

Software Highlights and Review Preparation

Graeme Stewart (CERN EP-SFT) and Liz Sexton-Kennedy (FNAL)

LHCC Referees, 2021-08-31



Google Summer of Code

- This year Google changed the format
 - Only 175 hours of coding (roughly half that of previous years)
 - Lower level of interest from potential students, doubtless also a COVID effect
- However, still an impressive amount of projects for CERN-HSF Organisation
 - 36 proposals
 - 27 slots granted
 - 25 students successfully completed (93%)
- Highlights
 - Lots of work on automatic differentiation via CLAD
 - GPU abstraction evaluation and open source GPU partitioning
 - Data management plugins for Rucio with Science Mesh Cloud



CERN-HSF GSoC participation





CernVM

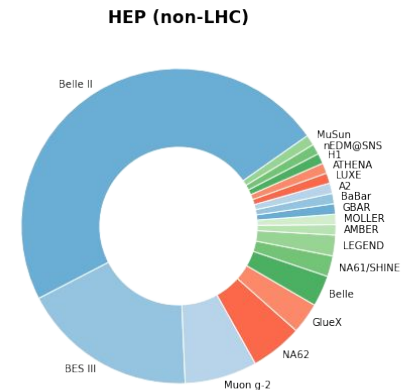
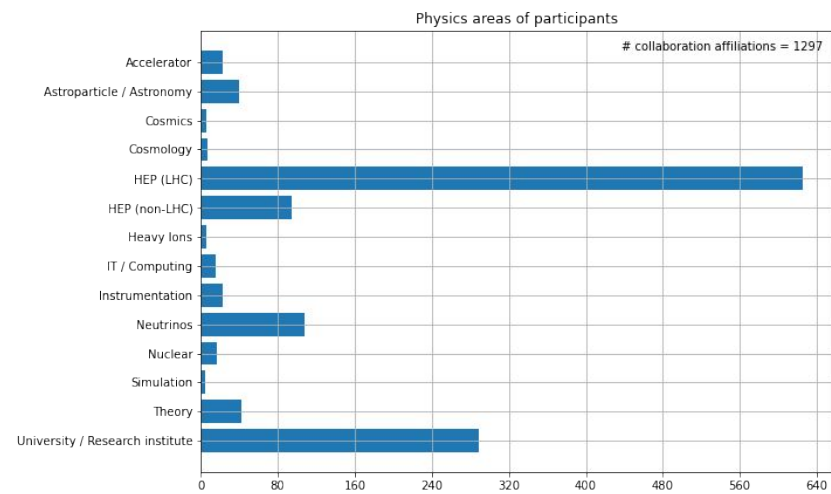
- Preparation of the 2.8.2 patch release (ETA September 2021)
 - Better diagnostics and client fixes for certain container workflows
- Preparation of the 2.9 feature release (ETA October 2021)
 - Various improvements to publishing speed and distributed publisher deployments
 - Various code consolidation works, e.g. refactoring of server-side code, additional code analysis in the CI, etc.
 - Support for S3 over HTTPS
 - The secure transport is not required by cvmfs but some S3 deployments only offer HTTPS, e.g. at RAL
 - In collaboration with Linux kernel developers: addition of kernel functionality necessary for certain containerized batch jobs using CernVM-FS, expected to be merged in the EL8 kernel ([details](#))
- GSoC project on pre-caching of files for faster cold-cache startup times of complex applications, e.g., TensorFlow
- Participation at the [DPHEP collaboration workshop](#)
 - With CernVM and CernVM-FS being enabling technologies for the long-term preservation of software and data analysis environments
- Challenging situation with developer effort, < 1 FTE of non-student effort, managing support, development and releases
 - Some easing from Autumn, new staff member starts, but certain items are delayed (improved OS X support, distributed publishing features, CernVM 5)

Software Process and Infrastructure

- Several LCG_100 (ROOT 6.24/00) based layers for ATLAS and LHCb; installations for MC Generator validation
- Preparing for LCG_101: gcc11, clang12, Python 3.9, ROOT 6.24/04
 - Plan to have a parallel installation created with spack
- Plan of Work
 - Progress on investigations of spack as build coordinator replacement
 - Use of kubernetes for build machines ([presentation](#))

PyHEP

- PyHEP 2021 workshop took place on 5-9 July, running for the 2nd consecutive year as an event online.
 - It attracted approximately 1300 registrants from a very broad spectrum of communities
 - LHC to neutrinos and astroparticle physics
 - Undergraduates and early-career graduates
 - Presentations and captioned recordings are available from [Indico](#)
- Having topical meetings on a monthly basis, modulo during the Summer break. The next meeting is already scheduled
- Plans for the year are all on track!



Data Analysis WG

- Metadata document to “motivate and codify a set of requirements on systems for storing analysis metadata for future experiments in High Energy Physics (HEP)”
 - Many contributions and comments received - finalising technical summary which we plan to publish
- Declarative analysis & analysis frameworks
 - Dedicated meeting with IRIS-HEP on analysis benchmarks (c.f. <https://github.com/iris-hep/adl-benchmarks-index>).
 - Will look to expand on current benchmarks including testing interfaces between parts of analysis workflow and more specific analysis techniques e.g. multi-dimensional fits that should be well received by experiments
- Overviews from non-LHC experiments
 - Four meetings in the “Analysis software in the wider HEP/nuclear community” series so far with contributions from the non-LHC experiments: SBN/ DUNE/ JeffersonLab/ sPHENIX/ FAIR/GSI/ LZ/ Xenon1T/ SuperNEMO/ BELLE II/ EIC
 - Series has been very informative and well received

ROOT



- Involved in a vivid student program: GSoC + summer student + sponsored non-member state students, follow-up from 2018 ROOT workshop in Sarajevo
- Significant training load, e.g., asked by "software carpentry" for ROOT training, asked by HSF for C++ training, ROOT training at Pakistan HEP School, CERN academic training; CERN technical training coming up
- RNTuple support for analysis data has been merged into CMSSW - a milestone for HEP's future I/O format! Significant development effort by CMS itself: very much appreciate investment and coordinated development.
- R&D progressing, for instance RNTuple data layout and object store support (DAOS), distributed analysis (distRDF); aiming to showcase results at ACAT, Nov 2021
- Considerable delays in developments related to machine learning, due to lack of (non-student) developers; hope to cover with external funding from ROOT / finance collaboration (HighLO project)
- Currently soliciting experiments' feedback on ROOT's review input documents
- ROOT will be revisiting (reminding and re-adjusting) 2021's plan of work next month; Sept 13-17, ROOT development will be put on hold for a project-wide focus week on improving ROOT's documentation

Reconstruction WG

- Survey in the community about interest for different topics (with feedback from non-LHC experiments)
 - Clear interest in heterogeneous computing, machine learning and 4D reconstruction
 - Also interest in other areas (see this [presentation](#) for details and plots)
- Organized short topical series on 4D reconstruction:
 - First topical meeting in June focusing on future upgrades with timing detectors for reconstruction
 - Follow-up topical meeting on algorithms in September / October
- Plan to organize a mini-workshop on heterogeneous computing aspects for trigger / reconstruction in the fall
 - Focus on both presentations of state-of-the art applications and on training & discussion sessions
 - Possibly co-organized with the training and framework WGs

Detector Simulation WG

- Since January we have held five topical sessions on:
 - Muon Collider & Detector Simulation
 - NEST (Noble Element Simulation Technique) software package
 - Using GPUs in detector simulations
 - Simulating solid-state microphysics with Geant4 and G4CMP
 - pyg4ometry: a package to create, load, write and visualise solid geometry for particle tracking simulations; GDML - Status and Plans
- Total of eight talks/speakers and a good attendance and lively discussions at each of the sessions
- For the near future, and as speakers become available, plan on
 - covering additional topics, e.g., related to AI and future colliders
 - expanding / updating for some of the more active ones already touched on

Simulation - Geant4

- Geant4 11.0-Beta released; final release expected in December
 - Consolidated Tasking system based on PTL v2.0.0 (or Intel/TBB)
 - Switch to C++17 as minimum standard
 - Configuration system supporting modular builds
 - Quantum entanglement of e+ annihilation
 - Updated electromagnetic integral approach
 - Deprecation/removal of legacy code/interfaces
 - Notes: <https://cern.ch/geant4-data/ReleaseNotes/Beta4.11.0-1.txt>
- New release VecGeom v1.1.16
 - New navigation optimization structure and GPU API extensions

Event Generators

- Event generator paper published: "Challenges in Monte Carlo event generator software for High-Luminosity LHC", [arXiv:2004.13687](https://arxiv.org/abs/2004.13687); [CSBS 5 \(2021\) 12](#)
- [WG meetings](#) concentrated on discussing the work detailed in the CSBS paper and for LHCC Review
 - [General](#) - 25 March
 - [Sherpa](#) - 22 April
 - [EvtGen and MG5_aMC GPU](#) - 6 May
 - [Pythia8](#) - 20 May
 - [MG5_aMC](#) - 27 May
 - [Herwig](#) - 17 June
 - [Powheg, LHAPDF/ Rivet / etc.](#) - 24 June
- Based on these discussions the draft event generators for HL-LHC computing review stage-2 is written and circulated for comments
 - Updated version of the paper is in preparation
- Significant improvement in the determination of CPU costs in CMS and ATLAS but work is ongoing ([initial report](#) from CMS)

Training WG

- 1100 people trained since inception of the training group in 2019
 - Partners with [IRIS-HEP](#), [FIRST-HEP](#), [the Carpentries](#), [ROOT](#)
 - Very grateful to all trainers, tutors and helpers for giving time to these important activities!
 - Nuclear Physics and Neutrino communities now actively participating
- Modules on the HSF curriculum being improved/updated
 - 2 hackathons in June/July 2021 - [Training Challenge](#), [C++ modules](#).
 - Training challenge focussed on scalability and sustainability of the training events, proposal to expand the training scope next 3 years, standardise surveys for feedback
 - C++ event focussed on partition C++ training into Basic, Intermediate and Advanced
- Software Carpentry workshops (SCW) will be organized every ~3 months.
 - Virtually organised the first [Software Carpentry workshop](#) 25-27 Aug 2021 (75 participants registered)
- C++ Training - [3rd Iteration of Course](#), 30 August - 3 September
 - 75 students, with 90 on waiting list; Pretty complete C++, intermediate level
- First publication on Software training (based on vCHEP 2021 contribution) submitted to Springer [Computing and Software for Big Science](#)
 - arXiv version is here - <https://arxiv.org/abs/2103.00659>

HL-LHC Review Documents

- Overall preparations are well underway
- Each of the proposed documents has an editorial team, as a reminder:
 - Event Generators
 - Detector Simulation
 - Foundation and Core Tools
 - Analysis
 - ROOT
 - Python Data Science
 - DOMA
- All documents have been delivered and are/have been commented on by experiments and community
- Final documents to be delivered to the LHCC Reviewers by 1 October

Summary

- Summer break slows down some activities
 - Especially communication and meetings
- But can be a good time for coding!
 - GSoC student activity was lively, despite changes
- LHCC document drafts are in decent shape
 - No showstoppers to delivering them on time
- Plans of Work are a good guide
 - Mostly on track and they always provide a useful compass for work done throughout the year