

Analysis Grand Challenge

Analysis Grand Challenge summary

What is it:

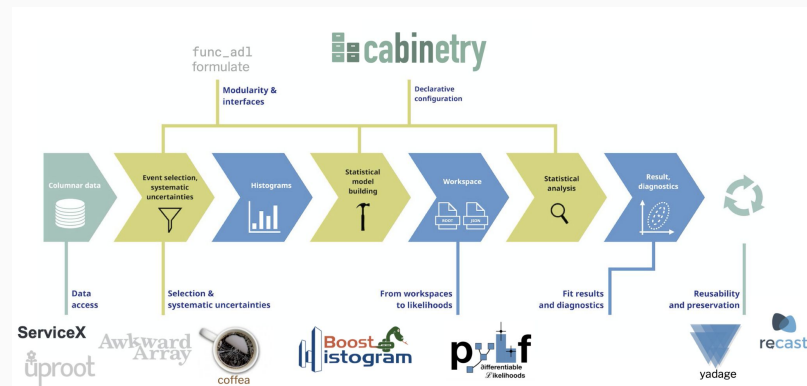
- A technical demonstration of an analysis system executing a representative HL-LHC analysis using new tools + enhanced functionality
- Process ~200TB of data starting with ServiceX, using columnar analysis tools, creating and fitting pyhf statistical model with fitting service to get results
- Include optimization of the analysis (multiple iterations over the data with auto-diff.)
- Include preservation and reinterpretation.

Where do we stand:

- The analysis tools to be used have been identified, preliminary versions are ready, tools are being integrated into vertical slice
- Plan to use reformatted CMS Open Data for challenge.
- Agreement on primary types of functionality that we want to demonstrate
- Prototype analysis facility with some functionality (coffea-casa) exists. Also:
 - Various tests of ServiceX being performed
 - funcX-based fitting service prototyped
- **New:** AGC Coordinators

Preparation for AGC

- Develop notebooks demonstrating interactions between projects in AS pipeline
 - *coffea + cabinetry + pyhf*
 - e.g. *deployed on coffea-casa*
 - *also integrate with fitting service*
 - *ServiceX + funcADL + coffea*
 - *also add Skyhook if it would be possible*
- Expand notebooks to represent realistic analysis workflows
- We have IRIS-HEP fellows (Summer 2021) working on related projects



Analysis Grand Challenge coordination

Coordinators: Alex Held (NYU), Oksana Shadura (UNL)

- *Closely connected to and interacting with AS (Alex) and DOMA (Oksana)*

Scope and responsibilities:

- *Develop milestones to measure progress and converge on Grand Challenge*
- *Coordinate with focus areas to achieve milestones*
- *Dedicated Grand Challenge meetings as needed*
- *More detailed job description under development*

Data processing Grand Challenge (DGC)

- DGC is organised in collaboration with **US LHC Operations programs, the ATLAS and CMS global collaborations, and the WLCG**
 - *Organised as a series of data challenges for the next several years (2021, 2023, 2025, 2027)*
- DGC has a well defined plan with milestones (data challenge target rates):
https://docs.google.com/document/d/1IMG4dfiPo9bPf-tAO0bINDAuEUllC45Y-vwu1E9_Xw/edit

Analysis Grand Challenge (AGC) milestones

Milestones in PEP:

- **G2.12** Fully differentiable fitting as a service on remote sites with GPUs ([Jun 1, 2022](#))
- **G2.14** Functionality and integration demonstration for Analysis System components ([March 1, 2022](#))
- **G2.15** Coordinate with DOMA, SSL, and operations programs to benchmark performance of prototype system components to be used for Analysis Grand Challenge ([Jun 1, 2022](#))
- **G2.16** Coordinate with DOMA, SSL, and operations programs to execute the Analysis Grand Challenge ([Mar 1, 2023](#))

Planning Analysis Grand Challenge

- Collect feedback and coordinate with *AS*, *DOMA*, and *SSL*
- Write a working plan similar to DGC
 - Based on feedback from IRIS-HEP review:
 - it should be easier to digest
 - divided in a subset of challenges
 - ***with clear target numbers***
- Collect feedback US LHC Operations programs, the ATLAS and CMS global collaborations

Preliminary work items and milestones

- **Milestone 1 [09/2021]:** ATLAS and Open Data deployments of coffea-casa
 - *DOMA milestone Y4: deploy 5 different instances of coffea-casa in different sites*
- **Milestone 2 [10/2021]:** develop a training material inspired by Columnar Analysis Tools HATS@LPC (<https://github.com/CoffeaTeam/coffea-hats>)
 - OpenData version (including ServiceX, possibly pyhf and cabinetry)
 - Make an ATLAS version (including ServiceX, possibly pyhf and cabinetry)
- **Milestone 3 [11/2021]:** HATS-like event in October-November with OpenData + ATLAS and CMS specific material