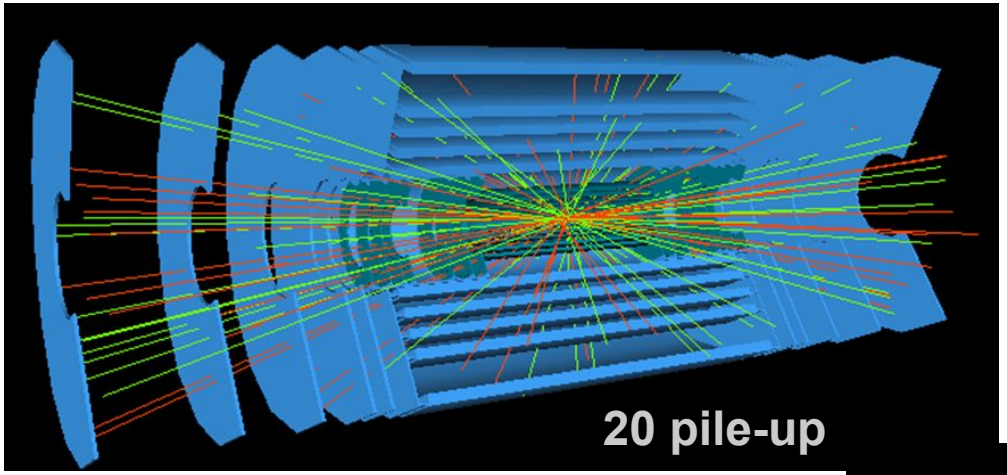




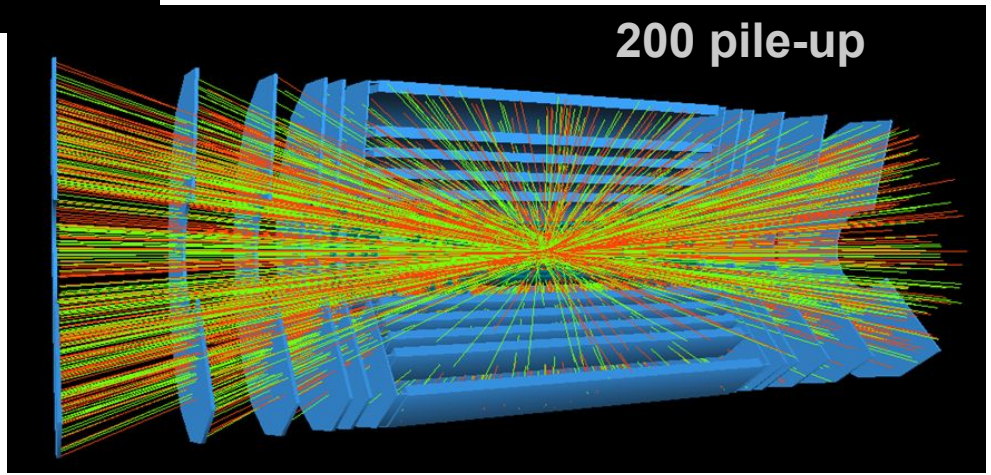


The HL-LHC Challenge (2029)





Multiple proton-proton interactions per beam bunch crossing ("pile-up") as seen in a simulation of the ATLAS Inner Tracker

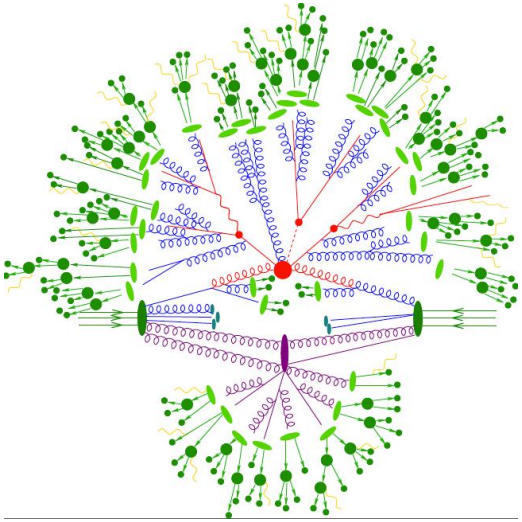


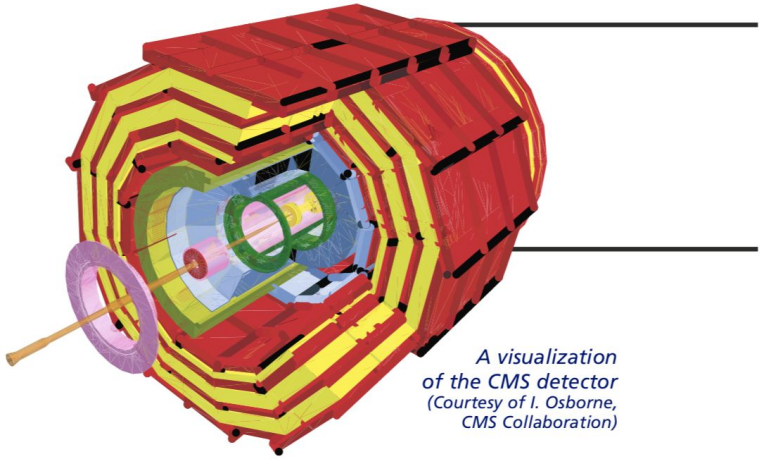
Higher probability of an interesting interaction, but with consequences: detectors/electronics need to handle the higher rate, higher radiation dose and significantly more complex events

Monte Carlo Simulation

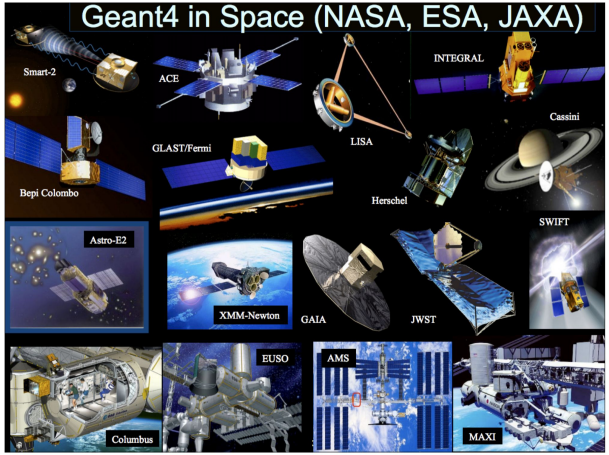
We use a Monte Carlo simulation of the underlying physics of the collisions, then the passage of the resulting long-lived particles through the detector apparatus. This includes the best approximations of all of the micro-physics associated to the electromagnetic and nuclear interactions of the particles with the material in the detectors across a broad range of energies.

Geant4 (https://geant4.web.cern.ch) is a toolkit for the simulation of the passage of particles through matter. Its areas of application include high energy, nuclear and accelerator physics, as well as studies in medical and space science.

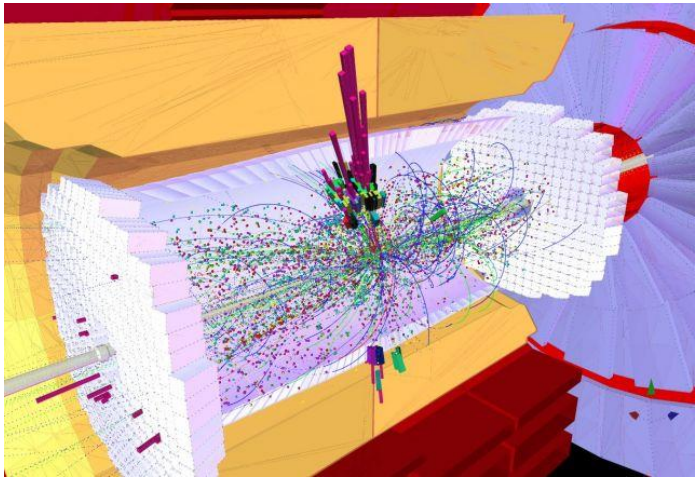




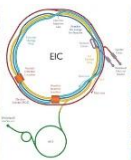
A visualization of the CMS detector (Courtesy of I. Osborne, CMS Collaboration)

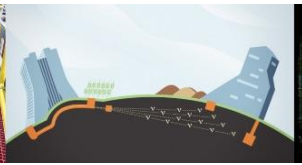


Software as an infrastructure and as an intellectual product



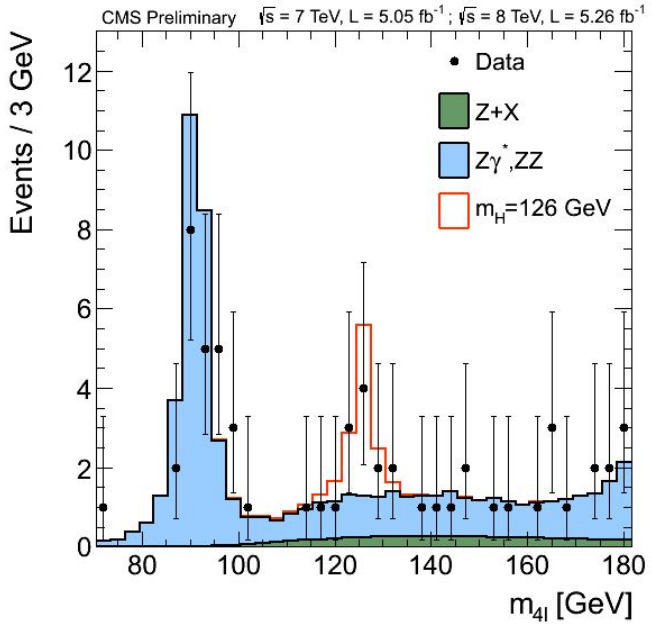
GEANT4
A SIMULATION TOOLKIT



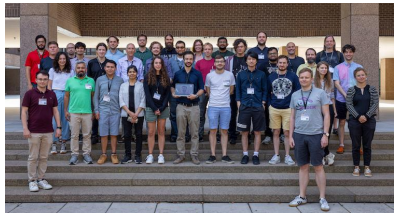
Data Analysis - “Data Science” in HEP



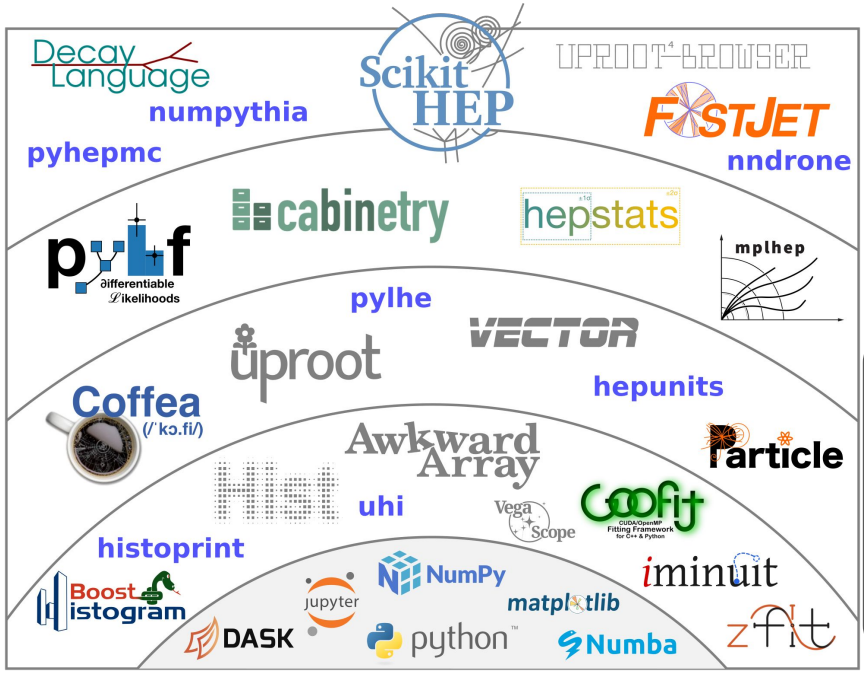
Python - not just a programming language useful for this kind of activity (see talk of Jim), but also a “lingua franca” with the larger data science and machine learning world….

PyHEP Ecosystem (Scikit-HEP)

[PyHEP.dev 2023 workshop](#)
40 participants



Our scientific Python development vision/ecosystem



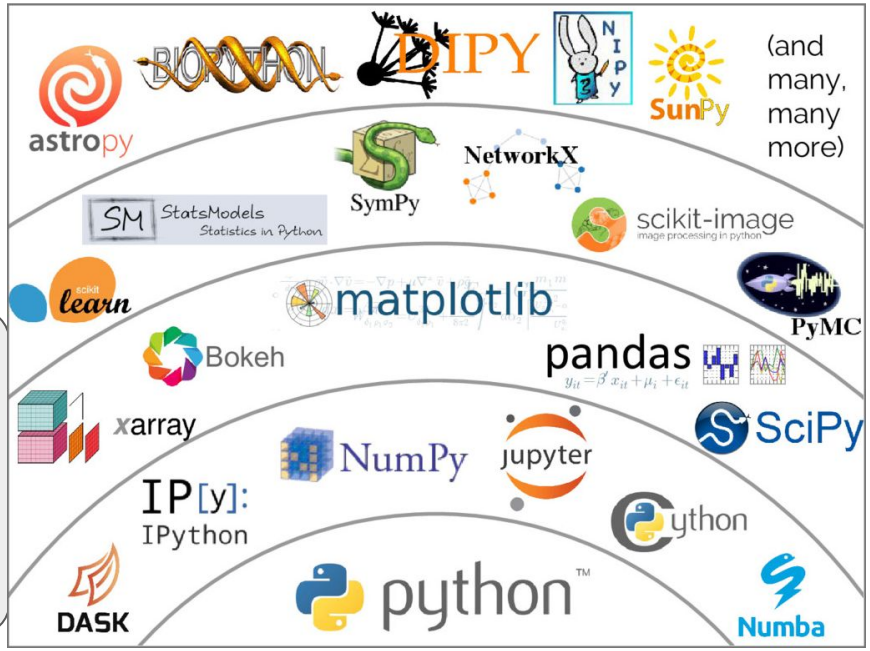
Application Specific

Domain Specific

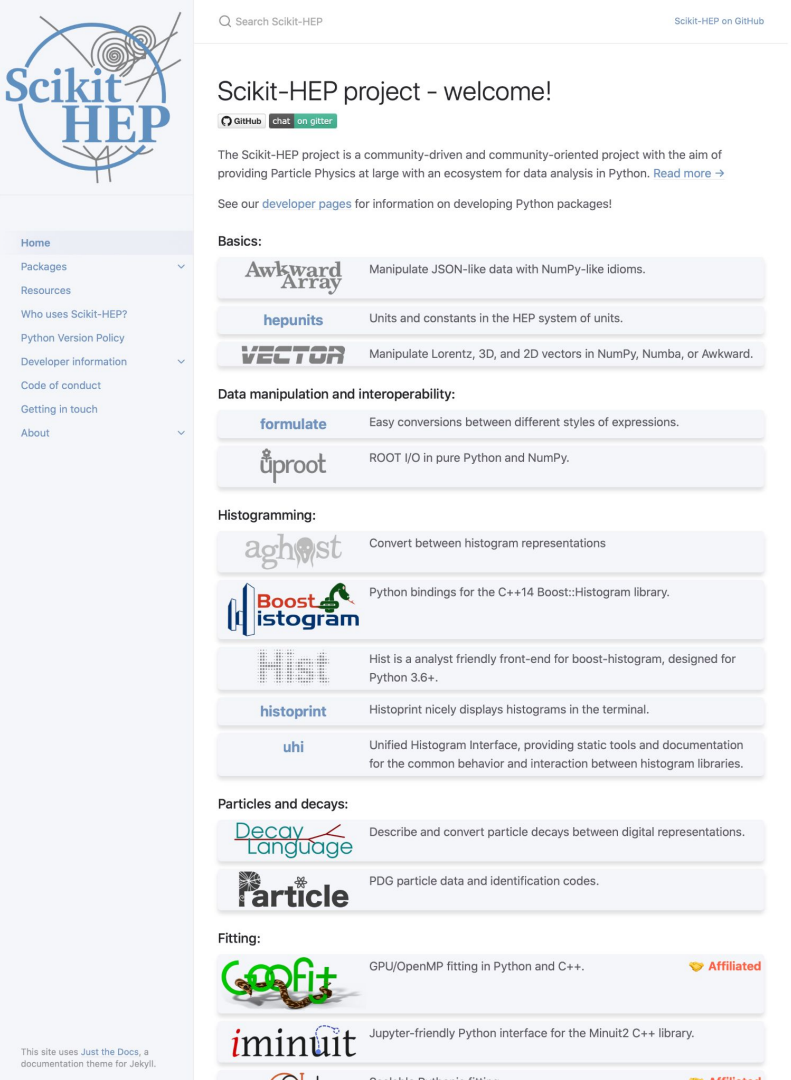
Technique Specific

Foundational

Scientific Python / PyData vision/ecosystem



Jake Vanderplas 2017 keynote-style Python ecosystem, but with layer labeling a la "Array programming with NumPy", *Nature* , 585, 357-362 (2020)



Scikit-HEP

Umbrella for community collaborations for high energy physics (HEP) in the python ecosystem. Inspired in part by highly successful AstroPy ecosystem in the astronomy community

An open ecosystem of tools, not a single framework, with a role as a focal point for development of tools outside of any given single experiment. Python is not just a “programming language”, but an opportunity for connecting to the larger scientific and data science communities.

Some tools (e.g. Awkward Array) are in fact more generally useful than just in HEP and are developing wider scientific and industry collaborations (e.g. Anaconda, NVIDIA)

Analysis Grand Challenge - Aiming for the HL-LHC

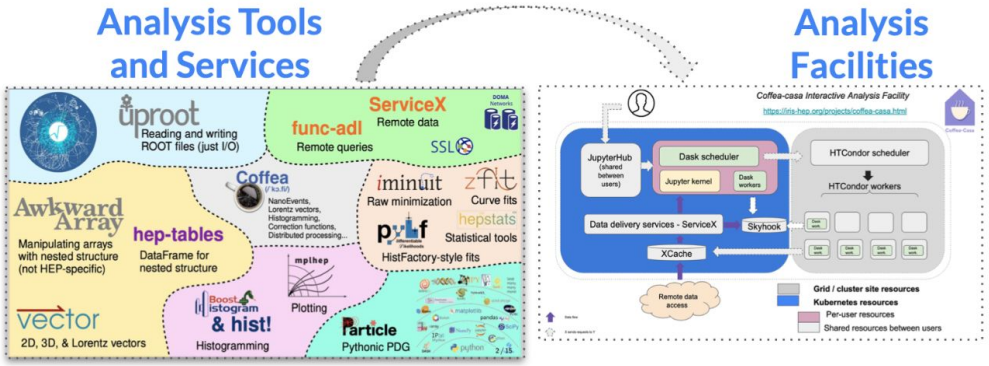
The Analysis Grand Challenge (AGC) is about performing the last steps in an analysis pipeline at scale to test workflows envisioned for the HL-LHC. This includes

- **columnar data extraction** from large datasets,
- processing of that data (event filtering, construction of observables, evaluation of systematic uncertainties) into histograms,
- statistical model construction and statistical inference,
- relevant visualizations for these steps,

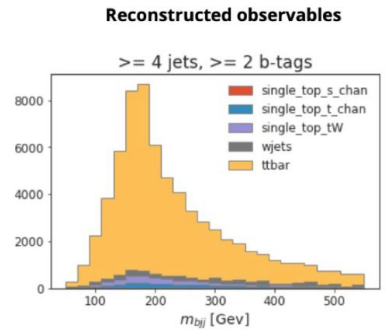
all done in a reproducible & preservable way that can scale to HL-LHC requirements. Begun as an integration exercise for IRIS-HEP, but has evolved as a community project and benchmark reference bringing together different groups and experiments.

Timeline	Fraction of HL-LHC dataset processed in 1h
Year 2	20% (40 TB)
Year 3	50% (100 TB)
Year 4	75 % (150 TB)
Year 5	100% (200 TB)

See also "Demonstrator Analysis 200 Gb/s" (B.Bockelman) at WLCG/HSF



Execution of AGC analysis benchmark



example output of analysis notebook

Other significant application software components

Application frameworks - software systems to organise and run various algorithms over the data

Visualization - visualisation tools are key to reconstruction, simulation and analysis for both debugging and insight into how things are working

Geometry - the geometrical description of the locations, positioning and size of all elements of the experimental apparatus, needed for reconstruction and simulation

Material Descriptions - descriptions of the material, needed both for reconstruction and simulation

Calibrations - systems to calibration and align the detectors, with outputs that vary in time

Data Quality Monitoring - systems needed to check that data acquired is ok, to recognise problems/failures and to validate that software changes (in the trigger/reconstruction and simulation, for example) do not compromise the physics quality of the data

Software Development Tools - tools to support the development activities, debugging, etc.

Software as a “cyberinfrastructure” - 1990s/early-2000s software

Geant4 (RD44 in 1994, V1.0 in 1998) began 30 years ago and just passed the 25th anniversary of its first release, and **ROOT** (1994) is at or near its 30th anniversary. Both are ubiquitous in particle, nuclear and astroparticle physics. (Geant4 also medical and space physics.)

RooFit (D.Kirkby, W.Verkerke) is now nearly 25(?) years old and played a key role in the Higgs discoveries and LHC (and other) physics.

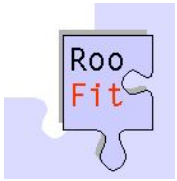
EvtGen (A.Ryd, D.Lange), originally developed at CLEO and developed in BaBar, lives on in many experiments.

xrootd (A.Hanushevsky, many others) is now around 20 years old. Initially planned to fix issues with the Objectivity AMS, it was repurposed into a next generation file server for BaBar with effort from INFN (A.Dorigo, F.Furano) and (later) CERN.
xrootd has grown into a collaboration of SLAC/CERN/UCSD and others, It is not only widely used by the LHC experiments, but it is a key element of the OSG’s “Open Science Data Federation (OSDF)” service broadly supporting science in the US (next slide) that needs high throughput computing.

High Energy Physics Software Ecosystem at LHC startup

IgProf

FroNTier



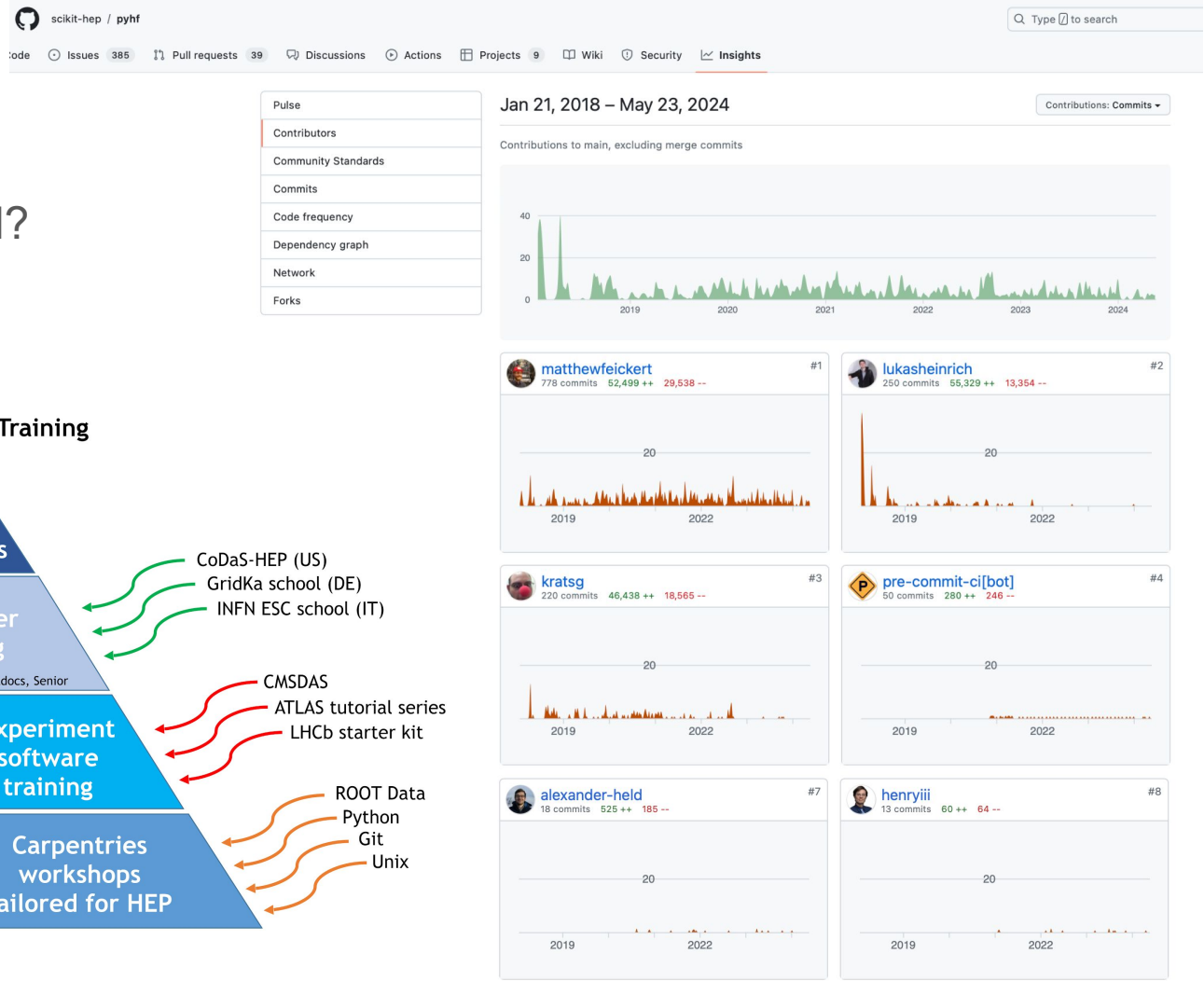
EvtGen

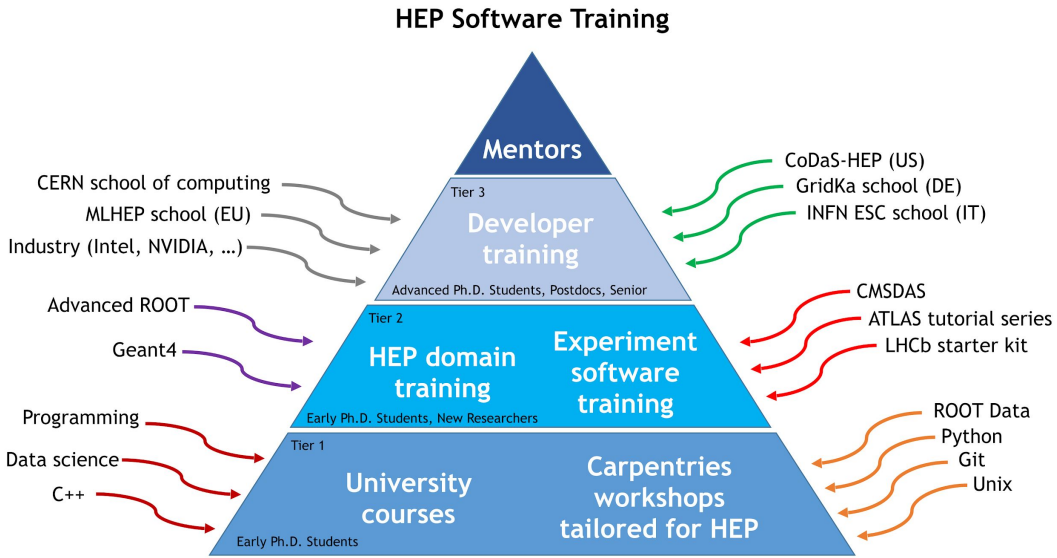
FastJet

RooStats

Examples, definitely incomplete!

How do I get involved?





henryiii

Type / to search

Overview | Repositories 391 | Projects | Packages | Stars 553

Henry Schreiner

henryiii · he/him

Follow

Particle physicist and software architect working with **@iris-hep**, **@scikit-hep**, **@pybind**, and **@scikit-build** admin. **@pypa** member.

640 followers · **8** following

- Princeton University
- Princeton, NJ
- 04:08 - 6h behind
- HenrySchreinerIII@gmail.com
- iscinumpy.dev
- https://orcid.org/0000-0002-7833-783X
- @henryschreiner3
- @henryiii@fosstodon.org

Achievements

x3 x4 x4 x3

Beta Send feedback

Highlights

PRO

Organizations

Block or Report

henryiii / README.md

Favorite posts and series

C++ 11 14 17 20 23 · macOS Setup (AS) · Azure DevOps (Python Wheels) · Conda-Forge ROOT · CLI11 · GooFit · cibuildwheel · Hist · Python Bindings · Python 2→3, 3.7, 3.8, 3.9, 3.10, 3.11, 3.12 · SSH

My classes and books

Modern CMake · CompClass · se-for-sci

My workshops

CMake Workshop · Python CPU, GPU, Compiled minicourses · Level Up Your Python · Packaging

My projects

pybind11 (python_example, cmake_example, scikit_build_example) · cibuildwheel · build · pipx · pyproject-metadata · nox · scikit-build (core, cmake, ninja, moderncmakedomain) · boost-histogram · Hist · UHI · Vector · GooFit · Particle · DecayLanguage · Conda-Forge ROOT · Jekyll-Indico · uproot-browser · Scientific-Python/cookie · repo-review · CLI11 · meson-python · Plumbum · validate-pyproject · pytest GHA annotate-failures · flake8-errmsg · check-sdist · beautifulhugo · POVM · hypernewsviewer

My sites

ISciNumPy · Scientific-Python Development Guide · IRIS-HEP · Scikit-HEP · CLARIPHY

Pinned

level-up-your-python Public
Course over intermediate Python
Jupyter Notebook ☆ 105 ⑂ 31

se-for-sci Public
Software Engineering for Scientific Computing
Jupyter Notebook ☆ 23 ⑂ 7

flake8-errmsg Public
Flake8 checker for raw literals inside raises.
Python ☆ 10 ⑂ 2

aoc2023 Public
Advent of code in Rust
Rust ☆ 6

check-sdist Public
Check to see if an SDist matches Git
Python ☆ 4 ⑂ 2

validate-pyproject-schema-store Public
Weekly mirror of SchemaStore for validate-pyproject
Python ☆ 1

3,857 contributions in the last year

2024 2023 2022 2021

May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May

Mon Wed Fri

Learn how we count contributions Less More

jpivarski

Type / to search

Overview | Repositories 66 | Projects | Packages | Stars 55

Jim Pivarski

jpivarski

Follow

pivarski@princeton.edu

287 followers · **0** following

Princeton, IRIS-HEP, PyHEP, Scikit-HEP

Achievements

x3 x4 x3 x4

Beta Send feedback

Organizations

Block or Report

Popular repositories

hyperbolic-storage-space Public
Python ☆ 15

pyminuit Public
Automatically exported from code.google.com/p/pyminuit
C++ ☆ 14 ⑂ 3

svgfig Public
Automatically exported from code.google.com/p/svgfig
Python ☆ 12 ⑂ 10

could-have-would-have-should-have Public
Could have, would have, should have
☆ 10

JupyterLab-presentation-cell Public
A simple theme, based on JupyterLab light, to show only the current cell.
CSS ☆ 9

rootspec Public
An attempt to describe ROOT I/O with a language-neutral specification.
Python ☆ 5

1,593 contributions in the last year

2024 2023 2022 2021 2020 2019 2018 2017 2016 2015 2014 2013 2012

May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May

Mon Wed Fri

Learn how we count contributions Less More

@scikit-hep @jpivarski-talks @hsf-india-decembe... More

Activity overview

Contributed to **scikit-hep/awkward**, **scikit-hep/uproot5**, **JuliaHEP/AwkwardArray.jl** and 78 other repositories

24% Code review
58% Commits
9% Issues
9% Pull requests

Contribution activity

May 2024

Created 143 commits in 35 repositories

Created 9 repositories

Created a pull request in copier-org/copier that received 9 comments May 8

feat: Allow copier-templates-extensions as an extra

Former IRIS-HEP Fellow, now doing a research software engineering internship at CERN

ManasviGoyal

Type / to search

Overview | Repositories 37 | Projects | Packages | Stars 23

Manasvi Goyal
ManasviGoyal

Follow

9 followers · 4 following

10:12 - same time

Achievements

x2

Beta Send feedback

Highlights

PRO

Block or Report

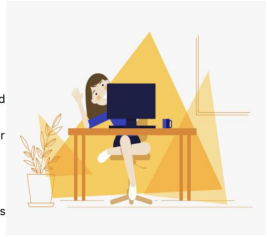
ManasviGoyal / README.md

Hi there 👋

Hello! I am Manasvi 😇

I love learning new things and everything about Technology. I keep exploring new things. I am passionate about doing something to solve real-world problems using technology and making this world a better place.

- 🎓 I pursued Production and Industrial Engineering from Delhi Technological University with a minor in Computer Science
- 💻 I am currently working at CERN as a Research Software Engineer. Previously, I have interned at Princeton University, The University of Auckland, INSEAD, and EY.
- 🌱 I'm fascinated with Software Development, Research, Open Source, and Machine Learning.
- 🌳 I actively work towards empowering women in technology and volunteer for various social causes
- 🔭 I love reading about Space and Astronomy.
- 📽️ I love watching movies and shows in every language in my free time.
- 📫 Feel free to contact me if you want to work on some interesting projects or to discuss technology.



Tech Stack

Python | C | C++ | Java | Scala | HTML | CSS | JavaScript | Git | GitHub | PostgreSQL | Power BI | Machine Learning | Deep Learning | Jupyter Notebooks | Markdown | Visual Studio Code | PlotLy | Microsoft Office | ScikitLearn | NumPy | Pandas | Data Structures | Algorithms | Operating Systems | Object Oriented Programming | Selenium

Reach out to me on

LINKEDIN | GITHUB | GMAIL | INSTAGRAM

Pinned

det-lab/kaitai_struct_awkward_runtime Public
Kaitai Struct: runtime for Awkward Arrays
Kaitai Struct ☆ 2 ⑂ 2

PyHEP-2023-Awkward-Target-for-Kaitai-Struct Public
Code Repository for PyHEP-Users-Workshop-2023 talk - "Awkward Target for Kaitai Struct"
Jupyter Notebook ☆ 1

ACAT2022-The-Awkward-World-of-Python-and-CPP Public
ACAT2022 Presentation on 27th October in Bari, Italy

Enron-Classification Public
Exploratory Analysis of Enron Dataset and Classification using multiple algorithms
Jupyter Notebook ☆ 4 ⑂ 4

Gmail-Classification Public
Extract Emails from Gmail account, convert to Excel file and classify using various classification algorithms.
Jupyter Notebook ☆ 13 ⑂ 5

Restaurent-Management-System Public
A Restaurant Management System using OOPs concepts
C++ ☆ 2 ⑂ 1

294 contributions in the last year

May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May
Mon
Wed
Fri

Learn how we count contributions — Less More

2024 2023 2022 2021 2020 2019

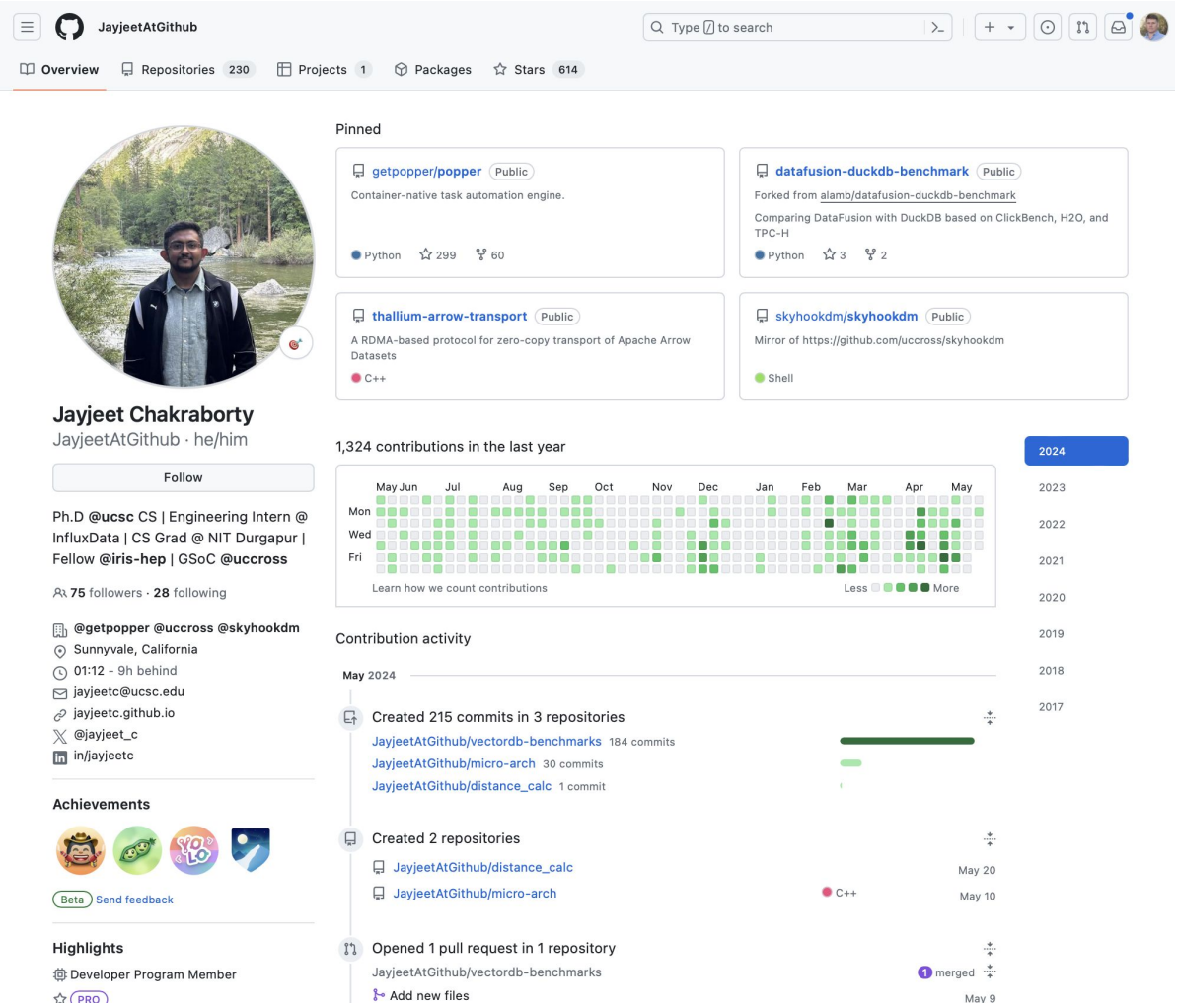
Contribution activity

May 2024

Created 1 commit in 1 repository
scikit-hep/awkward 1 commit

Opened 2 pull requests in 1 repository
scikit-hep/awkward — 2 open
chore: trying atomics and tree reduction for CUDA reducer kernels — May 16
feat: add reduce CUDA kernels — May 2

Reviewed 1 pull request in 1 repository
scikit-hep/awkward — 1 pull request
fix: skip tests that require pyarrow when it is not installed — May 13



Former GSoC and (pre-HSF-India) IRIS-HEP Fellow, now PhD student in Computer Science

HSF-India and HSF/GSoC student programs

Research Software Collaborations — About — Connect — Activities

Research Software Trainees

People are the key to successful software. The HSF-India project aims to promote the development of international research software collaborations in physics by providing opportunities for undergraduate and graduate (both masters and PhD candidate) trainees to connect with expert mentors from the particle physics, nuclear physics and astroparticle physics communities as well as from the Computational/Data Science community. Our program aims to help students to gain experience working in a unique scientific and collaborative environment and to learn new programming, research and analysis skills that are important for future careers in science and technology.

We maintain a list currently open available ideas can be found here. Get inspired by these and contact us with questions. To add your own project idea, please see the instructions in our GitHub.

Are you interested in building research software skills and gaining experience working as part of a research team and contributing to cutting edge open source research software projects? Then the HSF-India program could be for you. Find more information about our program and how to appy here.

Current and Former Trainees

Sanjeev Kumar	Juhi Poddar	Durbar Chakraborty	Chitrakshee Yede	Ananya Gupta
TIFR, Mumbai	Tata Institute of Fundamental Research	National Institute of Technology Durgapur	Savitribai Phule Pune University	Indira Gandhi Delhi Technical University for Women
Mar – Jul, 2024	*Feb – Jul, 2024*	*Jul – Nov, 2023*	*Jun – Nov, 2023*	*Jun – Sep, 2023*

http://research-software-collaborations.org/trainees.html

HSF — Working Groups — Activities — Meetings — Communication — Projects & Support — About

Google Summer of Code 2024

Introduction

Google Summer of Code is a program that allows students to contribute to the development of open-source projects, mentored by participating organizations.

Particle physics is an exciting field where large collaborations of scientists collect and analyze petabytes of data from high-energy physics experiments, such as those at the Large Hadron Collider, hosted at the CERN laboratory in Geneva, Switzerland. Some of the questions that we collectively ask are:

- what are the fundamental blocks that make up our Universe?
- what is the nature of dark matter and dark energy?
- what is the nature of the asymmetry between matter and antimatter?
- what was early Universe like?

To answer these questions, particle physicists build software to simulate and analyze what happens in particle physics detectors.

The CERN software for experiments (CERN EP-SFT) group has participated in the GSoC since 2011. Since 2017 the program has expanded to involve the high-energy physics community under the umbrella of the HEP Software Foundation.

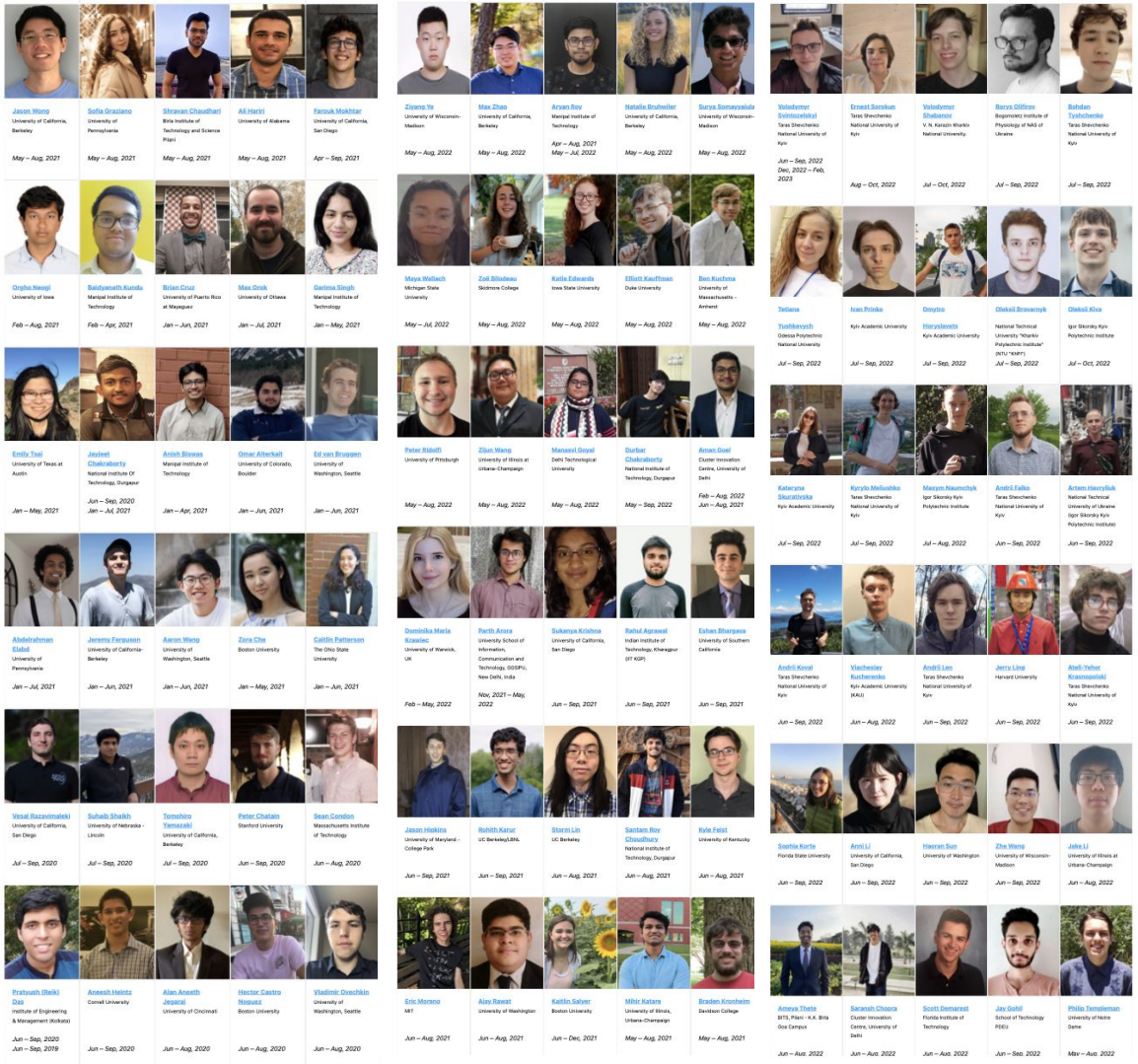
Information from last year's GSoC can be found here.

GSoC 2024

In 2024 HSF is participating in the program as a GSoC umbrella organization, under the name CERN-HSF. The HSF project ideas are listed below. Please have a look at the general program rules and the timeline provided by Google.

The selection process in our organization will be split into two phases. All candidates will have to pass pre-selection evaluation tests prepared by mentors, demonstrating the skills needed for the respective projects. The successful candidates will have a detailed exchange with mentors on a given project idea, and their proposals are then evaluated for the final student selection. The detailed timeline of these phases is shown below.

https://hepsoftwarefoundation.org/activities/gsoc.html



IRIS-HEP Fellows

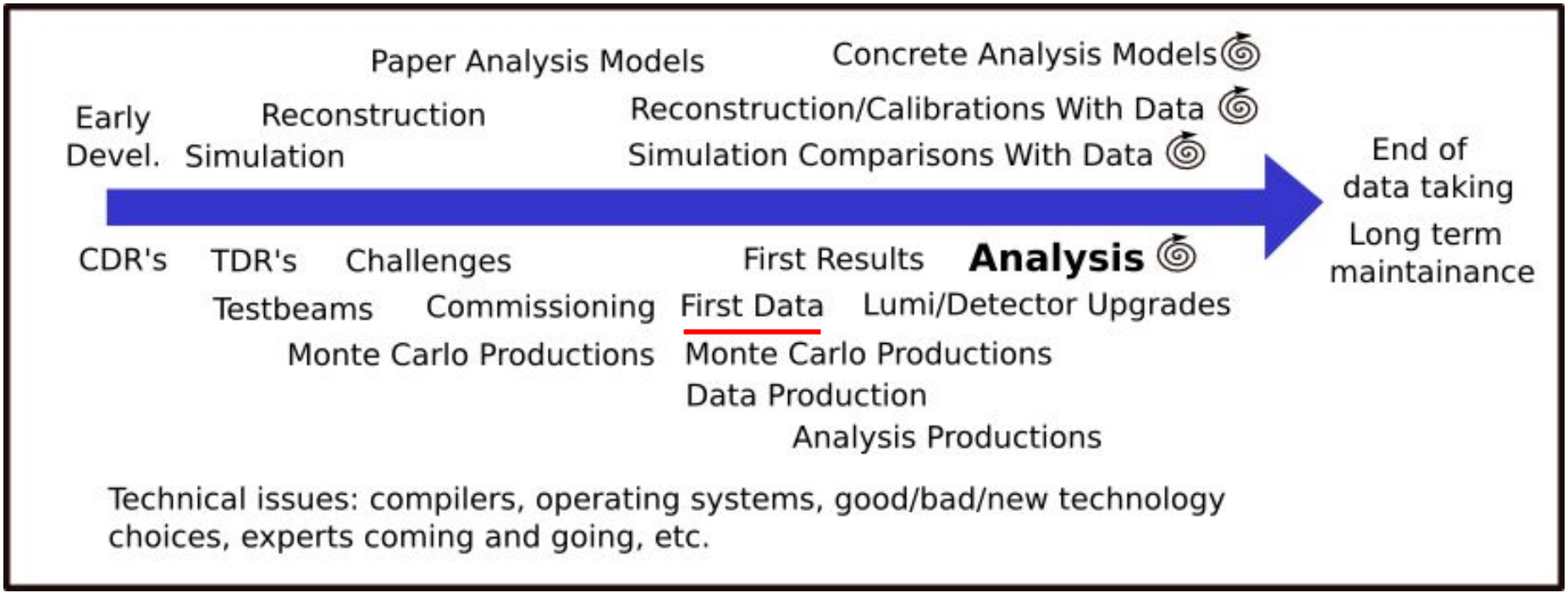
Connects undergraduates in physics and computer science with HEP mentors active in developing HEP research software.

They work mostly in remote with their mentors on cutting-edge software-centric summer research projects

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

What is the “Future of (Software and) Computing for HEP?

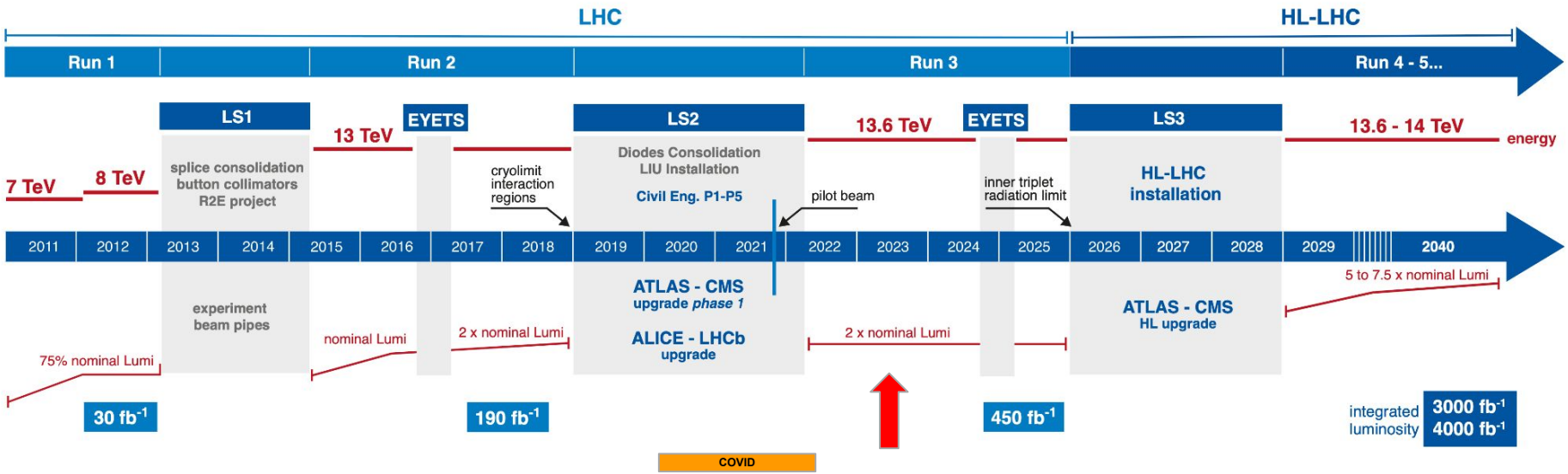
Software Lifecycle in High Energy Physics



The Life Cycle of HEP Offline Software, CHEP 2007

(January, 2022 version)

LHC / HL-LHC Plan



S2I2-HEP

IRIS-HEP Institute	IRIS-HEP Institute
Design	Execution

Institute Conceptualization and Community White Paper Process

Snowmass

U.S. HEP Community Planning Process

HL-LHC “Software Upgrade” - Developing a Global R&D Roadmap

Community charge from the Worldwide LHC Computing Grid to the (then nascent) HEP Software Foundation in July 2016:

- Anticipate a “software upgrade” in preparation for the HL-LHC
- Identify and prioritize the software research and developments investments
  1. to achieve improvements in software efficiency, scalability and performance and to make use of the advances in CPU, storage and network technologies
  2. to enable new approaches to computing and software that could radically extend the physics reach of the detectors
  3. to ensure the long term sustainability of the software through the lifetime of the HL-LHC

NSF funded the S2I2-HEP Conceptualization Project (s2i2-hep.org/) in July 2016

2016-2017 - Software Roadmap

January 2017
UCSD

June 2017
Annecy

[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

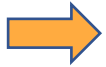
- International participants
- Computing Management from the Experiments and Labs
- Individuals interested in the problems
- Members of other compute intensive scientific endeavors
- Members of Industry
- http://s2i2-hep.org/
- https://hepsoftwarefoundation.org/

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

arXiv 1712.06982 arXiv 1712.06592



IRIS-HEP

Growing a Global Collaboration

JLab
March, 2019
HSF/OSG/WLCG

UCSD/SDSC
January, 2017
HSF CWP

Annecy
June, 2017
HSF CWP

Naples
March, 2018
WLCG/HSF

HEP Software Foundation (HSF)

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 common projects (new resources!)
- Provide a structure to set priorities and goals for the work

The HSF is an unfunded, volunteer organization, with a “bottoms-up” structure.

Conceptual motivations behind the HEP Software Foundation

Computer hardware is a consumable. Software is the actual "cyberinfrastructure".

Importantly, software is also an *intellectual product* of our research, not just a tool.

We can seed and build collaborations around software in a similar fashion to our experimental collaborations.

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

News and Featured Stories:

IRIS-HEP Receives $25M Funding for Another Five Years of Research

"IRIS-HEP received funding from the Office of Advanced Cyberinfrastructure and the Physics Division at the National Science Foundation for five years."

Read more

Out of harm's way: Physics research program supports Ukrainian students displaced by war

"Ukrainia students escape the war and pursue research at the Large Hadron Collider (LHC), under supervision from Princeton University faculty."

Read more

Upcoming Events:

May 24, 2024	*CERN*
IRIS-HEP / AGC Demo Day #5	
Jun 20–21, 2024	*Princeton University*
USCMS/IRIS-HEP Software Training	
Jul 8–14, 2024	*Tacoma, Washington*
Scientific Computing with Python (SciPy) 2024	
Jul 18–19, 2024	*University of Washington*
USATLAS/IRIS-HEP Software Training	
Jul 22–26, 2024	*Princeton University*
CoDaS-HEP 2024 - Computational and Data Science Training for High Energy Physics	
Aug 26–30, 2024	*Aachen, Germany*
PyHEP.dev 2024 - "Python in HEP" Developer's Workshop	
Sep 4–6, 2024	*University of Washington*
IRIS-HEP Institute Retreat	
Sep 23–25, 2024	*Valencia (Spain)*
Fourth MODE Workshop on Differentiable Programming for Experiment Design	

View all past events

http://iris-hep.org

Conceived as a **"software upgrade"** project and guided initially by the "Community White Paper" roadmap developed in 2016-2017: it involves 21 universities, spanning ATLAS, CMS and LHCb.

IRIS-HEP is supported by the U.S. National Science Foundation through the **Office of Advanced CyberInfrastructure** in the Directorate for Computer and Information Science and Engineering and the **Division of Physics** in the Directorate for Mathematical and Physical Sciences.

10-year project: Originally funded in 2018 as OAC-1836650 and renewed in 2023 through 2028 as PHY-2323298.

iris hep

About Connect Activities Fellows Jobs

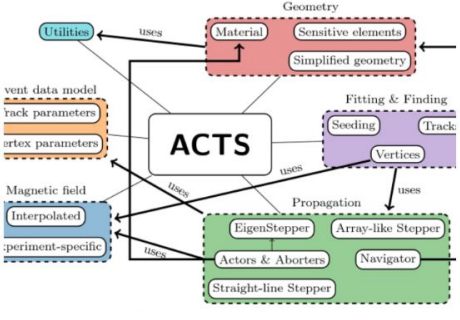
Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)

https://iris-hep.org/

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

News and Featured Stories:



Upcoming Events:

May 3–5, 2023 *University of Wisconsin–Madison*
IRIS-HEP AGC workshop 2023

Jul 12–14, 2023 *Virtual*
HSF/IRIS-HEP Software Basics Training (Virtual)

Jul 17–21, 2023 *Princeton University*
CoDaS-HEP 2023 - Computational and Data Science Training for High Energy Physics

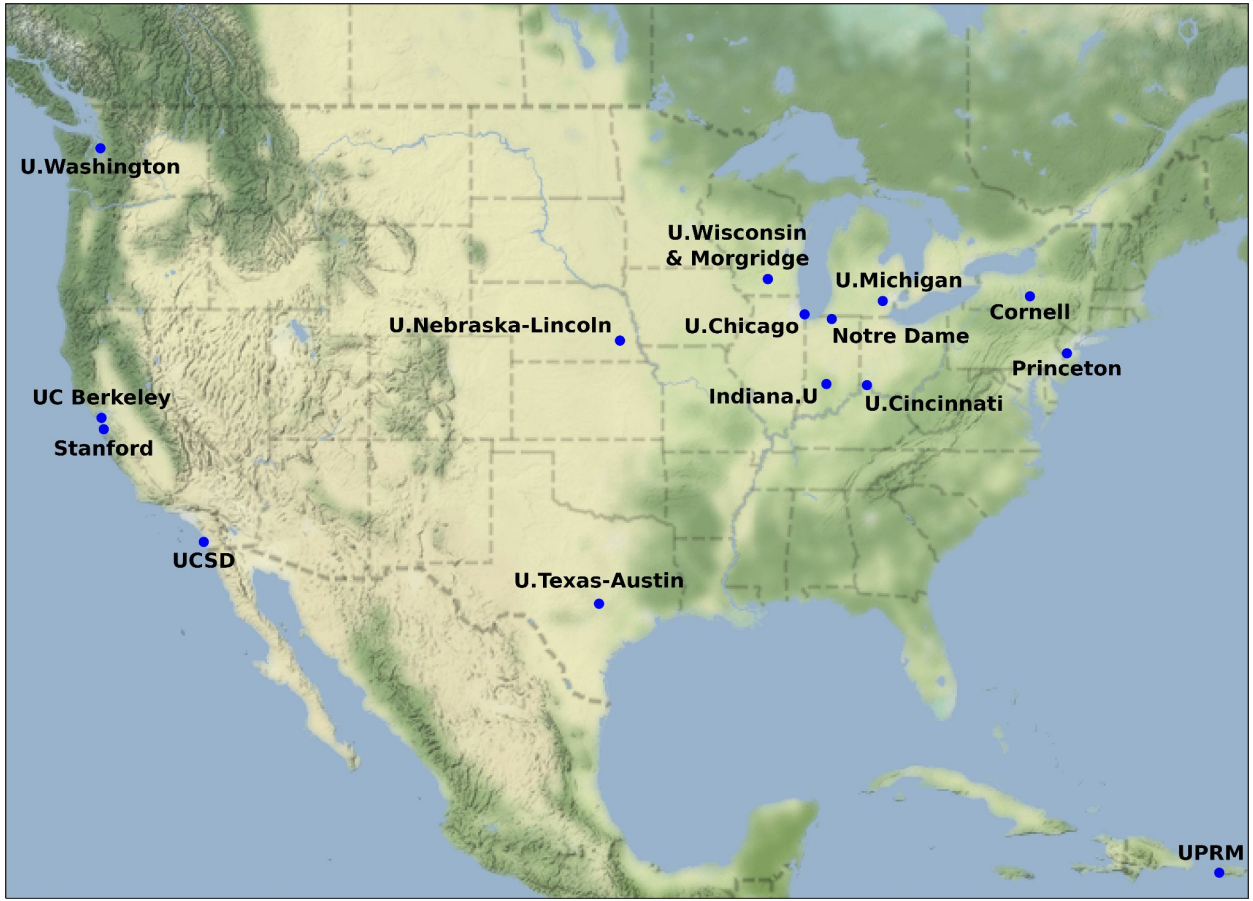
Jul 24–26, 2023 *Princeton University*
3rd MODE Workshop on Differential Programming for Experiment Design

Jul 25–28, 2023 *Princeton University*
PyHEP.dev 2023 Developers Workshop

View all past events

Upcoming Topical Meetings:

IRIS-HEP - A Virtual (Distributed) Institute



HSF-India is a US National Science Foundation (NSF) funded project that began on October 1, 2022. HSF-India aims to join collaborative networks in India to networks in the U.S. and Europe in order to build the international research software collaborations required to reach the science goals of next-generation particle, nuclear and astroparticle physics experiments including the High-Luminosity Large Hadron Collider at CERN, the Deep Underground Neutrino Experiment at Fermilab, and the Electron Ion Collider at Brookhaven National Laboratory. To fully realize their discovery potential a new generation of software algorithms and approaches is required. Building these research software collaborations is challenging and inherently international matching the international nature of the experimental undertakings themselves. HSF-India will provide students, postdocs and early career personnel significant experience in international team science through engagement in a diverse research community.

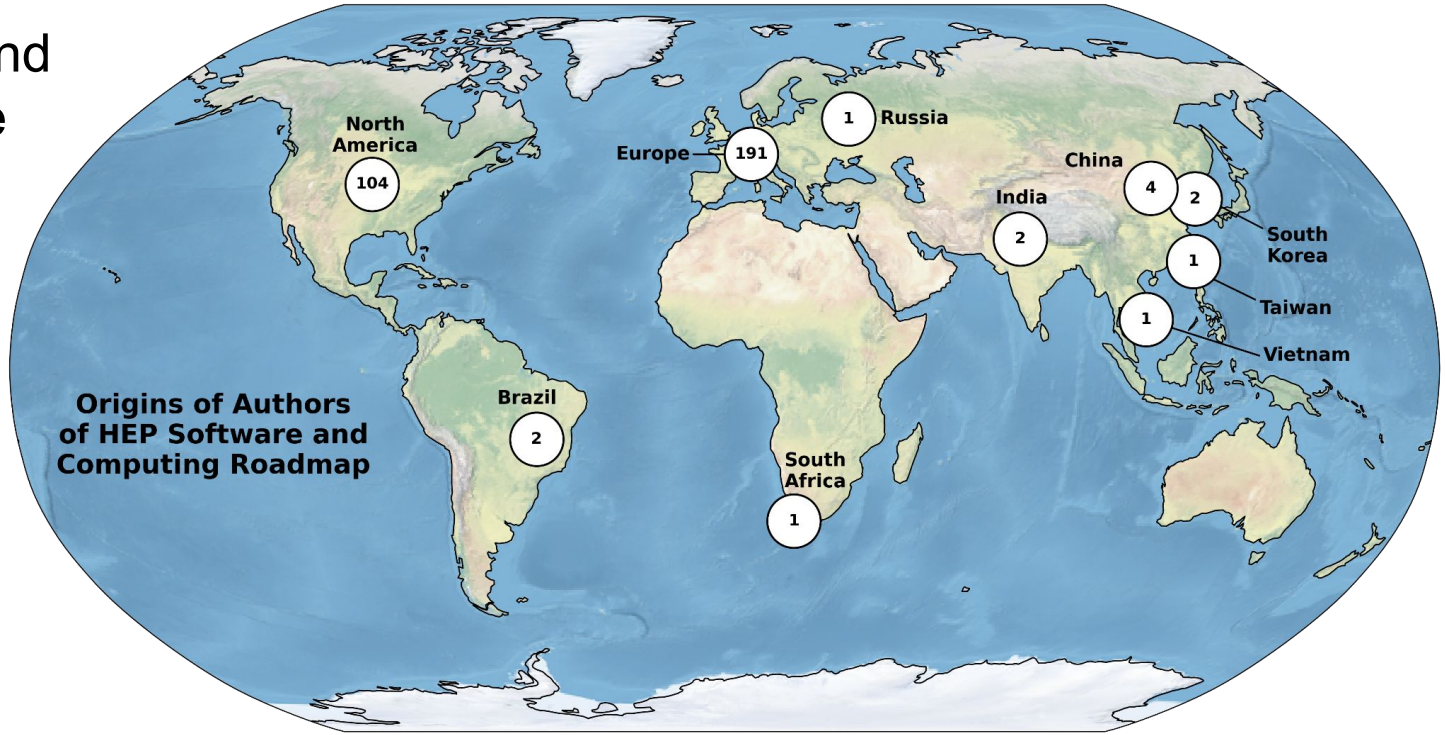
The HSF-India network of networks aims to extend the growing research software collaborations between physics and computer science researchers in the U.S. and Europe to include researchers in India. By exploiting national capabilities and strengths, an immediate mutual benefit of the international collaboration will be a training network that enables early-career researchers to pursue impactful research software initiatives in ways that advance their careers in experimental data-intensive science. HSF-India will build international collaborative structures which will enable innovation in critical software areas for discovery in particle, nuclear and astroparticle physics and lay the foundations for similar international collaborations across the scientific spectrum. It is intended to be a long-term investment in team science in physics.

Questions? Get in touch with us by email

Stay in touch: Get notified about program opportunities and announcements by joining the **rsc-announcements** google group (or contact us to be added)

HSF-India project

Building global collaborations around common software



Although participation in experiment software is international, much of the core software development comes primarily from the US and Europe.

HSF-India project: 5-year project: software training, researcher visits and seeding of software collaborations

http://research-software-collaborations.org/ (OISE-2201990)

Technologies, Software, Systems

1990s Technology Context

Dec 1991 - first web server (later first DB) deployed in the US, at SLAC

Sep 1992 - first public Linux version

Jan 1993 - Mosaic browser released

Aug 1995 - Windows 1995

Yahoo, AltaVista, …. Google still to come.

Windows NT was the “strategic platform” at CERN in the late 1990s. Browser wars!

Carrying around a laptop really only became a “thing” over the course of the 1990s. Cell phones were starting to arrive on the scene, for calling/messaging. Pagers were a thing…

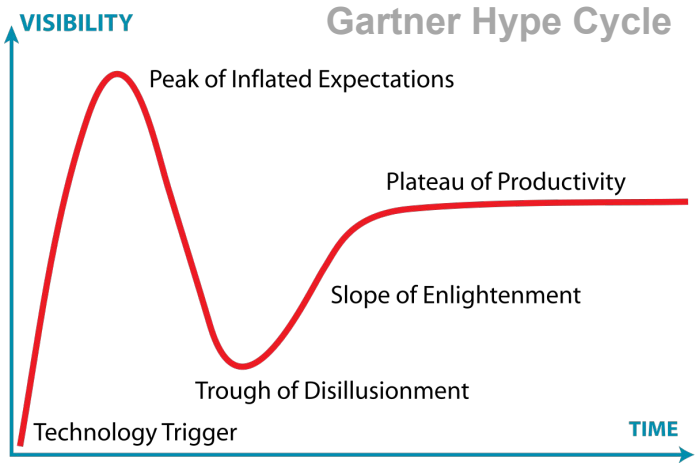
International dialing at reduced rates to participate in meetings with phone cards

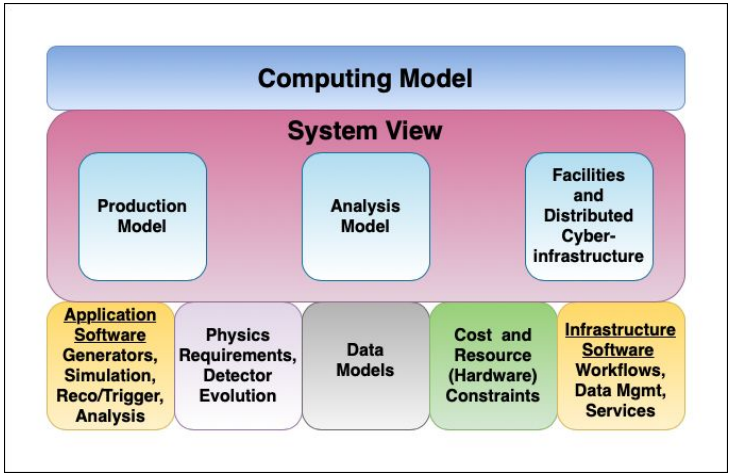
Modem Dial-up Internet (Geneva)
Local phone calls were 4CHF/hr ($2.65/hr in 1998) during the day, 2CHF/hr in the evening and 1CHF/hr in the night —> CERN Automatic Call Back (ACB) system

Future of Computing?

Technologies suffer from the hype cycle, then often are subsumed into our everyday work, or just disappear.

In addition, we build “systems” that support “computing models”, so each technology needs a context...





Ask ChatGPT - Tell me about the future of computing in high energy physics.

You

Tell me about the future of computing in high energy physics

ChatGPT

The future of computing in high-energy physics holds exciting prospects driven by technological advancements and the evolving needs of scientific research. Here are some trends and areas of development that could shape the future of computing in this field:

1. **High-Performance Computing (HPC)**: …
2. **Distributed Computing and Grid Computing**: …
3. **Machine Learning and Artificial Intelligence**: …
4. **Quantum Computing**: …
5. **Data Management and Storage**: …
6. **Open Science and Collaboration Platforms**: …

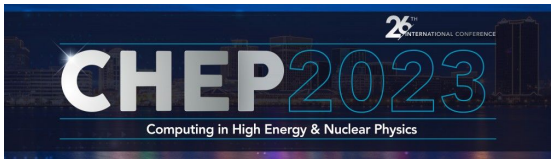
Overall, the future of computing in high-energy physics is likely to be characterized by a combination of technological innovation, collaborative research efforts, and a continued focus on addressing the computational challenges posed by large-scale experiments and datasets.

26TH INTERNATIONAL CONFERENCE ON COMPUTING IN HIGH ENERGY & NUCLEAR PHYSICS

CHEP is a long running conference (every 1.5 years) on computing and software in high energy and nuclear physics. The first CHEP was in 1985 in Amsterdam.

CHEP 2023 Scientific Program

Track 1 - Data and Metadata Organization, Management and Access
Track 2 - Online Computing
Track 3 - Offline Computing
Track 4 - Distributed Computing
Track 5 - Sustainable and Collaborative Software Engineering
Track 6 - Physics Analysis Tools
Track 7 - Facilities and Virtualization
Track 8 - Collaboration, Reinterpretation, Outreach and Education
Track 9 - Artificial Intelligence and Machine Learning
Track 10 - Exascale Science
Track 11 - Heterogeneous Computing and Accelerators
Track 12 - Quantum Computing



https://www.jlab.org/conference/CHEP2023

CHEP 2019 - Scientific Program

Track 1 - Online and Real-time Computing
Track 2 - Offline Computing
Track 3 - Middleware and Distributed Computing
Track 4 - Data Organisation, Management and Access
Track 5 - Software Development
Track 6 - Physics Analysis
Track 7 - Facilities, Clouds and Containers
Track 8 - Collaboration, Education, Training and Outreach
Track 9 - Exascale Science

http://chep2019.org/

CHEP 2007 Scientific Program

Track 1 - Online Computing
Track 2 - Event Processing
Track 3 - Software components, tools and databases
Track 4 - Computer facilities, production grids and networking
Track 5 - Grid middleware and tools
Track 6 - Distributed data analysis and information management
Track 7 - Collaborative initiatives with other sciences
Track 8 - Collaborative tools

CHEP'07
VICTORIA, BC

https://indico.cern.ch/event/3580/

CHEP 2000 - Scientific Program

Track 1 - Data Analysis: Algorithms and methods
Track 2 - Data Acquisition and Control System
Track 3 - Object Persistency and Data Handling
Track 4 - Network: Applications and Services
Track 5 - Commodity Hardware and Software and Integration in Farm and Large Systems
Track 6 - Data Analysis: Technology and Presentation

COMPUTING IN HIGH ENERGY AND NUCLEAR PHYSICS
CHEP 2000
International Conference on
COMPUTING IN HIGH ENERGY AND NUCLEAR PHYSICS
February 7 - February 11, 2000
Teatro Antonianum, Padova (Italy)

https://chep2000.pd.infn.it/

Computing in High Energy Physics
CHEP '97
B E R L I N

International Conference on

Computing in High Energy Physics

Berlin, Lichtenberger Congress Center
April 7-11, 1997

Computing in High Energy Physics (CHEP) is a major series of international conferences for physicists and computing professionals from the High Energy Physics community, Computer Science, and IT industry. The CHEP conference provides an international forum to exchange the experience and needs for the community, and to review recent, ongoing, and future activities.

The CHEP conference was organized by Deutsches Elektronen-Synchrotron (DESY) which has locations in Hamburg and in Zeuthen (near Berlin).

Over the last 8 years, this conference has been held in different countries around the world:
Oxford, UK (April 89), Santa Fe, USA (April 90), Tsukuba, Japan (March 91), Annecy, France (September 92), San Francisco, USA (April 94), Rio de Janeiro, Brazil (September 95).

Conference Secretariat

DESY-Zeuthen
Platanenallee 6
D-15738 Zeuthen
Germany

Fax: +49 33762 77-501
Phone: +49 33762 77-500
Email: chep97@desy.de
URL: http://www.desy.de/chep97

last update Aug 25, 1997

CHEP´97

Homepage

Notes

Information

Program

Documents

Multimedia

Location

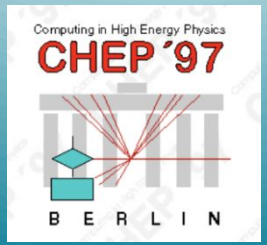
Sponsors

Committees

https://www.zeuthen.desy.de/CHEP97/chep97.htm

Conference Schedule

Conference Schedule					
Sunday	**Monday**	**Thuesday**	**Wednesday**	**Thursday**	**Friday**
Workshop: HEP and the Internet 10:00-12:00	Opening Remarks at 9:00 Plenary Talk: Java and Internet Computing 9:30-10:15	Parallel Talks 9:00-10:30	Object Technology and Software Processes Panel: Transition to OO 9:00-11:00	Plenary Talk: PC's: Facts, Figures and Forecasts 9:00-10:30 Summary Session C: 9:00-10:30	Conference Summaries E, F, A: 9:00-10:30
Workshop: HEP Internet Monitoring 13:30-17:30	Panel: Collaborative Tools in the Internet 10:45-12:15 Plenary Talk: HPC and Archival Storage Systems 12:15-13:00	Parallel Talks 11:00-12:30	Plenary Talk: A Software Engineering Service Center for Scientific Software Production 11:30-12:15	Plenary Talk: Models of Multiprocessor Computing Plenary Talk: Overview on LQCD Computing 11:00-12:30	Conference Summaries B, D, G: 11:00-12:30
Registration 14:00-20:00	Parallel Talks 14:30-15:30 Vendor Talks, Parallel Talks A, B 15:30-16:30 17:00-18:00	Panel: Computing at Major HEP Sites 14:00-15:15 Panel: Computing Models of Major HEP Experiments 15:45-18:00	Parallel Talks 13:45-14:45 15:15-16:30 17:00-19:20	Parallel Talks 14:00-15:30 16:00-18:00	Panel: Future of HEP Computing 14:00-15:30
Welcome Party 18:00-20:00	Video Conference: Tutorial from WWW6 18:00-21:30	Poster Session 18:00-20:00	Video Conference: Talks from WWW6 18:45-20:00	Conference Dinner Deutsches Technikmuseum 19:30-22:30	Closing Remarks 15:30-16:00



- **What do you anticipate to be the impact of the changes in the wide world of computing in each of these technological areas on the HEP challenges we have identified?**
- **What are the problems that HEP computing will need to address by its own efforts (beyond the usual integration of commercially available components) to meet the challenges we have identified?**
- **What will the data geographical model be?**
  - **Centralized data vs partially centralized date (Jürgen Knobloch)**
  - **Impact on network and computing iron/storage requirements**
  - **What is a regional center?**
- **Will OO DBs be lightweight enough to be used throughout data cycle?**
- **Java und C++**
  - **What will be (should be) the place for each?**
    - **(When) will C++ code become legacy?**
      - **Will Java always be slower than C++?**
    - **Grad students should learn both (to be more employable)?**
  - **(When) are learning curves worth the pay back?**
  - **Should physicists learn (computing) analysis skills?**
  - **Role of computer scientists? Consultants?**
    - **How do you know you have found a good one?**
  - **Requirements? When in process?**
    - **Waterfall vs iterative development?**
    - **Rapid prototyping?**
  - **Reviews and checkpoints: electronic only?**
- **Process?**
  - **Let them code first? Will they ever design?**
  - **Daily, weekly, monthly build?**
    - **(Why) are we slower than Netscape, Microsoft release cycles?**
- **Will complex, do-everything programs die, evolve to Component software?**
  - **Word, GEANT, Experiment analysis packages**
  - **Chosen, focused functionality**
  - **What is appropriate level of component granularity?**
- **Future (appropriate roles) (positions) of NT *vs* Unix *vs* (Java + Browser)?**
- **Integration of HEP computing into HEP**
  - **What is trend of % cost for computing in experiments?**
  - **Should (can, will) computing be included in TPC (total project cost)?**
    - **in project work break down structure (WBS)?**
  - **What should we be saying about computing issues as leaders in CHEP to leaders in HEP?**

CHEP 1997 - “Future of HEP Computing” Panel

https://www.zeuthen.desy.de/CHEP97/slide/p518/p518.htm

Some conclusions from the exploring the CHEP history

General trend is towards higher level “system” concepts, while earlier technology tracks now “subsumed” into the system tracks. Some new “technology” tracks appear (e.g. AI/ML, Quantum Computing). ChatGPT suggested more technology than systems.

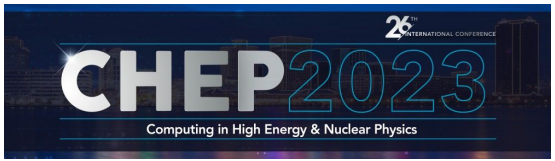
“Conway's Law” effect? - the way an organization communicates and is structured will be directly reflected in the systems and software it creates

One consistently present track is “data analysis” and/or “analysis tools”. ChatGPT missed this one.

ChatGPT: “Overall, the future of computing in high-energy physics is likely to be characterized by a combination of **technological innovation**, **collaborative research efforts**, and a continued focus on addressing the **computational challenges posed by large-scale experiments and datasets**.”

CHEP 2023 Scientific Program

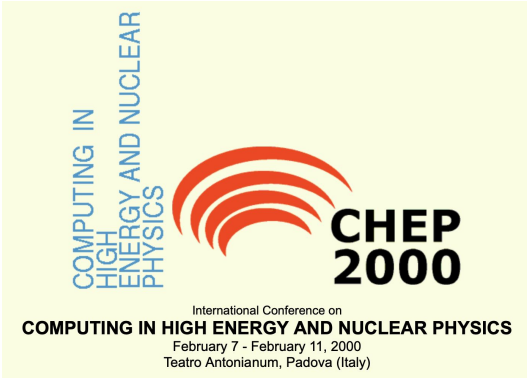
Track 1 - Data and Metadata Organization, Management and Access
Track 2 - Online Computing
Track 3 - Offline Computing
Track 4 - Distributed Computing
Track 5 - Sustainable and Collaborative Software Engineering
Track 6 - Physics Analysis Tools
Track 7 - Facilities and Virtualization
Track 8 - Collaboration, Reinterpretation, Outreach and Education
Track 9 - **Artificial Intelligence and Machine Learning**
Track 10 - Exascale Science
Track 11 - Heterogeneous Computing and Accelerators
Track 12 - **Quantum Computing**



https://www.jlab.org/conference/CHEP2023

CHEP 2000 - Scientific Program

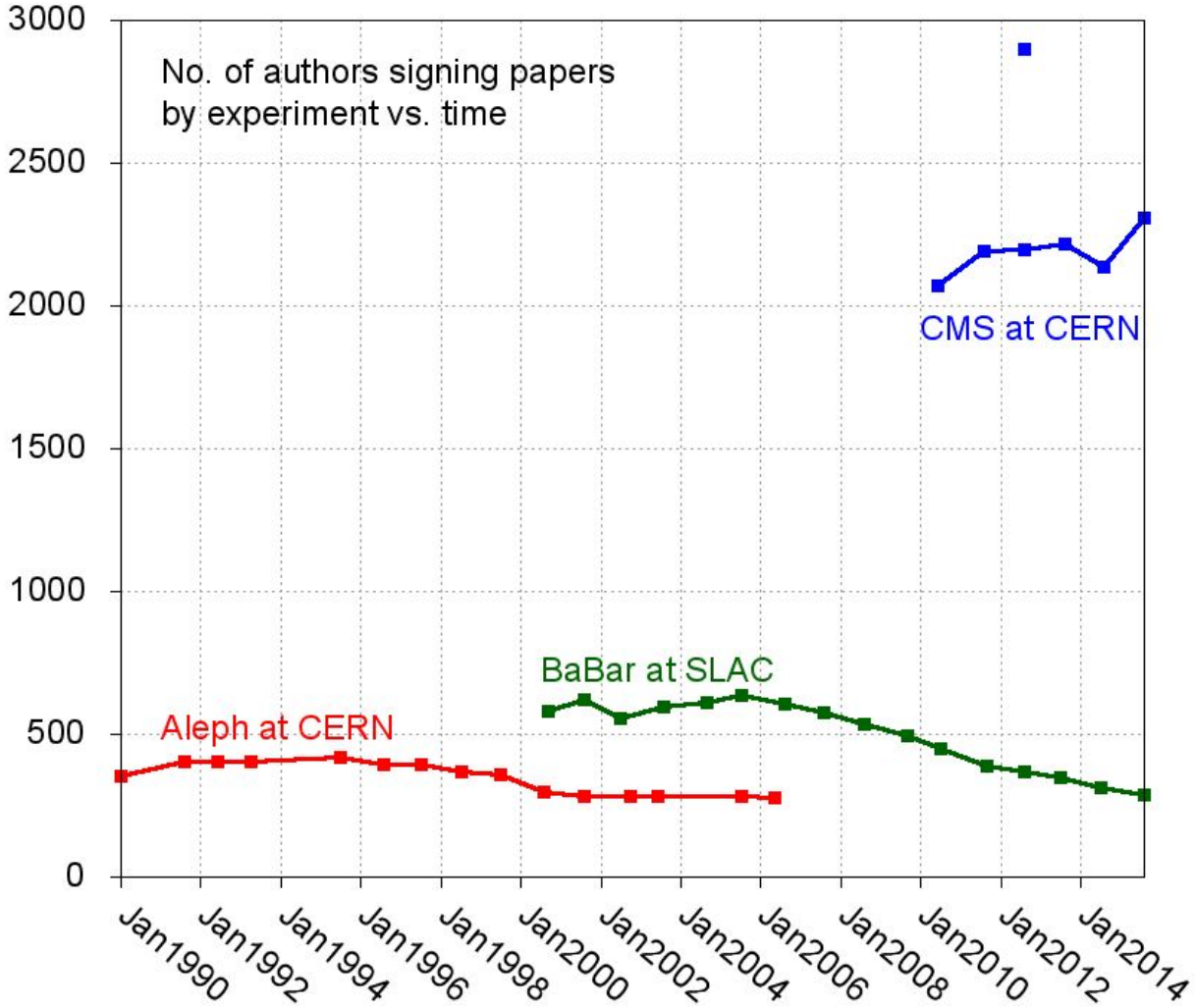
Track 1 - Data Analysis: Algorithms and methods
Track 2 - Data Acquisition and Control System
Track 3 - **Object Persistency** and Data Handling
Track 4 - **Network: Applications and Services**
Track 5 - **Commodity Hardware and Software** and Integration in Farm and Large Systems
Track 6 - Data Analysis: Technology and Presentation



https://chep2000.pd.infn.it/

Particle Physics Experiments as "collaborative research efforts"

Project Size Over Time



Community Collaborative Structures Reflect Our Community Evolution/Needs

The **Worldwide LHC Computing Grid (WLCG)** project is a global collaboration of around 170 computing centres in more than 40 countries, linking up national and international grid infrastructures. The mission of the WLCG project is to provide global computing resources ... **[2000's era]**

The **HEP Software Foundation** facilitates cooperation and common efforts in High Energy Physics software and computing internationally. **[2010's era]**

WLCG/HSF 2024 (13-17 May) - https://indico.cern.ch/event/1369601/

And the “Future of Computing”?

The **Worldwide LHC Computing Grid (WLCG)** project is a global collaboration of around 170 computing centres in more than 40 countries, linking up national and international grid infrastructures. The mission of the WLCG project is to provide global computing resources ... **[2000’s era]**

The **HEP Software Foundation** facilitates cooperation and common efforts in High Energy Physics software and computing internationally. **[2010’s era]**

What collaborative research efforts will the rest of the 2020’s and the 2030’s produce for the “future of computing” in the HEP community?