Re-analysis Exercise: Purpose and Context

STAR Re-analysis Verification: Early Observations
As part of the RHIC Data and Analysis Preservation effort, one practical step is to verify whether past STAR analysis notes contain sufficient information to reproduce or re-run selected analyses in today's computing environment.

This exercise is not intended as an audit of individual analyses. Rather, it is a stress test of long-term preservation: can an analysis described many years ago still be reconstructed after changes in software repositories, operating systems, build environments, storage paths, and institutional computing practices?

Early observations show that STAR analysis notes vary significantly in style, level of detail, and assumptions about the surrounding computing environment. Some notes provide enough conceptual and procedural guidance to restart the workflow, while others rely on external files, user-area code, implicit conventions, or software environments that have evolved since the original work.

The value of this exercise is that it identifies where preservation practices can be strengthened going forward, especially for future re-use of analysis workflows by people who were not part of the original analysis team.

Issues Observed and Lessons for Preservation

Common Reproducibility Friction Points

The first re-analysis tests showed several recurring issues that are typical when revisiting legacy analyses many years later.
Issues observed:

- Some analysis notes assume access to files, scripts, or macros that were available at the time but are no longer in stable shared locations.
- Certain input datasets are referenced conceptually, but the exact production, location, or access path is not always fully captured.
- Software and build instructions sometimes depend on older computing environments and require translation to current systems.
- Repository transitions, such as moving from older code-management systems to newer ones, introduce small but important adjustments (how to document?)
- Some workflows contain implicit expert knowledge that was obvious to the original analysts but not fully documented for future users.

Preservation lessons:

- The exercise demonstrates that analysis preservation requires more than saving a final note or paper.
- A robust preservation package should include stable references to input data, analysis code, configuration choices, environment requirements, and enough procedural context
  for a knowledgeable but non-original analyst to restart the workflow. Best attempt was made but not always consistently applied across notes.