

RHIC Data and Analysis Preservation Round Table

04/17/2025

Introduction

Notes from previous meetings

- Notes from previous meetings are available on individual indico pages
- Please have a look at the notes and provide comments if needed.

Update on DOIs

- InvenioRDM, the document repository software, supports DataCite Digital Objects Identifiers (DOIs)
- [DataCite](#) is a non-profit organization that provides DOIs for research
 - It has 247 members worldwide and
 - 65 consortia
- The DOE Office of Scientific and Technical Information (OSTI) leads one of the consortia that includes DOE
- Contacts have been established with OSTI, which confirmed that we can issue DOIs for scientific research at BNL under the OSTI umbrella.
- OSTI has developed a new version of their application [E-link](#), which includes APIs for generating DOIs.
 - Tests are ongoing with a dedicated InvenioRDM instance at BNL to generate (test) DOIs for public documents

Which data volume needs to be preserved?

- A detailed estimate of the total data volume to be preserved must be provided in preparation for the DAP Plan
- Which data will be ‘preserved’ ?
- The anticipated access requirements (frequency, users, permissions) need to be defined.
- A storage infrastructure should be designed, aligned with the expected access patterns and preservation needs

Data volume to be preserved (1 copy)

[PB]	PHENIX	sPHENIX	STAR
RAW	25	200	100
Analysis Objects	5	15	40

Data volume by the end of 2025:

- These are **PRELIMINARY** estimates
 - Order of magnitude
 - Analysis Objects: volume for one processing
-
- Analysis Objects: are data format used for analysis
 - PHENIX: DST
 - sPHENIX: DST
 - STAR:pico-DST

Check these PRELIMINARY numbers - Please

Questions:

- Should more than one processing be preserved?
- Other data formats?
- Simulated events?

Data volume and preservation levels

[PB]	PHENIX	sPHENIX	STAR
RAW	25	200	100
Analysis Objects	5	15	40

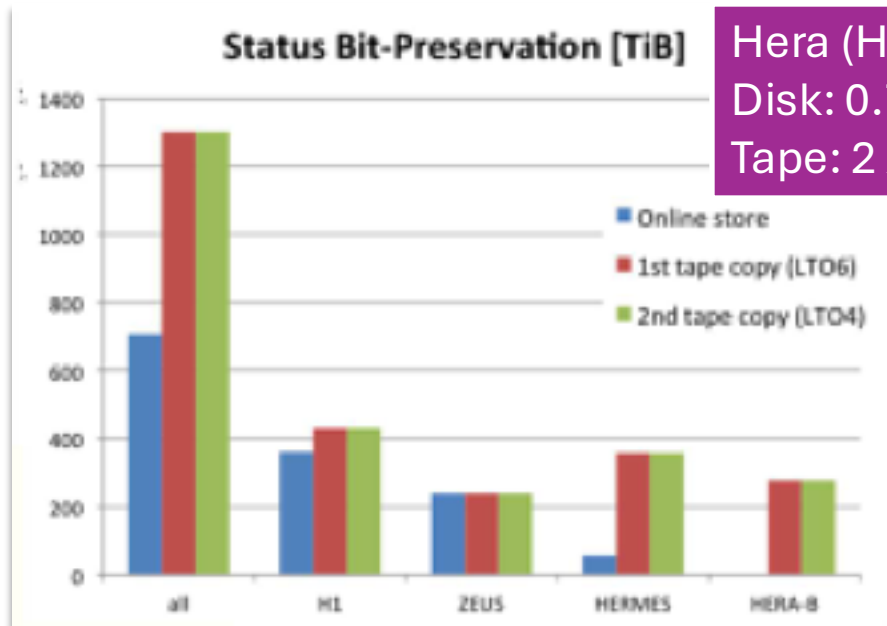
DPHEP preservation :

- Level 3: Ability to reproduce analyses and perform analyses
- Level 4: Ability to redo reconstruction, to reproduce analysis and to perform analysis

[PB]	PHENIX	sPHENIX	STAR	Sum
Level 3 (AO only)	5	15	40	60
Level 4 (RAW + AO)	30	215	140	385

Compared to other experiments

[PB]	PHENIX	sPHENIX	STAR	Sum
Level 3	5	15	40	60
Level 4	30	215	140	385



Hera (H1 + Zeus + ...)
Disk: 0.7 PB
Tape: 2 x 1.3 PB

Babar
1.5 PB

What is “Everything”?

- **Data**

- collected collision data and generated MC events (~1.5PB)
- all in root files

RHIC is more than two orders of magnitude larger

Data volume and access

[PB]	PHENIX	sPHENIX	STAR	Sum
RAW	25	200	100	325
Analysis Objects	5	15	40	60

Storage options:

- Tape (or other cold storage) for archival and scheduled access
- Disk (or other fast storage) for timely and random access

Storage by Data type to be defined.

- RAW data on tape (or other cold storage)
- Analysis Objects on disk (or other fast storage)

Data volume and access

[PB]	PHENIX	sPHENIX	STAR	Sum
RAW	25	200	100	325
Analysis Objects	5	15	40	60

Storage options:

- Tape (or other cold storage) for archival and scheduled access
- Disk (or other fast storage) for timely and random access

Storage by Data type to be defined.

- RAW data on tape (or other cold storage)
- Analysis Objects on disk (or other fast storage)

Your inputs and feedback are needed

Today

1. RHIC Data Management Plan – Jamie
2. RHIC OpenData – Vincent
3. DAP repository - Eric

- Next meeting: Thursday 4/24