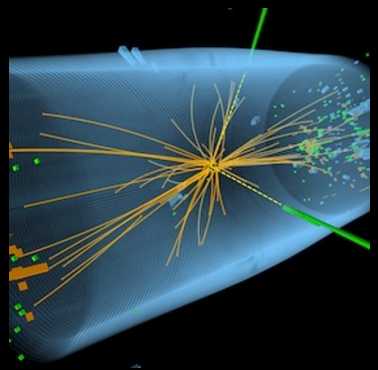
A Framework for Integrated Research Software Training in HEP

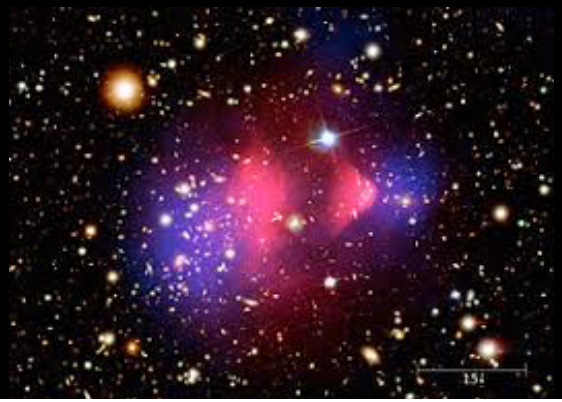
Sudhir Malik - Univ. of Puerto Rico Mayaguez

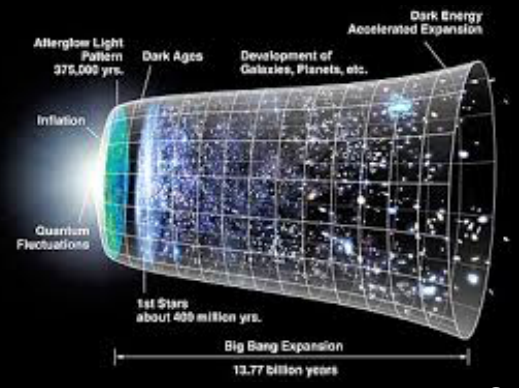
New Technologies for Discovery IV: The 2018 CPAD Instrumentation Frontier Workshop, 9-11 December 2018, Providence, RI

Science Drivers

- Use the Higgs boson as a new tool for discovery
- Pursue the physics associated with neutrino mass
- Identify the new physics of dark matter
- Understand cosmic acceleration: dark matter and inflation
- Explore the unknown: new particles, interactions, and physical principles
- Bases of operation of many current experiments as well as the design of the large, next-generation, facilities in HEP HL-LHC, LBNF at Fermilab, Super KEK-B and experiments - ALICE, ATLAS, CMS, LHCb, DUNE, Belle-

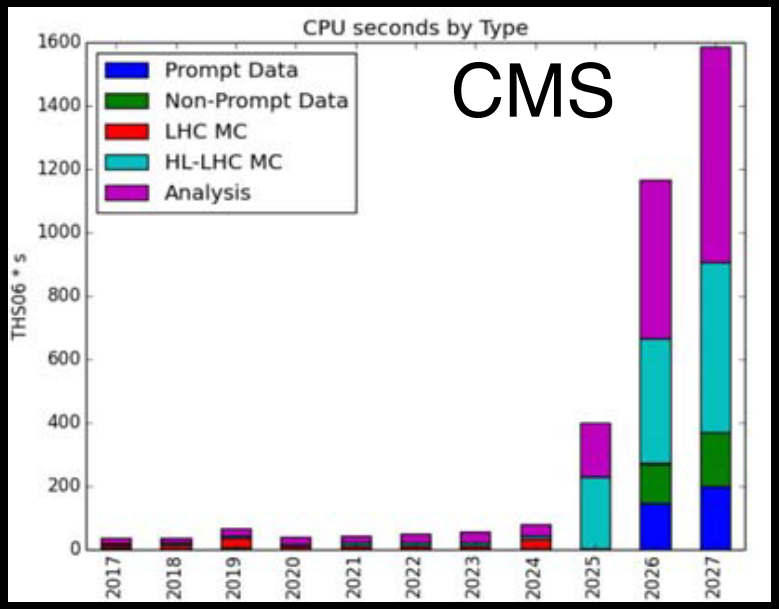


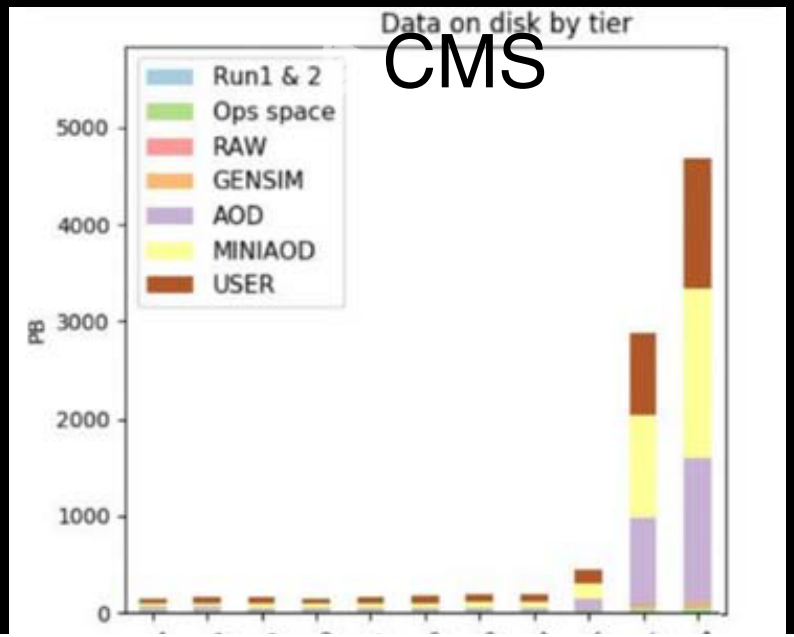




Computing Challenges in HEP

- Large data-intensive HEP experiments rely on
  - Significant data storage
  - High throughput computing
  - LHC Experiments - ~170 computing centers, nearly an exabyte of disk and tape storage and 750,000 CPU cores
  - HL-LHC
    - 100 billion proton-proton collisions per year
    - 10x or more computing needs, 100 times the data of (unto 2030s)
    - Other HEP facilities are planning similar increases in data volume

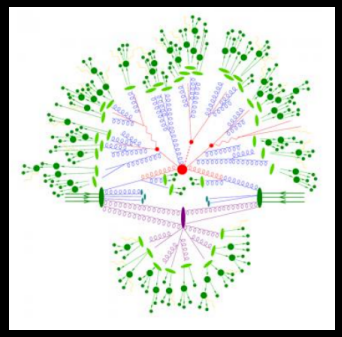




Questions guiding software & computing R&D

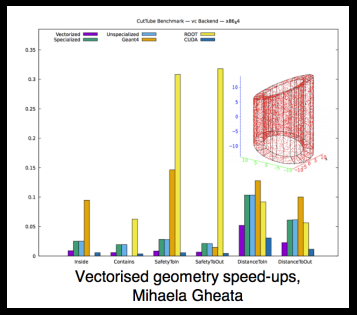
- **Physics:** Will efforts in this area enable new approaches to computing and software that maximize, and potentially radically extend, the physics reach of the detectors?
- **Resources/Cost:** Will efforts in this area lead to improvements in software efficiency, scalability and performance and make use of the advances in CPU, storage and network technologies, that allow the experiments to maximize their physics reach within New tools requires investing in training
- **Sustainability:** Will efforts in this area significantly improve the long term sustainability of the software through the lifetime of the HL-LHC?

HEP software ecosystem and priorities



Physics Event Generators

Data, Software,
Analysis Preservation

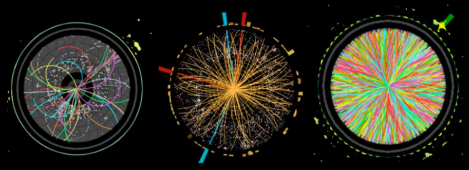


Detector Simulation

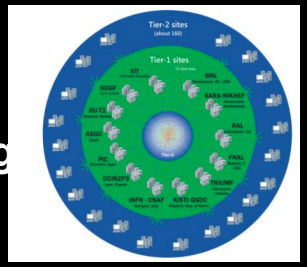
Security

Software Development

Trigger,
Event Reconstruction



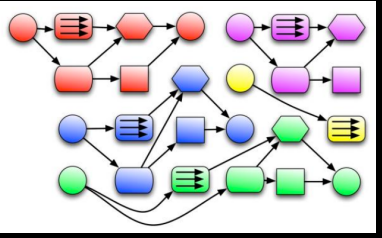
Facilities,
Distributed Computing



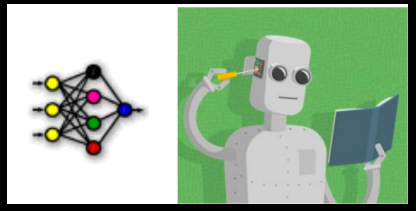
Data Analysis
Interpretation

Visualization

Data Processing
Frameworks



Machine Learning



Data Management
Organisation Access

Sustainability

- Successful evolution of this ecosystem to meet the challenges, new tools etc requires a workforce with a combination of HEP domain knowledge and advanced software skills
- Investment in SW critical to match HL-LHC requirements of “flat budget” scenario
- Investing in training leads to preservation and propagation of knowledge
- Investing software skills is not only important to actually build the requisite software infrastructure, but will also change community norms, create role models and promote career paths
- Computation is a central element of 21st century science, and clearer career paths will provide a virtuous cycle of feedback to enhance the vibrancy of the training and workforce development activities

Current Training Practices

- Training activities today are fragmented and partially redundant
- Each project (experiment, laboratory, etc.) is left to reinvent most aspects of research software training from scratch, the result is duplicative and often incomplete
- Most training activities are carried out “locally”, with specific objectives in the
- The modest effort devoted to training is not always positioned for maximum impact.
- The resulting activities are also quite difficult to sustain over time
- They are too often critically dependent on specific individuals whose careers evolve
- The effort to keep training materials up-to-date is too often lacking
- No single entity has a mandate to organize these disparate efforts into a collective effort whose impact would be much greater than the sum

Examples of Current Training

- CMS Data Analysis School (CMSDAS) by CMS experiment help new collaborators learn hands-on the complex set of software tools needed to work with CMS data and the distributed CMS computing system
- LHCb and ALICE joined efforts for basic entry level software training called StarterKit, Belle II has its own flavor of StarterKit
- ATLAS has tutorials setup for beginner, intermediate, developer level training via talks/recorded videos/workbook
- Neutrinos, Muons, Astro,DES community at FNAL Currently have Online Software Exercises and 1-2 day workshops aligned with collaboration meetings
- INFN School on Architectures, Tools andMethodologies for Developing Efficient Large Scale Scientific Computing Applications (ESC)
- CoDaS-HEP: School on tools, techniques Data Science for HEP

Training Vision

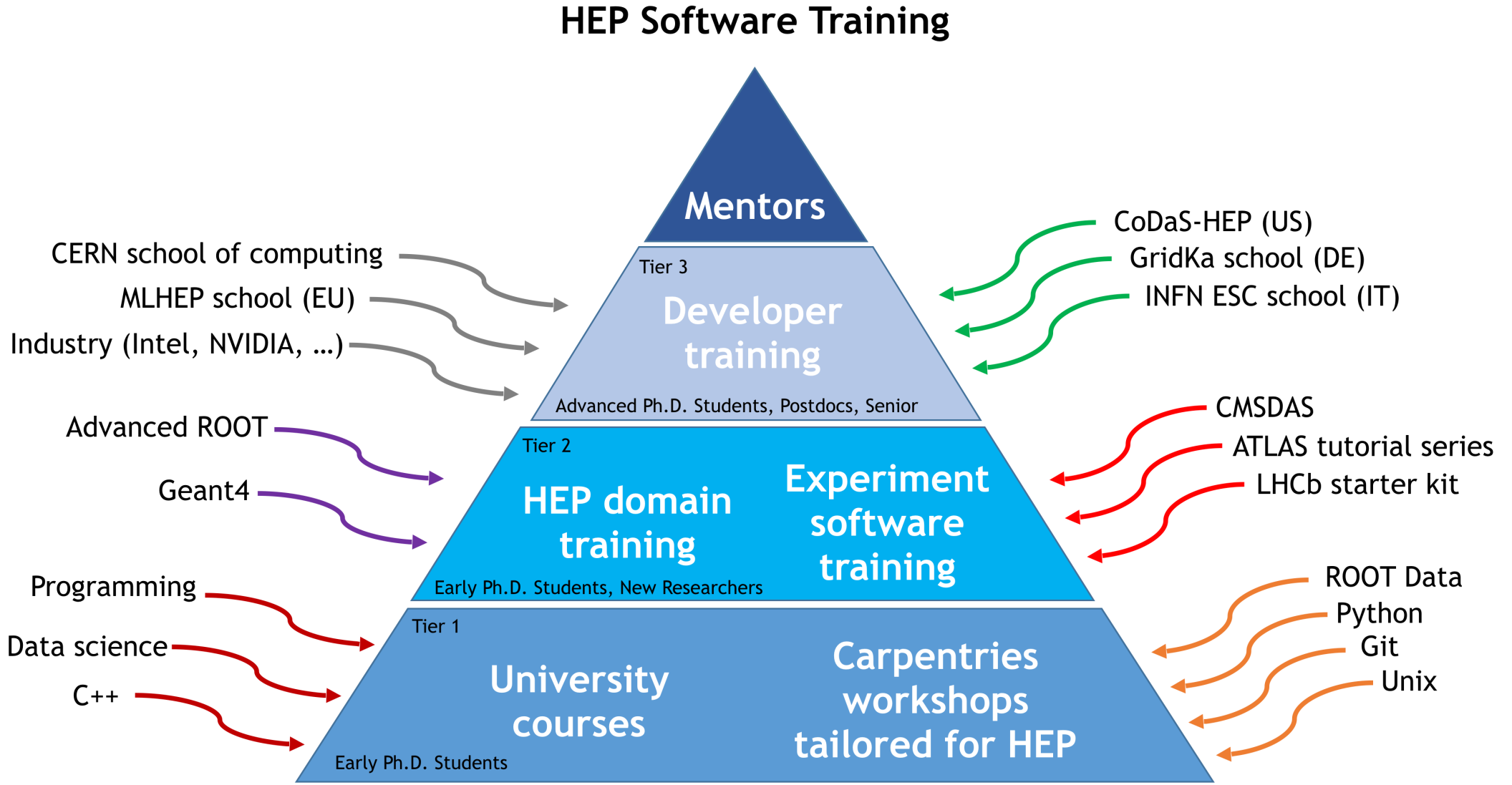
- Establish a community framework for software training in order to prepare the scientific and engineering workforce required for the computing challenges of HEP experiments
- For sustainability , not only build “users” community but also a “developer” community
- Not do significant new curriculum development itself, but will instead but leverage and build upon existing material from the HEP and larger research community
- Build training activities and material into a “common good” with a strong community of both instructors and participants, and with a feeling of community ownership

Training Goals

- To achieve this vision three major goals

  - To build a community consensus around this vision and a collaborative framework among national and international entities for implementing the vision
  - To develop a Carpentries-style introductory HEP software training curriculum, and a community of instructors, which is seen nationally and internationally as the entrance point into he overall training vision
  - To develop a more advanced training school within the U.S. for the relevant software skills (building on CoDaS-HEP) as well as an international federation of similar advanced training schools in HEP and beyond

Proposed Vision of Training Chain



Tier-1 (Basic curriculum)

- Establish training elements to provide a uniform set of basic skills for all Common basics (Linux/Shell/Python/C++/ Git/ROOT, Machine Learning) for all HEP graduate students and postdocs, and broadening participation from institutions lacking such courses

- Establish a collaboration with “The Carpentries” rather than develop its own curriculum from the ground up, assemble an introductory HEP curriculum built out of existing Software Carpentry material and augmented by HEP-specific material training material.

- The collaboration with The Carpentries has two aspects
  - Organize workshops to provide all U.S.-based HEP students and postdocs the opportunity to participate in a set of training course
  - To collaborate with senior HEP researchers to become instructors for The Carpentries workshop
- Lead to a community converging on a “common good” curriculum for the introductory material

- IRIS-HEP/ FIRST-HEP has funds used to train the instructor pool

  - Workshops will be organized at HEP institutions, either with a sufficient local audience to reach a large community or in conjunction with other community events, e.g. conferences, computing related workshops, experiment collaboration meetings, etc. Some fraction of the workshops will be held at the major HEP laboratories (Fermilab, CERN, etc.) and others at universities

Tier-2 (HEP-domain specific training)

- HEP-domain specific training exist already
  - Appropriate for Ph.D. students,new researchers, mastered the basic skills, starting their research activities
  - Work with partners to recognize and solve common problems
    - Variability of basic skills knowledge, evolve the “First Steps” curriculum
    - Gives natural pool of possible instructors
  - Organize “Birds of a Feather” sessions at HEP conferences
    - To increase the impact and sustainability of training activities
    - Can combine ROOT/ Geant4 training to more visible intermediate-level “boot camp
    - Make these tools more accessible outside of HEP?
    - Modest fund for a pilot intermediate “boot camp” in U.S. on above/HEP-related tools

Tier-3 (Developer Training)

- For researchers
  - Mastered basic and experiment specific skills
  - Ready for research contribution/activities
  - Develop innovative new methods
- Week long school
  - Like-minded students, similar interests in computational problems
  - Build collaborations, interact and find senior mentors professional development
  - Example - CoDas-HEP, DIANA project

http://codas-hep.org http://diana-hep.org

NSF Grants to Support the Vision

- "*Framework for Integrated Research Software Training in High Energy Physics (FIRST-HEP)*" - http://first-hep.org
  - Funding for Training Workshops/Participant Support/ Brain Storming Session

- "*Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)*" - http://iris-hep.org
  - Funding for Training Activities/Participant Support
  - Fellowships to work on Software and Computing Projects
  - Job Opportunities

HEP Software Foundation (HSF)

- The HEP Software Foundation formed in 2015 facilitates cooperation and common efforts in High Energy Physics software and computing internationally
- Sharing expertise; raising awareness of existing software and solutions; catalyzing new common projects; promoting commonality and collaboration in new developments to make the most of limited resources; aiding developers and users in creating, discovering, using and sustaining common software; and supporting career development for software and computing specialists
- It has strong training component to build HEP domain knowledge + advanced software skills+ strong connections to other related disciplines

CWP - Training

- https://arxiv.org/pdf/1807.02875v2.pdf

HEP Software Foundation Community White Paper Working Group – Training, Staffing and Careers

Dario Berzano,[1] Riccardo Maria Bianchi,[2] Peter Elmer,[3] Sergei V. Gleyzer[4] John Harvey,[1] Roger Jones,[5] Michel Jouvin,[6] Daniel S. Katz,[7] Sudhir Malik,[8] Dario Menasce,[9] Mark Neubauer,[7] Fernanda Psihas[10] Albert Puig,[11,a] Graeme A. Stewart,[1] Christopher Tunnell,[12] Justin A. Vasel,[10] and Sean-Jiun Wang[4]

[1] *CERN, Geneva, Switzerland*

[2] *Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA, USA*

[3] *Princeton University, Princeton, NJ, USA*

[4] *University of Florida, Gainesville, FL, USA*

[5] *Lancaster University, Lancaster, UK*

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

- https://arxiv.org/pdf/arXiv:1712.06982v3.pdf

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

HEP Software Foundation[1]

ABSTRACT: Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders

High Energy Software Foundation

- https://hepsoftwarefoundation.org

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

In 2017 the HEP Software Foundation, charged by the WLCG, produced a roadmap white paper on the software and computing challenges that will be faced during the next decade.

Community White Paper Reports

The roadmap summarised reports from fourteen working groups who studied the challenges in their sub-domains. All of the reports produced during the Community White Paper process are listed below. Working groups are in the process of finalising and uploading their work to arXiv.

Paper	*Report Number*	*Link*
CWP Roadmap	HSF-CWP-2017-01	arXiv
Careers & Training	HSF-CWP-2017-02	arXiv
Conditions Data	HSF-CWP-2017-03	Google Doc
Data Organisation, Management and Access	HSF-CWP-2017-04	arXiv
Data Analysis and Interpretation	HSF-CWP-2017-05	arXiv
Data and Software Preservation	HSF-CWP-2017-06	arXiv
Detector Simulation	HSF-CWP-2017-07	arXiv
Event/Data Processing Frameworks	HSF-CWP-2017-08	Github
Facilities and Distributed Computing	HSF-CWP-2017-09	Google Doc
Machine Learning	HSF-CWP-2017-10	arXiv
Physics Generators	-	No separate paper, see CWP Roadmap, section 3.1
Security	-	No separate paper, see CWP Roadmap, section 3.13
Software Development, Deployment and Validation	HSF-CWP-2017-13	arXiv
Software Trigger and Event Reconstruction	HSF-CWP-2017-14	arXiv - Executive Summary; arXiv - full document
Visualisation	HSF-CWP-2017-15	arXiv

Impact

- Coordinate training related activities
- Assemble and communicate coherent vision of a training program for HEP community - graduate students, postdocs, senior researchers in software and computing
- Develop a process with the community for implementing and updating this vision over time
- Build a “federated" view over the possible training opportunities
  - Experiments, labs, dedicated summer schools and other sources (HEP and non-HEP)
- Bring together the people organizing the training activities to articulate the vision, develop plans to enhance the sustainability, reusability and impact of the training activities
- Work with the community to build an assessment framework for the ensemble of activities that allows us to measure the impact of our activities

Related Links/Info

- IRIS-HEP website http://iris-hep.org/
  - Jobs on IRIS-HEP and Collaborating Projects http://iris-hep.org/jobs
  - General public announcement mailing list for IRIS-HEP events, talks, meetings, workshops, opportunities for training and job opportunities (subscribe to)
  - announcements@iris-hep.org

- HSF (HEP Software foundation) - https://hepsoftwarefoundation.org
  - General Information about HSF (subscribe to): hsf-forum@googlegroups.com
  - Discussions and activities in the HEP Software Foundation mailing lists can be found here (General and Dedicated Forums): https://hepsoftwarefoundation.org/forums.html
  - You can contribute https://hepsoftwarefoundation.org/cwp/cwp-working-groups.html
  - HSF Events/Workshops - https://hepsoftwarefoundation.org/events.html

- FIRST-HEP website http://first-hep.org
  - Funding for participants and lecturer support for Training