The HEP Software Foundation and Nuclear Physics

Torre Wenaus, BNL
Future Trends in Nuclear Physics Computing
Jefferson Lab
May 3, 2017

hepsoftwarefoundation.org

Check out the website!   hepsoftwarefoundation.org

HEP Software Foundation   Activities ▾   Communication ▾   Projects & Support ▾   Get involved!   About ▾

The HEP Software Foundation (HSF) facilitates coordination and common efforts in high energy physics (HEP) software and computing internationally.

The HSF is now beginning a community process to develop a consensus roadmap for HEP Software and Computing R&D for the 2020s. More infomation about this can be found on the Community White Paper (CWP) page on the HSF site.

📅 HSF Workshop on the HEP Analysis Ecosystem, Amsterdam, May 22-24 2017 (more info)

Meetings

All our activities and ideas are discussed weekly in our HSF meeting. Feel free to participate!

- HSF Weekly Meeting #90, April 27, 2017
- HSF Weekly Meeting #89, April 20, 2017
- HSF Weekly Meeting #88, April 6, 2017

Full list of meetings »

Newsletter

If you would like to stay updated, please subscribe to our newsletter:

- Community White Paper (CWP) Workshop, San Diego, 23-26 January 2017
- Third HSF Workshop
- Sharing ideas and code

Older newsletters »

Activities

Our plenty of activities span from our working groups, organizing events to supporting projects as HSF projects, and channeling communication within the community with discussion forums, technical notes and a knowledge base.

How to get involved »

HEP Software Foundation (HSF)

The HEP Software Foundation was created in 2014 as a means for organizing the HEP community to address current and future software challenges

- Why now? The software challenges as HEP computing scales up and adapts to new architectures are immense, while resources are tight
- "HSF facilitates coordination and common efforts in HEP software and computing internationally"
- A grass roots 'do-ocracy' (term borrowed from Apache Software Foundation) with activities driven by its members acting on community input
- Coordinated by a startup team led by P Mato (CERN), T Wenaus (BNL)
  - Startup team membership open to anyone interested & proactive
- Activities driven in a weekly meeting open to all
  - e.g. last week, meeting #90, 16 attendees planning an upcoming analysis tools workshop

HSF objectives

- Share expertise

  - White papers, peer reviews, topical workshops

- Promote commonality

  - Raise awareness of existing software and solutions

  - Catalyze new common projects

  - Promote collaboration on new projects to make the most of limited resources

- Support common software

  - Aid developers & users in creating, discovering, using, and sustaining common software

  - Act as a framework for attracting support to S&C common projects

- Support careers

  - Support career development for software and computing specialists

  - Serve as a training resource

- Facilitate wider connections with other sciences & communities

HSF scope

- HSF scope is what HSF participants make it

- Participants, and hence the activities, thus far weighted towards LHC

- But not exclusively, participation also from neutrino program, Belle II, ILC, Future Circular Collider (FCC)

- Some particle astrophysics, light source participation in workshops but haven't taken hold as participants

- The most natural extension of scope would be to our closest relation, nuclear physics
  - Already involved through ALICE

HSF events

- Formative workshop held at SLAC, January 2015, ~80 participants
  - Established objectives, priorities and activities subsequently launched
- Second workshop at LAL Orsay, May 2016, ~70 participants
  - Focused on active topics with the greatest community interest
  - Software performance, packaging, machine learning, community white paper
- First HSF-organized community meeting at CERN, October 2016, an assessment (a.k.a. review) of the GeantV R&D project
- HSF community white paper workshop at UCSD, January 2017, ~118 participants
- HSF visualization workshop at CERN, March 2017, ~34 participants
- HSF workshop on the HEP analysis ecosystem in Amsterdam, May 22-24 2017
- HSF community white paper workshop at Annecy, June 26-30 2017

HSF activities: Early priorities

Early efforts focused on the top priorities identified at the first workshop

- Software packaging and common project support
  - Working group surveyed nine packaging solutions from HEP, NP and open source
  - Came to unanimous agreement on one of them as the best basis for a software packaging tool for HEP: Spack
  - Originated in US HPC community (LLNL) to bring order to the great complexity of software configuration for HPCs (but not limited to them)
  - HSF WG members are now contributing to Spack main line development
  - Well suited to smaller experiments not bound to a legacy/in-house solution
- Software and computing knowledge base
  - To "Aid developers and users in creating, discovering, using and sustaining common software"
  - Operating at hepsoftware.org since early 2016, described in a talk at CHEP 2016
  - Offshoot of ATLAS 'data knowledge base' R&D
  - *Please contribute!*

# HSF activities: Supporting development

- Common project support
  - HSF produced a software best practices guide and a project starter kit with project templates
  - Activity will grow when more effort arrives -- CERN is making an HSF hire
- Software licensing
  - HSF wrote a document providing advice on licensing of HEP open source software, drawing on prior work and expertise at CERN, Fermilab and elsewhere
  - Cuts through a complex issue to provide relatively simple advice
  - e.g. ATLAS used it in deciding to use the Apache 2.0 license
- Training
  - Well suited to HSF since training has so much commonality
  - Many sources to filter, can draw on HSF expertise
  - Sadly not a lot of activity thus far; some work on developing WikiToLearn as a basis, with some embryonic collaboration with EP-SFT and ROOT
  - Some effort would have a very long lever arm to create value for experiments
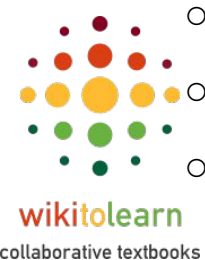
HSF activities: Planning and publishing

- HSF's [HEP software & computing white paper project]() is developing a roadmap for HEP S&C over the next decade
  - Serving as today's basis for long range LHC computing planning
  - Will serve HEP S&C funding proposals to NSF and DOE
  - Weighted towards LHC, but broader participation also: neutrino program, Belle II, linear collider so far. *NP would be a natural participant (e.g. via this workshop)*
- Participated in establishing Springer's new "[Computing and Software for Big Science]()" journal *now open to submissions*
  - Several HSF affiliated editors
  - Avenue for career recognition through peer reviewed S&C publications
  - Document our work and promote wider application
  - Welcomes papers at all scales, from big reviews to describing a clever algorithm
- HSF has its own [technical note series]() for its documents and reports

HSF activities: Sharing expertise

- Sharing expertise via peer reviews and topical workshops was identified early as a good role for the HSF
- HSF was asked last Spring by the GeantV simulation R&D project principals to conduct a peer review of GeantV
  - After 3 years of development, expose their work to the community, raise awareness, receive expert advice to guide the project's plans
  - HSF accepted and it took place in October, regarded as a big success by reviewees and reviewers
  - A panel of experts produced a detailed report
- That success catalyzed an HSF-organized workshop on the "HEP analysis ecosystem", which will take place later this month in Amsterdam
- More on both of these later

HSF affiliated activities - 1

- **Common tracking software forum**
  - To increase the exchange of experience in and software for track reconstruction
  - **Connecting the dots** workshop series
  - **ACTS**: ATLAS-initiated common project developing next-gen track reco software
- **Software technology R&D forum**
  - Open meeting series on new sw technologies, particularly concurrency
  - Recent topics: ReactiveX data-driven concurrent processing, streaming DAQ, compiled python
- **Google Summer of Code**
  - HSF as a GSoC sponsor (specifically CERN-HSF) is an avenue by which anyone in the community can propose a GSoC project
  - 39 organizations and projects participated via HSF, 36 project proposals
  - Awarded 26 slots (the maximum) by Google in 2017

HSF affiliated activities - 2

- **Inter-Experimental LHC Machine Learning Working Group**
  - Focused on development of state-of-the-art ML methods, techniques, practices… solutions, software and training beneficial to LHC and other HEP experiments

- **Gaudi framework** as a common project among LHCb, ATLAS, FCC, …
  - Sharing frameworks is difficult but sometimes successful
  - Choice quote from Amber at the January HSF workshop, describing the typical (not her!) attitude: "I'd rather use your toothbrush than your framework"

- **DIANA** (Data Intensive ANAlysis), US NSF funded project focused on analysis software

- Encouraging commonality on next-generation conditions database development
  - ATLAS + CMS + … LHCb, Belle II?
  - A second generation of the Frontier approach (REST web service fronted by squid caches and backed by a relational DB)
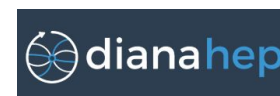
Common infrastructure components

- I was specifically asked to address common infrastructure components, the topic of this session -- what are they and how do they differ between NP and HEP
- Potentially, they are the full stack up to the experiment-specific analysis code
- Sociologically and practically, that is almost always impossible
- Past attempts to thicken the common part of the stack by fiat from above have not been very successful
  - Imposing common frameworks by fiat is even harder; cf. Amber's comment
- Successful efforts are seeded and grow from below, attracting a community
  - ROOT, Geant4, Gaudi live and prosper; SEAL and POOL do not
- HSF is directed at cultivating the bottom-up approach
  - Cultivate common efforts, expose them to light, the worthy ones will grow
- NP and HEP differences in common infrastructure components? There are none! (To a very good approximation)
  - Weight factors differ, in experiment size, data volume, etc, but both cover the spectrum, with event based analysis tasks that are fundamentally similar
  - When GlueX has a 10 PB/yr data volume, HEP and NP distinctions in data intensive analysis are pretty much gone

After that quick tour...

A closer look at three activities with NP relevance

- GeantV assessment

- HEP analysis ecosystem workshop

- S&C Roadmap community white paper

GeantV

geometry baskets

Basketizer

Filters | Geom (Fast) Phys

concurrent queues

in development

physics baskets

Scheduler

- An R&D project to re-engineer Geant4 for modern computer architectures, in particular to exploit vector processing

- Instead of Geant4's approach of tracking particles volume by volume through the detector…

- Gather baskets of tracks sharing locality criteria such as passing through the same volume and process them in parallel
  - Process many events concurrently to populate the baskets sufficiently for vectorization gains to outweigh the overheads

- While vectorized track transport has been the core objective of GeantV, the project has also developed modular components
  - starting with a new geometry modeller VecGeom
  - that can also be used within Geant4 (essential for validation)
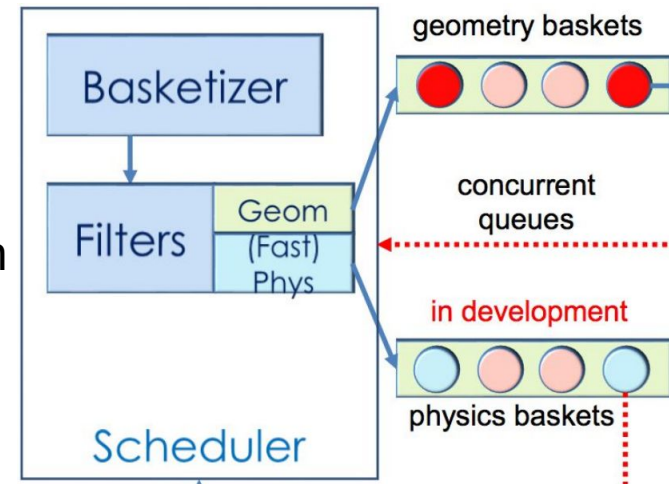
GeantV assessment

- HSF welcomed GeantV's proposal for a review because the relationship between GeantV and Geant4 has been confusing and sometimes contentious
  - (The US DOE HEP office wouldn't let us call it a review -- don't ask)
- Having GeantV assessed by experts primarily from the experiments was bound to bring some clarity and lots of good advice
- It was important that it was held outside the purview of a particular lab, agency, etc… but rather as a community activity: no 'institutional pulls'
- The outcome did clarify GeantV's status and role
- **Pre review**: GeantV's expressed goal was to have GeantV in the hands of the experiments as a distinct simulation toolkit from Geant4 in 2018, in time for experiments to evaluate it in LHC long shutdown 2
- **Post review:** the panel view was that *Geant4 will be the production simulation for the next ~10 years*

BROOKHAVEN

HSF

GeantV assessment - 2

- Panel found that vectorizing track transport by 'basketizing' tracks into geometry and (to come) physics baskets for parallel transport is far from being demonstrated as a capability or as having favorable cost/benefit
  - Devil is in the details: incorporating physics, complex magnetic fields, MC truth, ... while preserving vectorization gains is very difficult
- Panel concluded that GeantV's route to contributing value to the experiments is modules used via Geant4 as plugins (e.g. geometry, EM physics). GeantV should formalize a project plan for these components and be answerable for deliverables worked out collaboratively with Geant4.
- The review's conclusion clarifies the relative roles of Geant4 and GeantV, affirms the centrality of Geant4 for the experiments for the next decade, and provides guidance on where to put effort into simulation
- The CERN GeantV team presented a 2017 plan in January that was consistent with the review recommendations
- Panel will reconvene after 18 months for a follow-up

HEP analysis ecosystem workshop

- HSF is planning an [analysis ecosystem workshop](#) May 22-24 in Amsterdam
- ROOT sits at the heart of a rapidly evolving analysis ecosystem that is able to leverage ever more powerful open source tools
- How this ecosystem evolves will strongly impact the full processing chain
- ROOT itself is undergoing a major re-engineering with ROOT 7, leveraging the evolution of C++, with a major overhaul in its interfaces as seen both by its users and by satellite tools in the ROOT ecosystem.
- The workshop will examine the full ecosystem, ROOT's place in it, the roles of satellite projects around ROOT, and the roles of rapidly evolving open source software in data intensive analysis
  - Current ecosystem components, experiment analysis workflows, commonality
  - ROOT as the ecosystem hub, fulfilling that role effectively
  - Vision for the 5-10 year future, leveraging new technologies, analysis as a service
  - Missing pieces, R&D paths, learning from other fields

HEP analysis ecosystem workshop program

- Agenda as it is evolving here:
    - Today's analysis ecosystem landscape and toolset
        - Roles and plans for the tool set: ROOT, R, TMVA, XRootD, Diana, python ecosystem, browser based notebooks, ...
        - Analysis models of the experiments as they relate to a common ecosystem
        - Modularity, assimilating contributions
    - The 5-10 year vision
        - Analysis as a service: Browser based tools, streaming, storage & compute
        - Machine learning's role & inclusion in the ecosystem
        - Presentations on a range of analysis visions
    - Missing pieces
        - Where are the gaps, where is effort needed, R&D opportunities
        - Approaches, tools, ideas from other fields, industry
    - Conclusions, outcomes, actions, next steps
- Finish with a half day of writing and emerge with a ~complete report
    - Which will also be input to the community white paper...

BROOKHAVEN

HSF

Community white paper (CWP)

HSF is developing a [White Paper](#) on the strategy and roadmap for HEP S&C

- Initiated as a WLCG charge to the LHC experiments and HSF as a step towards the LHC experiment TDRs for HL-LHC

- Broadened to a community wide scope and effort

- Identify and prioritise the software R&D required to

  - achieve improvements in software efficiency, scalability and performance and to make use of the advances in CPU, storage and network technologies

  - enable new approaches that could extend the physics reach of the detectors

  - ensure sustainability of the software

- Develop the computing model vision and roadmap for 5-10 years out

- Promote commonality in the work, which is essential to securing funding

CWP - 2

- US funding agencies have taken a keen interest in the CWP
    - CWP is already explicitly directed at one potential funding channel, NSF S2I2, for a 'software institute' funded for 5 years if successful
- Participation in the CWP is in the material interest of anyone seeking support for HEP S&C
    - And NP S&C? You decide!
- It is the answer to admonitions from funding agencies and oversight bodies that the experiments should pursue planning and development in common wherever possible
- It bolsters the case for seeking S&C funding -- only common efforts will succeed

CWP - 3

- Kickoff workshop was at UCSD January 23-26

- Organized around topical working groups to write chapters, white papers

- Beyond the writing, use the process to catalyze common efforts

- Next (probably not final) workshop in Annecy June 26-30

Active groups:

Computing Models, Facilities, Distributed Computing

Detector Simulation

Event Reconstruction

Visualization

Data Access and Management

Security and Access Control

Machine Learning

Conditions Database

Event Processing Frameworks

Physics Generators

Monitoring

Not so active:

Triggering

Workflow and Resource Management

Data Analysis and Interpretation

Data and Software Preservation

Software Development, Deployment and Validation/Verification

Careers, Staffing and Training

Math Libraries

Technical Evolution (Software Tools, Hardware, Networking)

(Greater) NP participation is welcome!

BROOKHAVEN

HSF

Summarizing some of the potential for NP participation in HSF

- Community white paper: the S&C roadmap for the next 5-10 years
  - CWP will have NP relevance, and has some NP participation now (ALICE)
  - This workshop's report would be a useful contribution
  - Developing and documenting the roadmap will be an ongoing activity
- Broad topical/technical overlap makes technical workshops NP-relevant
  - e.g. visualization and analysis ecosystem workshops
  - NP-HEP commonality & collaboration could itself be a topical seed for workshop(s)
- Community peer reviews NP would like to see?
- Google Summer of Code project proposals are welcome
- Contribute to software packaging and common project support particularly for smaller experiments
- Contribute to the knowledge base!

BROOKHAVEN

HSF

Getting plugged in: mailing lists

- HSF open forum
  - http://groups.google.com/d/forum/hep-sf-forum, 274 members
  - Principal communication channel for activities, meetings, events
    - Also re-posts of jobs, conferences, schools etc.
- Specialized lists (created as needed)
  - Community white paper - hsf-community-white-paper
  - General HSF technical discussion forum - hep-sf-tech-forum
  - Packaging - hep-sf-packaging-wg
  - Training - hep-sf-training-wg
- HEP S&C community forum
  - http://groups.google.com/d/forum/hep-sw-comp, 345 members
  - General mailing list everybody involved in HEP S&C should subscribe to
- See the 'Get involved' page on the website for details

Finally

- HSF has advanced, slowly but materially, to a spectrum of activities that covers to varying degree the objectives laid out at its formation
- Running so far on the best-effort enthusiasm of a few but will gain its first dedicated FTE soon at CERN
    - The lightweight bottom-up organization is so far working well
- HSF activities and events attract a healthy level of participation
- Strongly endorsed by the LHC's oversight and scrutiny bodies, and supported by the experiments
- Small but growing participation beyond the LHC
- Ready to grow and ready to better engage NP with its similar if distinct S&C profile, if and when there is interest

Thank you

Thanks to everyone who is contributing their time to the HSF and/or supporting the HSF and its goals

Thanks to those who have contributed to the materials presented here

- Peter Elmer
- Benedikt Hegner
- Michel Jouvin
- Pere Mato

Supplementary

Startup team (as of recently)

Amber Boehnlein - Jefferson Lab
Peter Elmer - Princeton University
Daniel Elvira - FNAL
Frank Gaede - DESY
Benedikt Hegner - CERN
Michel Jouvin - LAL,IN2P3
Pere Mato - CERN, co-lead
Andrew McNab - Manchester
Dario Menasce - INFN
Elizabeth Sexton-Kennedy - FNAL
Graeme Stewart - Glasgow
Craig Tull - LBNL
Andrea Valassi - CERN
Brett Viren - BNL
Torre Wenaus - BNL, co-lead

HSF objectives revisited with their relevant activities

- Share expertise
  - CWP; GeantV review; analysis ecosystem workshop; topical workshops
- Raise awareness of existing software and solutions
  - Knowledge base; reviews; workshops
- Catalyze new common projects
  - Spack as common packaging solution; broadening of conditions DB participation
- Promote commonality and collaboration in new developments to make the most of limited resources
  - Project & packaging support; topical forums in tracking & concurrency; CWP
- Aid developers and users in creating, discovering, using and sustaining common software
  - Knowledge base; project & packaging support
- Support career development for software and computing specialists
  - Springer journal; training
- A framework for attracting support to S&C common projects
  - CWP as mechanism for funding proposal(s); possible ECFA directed action
- Facilitate wider connections with other sciences & communities
  - Springer journal; LHCC encourages us to do more here