

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



Thomas Boettcher



Kyle Cranmer



Tal van Daalen



Irina Espejo



Matthew Feickert



Ben Galewsky



Alexander Held



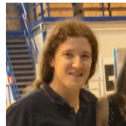
Mark Neubauer



Ianna Osborne



Jim Pivarski



Mason Proffitt



Henry Schreiner



Oksana Shadura



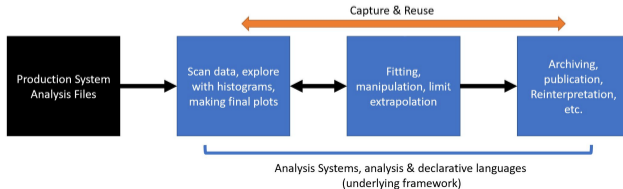
Mike Sokoloff



Gordon Watts

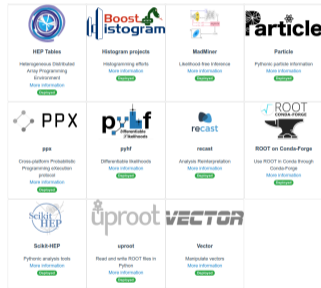
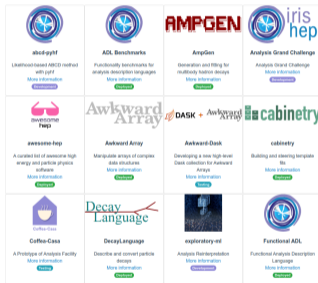
Analysis Systems Focus Area

- ▶ 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)



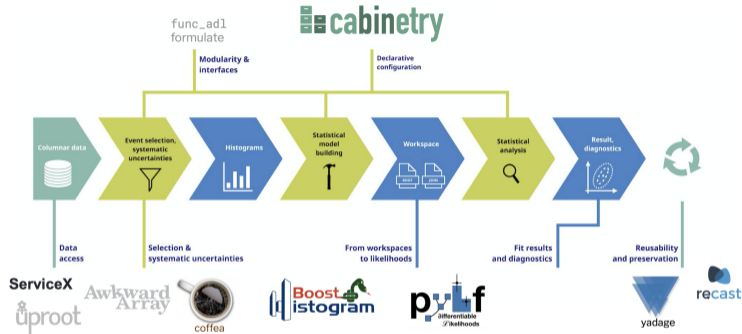
Analysis Systems Projects

- ▶ 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

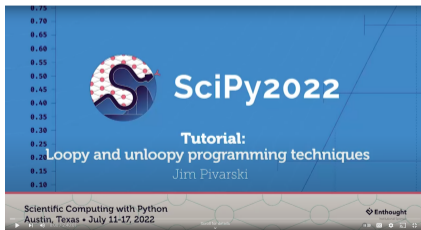
- ▶ Year 3 focused on **Integration and Adoption**
 - ▶ Getting tools to work coherently together and getting the community using them
- ▶ Year 4
 - ▶ **Project maturity / feature completeness**
 - ▶ (core) feature completeness of pyhf, cabinetry
 - ▶ Expanding high level API for hist for easier analysis 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

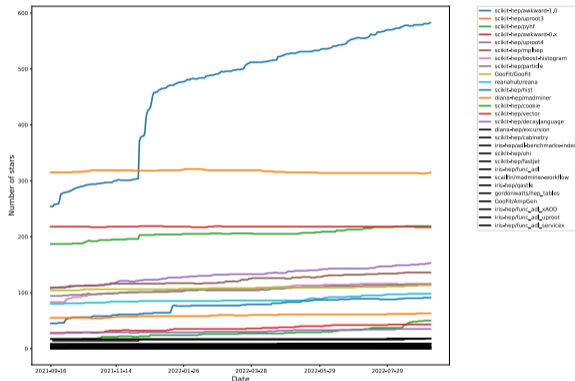
- ▶ Awkward is used outside of HEP, and very much appreciated



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.

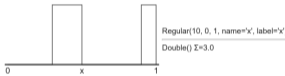
Area Project Highlights: hist

- ▶ 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)

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

```
Pyolite: A WebAssembly-powered Python kernel backed by Pyodide

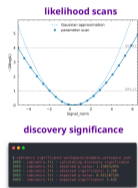
[1]: import micropip
[2]: await micropip.install(["hist"])
[3]: import hist
    from hist import Hist
[4]: h1 = Hist(hist.axis.Regular(10, 0, 1, name="x"))
    h1.fill([0.9, 0.3, 0.4])
[4]: 
```

([Pyodide](#) CPython port to WebAssembly/Emscripten powering [JupyterLite](#) kernel)
boost-histogram and hist were early trials for
Pyodide kernel improvements

Area Project Highlights: cabinetry

• Implementations for all **common inference tasks** exist

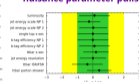
- includes associated **visualizations**
- results validated against **TRExFitter**



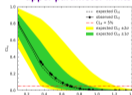
parameter correlations



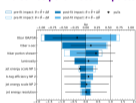
nuisance parameter pulls



upper parameter limits



nuisance parameter impacts



Alex Held, ATLAS SUSY Workshop 2021

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

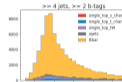
Implementation: ttbar analysis in a notebook

• From data delivery to statistical inference in a **notebook**

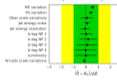
multiple supported processing schemes



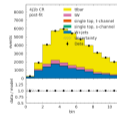
reconstructed observables



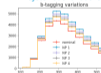
nuisance parameter pulls



post-fit distributions



systematic variations



IBS-HEP AGC

coffee processor

```

# Example code snippet from the notebook
import pyhf
import numpy as np

# Define the model
model = pyhf.Model('...')

# Define the data
data = ...

# Perform the inference
result = pyhf.infer(..., data=data)

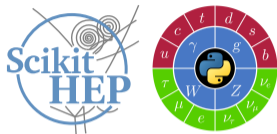
# Print the result
print(result)
    
```

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](#))



Area Project Highlights: Scikit-HEP

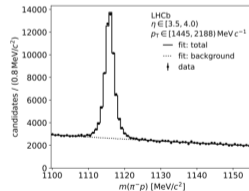
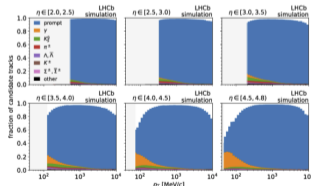
LHCb collaboration has published *JHEP* 01 (2022) 166 using only Scikit-HEP tools for the analysis

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



LHCb-PAPER-2021-010



Nate Grieser (IHEP, LHCb)

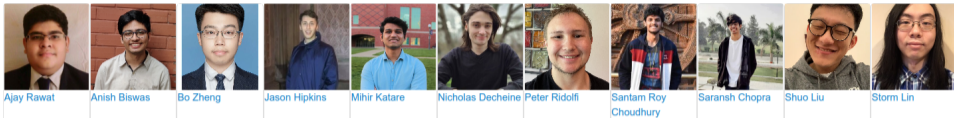
LHCb Python HEP Analysis Feedback

26-04-2022

5 / 17

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