

# DPF-PHENO 2024

DPF-PHENO 2024 - University of Pittsburgh / Carnegie Mellon University

## Future of Computing

Peter Elmer - Princeton University  
15 May 2024



# DPF-PHENO 2024

DPF-PHENO 2024 - University of Pittsburgh / Carnegie Mellon University

# ~~Future of Computing~~

**A story of technologies, collaborations and science driven challenges**

Peter Elmer - Princeton University  
15 May 2024



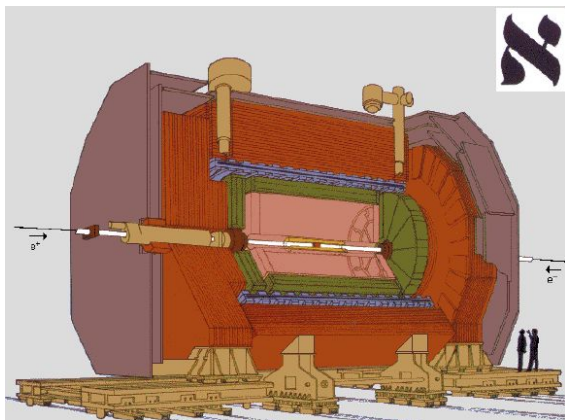
# Introductions.... Who am I?

I am an experimental particle physicist focused on computational and data science problems in my field, along with the software/computing systems to solve them. I am a senior research scientist with Princeton, but based at CERN.



I have primarily collaborated on collider experiments, which will likely color some of the perspectives in this talk. These include:

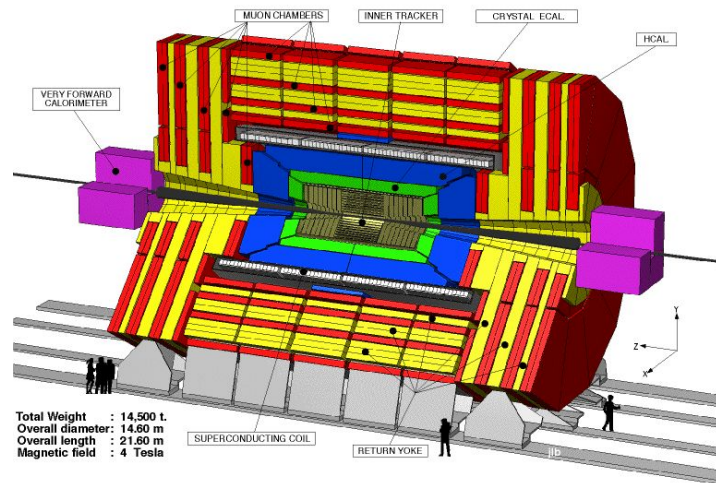
## Aleph@CERN



## BaBar@SLAC



## CMS@CERN



# 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



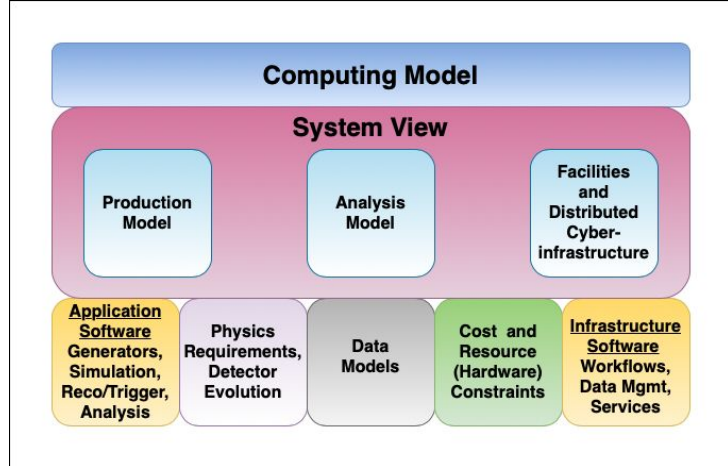
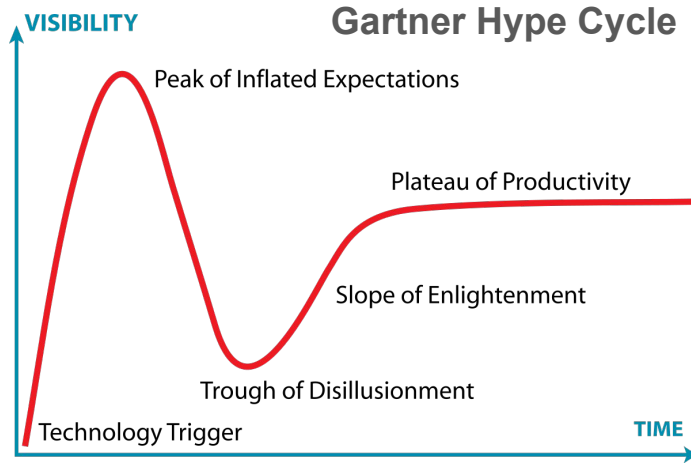
## Modem Dial-up Internet (Geneva)

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.

PE

**You**

Tell me about the future of computing in high energy physics









**ChatGPT**

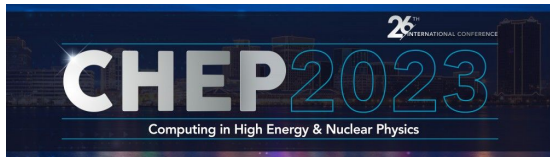
The future of computing in high-energy physics holds exciting prospects driven by technological advancements and the evolving needs of scientific research. Here are some trends and areas of development that could shape the future of computing in this field:

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.





# 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



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

# CHEP 2000 - Scientific Program

Track 1 - Data Analysis: Algorithms and methods

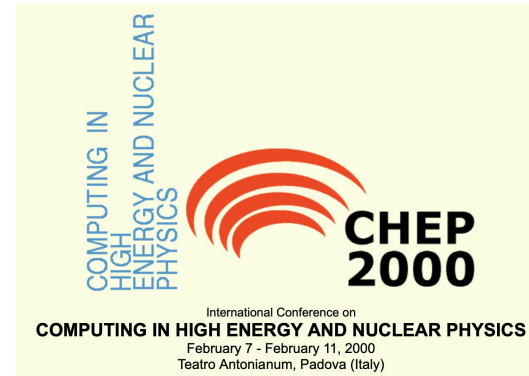
Track 2 - Data Acquisition and Control System

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

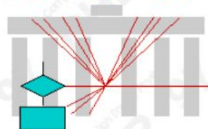


<https://chep2000.pd.infn.it/>



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](mailto:chep97@desy.de)  
URL: <http://www.desy.de/chep97>

*last update Aug 25, 1997*



CHEP'97

Home page

Notes

Information

Program

Documents

Multimedia

Location

Sponsors

Committees

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

# Conference Schedule

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

## Conference Program

[Homepage](#)

[Schedule](#)

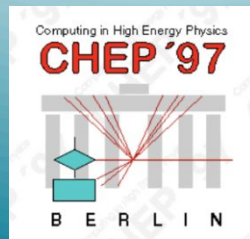
[Plenary](#)

[Parallel](#)

[Summary](#)

[Poster](#)

[Vendor](#)



- **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 data (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>

# CHEP 2023 Scientific Program

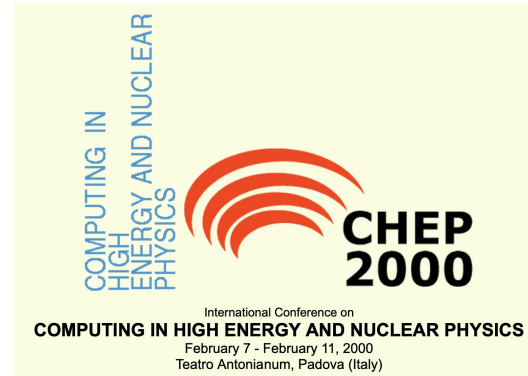
- Track 1 - Data and Metadata Organization, Management and Access
- Track 2 - Online Computing
- Track 3 - Offline Computing
- Track 4 - Distributed Computing
- Track 5 - Sustainable and Collaborative Software Engineering
- Track 6 - Physics Analysis Tools
- Track 7 - Facilities and Virtualization
- 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/>

# 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.

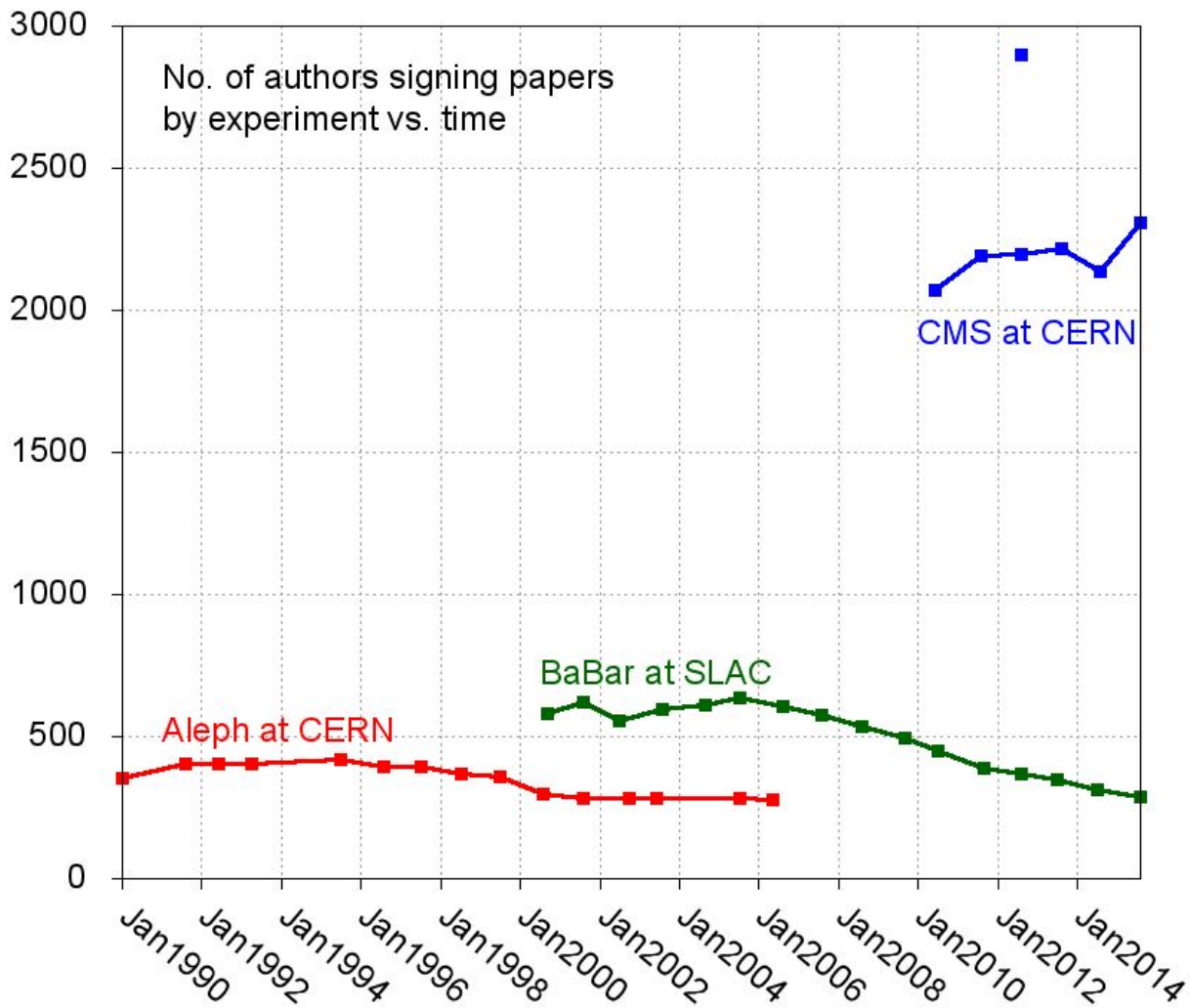
“Conway’s Law” effect? - the way an organization communicates and is structured will be directly reflected in the systems and software it creates

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.**”

Particle Physics  
Experiments as  
“collaborative  
research efforts”

Project Size  
Over Time



# Community 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]**

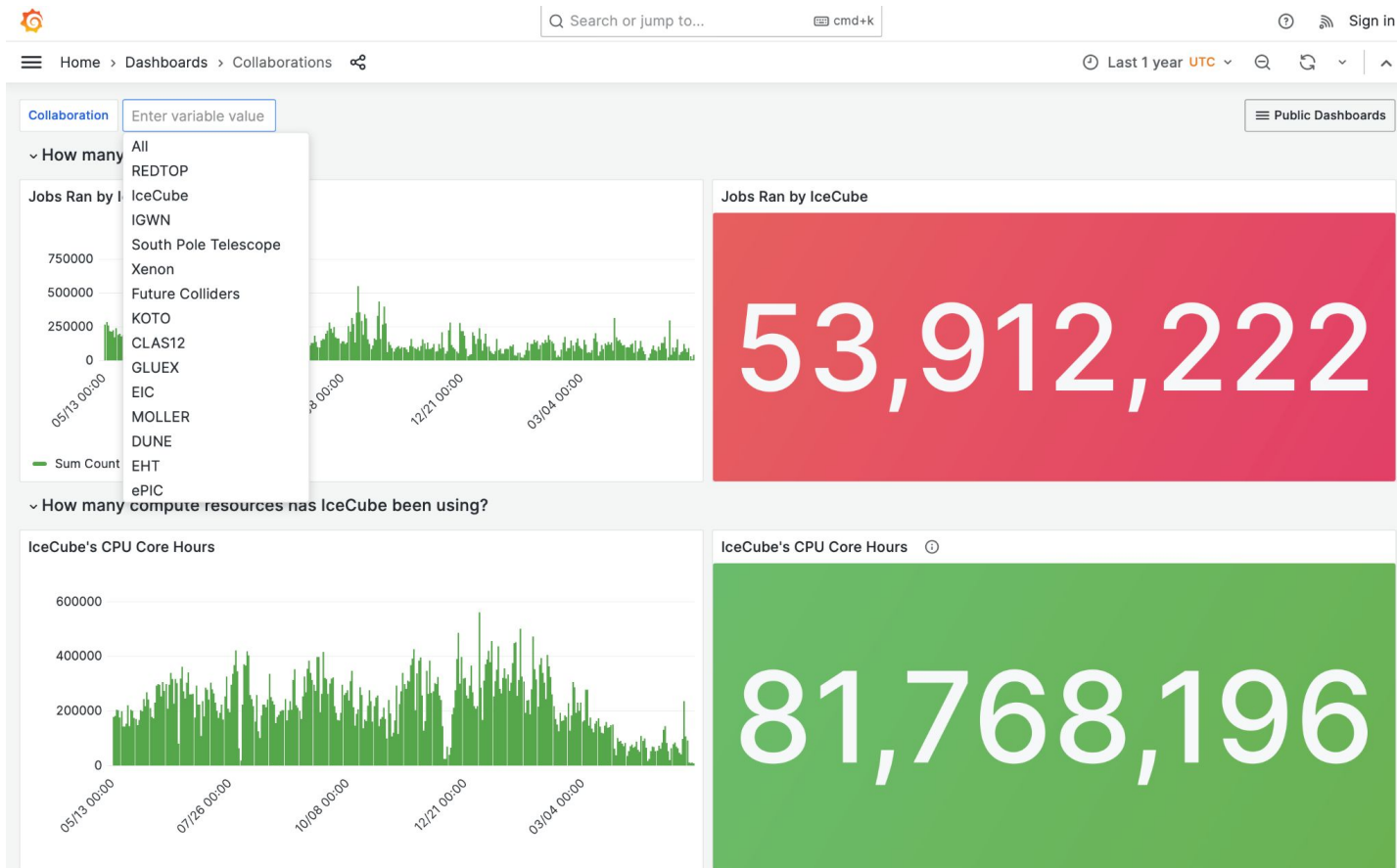


The **HEP Software Foundation** facilitates cooperation and **common efforts** in High Energy Physics software and computing internationally. **[2010's era]**

A banner for the WLCG/HSF Workshop 2024. The background is a blurred image of a city street with buildings and a church spire. On the left, the HSF logo is displayed. In the center, the WLCG logo is shown. To the right of the logos, the text 'WLCG/HSF Workshop 2024' is written in a large, bold, dark font. Below the main text, on the left, the dates '13-17 May 2024', the location 'DESY', and the time zone 'Europe/Zurich timezone' are listed. On the right, there is a search bar with the placeholder text 'Enter your search term' and a magnifying glass icon.

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

# OSG Support for non-LHC experiments

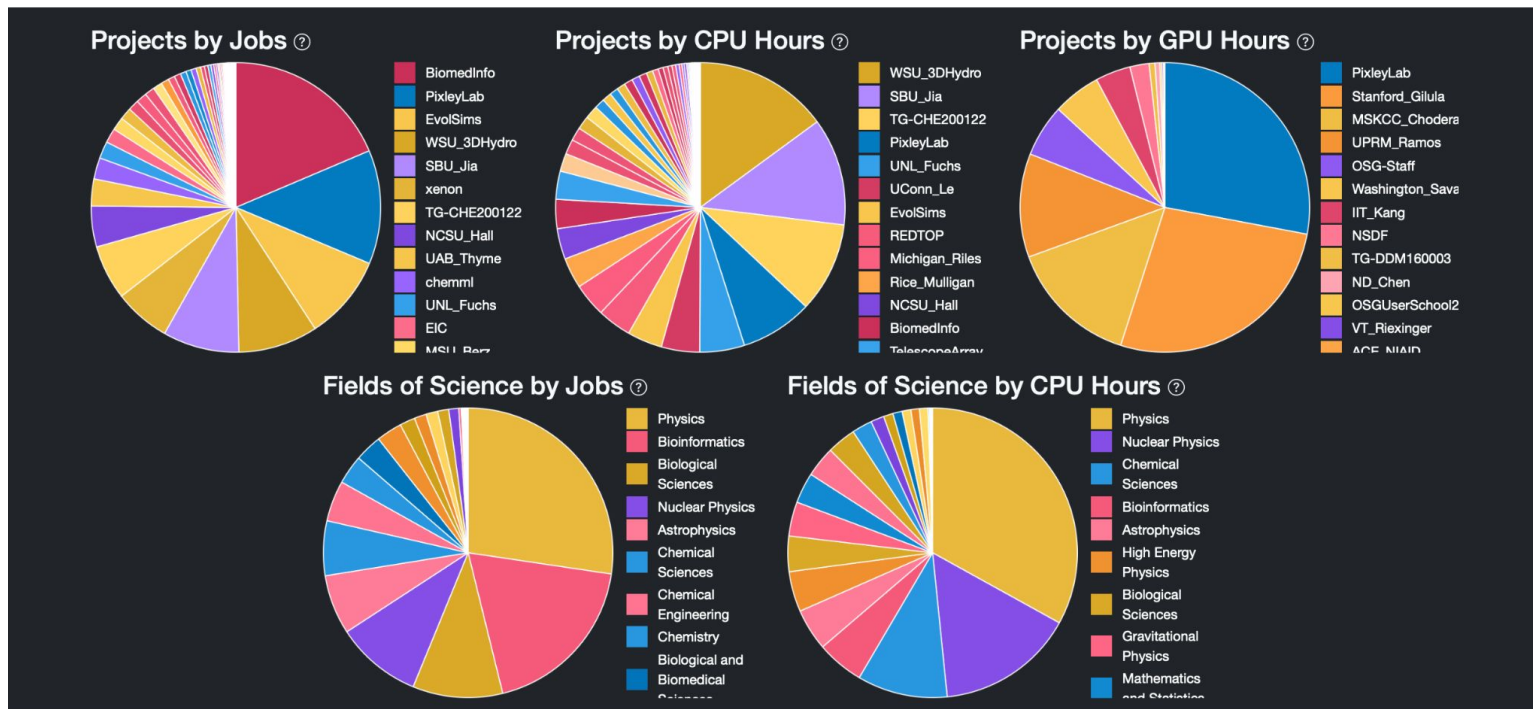




# OSPool Active Projects 104

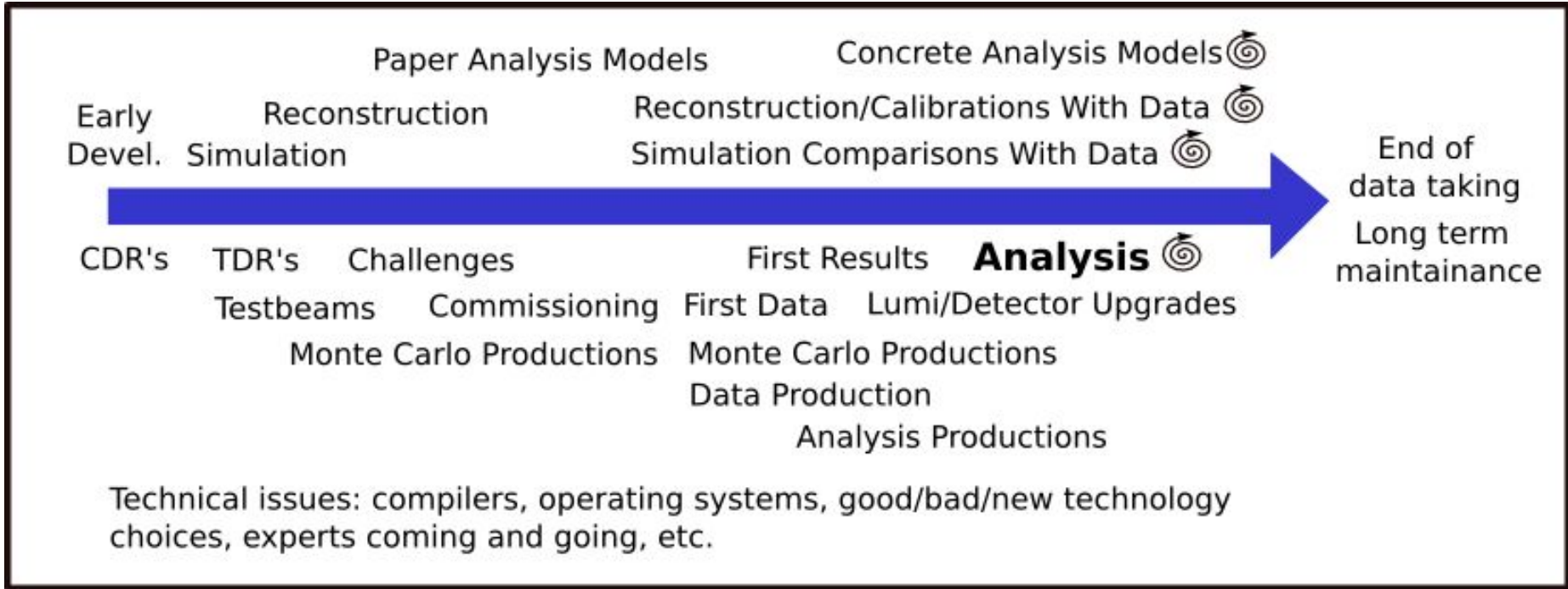
Data updated: 07/03/2024, 19:08:52

The below projects used OSPool resources to advance their research in the past year and ran more than 100 jobs. To run your own research on the OSPool sign up now on the [OSG Portal](#).



The [Open Science Data Federation](#) is built on the [Pelican Platform](#) (which uses xrootd), part of the NSF-funded National Discovery Cloud for Climate (NDC-C)

# Software Lifecycle in High Energy Physics

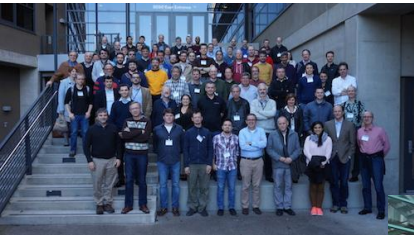


Community charge from the Worldwide LHC Computing Grid to the (then nascent) HEP Software Foundation in July 2016:

- 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/](https://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](https://arxiv.org/abs/1712.06982)

[arXiv 1712.06592](https://arxiv.org/abs/1712.06592)



IRIS-HEP

# Conceptual motivations behind the HEP Software Foundation

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

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.



# 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.

# APS DPF and Coordinating Panel for Software and Computing

## Snowmass CompF recommendation

We recommend the creation of a standing **Coordinating Panel for Software and Computing (CPSC)** under DPF, mirroring the panel for advanced detectors (**CPAD**) established in 2012.

*Purpose: Promote, coordinate, and assist the HEP community on Software and Computing, working with scientific collaborations, grassroots organizations, institutes and centers, community leaders, and funding agencies on the evolving HEP Software and Computing needs of experimental, observational, and theoretical aspects of the HEP programs. The scope should include research, development, maintenance, and user support.*

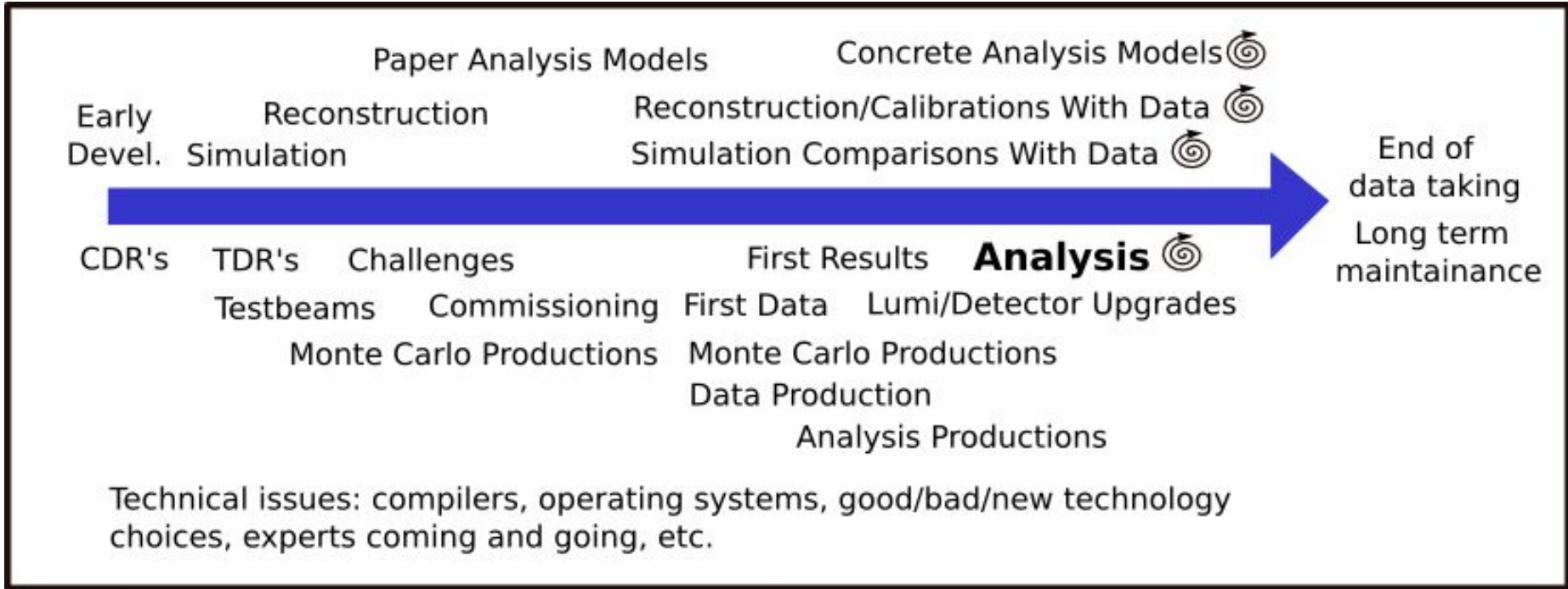
Further details of the community vision for the CPSC can be found in the body of this report.

A “Formation Task Force” has been appointed as an ad-hoc committee by the DPF EC, with the goal of producing a document similar in purpose to the one that defines CPAD’s charge, governance, internal organizational structure, and some initial activities, including some awards programs and community meetings.

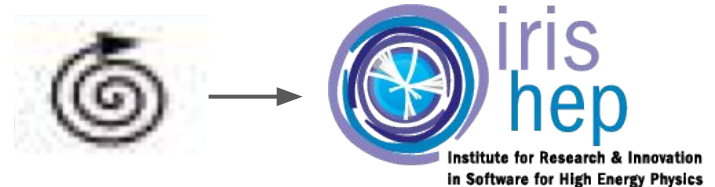
See [status report](#) on the FTF by Ian Fisk yesterday



# Software Lifecycle in High Energy Physics



[The Life Cycle of HEP Offline Software](#), CHEP 2007





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



<http://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:



#### 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](#)

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.

## HL-LHC Software and Computing Gaps

The four software and computing gaps are:

- G1. **Raw resource gaps:** The HL-LHC dataset will be enormous. Event complexity and count will each go up by about an order of magnitude. If no improvements to algorithms or resource management techniques are made, the HL-LHC experiments will simply be unable to process and store the data necessary for the science program.
- G2. **Scalability of the distributed computing cyberinfrastructure:** It is insufficient to buy cores and disk alone – the cyberinfrastructure used by the experiments must also scale to support the volume of hardware. This challenge is especially acute when it comes to data transfers: both the software must be ready and the shared networking resources (e.g., ESNet in the US) must be appropriately managed.
- G3. **Analysis at scale:** Analysis at the HL-LHC will be markedly different for two reasons: (a) the scale of the datasets involved and (b) the use of next-generation techniques (such as the latest machine learning techniques) to increase the scientific reach of each result. The former will require users to heavily utilize dedicated ‘analysis facilities’, optimized for high data rate I/O and the latter will require new services and data management techniques to be developed.
- G4. **Sustainability:** HEP is a facilities-driven science - the cyberinfrastructure assembled for an experiment must last or evolve on the decadal scale. This limits some strategies to cyberinfrastructure - for example, it is impossible for LHC to “do it yourself” and own the entire software stack. Specific sustainability strategies must be implemented even at the R&D phase to ensure that the cyberinfrastructure put in place at the beginning of the experiment is one the community can afford.

# Analysis Grand Challenge

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

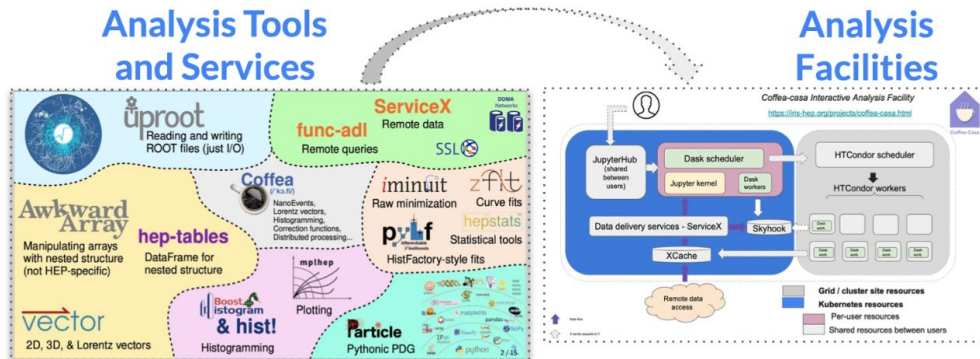
- **columnar data extraction** from large datasets,
- processing of that data (event filtering, construction of observables, evaluation of systematic uncertainties) into histograms,
- statistical model construction and statistical inference,
- relevant visualizations for these steps,

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.

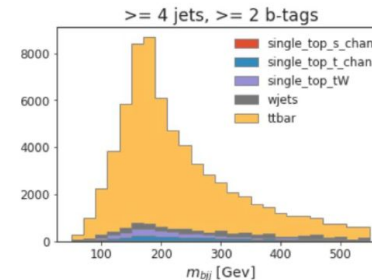
Timeline	Fraction of HL-LHC dataset processed in 1h
Year 2	20% (40 TB)
Year 3	50% (100 TB)
Year 4	75 % (150 TB)
Year 5	100% (200 TB)

See also [“Demonstrator Analysis 200 Gb/s” \(B.Bockelman\)](#) at WLCG/HSF



Execution of AGC analysis benchmark

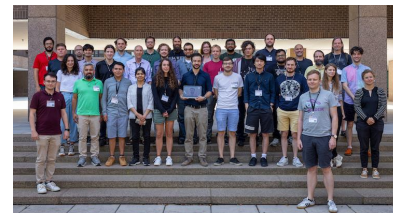
Reconstructed observables



example output of [analysis notebook](#)

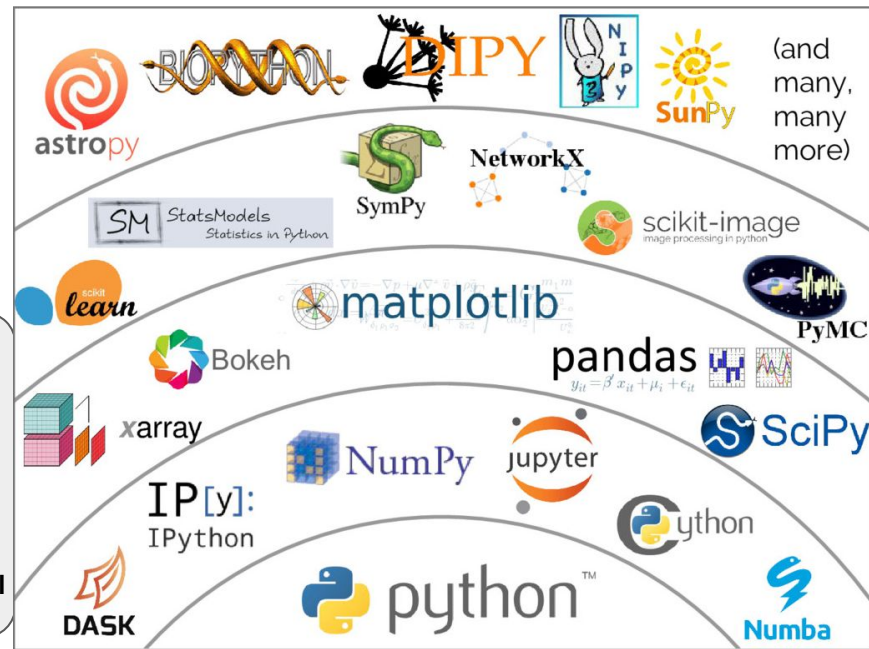
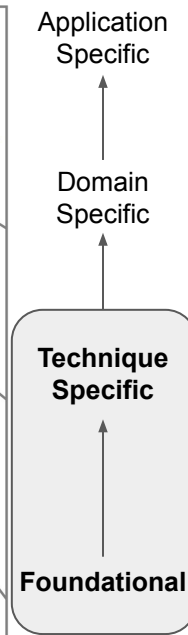
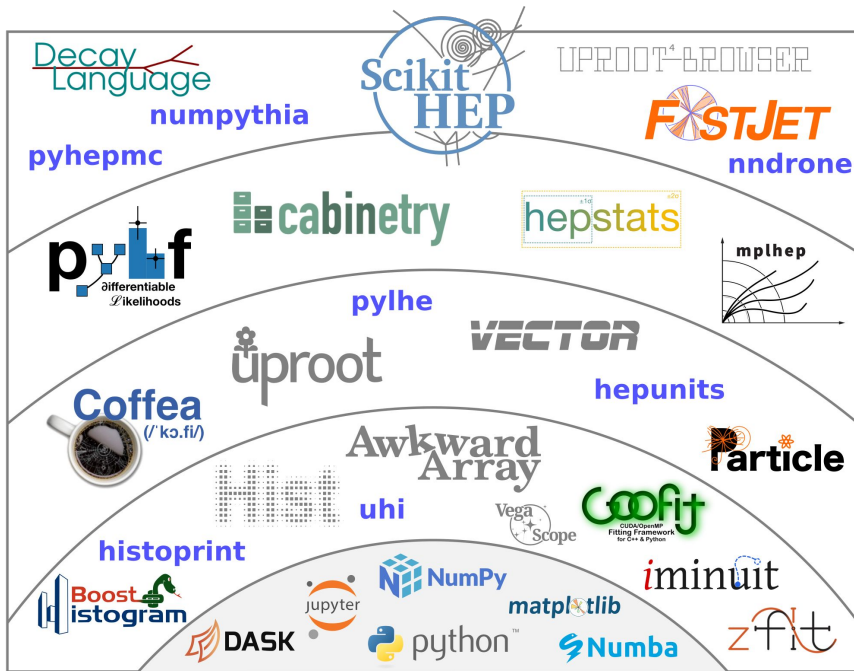
# PyHEP Ecosystem (Scikit-HEP)

[PyHEP.dev 2023 workshop](https://pyhep.dev)  
40 participants



Our scientific Python development vision/ecosystem

Scientific Python / PyData vision/ecosystem

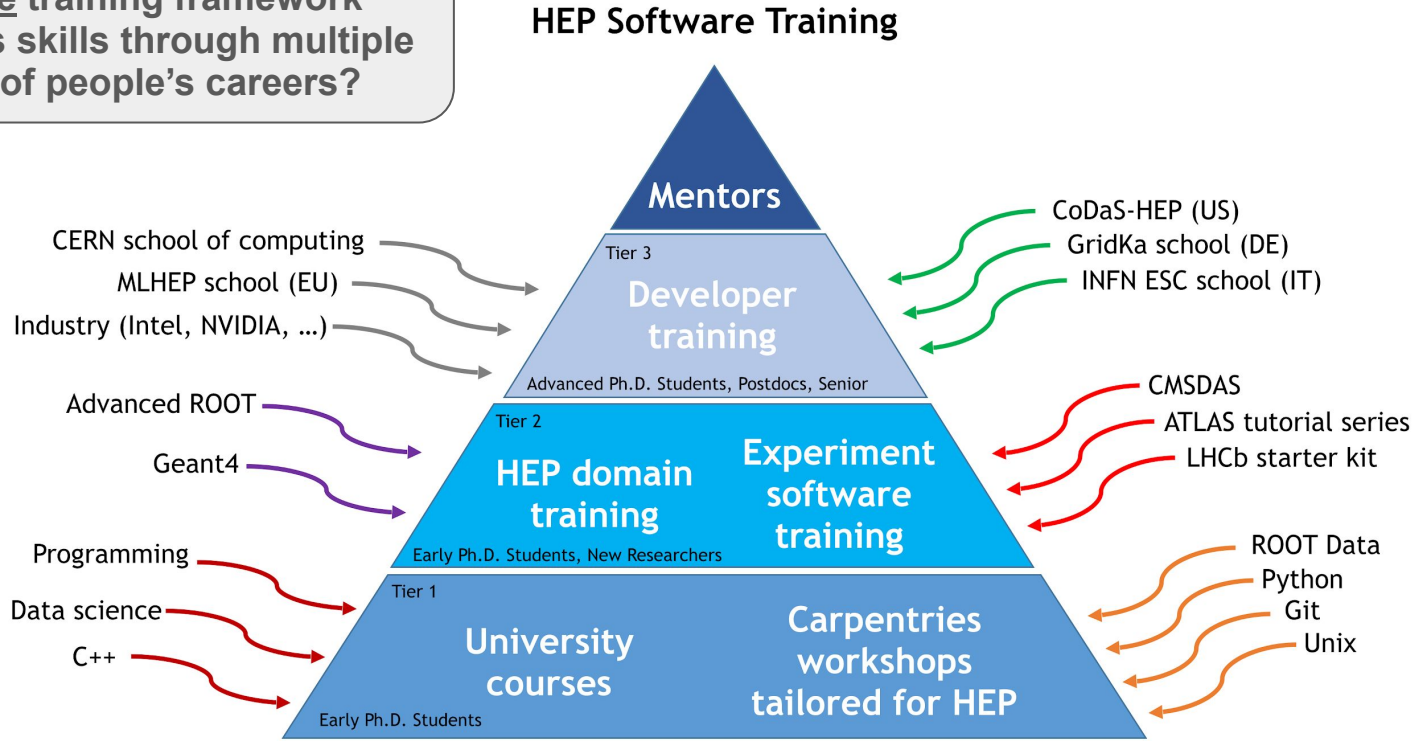


The “Future of Computing”  
is of course people to do  
it: software training!



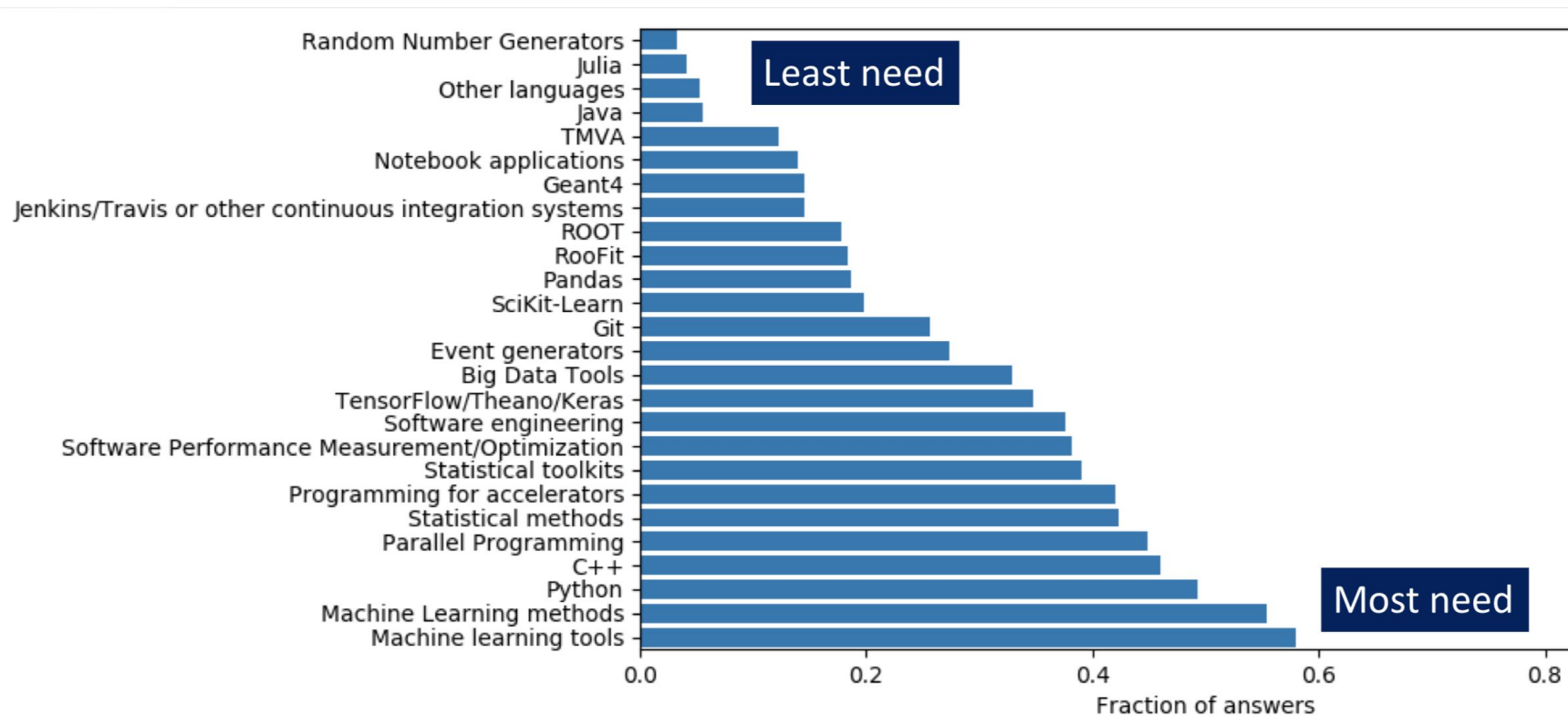
# Training and Education Model

How to build a sustainable and scalable training framework that grows skills through multiple stages of people's careers?



# Training Survey

In early 2019, we did a survey of training needs ([link for results summary](#)), 334 people responded!



# Multiple software, computing, data science and AI/ML programs aiming at different career stages

- Undergraduate summer programs: [US-CMS PURSUE](#) program, [US-ATLAS SUPER](#) program, [IRIS-HEP Fellows program](#)  
- [HSF/IRIS-HEP Training activities](#) (materials and events) - 1600 students and 50 educators in the past few years 
- [A3D3 PostBac program](#) 
- Summer schools: [IAIFI](#), [CoDaS-HEP](#) 
- US-ATLAS and US-CMS postdoc R&D programs  
- DOE CompHEP Traineeship projects: [TAC-HEP](#), [WATCHEP](#), [C2-The-P2](#) 
- International: Fellows/mentoring in NSF-funded [HSF-India](#) project, [HSF Google summer of code](#), CERN summer students, etc.   



# HSF Software Training Center

## Idea

Training in software and computing are essential ingredients for the success of any HEP experiment. As most experiments have similar basic prerequisites (Unix shell, Python, C++, ...) we want to join our efforts and create one introductory software training curriculum that serves HEP newcomers the software skills needed as they enter the field, and in parallel, instill best practices for writing software.

The curriculum is comprised of a set of standardized *modules*, so that students can focus on what is most relevant to them.

## The modules

### Basics

#### The UNIX Shell

A guide through the basics of the file systems and the shell.

📖 Start learning now!

🔧 Contribute!

#### SSH

Introduction to the **Secure Shell (SSH)**

⚠️ Status: Early development

📖 Start learning now!

🔧 Contribute!

#### ROOT

The most famous data analysis framework used in HEP

#### Version controlling with git

Track code changes, undo mistakes, collaborate. This module is a must.

📖 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!

#### Programming with python

Get started with an incredibly popular programming language.

📖 Start learning now!

🔧 Contribute!

#### Matplotlib for HEP

Make science prettier with beautiful plots!

⚠️ Status: Beta testing

📖 Start learning now!

🔧 Contribute!

# Central repository of training materials + instructor community - modeled on The Carpentries

## 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: Beta testing

📖 Start learning now!

📺 Watch the videos!

🔧 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!

Plus more...  
And growing...

## C++ corner

### HEP C++ Course

A full introduction to C++ based on a series of slides and exercises.

### Build systems: **cmake**

Building code is hard. **CMake** makes it easier.



1500 students/50 instructors to date.

CoDaS-HEP 2017



# Computational and Data Science (CoDaS-HEP) Summer School

<http://codas-hep.org>

In-person summer school with Lectures & hands-on exercises:



- Parallel Programming
- Data Science Tools and Techniques
- Machine Learning - Technology and Methods
- Practical skills: performance evaluation, use of git for version control

CoDaS-HEP 2018



CoDaS-HEP 2019



CoDaS-HEP 2022





# 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



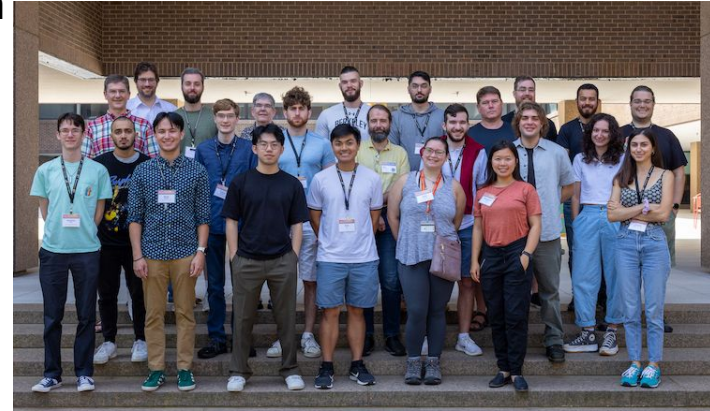
# DOE Traineeships in Computational HEP

New from 2023: 2 year traineeships for graduate students: coursework/training activities in software/computing plus dedicated software/computing R&D projects in collaboration with DOE lab staff, with an aim to develop the next generation of computational scientists and engineers.

4 funded projects: [TAC-HEP](#), [WATCHEP](#) and [C2-THE-P2](#) and [LGT4HEP](#) (see individual websites for universities and labs involved).

R&D topics span the experimental frontiers (energy, cosmic, intensity) plus lattice gauge theory.

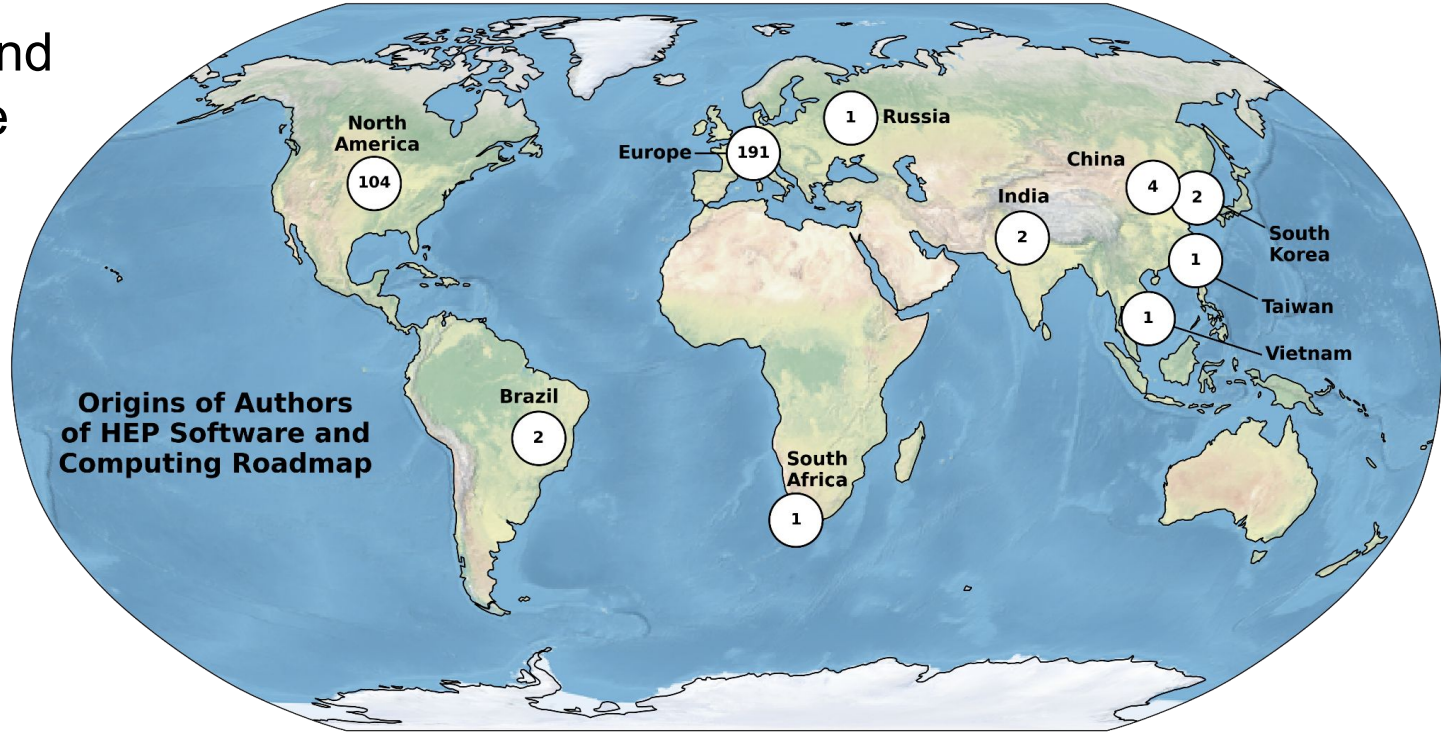
Beginnings of a cohort experience: Computational HEP Traineeship Summer School ([Princeton 2023](#), [FNAL 2024](#))



Princeton school, 2023

# 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)

# And the “Future of Computing”?

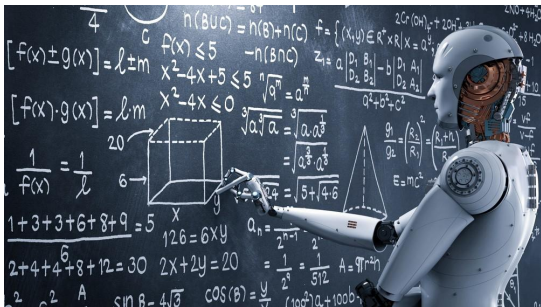


The **Worldwide LHC Computing Grid (WLCG)** project is a global collaboration of around 170 computing centres in more than 40 countries, linking up national and international grid infrastructures. The mission of the WLCG project is to provide global computing resources ... **[2000's era]**



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?