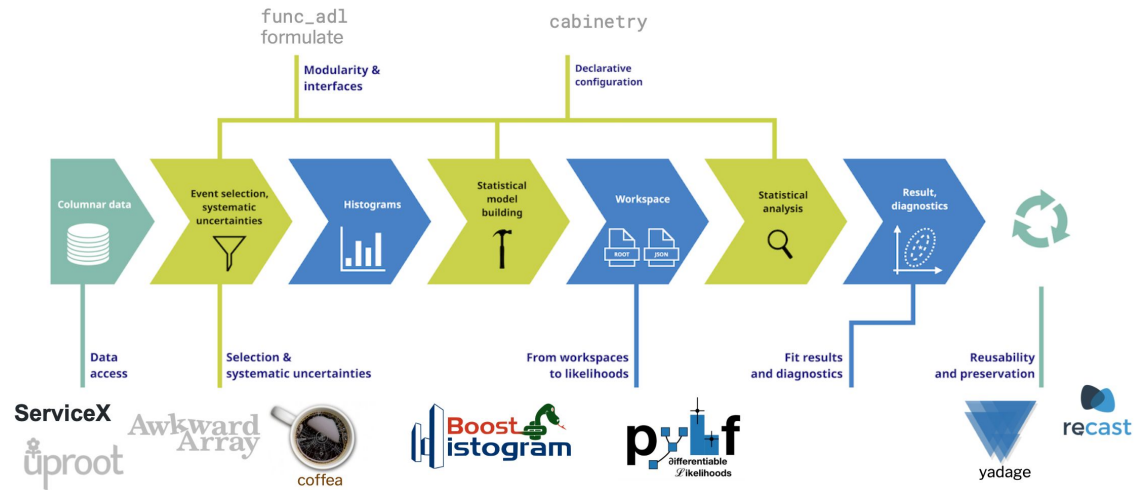# Outcome of Day 1

Kyle Cranmer

Brian Bockelman

Rob Gardner
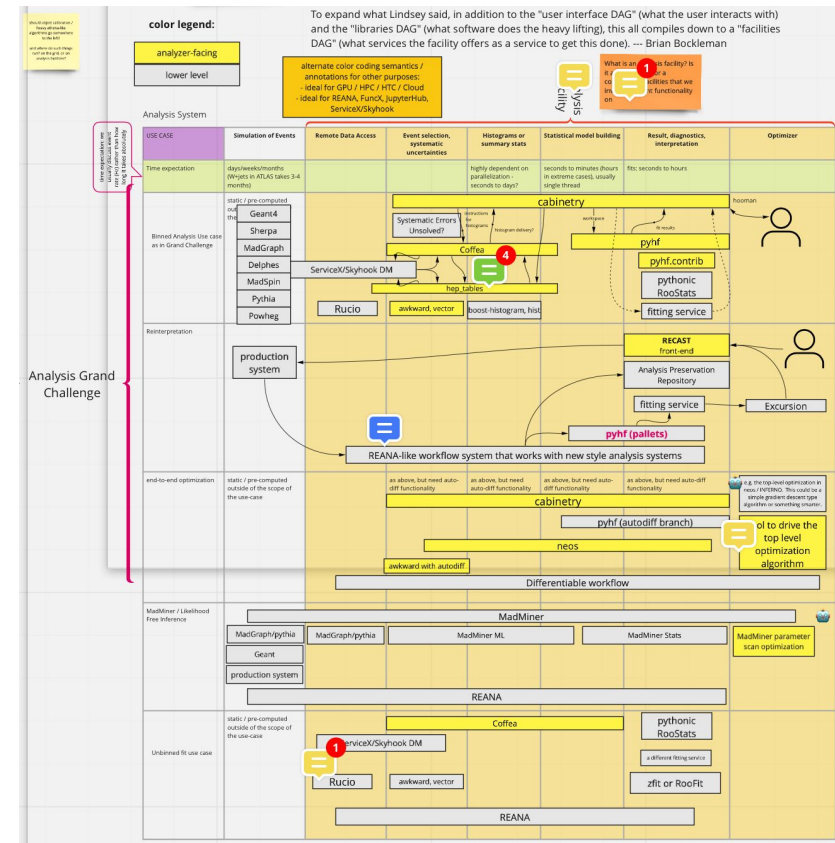
Day 1

How it started:

How it ended:

Miro board

We identified 5 use-cases (rows) with corresponding Analysis Systems.

- 3 of the rows refer to different aspects of the **Analysis Grand Challenge**
  - *binned analysis*
  - *reinterpretation*
  - *end-to-end optimization*
- + MadMiner / Likelihood Free Inference use case
- an unbinned analysis (similar conceptually to the binned analysis, but set of tools and libraries is not as well developed)

This view is oriented towards libraries, not services or infrastructure
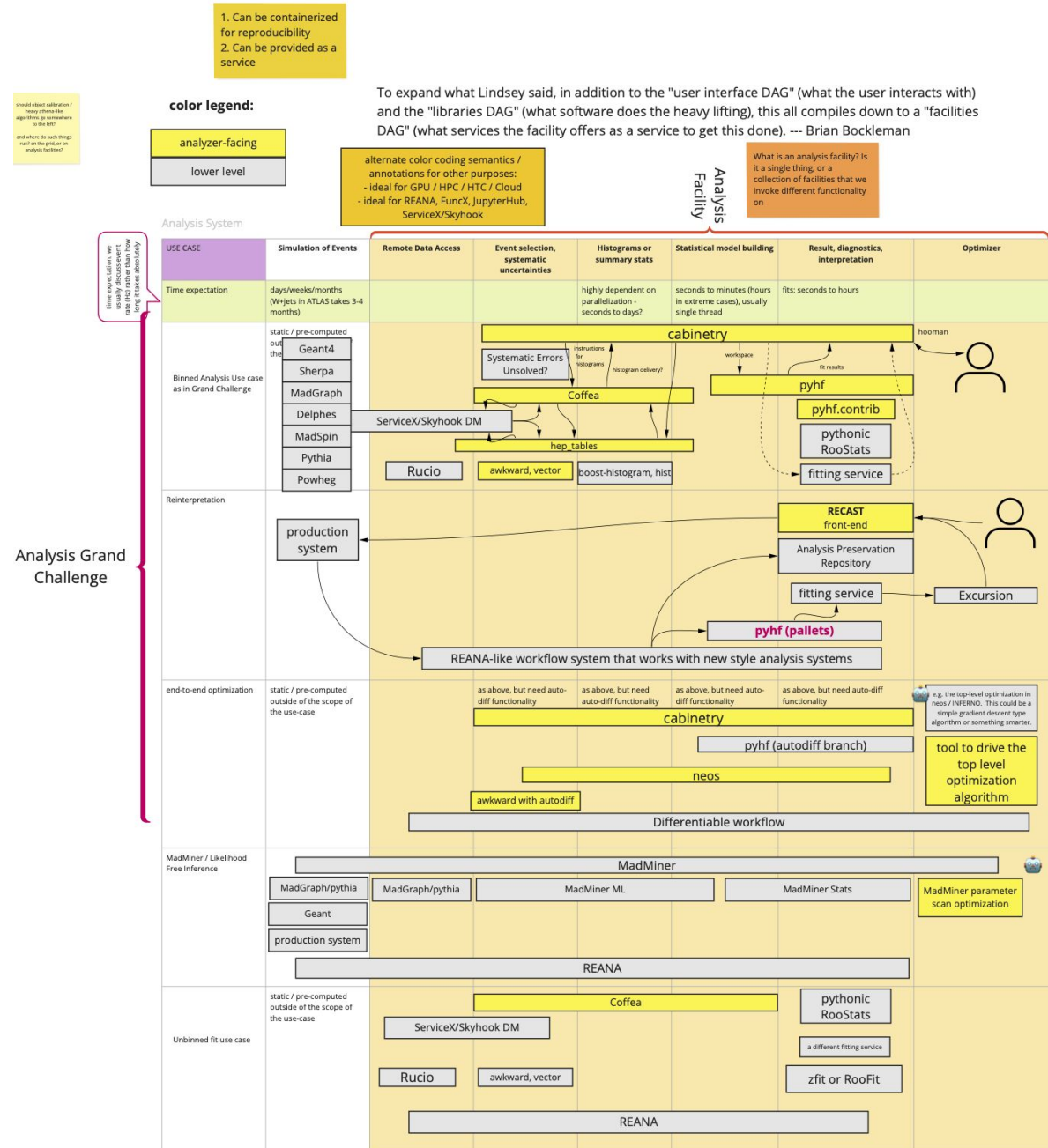
User Interface > Libraries > Infrastructure

We spent a lot of time yesterday talking about user interfaces and software libraries - but these are sometimes less-relevant to facilities. We likely have something like 4 layers:

- **User interfaces**.
- **Software libraries**.
- **Infrastructure** (Software services): Data Delivery (ServiceX), Column store (SkyHook), Task execution (Dask, FuncX), workflow-specific microservices.
- **Infrastructure** (Fundamental site services): AAI, Batch system, object store, DTNs, Service orchestration.

(Probably need better names for the layers)! The third layer – and its interaction with the second - is probably a good topic for today.

We had discussion about access control, resource allocations, HPCs, latency issues, etc.

Infrastructure / Software Services

Initial reaction: it seems that most of what was listed on the Miro board can be satisfied with:

1. ServiceX / SkyHook

2. FuncX for FaaS

3. REANA

4. and a generic JupyterHub entry point that can access customized containers with the necessary software
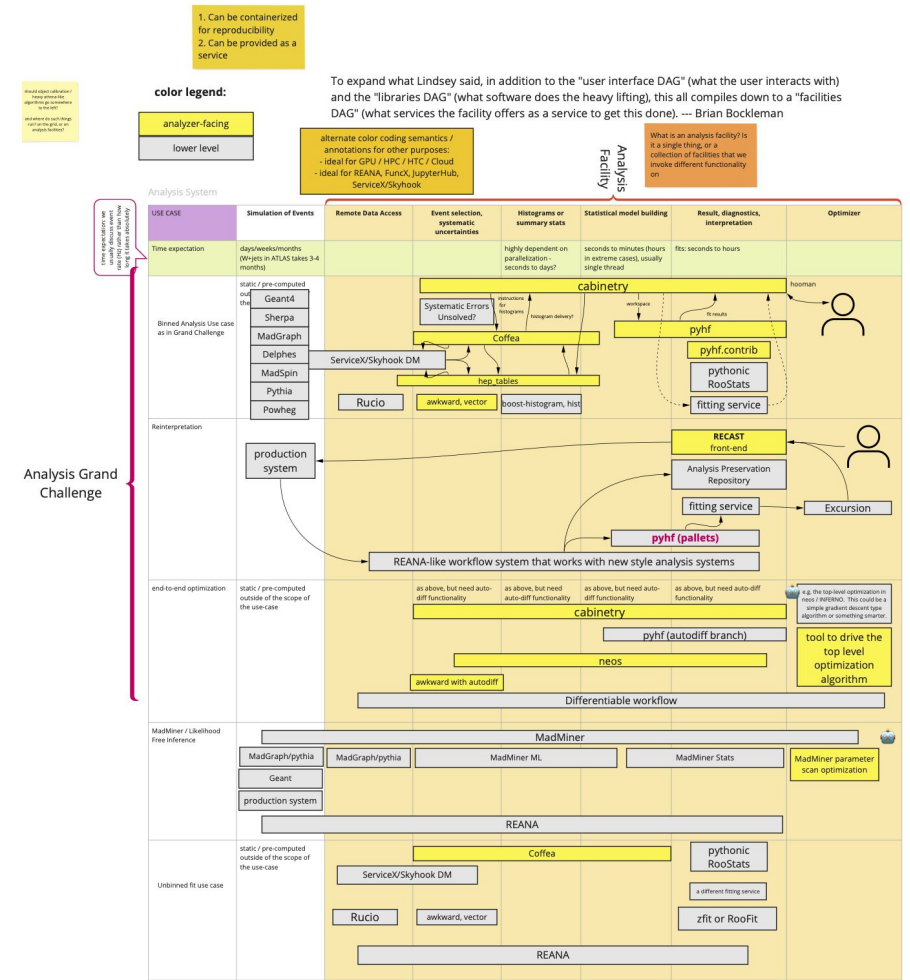
Comments & Questions

Not entirely clear what is happening within facilities vs. outside (eg. laptops / local compute, etc.) and how an analysis workflow would egress / ingress the facilities

Anticipate multiple Analysis Facilities with different capabilities.

- Possible that infrastructure components needed to implement one of these analysis systems aimed at a particular use cases may be satisfied by different facilities

What is the right abstraction for talking about the capabilities of the facilities? E.g. at QoS attributes for storage systems, or in future advanced features such as "programmable storage"

Our experience with reproducible analysis workflows is mainly based on current analysis models, not the analysis systems we are envisioning

- not yet clear what will break and need to be revisited

Costs

How can the infrastructure communicate expected "costs" to analysis system such as latency & time-to-completion

Some deployments will extend beyond a site or even a region, such as for remote data access or distributed processing.