



IRIS-HEP Focus Area Summaries: Analysis Systems

Matthew Feickert, Kyle Cranmer
(University of Wisconsin-Madison)

[September 11th, 2023](#)

<http://iris-hep.org>

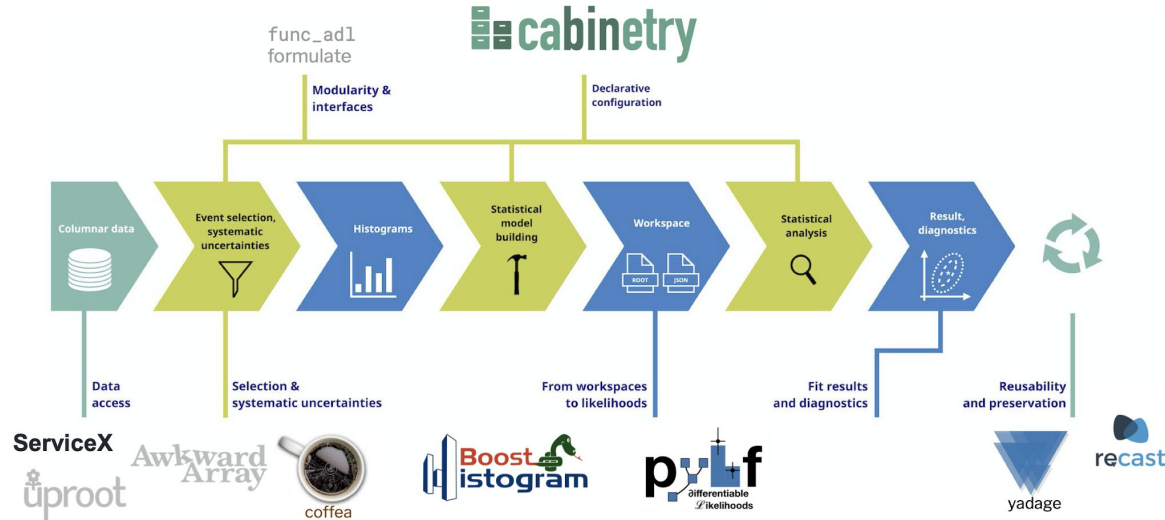


OAC-1836650



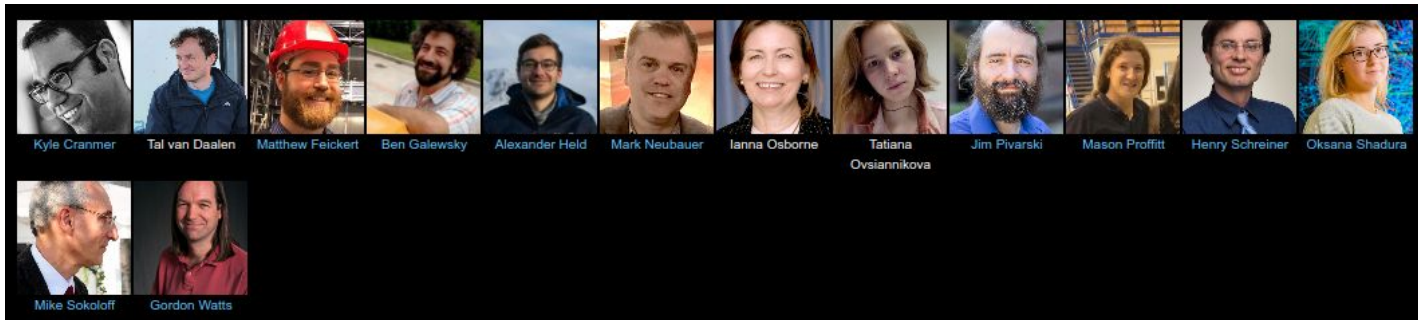
Analysis Systems in IRIS-HEP v2.0

- Productionizing the Analysis Systems pipeline (performance and UX)
- Using Automatic Differentiation as a unifying technology across Analysis Systems tooling



Productionizing the Analysis Systems pipeline

- Need to set out a realistic timeline for making Analysis Systems components useable in production (real analysis)
 - by both power users and new users alike
 - Should plan this timeline around current personnel FTE
- Would greatly benefit from additional personnel time (at least 1 FTE), given current listed personnel FTE is reduced



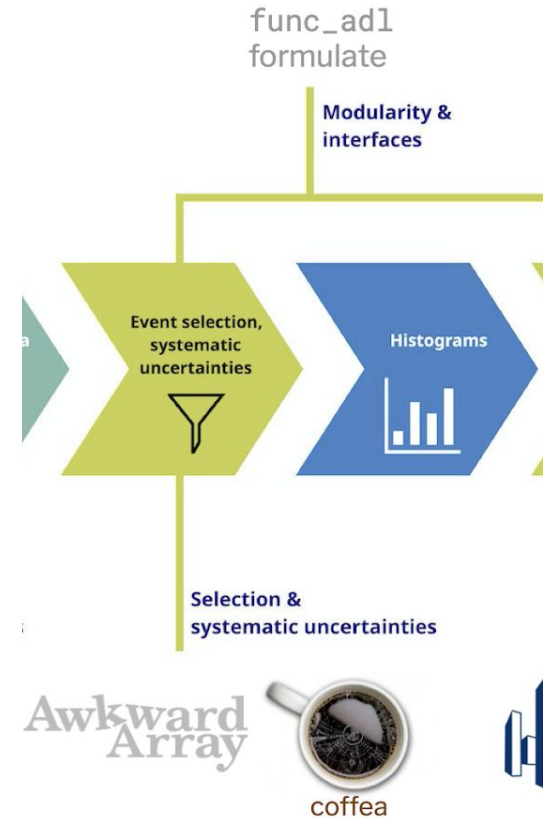
Productionizing: ServiceX frontend

- For production focus on UX in addition to performance
 - Improving func_adl query to result experience
- Communication between AS and DOMA on frontend-backend development would benefit from personnel
 - Idea: Pair physics- / user-oriented person with core technical developer
 - Metadata access
 - Error message propagation
 - Consistent APIs between different pieces of functionality (running locally vs dispatching to remote transformer)



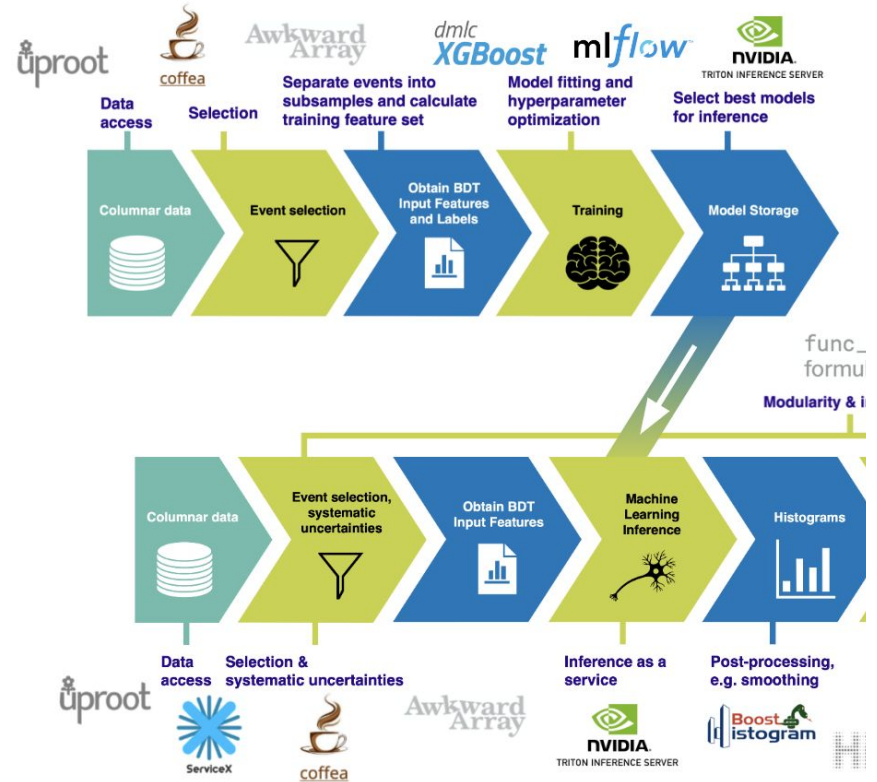
Productionizing: Coffea

- Coffea undergoing large improvements during the Daskification lead by coffea-team
 - For the timescales of IRIS-HEP v2.0 this is done (currently release candidate) but will need to continue to integrate it and add development (e.g. AD, though we can target coffea dependencies)
- As coffea is a core component of the Analysis Systems pipeline should also include IRIS-HEP supported people helping to productionize it for our goals



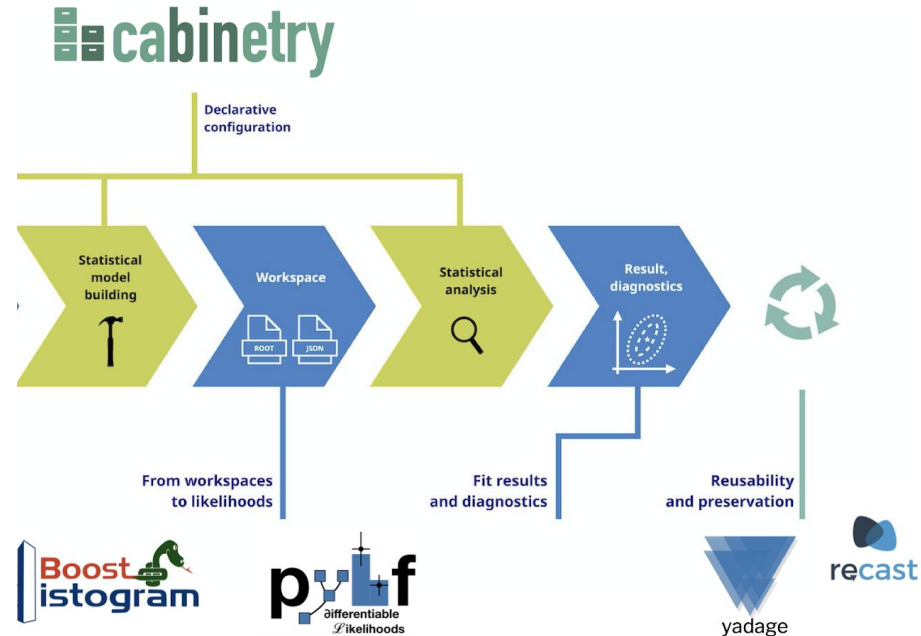
Productionizing: Machine Learning

- Adding productionized machine learning workflows is a IRIS-HEP v2.0 goal
- Currently uncovered in Analysis Systems
 - All work has been done in Analysis Grand Challenge
 - Facility side on SSL that will require coordination
- Scope and responsibilities need formalizing and FTE



Automatic Differentiation for Analysis Systems

- From the goals of a fully differentiable analysis incorporating automatic differentiation at the level of each tool becomes a requirement
 - Awkward already started (development at PyHEP.dev 2023)
- This provides an opportunity to have automatic differentiation be a unifying approach that reinforces value of Analysis Systems as a focus area beyond the individual tool level
- Components of productionizing statistical analysis components (e.g. gradient passing out of pyhf)



Timelines (in context of AGC) from NSF Review: Year 1

- **Setting up** facilities, services, and analysis task for the **next generation of the AGC**

Focus Area	Activity Target	Target Year
AS	All components of AS pipeline are fully supporting distributed analysis	Year 1
AS	Define analysis tasks for the top quark mass and di-Higgs measurement, create implementations	Year 1
AS	New version of AGC analysis with incorporated ML techniques	Year 1

Timelines (in context of AGC) from NSF Review: Year 2

- **Checkpoint for performance** and throughput & **functionality** demonstration

Focus Area	Activity Target	Target Year
AS	Benchmark analysis on dataset 20% HL-LHC scale to be completed in 1 hour	Year 2
AS	Demonstration of running full analysis able to use statistical models defined in unified HS3 serialization format	Year 2
AS	All core components of Analysis System pipeline support integration of differentiable operations and passing of gradients	Year 2
AS	Demonstrate AOD extraction prototype	Year 2

Timelines (in context of AGC) from NSF Review: Year 3-5

- New **functionality**: column joining, reinterpretation, differentiable analysis
- **Scaling** to full HL-LHC requirements

Focus Area	Activity Target	Target Year
AS	Execute AGC analysis with column extraction and joining workflow	Year 3
AS	Execute fully differentiable analysis	Year 4

Summary

- Transitioning IRIS-HEP Analysis Systems work plan to focus on reaching production readiness will require some change in focus for current effort and calls for increased effort (FTEs) for new themes.
- Planning and coordination this week to reach the needed information for writing the award PEP
- For both sustainability beyond IRIS-HEP and development continue alignment with data science and broader open source communities (e.g. dask-awkward, NumFOCUS affiliated/sponsored projects)