“Content” → how to use the preserved data

Data and **Analysis** Preservation
→ two sides of the same coin, equally important → ***neither makes sense preserving without the other***

Analysis → to get from calibrated subdetector data (DST) to meaningful physics observables
Level 3 → being able to mechanically replicate a published analysis → relatively easy
Level “3.5” → being able to trace (debug) a published analysis → much harder
Level 4 → preserving knowledge such that a new analysis can be done on old data → much-much harder

Requires vast amount of insider knowledge, e.g.
-- innards of the detectors, awareness of possible malfunctions (tell apart rare signal from rare noise)
-- understanding of methods, algorithms, why were they used (to recognize when they might fail)
-- understanding not just **what** the cuts, conditions, selections are, but **why** are they what they are
-- understanding the source of backgrounds, combinatorics
-- etc.
-- this kind of expertise evaporates quickly without a serious, dedicated effort, and can not be recovered later
→ preserved data essentially useless

Analysis preservation also needs (equivalent?) support

“Content” → who’s responsibility?

Clearly most of the responsibility goes to the Collaborations

But the host, BNL (DOE) also has a stake, exercises (some) control, and with a good reason
- → formal approval procedure before something can be published
- → putting the procedure on hold if needed (as happened e.g. with PHENIX/STAR discrepancies)
- → setting up “task forces” to clarify contradictory results (happened a few times, and rightly so)

For the very same reasons such control should also be applied to “analysis preservation”
- → but this also means some level of BNL involvement is inevitable
- → my “hobby-horse”: make analysis preservation part of some people’s job description

Metrics: successful test of the preserved analysis by a 3d (non-collaboration) person

My (biased?) view:
- → any BNL Data and Analysis Preservation plan without a strong “Analysis” part is unsatisfactory
- → preserving data without the capability to reliably analyze them makes no sense (and vice versa)

No, I don’t have a magic bullet, but we better start thinking how to make one
- → Bear with me: I was first seriously burnt 2008, and the wounds didn’t heal, in fact, multiplied since…

Summary – the analysis preservation side (“content”)

- Data and analysis preservation equally important
- Collaborations first, but BNL (DOE) also has a stake (and responsibility) in it
- Analysis preservation needs dedicated resource(s) in each experiment
- Scope needs to be discussed and defined
- Scope may be different for the three experiments
- Metrics: test by 3d person
- FTE estimates once scope is defined
- Provide early input for EIC