RHIC Data and Analysis Preservation Round Table

01/30/2025

Introduction & some notes from previous meetings

Welcome and congratulations to
Megan Connors and Jin Huang
the new sPHENIX elected spokespeople

Notes from previous meetings - I

- Within the next few months, prepare a preliminary Data and Analysis Preservation Plan (DAPP) that includes resource estimates for upcoming reviews and institutional events.
  - NP Budget Briefing:
  - ONP site visit:
  - PAC:

Jamie?

- Evaluate the resources and effort required to achieve level 3 and level 4 preservation.
- Start preserving knowledge now, as it is both the most important and challenging aspect
- Identify tools and practices that can benefit the various experiments

Notes from previous meetings - II

- Provide future users with a comprehensive understanding of RHIC experiments by combining published data with validated, unpublished data.
- Create a dedicated knowledge-based portal with advanced search capability to centralize and streamline access to RHIC information and data.
  - The chosen technology should be simple to ensure longevity and ease of maintenance. It should also offer agility to adapt to new solutions as they emerge.
  - Ensure the portal supports all RHIC experiments.
  - The portal can be an interface to repositories or a repository, depending on the information type.
  - Different access level rights may need to be implemented.

Level 3 vs Level 4
Data and Analysis Preservation

- Level 3 allows for re-analysis only.
- Level 4 allows for simulation and data reconstruction.

	Preservation Model		Use Case	
	1	Provide additional documentation	Publication related info search	**Documentation**
	2	Preserve the data in a simplified format	Outreach, simple analyses	**Outreach, reanalysis**
	3	Preserve the analysis level software and data format	Full scientific analysis, based on the existing reconstruction	**Technical Preservation Projects**
	4	Preserve the reconstruction and simulation software as well as the basic level data	Retain the full potential of the experimental data	

Data Preservation Levels defined by the Data Preservation in HEP (DPHEP) Collaboration

Today

- Walk through STAR answers to the questionnaire highlighting differences between Level 3 and Level 4
- Implementing Standards for Data Repositories
- Proposed date/time for next meetings:
  - Tuesday 2/4 – 10:00 AM
  - Thursday 2/13 – 8:00 AM
- What is the best day of the week for scheduling a regular meeting (Tuesday? or Thursday?)

Topics for future meetings

- sPHENIX publication policy
- PHENIX answers to the questionnaire
- Strategy for preserving software and workflows?
- Technologies for the Web portal and the repository

- Suggestions are welcome

Thank you

First steps - Questionnaire

A questionnaire has been circulated

The goal is to gain an overview of each experiment's data and analysis environment, helping identify gaps, commonalities, and needs and establishing priorities for the RHIC DAP.

Inventory and Best Practices

Document good practices and ongoing efforts related to DAP within experiments that can benefit others.

Evaluate Commonalities and Synergies

Evaluate commonalities in procedures and tools across experiments *(e.g., Invenio for document repository, container repositories, LLM-based search engines, Reana)* and possibly identify synergies with other programs and labs.

(Long) Questionnaire

- Data Volume, Organization & Storage
  - understand the current data management practices and identify areas for improvement.
- Data Management
  - understand the current data management protocols and identify areas for improvement.
- Metadata
  - understand how datasets are labeled, described, and managed, ensuring they are easily accessible and interpretable.
- Conditions Data
  - availability and relevance of condition data in the experiment
- Software
  - information about the software and its management
- Workflows
  - understand current workflow management practices
- Preservation – Documentation
  - understand how data, software, and workflows are documented, preserved, and accessible for future use
- Data Sharing
  - understand how data is shared with external collaborators and the public
- Engagement and Outreach
  - understand the current engagement and outreach practices
- Impact, Challenges, and Futures Plans

A complete answer is not expected for today.