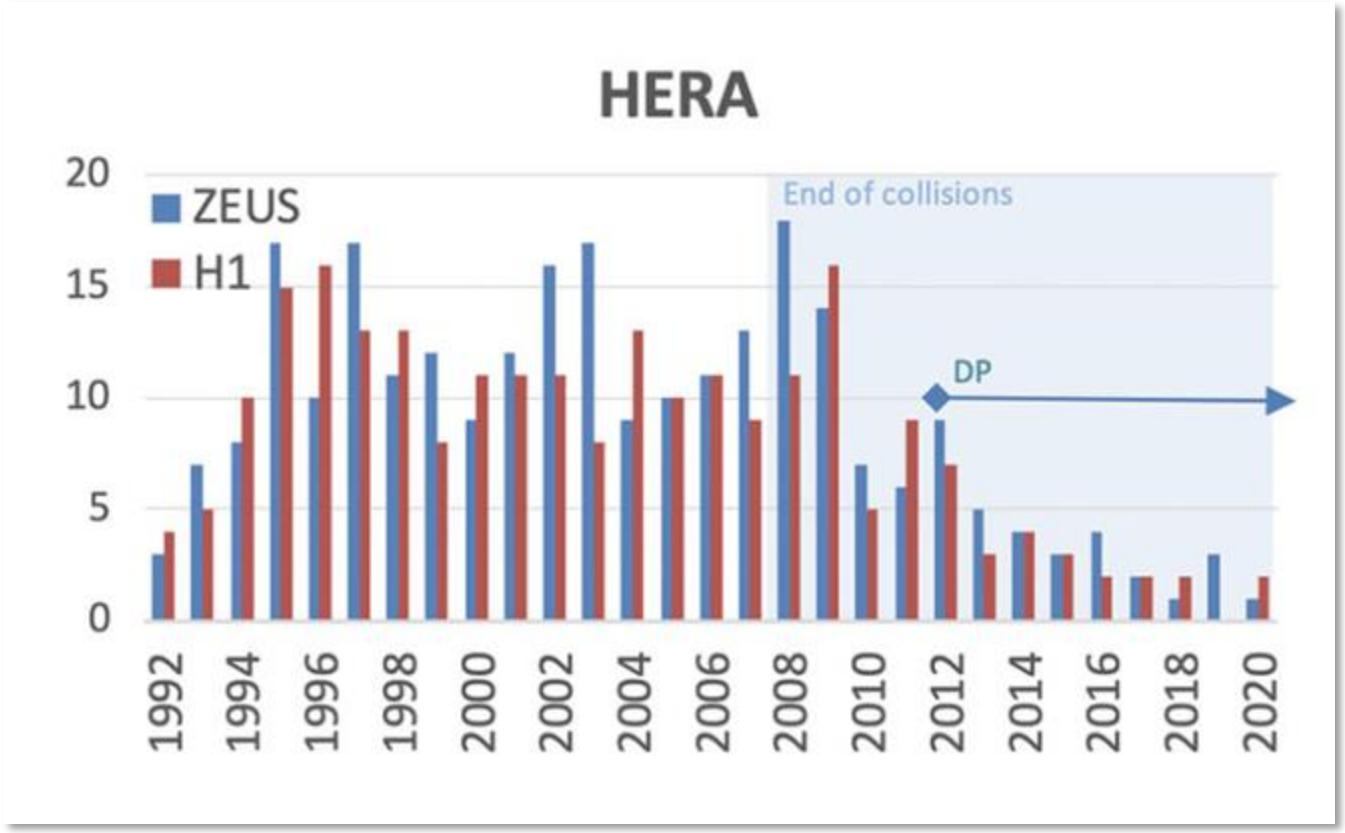
RHIC Data and Analysis Preservation Round Table

01/22/2025

Introduction & some notes from previous meeting

Experiment publications extend well after the end of the data-taking



Data Preservation mode begins when **software is no longer being actively developed**. A few years after end of data taking?
Data and Analysis Preservation planning should start before the end of data taking and should be **designed for long-term**.

Notes from previous meeting

- Prepare a preliminary Data and Analysis Preservation Plan (DAPP) within the next few months, including resource estimates for upcoming reviews and institutional events.
- Evaluate the resources and effort required to achieve level 3 and level 4 preservation, considering sharing between the collaboration and host lab.
- Start reserving knowledge, as it is both the most important and challenging aspect.
- Identify common tools and practices that can benefit various experiments.
- Developing a DAPP for RHIC could be beneficial for the EIC.

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

- Thoughts about a future RHIC DAP repository
- Feedback from experiments:
  - Questionnaire and perspective on DAP,
  - Highlights of level 3 vs level 4 in answers
- Schedule the next round table at a time that accommodates different time zones : https://doodle.com/meeting/participate/id/avR7wQra

JAN	JAN	JAN	JAN	JAN	JAN
27	28	30	30	31	31
MON	TUE	THU	THU	FRI	FRI
8:30 AM - 10:00 AM 1 h 30 min	8:30 AM - 10:00 AM 1 h 30 min	8:30 AM - 10:00 AM 1 h 30 min	9:00 AM - 10:30 AM 1 h 30 min	8:30 AM - 10:00 AM 1 h 30 min	9:00 AM - 10:30 AM 1 h 30 min

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.

On a RHIC DAP plan

- Several topics must be considered.
  - **Software Preservation**: Maintain operability of older software using containerization.
  - **Knowledge Management**: Capture and preserve knowledge through detailed documentation.
  - **FAIR Principles**: Ensure data is Findable, Accessible, Interoperable, and Reusable.
  - **Repositories**: Provide comprehensive access to RHIC data and analyses for non-RHIC users.
- Each component alone is insufficient; together, they form the complete DAP plan.
- Continuous Improvement: Plan will evolve through input and feedback from the RHIC user community.

Challenges for Preserving Analyses

Challenges in Preserving (each) Analysis

- Navigating complex analysis methods and workflows,
- Understanding the computing environment, detector, and accelerator conditions,
- Preparing data, analysis artifacts, and documentation for use by non-collaborators.

This requires:

- An early and sustained effort,
- Contributions from experts in analysis, computing, and software,
- Capturing knowledge before it fades.