

HSF DAWG ESPPU input



Alexander Held (University of Wisconsin–Madison)
Jamie Gooding (TU Dortmund)
Nick Smith (Fermilab)

Dec 16, 2024

<https://indico.cern.ch/event/1486671/>



This meeting

- HSF plans to provide input to the **European Strategy for Particle Physics**

Update (ESPPU), as an update to the 2017 Community Whitepaper (“CWP”)

- Similar to the US **Snowmass** process resulting in the **P5 report**

<https://www.usparticlephysics.org/2023-p5-report/>

- Want to discuss our plans for contributions on data analysis

The timeline: ESPPU

- See <https://europeanstrategyupdate.web.cern.ch/process-0>



The timeline: HSF input

- HSF will submit a **community input document** by **March 2025**
 - Will cover event generation, detector simulation, reco & trigger, data analysis, training & careers
- Will be short: **10 pages total** , so only about 2 pages of content for data analysis
- Aim is to provide an update wrt. the 2017 CWP
- HSF DAWG conveners will provide draft for feedback & subsequent endorsement in January
- This meeting today is an opportunity to identify topics of interest to our community
 - Please share your views!
 - Can still provide input to us by end-of-year via email: nick.smith@cern.ch, jamie.gooding@cern.ch, alexander.held@cern.ch

Related documents

- pyhep.dev summary <https://arxiv.org/abs/2410.02112> (2024)
- AF white paper: <https://arxiv.org/abs/2404.02100> (2024)
- Julia white paper: <https://arxiv.org/abs/2306.03675> (2023)
- P5 report: <https://www.usparticlephysics.org/2023-p5-report/> (2023)
- AEWII workshop summary: <https://arxiv.org/abs/2212.04889> (2022)
- CWP <https://hepsoftwarefoundation.org/organization/cwp.html> (2017)

Topics for discussion

Software evolution since 2017

- Major developments / overall direction in **ROOT**
 - RDataFrame, RNTuple, RooFit backends (including autodiff)
- **Python** ecosystem
 - Significant growth over past years in Scikit-HEP project and beyond
- Role of **Julia**
 - Active community & workshops, should keep monitoring
- **Interoperability** (“bridges and ferries” in CWP)
 - Lots of progress here building bridges between different approaches
 - e.g. HS3 standard ([github link](#)) for serialization of statistical models

Machine learning, automatic differentiation, etc.

- **ML is everywhere** and being used in an increasingly number of ways beyond classifying events
- **Targeted applications** of automatic differentiation to help with:
 - Minimization of statistical models
 - Systematics-aware NN training
 - Mixed classical/ML modeling (“digital twin”)
- Should **monitor this active field** , potential for big changes in the way we do physics analysis
 - e.g. recent simulation-based inference results from ATLAS + CMS
 - Role of foundation models in the future
 - Completely new ideas?

Sustainability

- Availability of **long-term funded positions** for software development & maintenance remains limited but is crucial for long-term sustainability
 - See also P5 recommendation
 - *Area Recommendation 19: Research software engineers and other professionals at universities and labs are key to realizing the vision of the field and are critical for maintaining a technologically advanced workforce. We recommend that the funding agencies embrace these roles as a critical component of the workforce when investing in software, computing, and cyberinfrastructure.*

Analysis facilities

- Dedicated recent white paper summarizing status
- Remains area of **active developments & discussions**
 - See also related LHCC [questions to experiments](#)
- Encourage work to continue with goal of providing convenient interfaces for physicists to accomplish their goals
 - Topics of interest: environment handling, data access, collaboration with teams, turnaround time, transition between local / AF resources

Additional topics & community input

- **Analysis description languages** : no major evolution in past years (?)
- **Documentation** / onboarding new users / tutorials / where to find support
 - Success stories like LHCb starterkit lessons, but should keep investing in this
- Which things deserve discussion that **haven't been raised yet** ?
 - What are the biggest challenges in user data analysis today?
 - Anything related to those challenges we should mention?

Organizational aspects

- Join our mailing list to stay in touch: hsf-analysis-wg@googlegroups.com ([sign-up](#))
- Ideas or requests for future meetings? Please get in touch with your suggestions!