Analysis Systems Updates

Matthew Feickert

University of Wisconsin-Madison

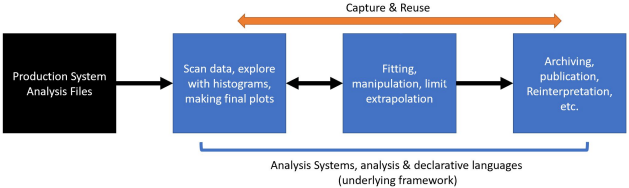
IRIS-HEP Steering Board Meeting #15
September 13th, 2022

Analysis Systems Team

- Wisconsin-Madison
  - Kyle, Alex, Matthew
- Washington
  - Gordon, Mason, Tal
- Princeton
  - Jim, Henry, Ianna
- University of Cincinnati
  - Mike, Thomas
- Illinois
  - Mark, Ben
- NYU
  - Irina
- University of Nebraska-Lincoln
  - Oksana

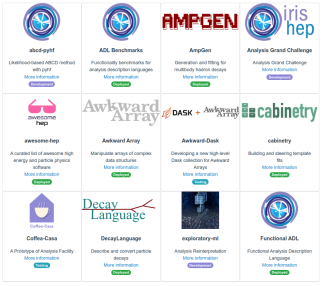
Analysis Systems Focus Area

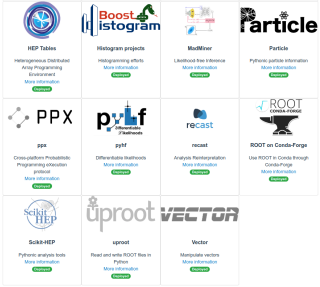
- Goal: Develop **sustainable** analysis tools to extend the physics reach of the HL-LHC experiments
  - **creating greater functionality**
    (extending the reach and abilities of analysts)
  - **reducing time-to-insight**
    (reducing complexity and accelerating expensive computation)
  - **lowering the barriers for smaller teams**
    (robust, deployable tools and workflows)
  - **streamlining analysis preservation, reproducibility, and reuse**
    (making reproducible analyses straightforward)



Analysis Systems Projects

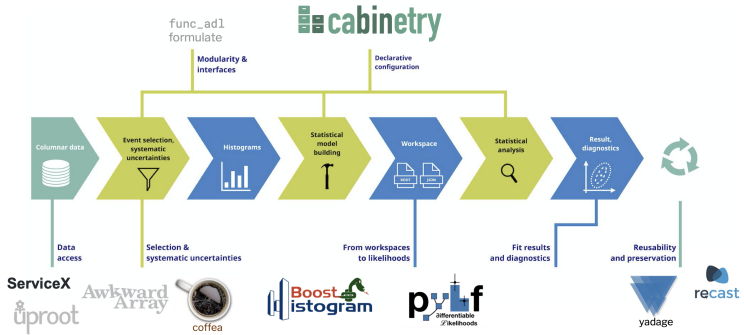
- Analysis Systems are connected to analysis use cases
- Systems are composed of components
- Large number of these projects refer to these components
  - Many projects include people beyond IRIS-HEP directly
- Milestones and activities are mainly oriented towards integration, evaluation, with a global overview of the vertical





Analysis Systems Projects

Analysis Systems Pipeline



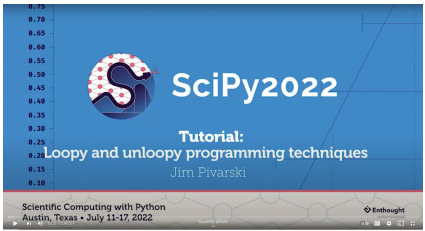
- Together these project compose to form a coherent **analysis pipeline**
- Interface with DOMA (ServiceX, Coffea-Casa, and friends) and Scalable Systems Laboratory (analysis facilities) areas, benefitting from Blueprint Activity process
- Underlying pipeline for the Analysis Grand Challenge (AGC)
  - c.f. Alex and Oksana's AGC update in Steering Board Meeting 12

Year 4 Themes

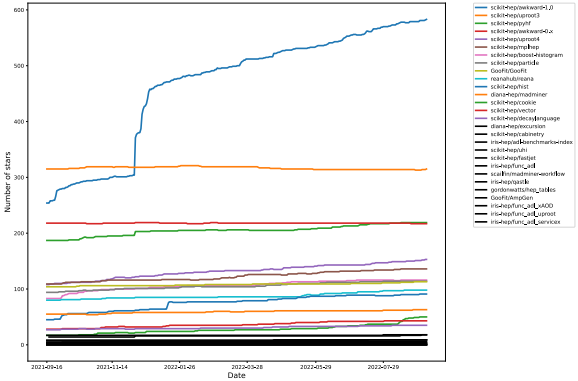
- Year 3 focused on **Integration and Adoption**
  - Getting tools to work coherently together and getting the community using them
- Year 4
  - **Project maturity / feature completeness**
    - (core) feature completeness of pyhf, cabinetry
    - Expanding high level API for hist for easier analysis tasks
    - Roadmaps for integration of Awkward to broader analysis communities (Awkward-Dask) and differentiable analysis
  - **Community socialization / supporting analysis use**
    - pyhf used in published analyses by ATLAS, Belle II, Belle, pheno community (35 use citations to date)
    - Scikit-HEP tools finding traction outside of HEP (Awkward, boost-histogram)
    - Scikit-HEP / IRIS-HEP tools and maintainers being invited to broader Scientific Python conferences and communities (SciPy Conferences, PyPA packaging summit)
    - Adoption of pieces of ecosystem making it easier for the whole (uproot ubiquitous, ATLAS investigating Daks for analysis workflows, coffea adopts hist, LHCb publishes first analysis using only Scikit-HEP tools)

Area Project Highlights: Awkward

- Awkward is used outside of HEP, and very much appreciated
- Awkward Array CSSI (OAC-2103945) has development of a new Dask container type representing Awkward Arrays



Part of Jim's SciPy 2022 Tutorial



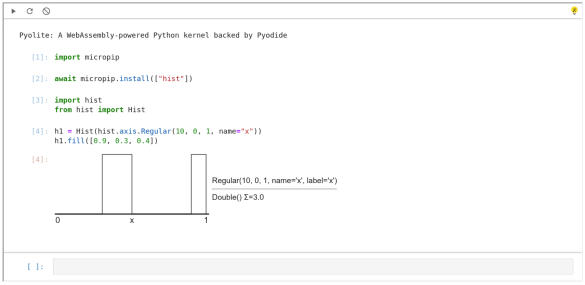
Cumulative GitHub stars of IRIS-HEP/Scikit-HEP projects vs. time. Awkward has the most by far.

Area Project Highlights: hist

- hist is being adopted as the histogramming library of coffea

> Additionally: this is the first release to include deprecation warnings for things that will go away with coffea v0.8.0 which is due early 2023. In particular `coffea.hist` will cross the rainbow bridge to Valhalla (having fought honorably these past 3 years). Please migrate to the scikit-hep `hist` package over the next 6 months! (edited)

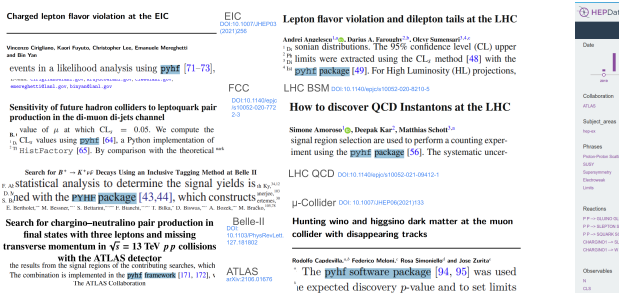
- Furthers idea of interoperable ecosystem where adoption of parts means better integration of the whole



(Pyodide CPython port to WebAssembly/Emscripten powering JupyterLite kernel)

boost-histogram and hist were early trials for Pyodide kernel improvements

Area Project Highlights: pyhf



Charged lepton flavor violation at the EIC

EIC
DOI:10.1007/JHEP03(2021)256

Vincenzo Cirigliano, Kaori Fuyuto, Christopher Lee, Emanuele Mereghetti and Bin Yan

events in a likelihood analysis using pyhf [71–73],

FCC
DOI:10.1140/epjc/s10052-020-7722-3

Sensitivity of future hadron colliders to leptoquark pair production in the di-muon di-jets channel

value of μ at which $\mathrm{CL_s} = 0.05$. We compute the $\mathrm{CL_s}$ values using pyhf [64], a Python implementation of HistFactory [65]. By comparison with the theoretical

LHC BSM DOI:10.1140/epjc/s10052-020-8210-5

Search for $B^+ \to K^+\nu\bar{\nu}$ Decays Using an Inclusive Tagging Method at Belle II

statistical analysis to determine the signal yields is ca ned with the pyhf package [43,44], which constructs

Search for chargino–neutralino pair production in final states with three leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector

the results from the signal regions of the contributing searches, which The combination is implemented in the pyhf framework [171, 172], v

The ATLAS Collaboration

Lepton flavor violation and dilepton tails at the LHC

Andrei Angelescu[1,a], Darius A. Faroughy[2,b], Oleve Sumensen[2,3,c]

sonian distributions. The 95% confidence level (CL) upper limits were extracted using the $\mathrm{CL_s}$ method [48] with the pyhf package [49]. For High Luminosity (HL) projections,

How to discover QCD Instantons at the LHC

Simone Amoroso[1], Deepak Kar[2], Matthias Schott[3,4]

signal region selection are used to perform a counting experiment using the pyhf package [56]. The systematic uncer-

LHC QCD DOI:10.1140/epjc/s10052-021-09412-1

Hunting wino and higgsino dark matter at the muon collider with disappearing tracks

Rodolfo Capdevilla,[a,f] Federico Meloni,[a] Rosa Simoniello[a] and Jose Zurita[a]

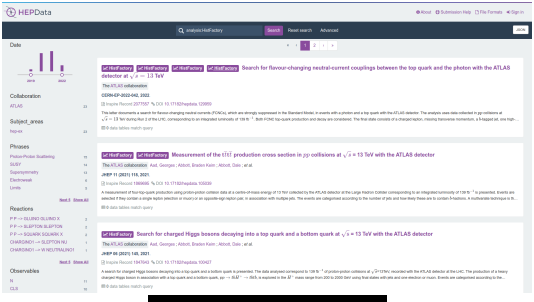
The pyhf software package [94, 95] was used

ie expected discovery p-value and to set limits

μ-Collider DOI: 10.1007/JHEP06(2021)133

Belle-II
DOI: 10.1103/PhysRevLett.127.181802

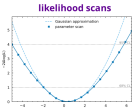
ATLAS
arXiv:2108.01676

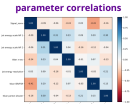


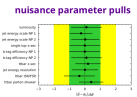
- pyhf adopted widely through the HEP community (35 use citations)
- HistFactory probability model publishing with pyhf JSON schema from ATLAS (23 published)
- Published HistFactory probability models get own DOI on HEPData (future: model render, interactivity) plus a nice badge

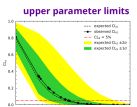
Area Project Highlights: cabinetry

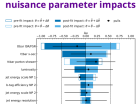
- Implementations for all **common inference tasks** exist
  - includes associated visualizations
  - results validated against **TRExFitter**





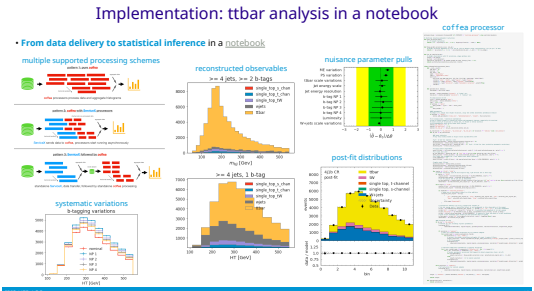






Alex Held, ATLAS SUSY Workshop 2021

- Uses pyhf as inference engine
- Provides common visualization for inference validation

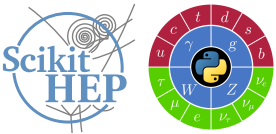


Alex Held, ICHEP 2022

- Steers inference and visualization of AGC

Area Project Highlights: Scikit-HEP

- IRIS-HEP supported community project with success in projects and maintainers
- PyHEP workshop series has drawn over 1000 participants annually since moving to Zoom
  - PyHEP 2022 Organizers: Eduardo Rodrigues, **Graeme A. Stewart**, **Jim Pivarski**, **Matthew Feickert**, Nikolai Hartmann, **Oksana Shadura**
- Forming bridges to broader scientific analysis communities (SciPy Conference, Scientific Python, NumFOCUS)



NUMFOCUS
2012-2022

Area Project Highlights: Scikit-HEP

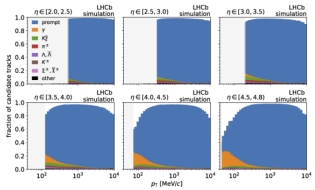
LHCb collaboration has published *JHEP* 01 (2022) 166 using only Scikit-HEP tools for the analysis

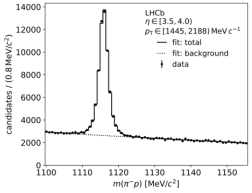
Scikit-HEP packages cover all aspects of analysis and working with IRIS-HEP to spread adoption

LHCb Publication Using Solely Scikit-HEP Tools

Post data-processing all performed with Python HEP tools!

- Uproot: Interfacing with input ROOT files
- boost-histogram: Replace classic TH* ROOT classes; Bonus: Multi-dimensional histograms!
- iminuit: User-friendly interface to minuit2 to process minimization
- PDF build in Python using SciPy library components





Nathan Grieser, IRIS-HEP AGC Tools 2022 Workshop

Area Project Fellows

Many excellent Analysis Systems Fellows! Focusing on two Summer 2022 projects:

Saransh Chopra: Improvements to vector API, docs, and benchmarking to bolster the quality of the library en route to first major release

Peter Ridolfi: Bidirectional translation between pyhf HistFactory JSON and CMS Combine DataCard + ROOT

Looking Toward Year 5 Goals

- Have ambitious goals of pursuing end-to-end differentiable Analysis Systems pipeline as part of AGC
  - Requires additional support be added to Awkward v2
  - Fully differentiable pyhf components (differentiable analogues to discrete features like bins)
  - Robust API for entire pipeline (ServiceX, func_ADL, HEP tables, Coffea, cabinetry, and pyhf)
- Foster software and maintainer sustainability and project life cycle plans
  - Legacy only works if there is a plan for maintenance and maintainers
  - Ongoing organization of Blueprint meeting on Software Citation and Recognition in HEP
- Work closely with AGC co-coordinators to have tight testing of workflows
  - Following IRIS-HEP Retreat in October aim to have better view of how AS can support AGC

Summary

- Analysis Systems projects are largely in a strong stage of software maturity
  - Have dedicated and diverse user base that help make the software battle tested!
  - Projects are supporting, interacting with, and being invited into the broader communities
- Looking to have software with longevity while also continuing development and support
  - Longevity in both technical stability and social support
- Goals of Analysis Systems and AGC are tightly aligned
- Similarly, continued close coordination with DOMA and SSL are necessary for success of AS pipeline