Working Towards Fully Differentiable Analysis

Matthew Feickert

matthew.feickert@cern.ch

PyHEP.dev Workshop 2023

July 25th, 2023

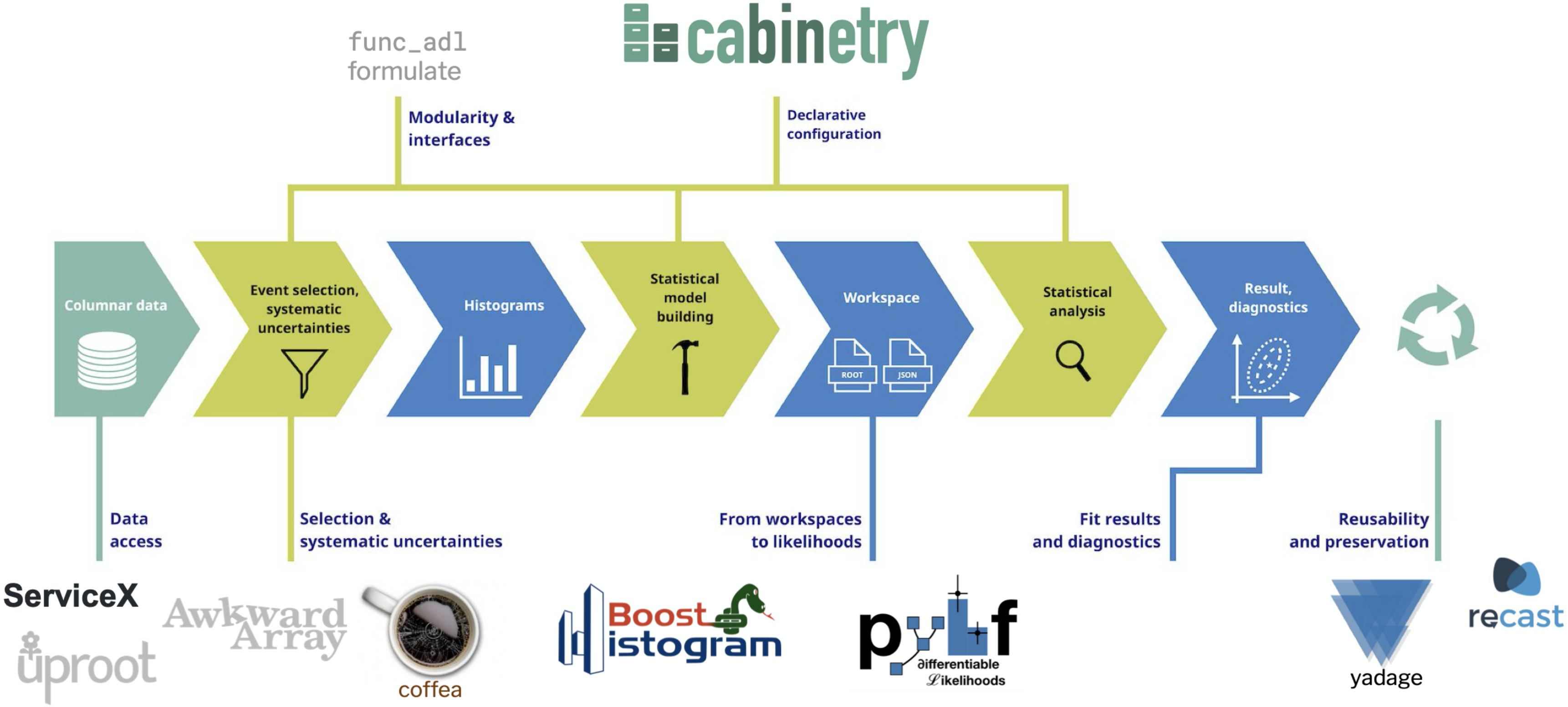
American Family Insurance
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON

iris hep

Introduction

- Postdoc on ATLAS at the University of Wisconsin-Madison Data Science Institute
- IRIS-HEP Analysis Systems focus area lead
- pyhf core developer and maintainer

IRIS-HEP Analysis Systems Pipeline



func_adl
formulate

cabinetry

Modularity & interfaces

Declarative configuration

Columnar data

Event selection, systematic uncertainties

Histograms

Statistical model building

Workspace

Statistical analysis

Result, diagnostics

Data access

Selection & systematic uncertainties

From workspaces to likelihoods

Fit results and diagnostics

Reusability and preservation

ServiceX

uproot

Awkward Array

coffea

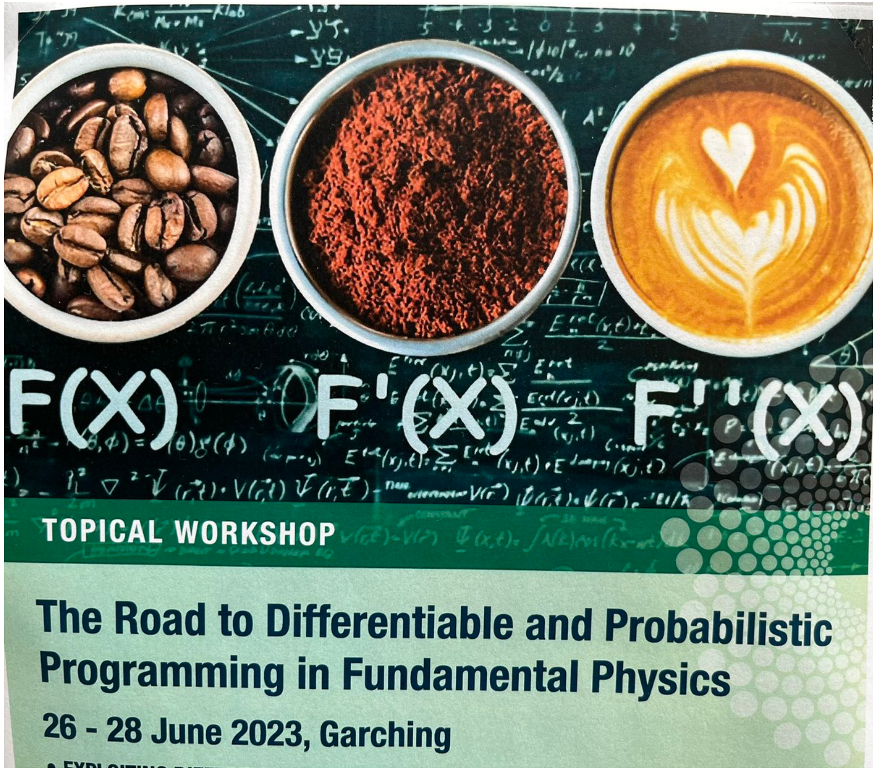
Boost histogram

pyhf
∂ifferentiable ℒikelihoods

yadage

recast

Looking towards differentiability of tools



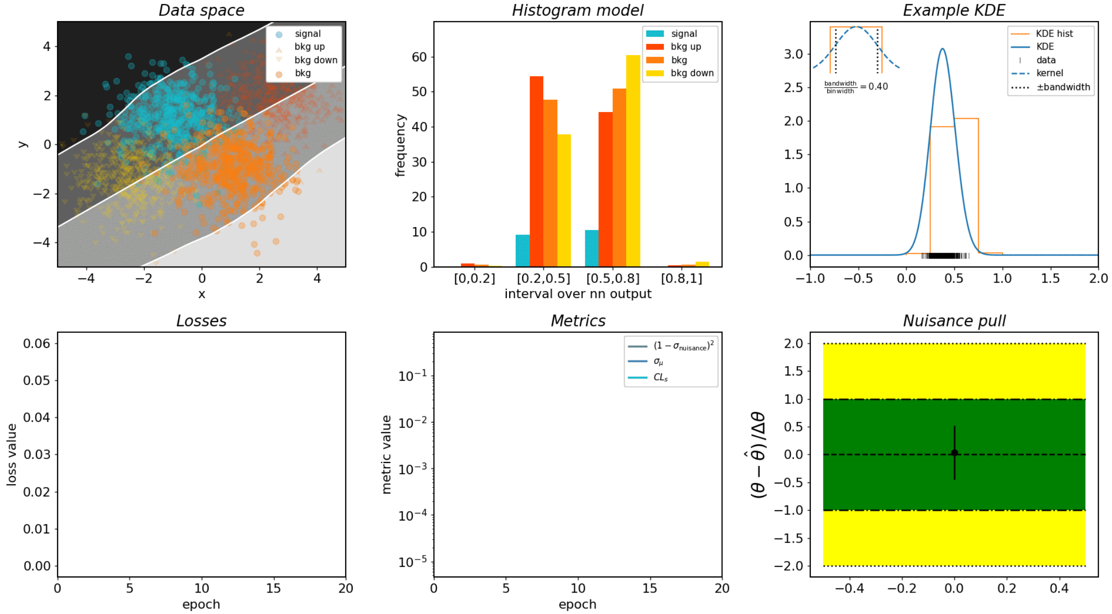
TOPICAL WORKSHOP

The Road to Differentiable and Probabilistic Programming in Fundamental Physics

26 - 28 June 2023, Garching

MIAPbP Differentiable and Probabilistic Programming for Fundamental Physics program

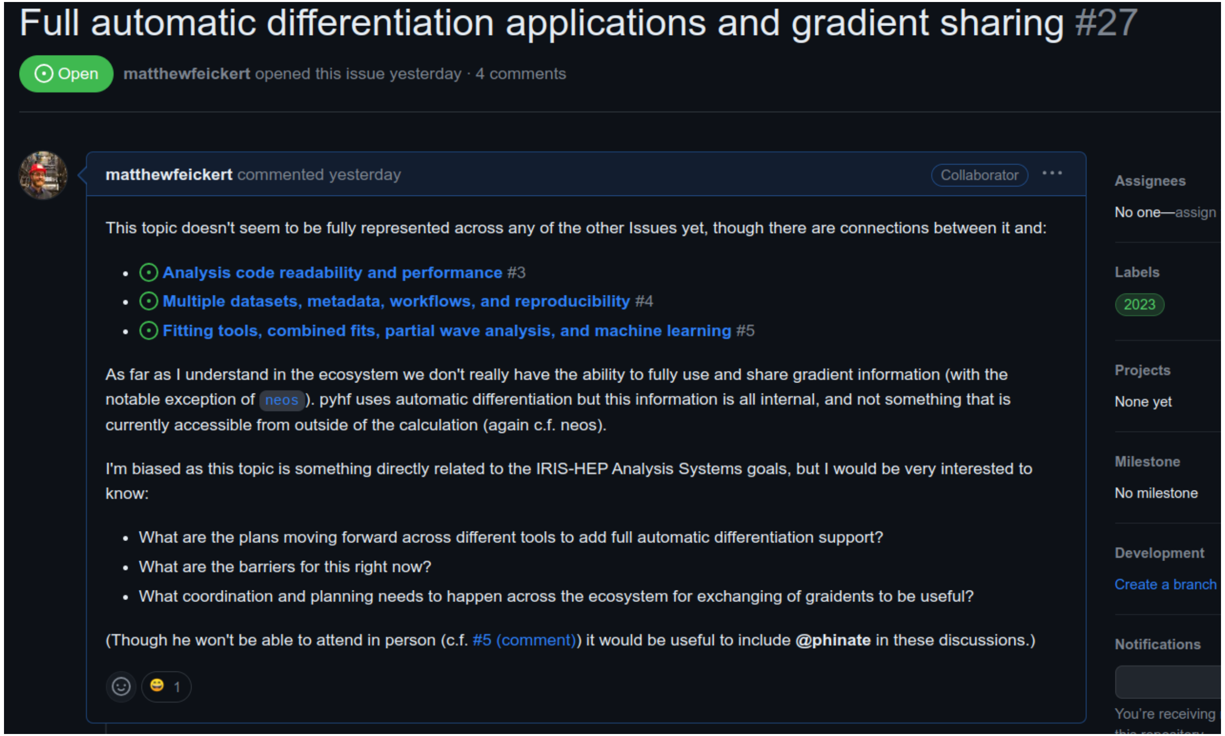
Prior art on this topic: neos



Nathan Simpson, Lukas Heinrich

"Leverages the shoulders of giants (jax and pyhf) to differentiate through a high-energy physics analysis workflow, including the construction of the frequentist profile likelihood." — neos docs

Would love to engage with you on this



Full automatic differentiation applications and gradient sharing #27

Open **matthewfeickert** opened this issue yesterday · 4 comments

matthewfeickert commented yesterday Collaborator

This topic doesn't seem to be fully represented across any of the other Issues yet, though there are connections between it and:

- **Analysis code readability and performance** #3
- **Multiple datasets, metadata, workflows, and reproducibility** #4
- **Fitting tools, combined fits, partial wave analysis, and machine learning** #5

As far as I understand in the ecosystem we don't really have the ability to fully use and share gradient information (with the notable exception of `neos`). pyhf uses automatic differentiation but this information is all internal, and not something that is currently accessible from outside of the calculation (again c.f. neos).

I'm biased as this topic is something directly related to the IRIS-HEP Analysis Systems goals, but I would be very interested to know:

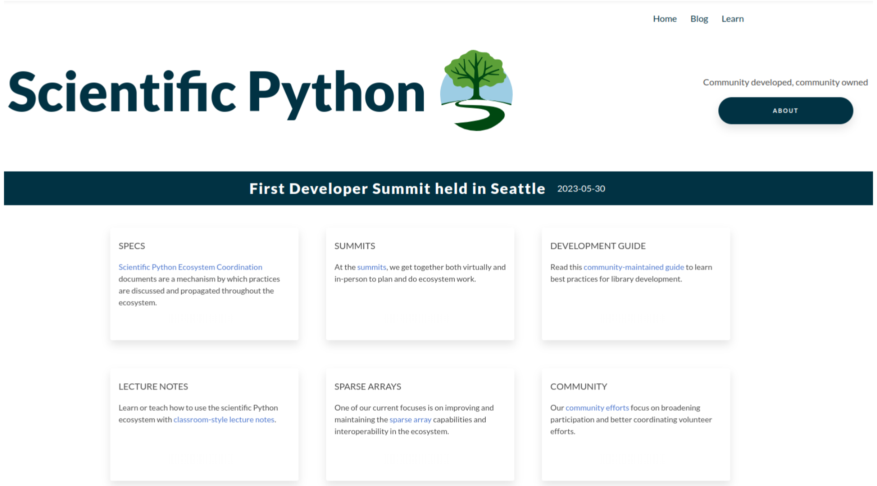
- What are the plans moving forward across different tools to add full automatic differentiation support?
- What are the barriers for this right now?
- What coordination and planning needs to happen across the ecosystem for exchanging of graidents to be useful?

(Though he won't be able to attend in person (c.f. #5 (comment)) it would be useful to include **@phinate** in these discussions.)

😄 1

PyHEP.dev 2023 Issue #27

Would also love for *us* to engage (more) with scientific Python



Scientific Python

Community developed, community owned

First Developer Summit held in Seattle 2023-05-30

SPECS

Scientific Python Ecosystem Coordination documents are a mechanism by which practices are discussed and propagated throughout the ecosystem.

SUMMITS

At the summits, we get together both virtually and in-person to plan and do ecosystem work.

DEVELOPMENT GUIDE

Read this community-maintained guide to learn best practices for library development.

LECTURE NOTES

Learn or teach how to use the scientific Python ecosystem with classroom-style lecture notes.

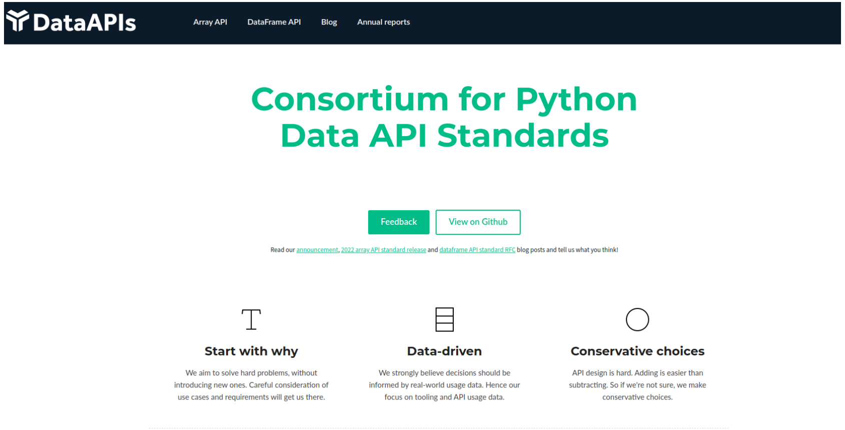
SPARSE ARRAYS

One of our current focuses is on improving and maintaining the sparse array capabilities and interoperability in the ecosystem.

COMMUNITY

Our community efforts focus on broadening participation and better coordinating volunteer efforts.

scientific-python.org



Consortium for Python Data API Standards

Read our announcement, 2022 array API standard release and dataframe API standard RFC blog posts and tell us what you think!

Start with why

We aim to solve hard problems, without introducing new ones. Careful consideration of use cases and requirements will get us there.

Data-driven

We strongly believe decisions should be informed by real-world usage data. Hence our focus on tooling and API usage data.

Conservative choices

API design is hard. Adding is easier than subtracting. So if we're not sure, we make conservative choices.

data-apis.org

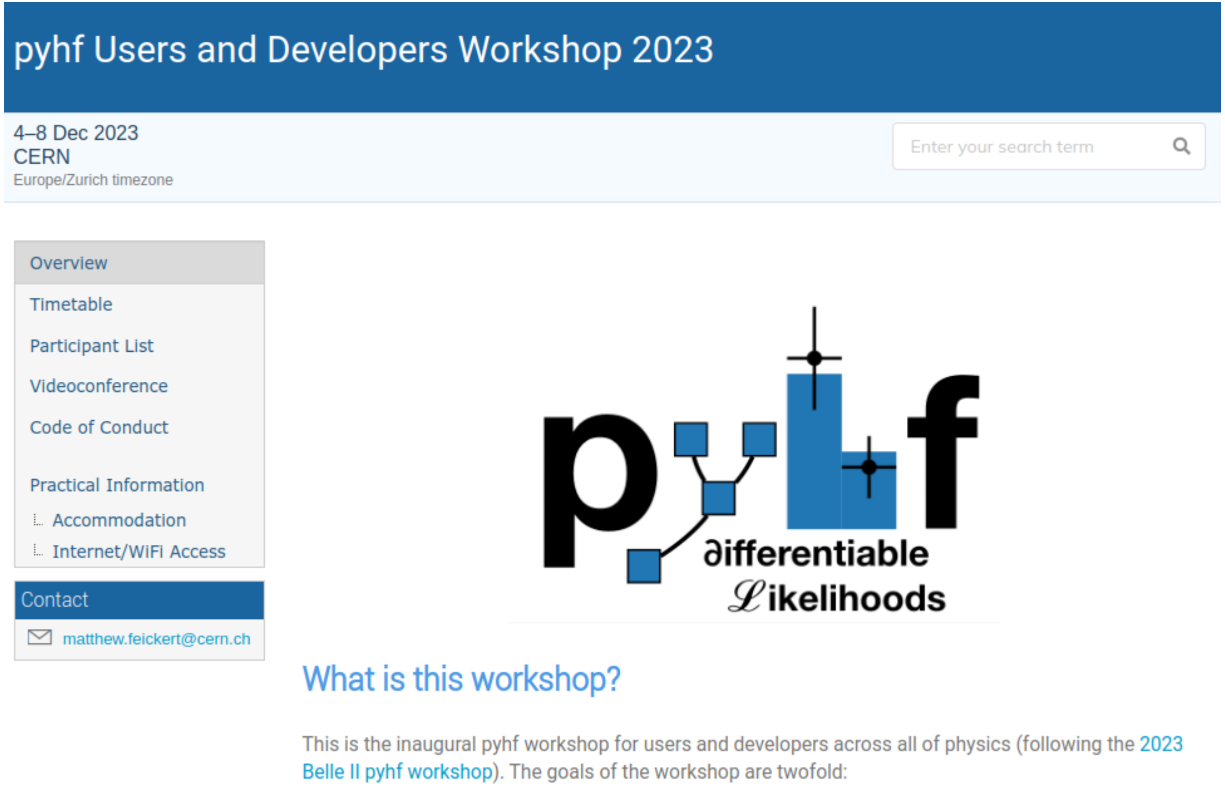
Goals for PyHEP 2023

- Working towards tool differentiability (personal focus on pyhf)
  - HSF/PyHEP.dev-workshops/ Issue #27
- Discussion and move towards tool adoption of HEP Statistics Serialization Standard (HS3)
  - HSF/PyHEP.dev-workshops/ Issue #5
  - HS3 GitHub
- Improve the onboarding experience for new contributors to Scikit-HEP/PyHEP ecosystem (personal focus on pyhf)
  - HSF/PyHEP.dev-workshops/ Issue #6
  - HSF/PyHEP.dev-workshops/ Issue #9

Let’s talk!

Looking forward to a good week

Plug for upcoming workshop



pyhf Users and Developers Workshop 2023

4–8 Dec 2023
CERN
Europe/Zurich timezone

What is this workshop?

This is the inaugural pyhf workshop for users and developers across all of physics (following the 2023 Belle II pyhf workshop). The goals of the workshop are twofold:

pyhf Users and Developers Workshop