RHIC Data and Analysis Preservation Round Table

05/29/2025

Introduction

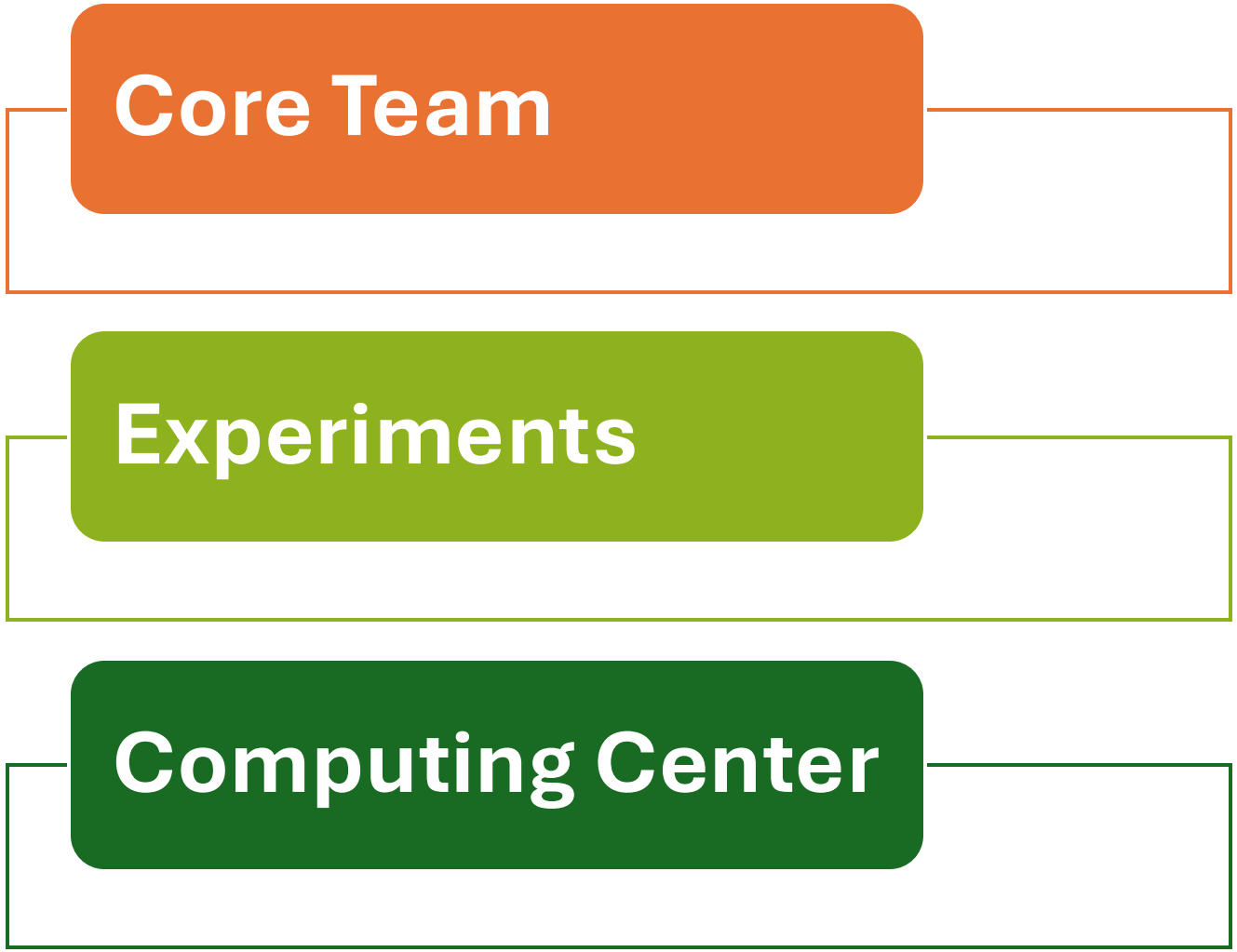
Content

1. Update on DAP team member roles within experiments
2. Estimated hardware resources required for DAP implementation

- On the second topic, key assumptions presented today require discussion and agreement

DAP Implementation

Success Through Collective Engagement



Roles of DAP team members within an experiment

- Updated descriptions incorporate comments received during and after the last meeting – Thank you!
  - Full descriptions available in shared document
  - Emphasis on collaborative roles among the experiments
- **Note:** These roles can be shared among multiple volunteers or combined based on team size and expertise.
- Goal: Ensure all essential functions are covered while maintaining flexibility for different collaboration structures.

Updated roles

Core Team

Experiments

- Preservation coordinator
- Scientific Knowledge Preservation
- Code Preservation
- Engagement & Training
- UX & Documentation Design
- Impact Analysis

Computing Center

RHIC Data Preservation Roles

Core Team

- DAP Manager
- Software & Workflow
- Repository Systems
- AI Integration
- Web Development
- Documentation & QA

Experiments

- Preservation coordinator
- Scientific Knowledge Preservation
- Code Preservation
- Engagement & Training
- UX & Documentation Design
- Impact Analysis

Computing Center

- Computing Center Liaison
- User Support
- Technology Watch

Preservation Coordinator

- **Role**: Provide strategic direction and ensure coordination between the RHIC DAP effort and the other experiments.
- **Focus**: Strategy, cross-collaboration alignment, and priority setting.
- **Example Duties**:
  - Collaborate with Core Team to represent experiment needs in DAP-wide planning discussions.
  - Coordinate with preservation, computing, and physics stakeholders.
  - Partner with teams to define preservation goals and timelines for Phases I and II.
  - Coordinate DAP activities with available resources and scientific priorities within the collaboration.

Scientific Knowledge Preservation

- **Role**: Safeguard essential knowledge necessary for interpreting, reusing, and preserving experiment data and context.
- **Focus**: Content selection, metadata enrichment, and scientific context.
- **Example Duties**:
  - Work with teams to identify and prioritize datasets, code, and documentation for long-term value.
  - Collaborate with domain experts to validate completeness and accuracy of preserved materials.
  - Work with collaborators to ensure metadata captures the scientific context of preserved data.
  - Work with teams to preserve relevant internal notes and calibration constants.
  - Work with Core Team to utilize AI tools for identifying patterns and gaps in scientific data.
  - Evaluate automated metadata generation to improve efficiency and coverage.

Code Preservation

- **Role**: Ensure that software environments remain functional and reproducible beyond the active operational phase of the experiment.
- **Focus**: Preserving analysis environments, containerization, and long-term software accessibility.
- **Example Duties**:
  - Work with teams to inventory critical code and software dependencies across projects.
  - Collaborate with developers to create and maintain container images for key workflows.
  - Partner with technical teams to document execution environments and build systems..
  - Collaborate with computing experts to ensure compatibility with evolving infrastructure.
  - Work with Core Team to investigate AI-assisted code analysis tools for documenting complex workflows.

Engagement & Training

- **Role**: Promote awareness of the DAP effort and allow future users to access and reuse preserved data effectively.
- **Focus**: Outreach, onboarding, user support, internal community engagement.
- **Example Duties**:
  - Work with teams to create online training materials and tutorials for collaboration members.
  - Collaborate with stakeholders to organize events and user onboarding sessions.
  - Partner with teams to promote best practices in documentation across the collaboration.
  - Work with collaboration members to encourage resource and expertise contributions to the DAP.
  - Coordinate with Core Team to integrate AI-powered chatbots into user onboarding and training resources.

UX & Documentation Design

- **Role**: Coordinate with the Core Team to improve access to preserved data and tools via centralized, user-friendly interfaces and clear, well-structured documentation..
- **Focus**: User experience, accessibility, technical writing.
- **Example Duties**:
  - Partner with Core Team to design intuitive navigation and search features for data repositories
  - Work with teams to standardize templates and formatting for collaborative materials
  - Collaborate with Core Team to improve accessibility and discoverability of preserved resources.
  - Test user workflows and iterate designs based on feedback.
  - Integrate AI-driven search tools with Core Team to enhance resource discovery.

Impact Analyst

- **Role**: Provide ongoing input to the Core Team on the scientific and operational impact of RHIC DAP efforts, with particular attention to internal activities and value not readily captured by centrally tracked metrics.
- **Focus**: Impact metrics, reuse tracking, stakeholder reporting.
- **Example Duties:**
  - Collaborate with Core Team to monitor citations, data reuse, and software downloads across projects.
  - Work with collaboration members to compile success stories demonstrating preservation value.
  - Produce summary reports and visualizations for stakeholders.
  - Work with training and UX teams

Resources projections

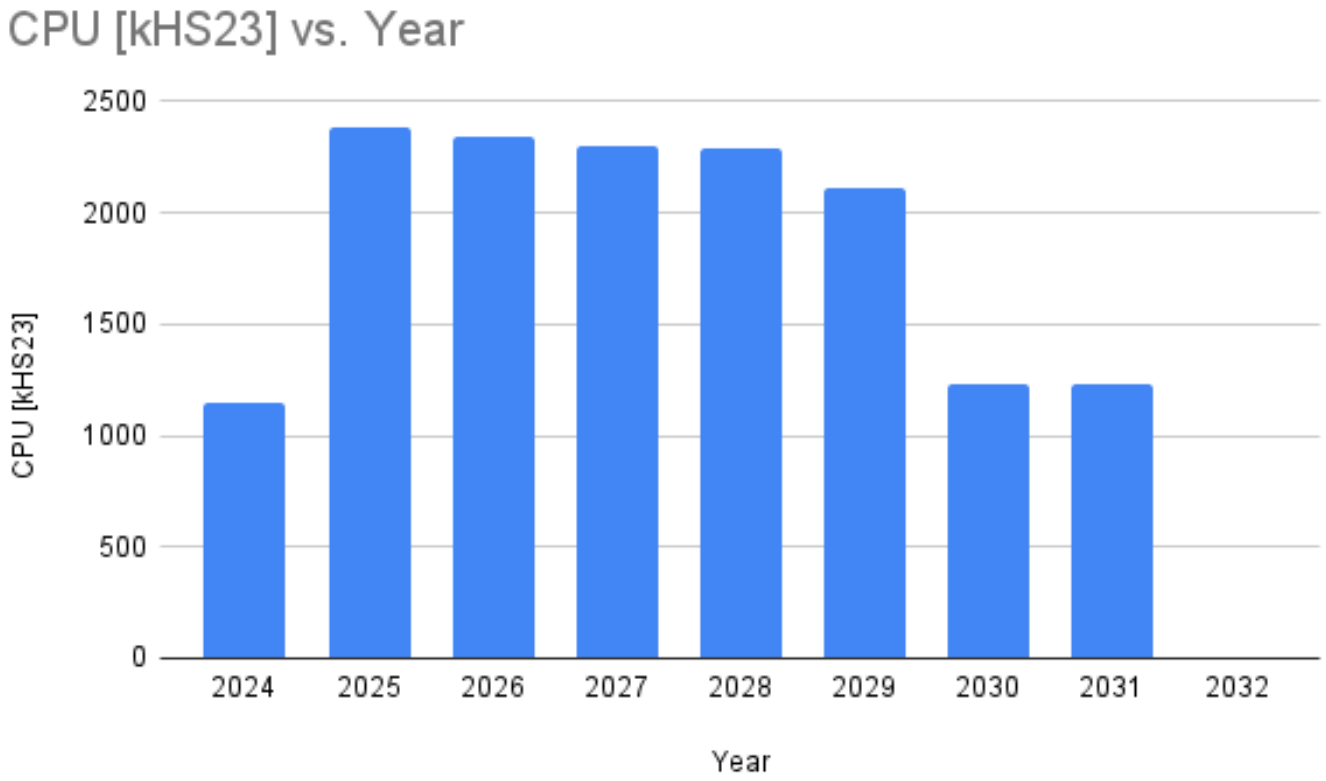
- **Main assumption**: Hardware resources currently deployed will not be renewed past their lifetime
- Hardware lifetime:
  - CPU: 7 years
  - Main Disk Storage / Tape Buffer: 5 years
  - Tape: repack and change of technology every 5 years
- **Challenge:** As lifetimes are different, the matching of resources (CPU processing speed vs disk/tape I/O) over time is not guaranteed.

Evolution of available RHIC CPU

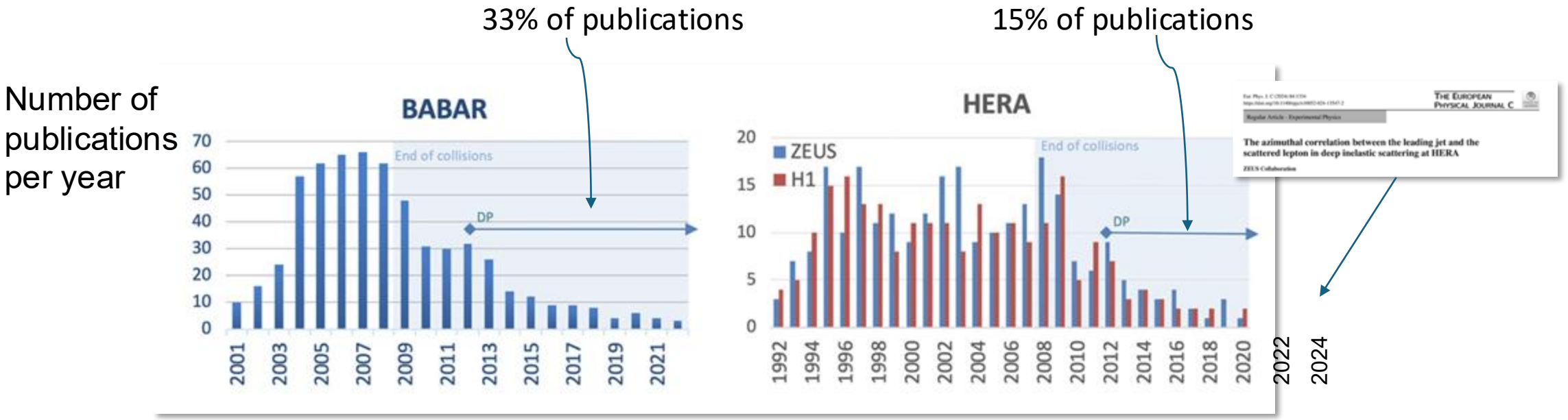
No issue until 2031

2025 sharing:

- sPHENIX: 89%
- STAR: 11%



Data Lives Long

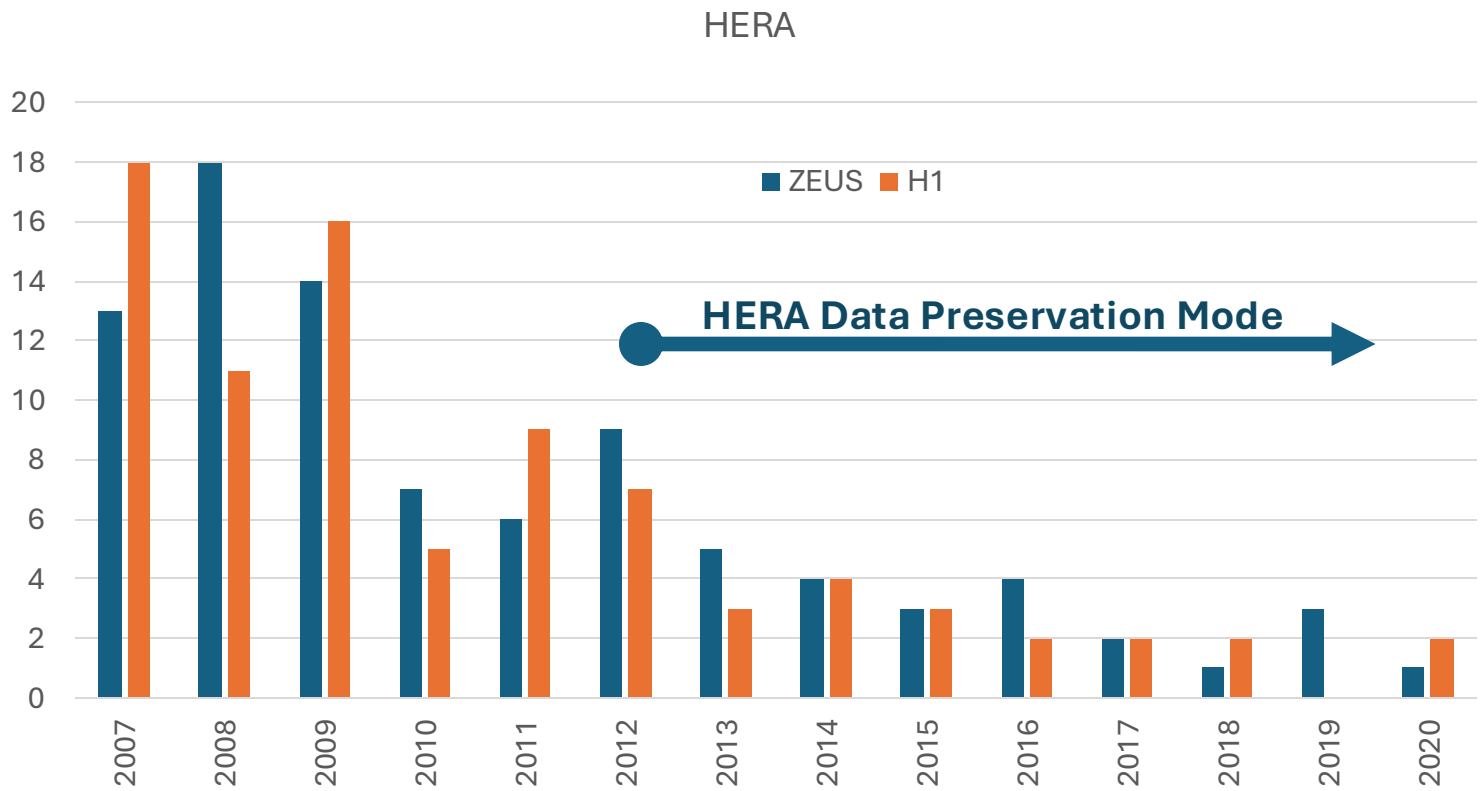


Evidence from other facilities: Significant fraction of publications in Data Preservation mode

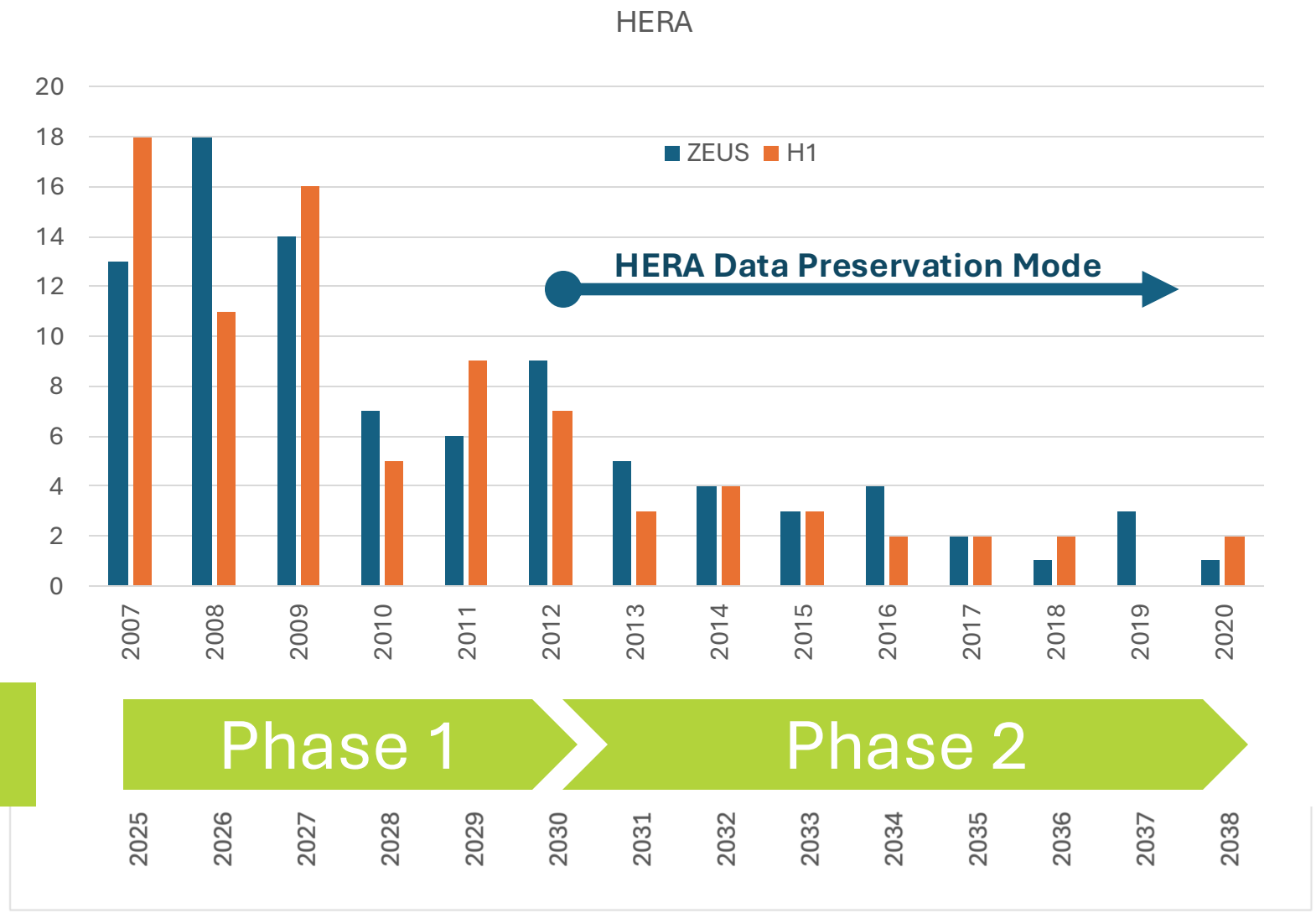
BaBar: More than 1/3 of total publications

HERA experiments: 17% more publications

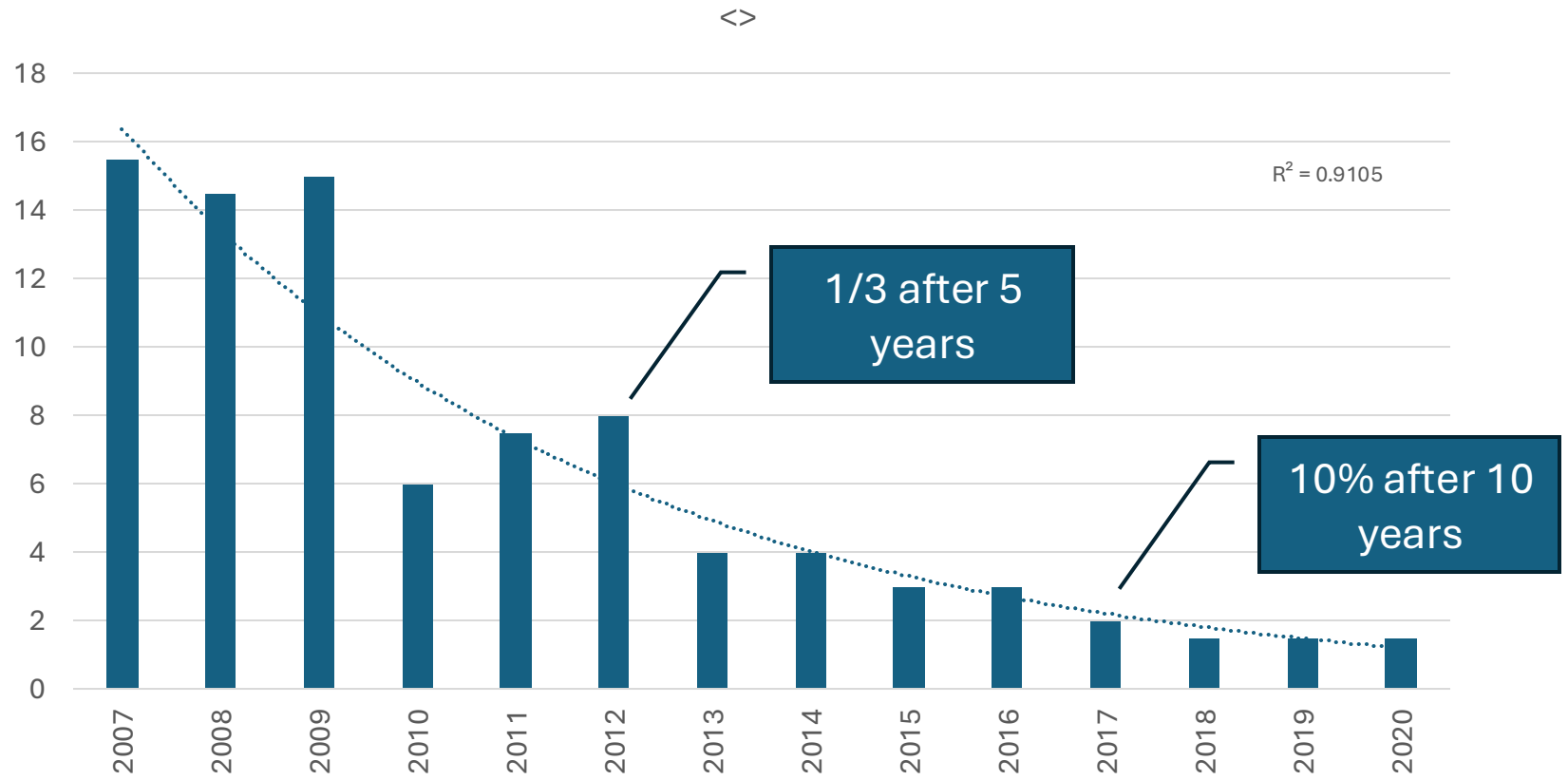
Number of publication per year after Data taking



Number of publication per year after Data taking



Average number of publication per year



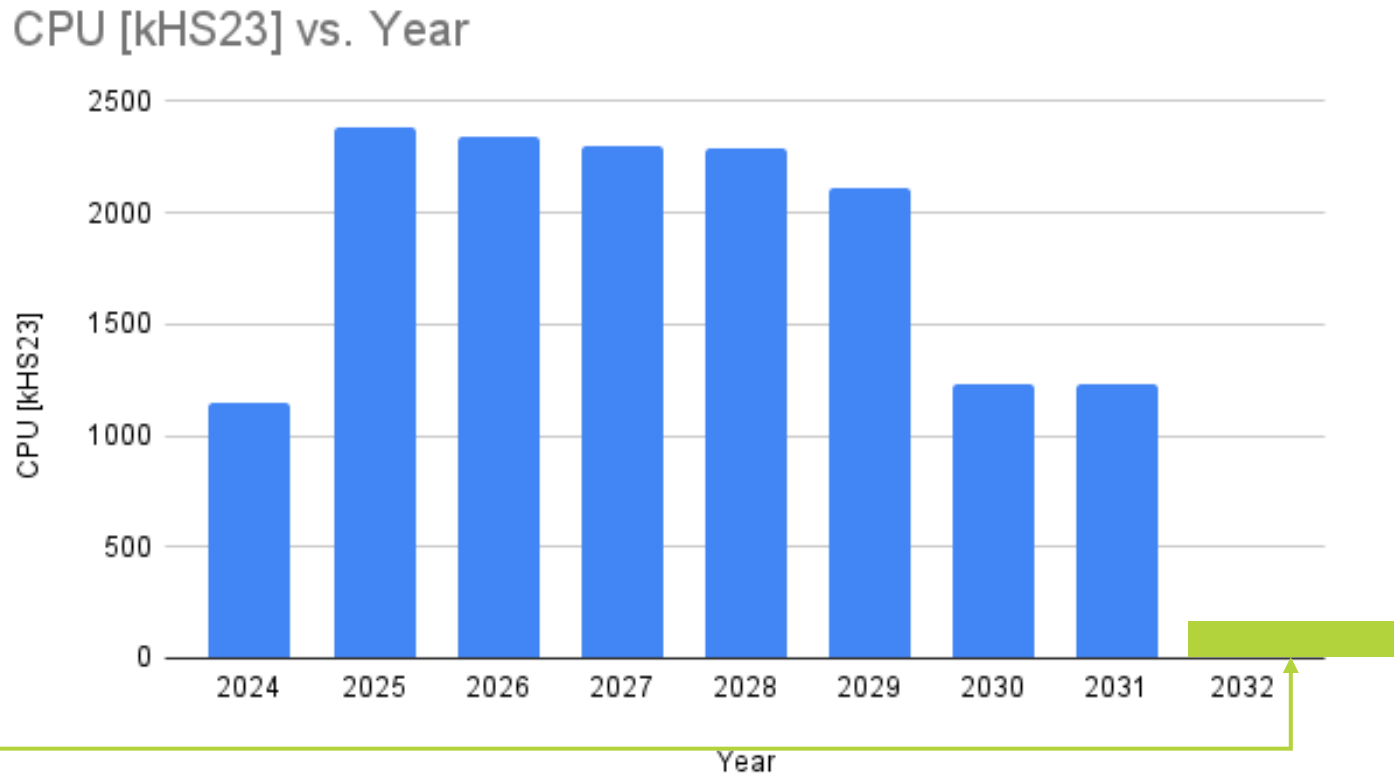
Evolution of available RHIC CPU

No reprocessing after 2030+

Remains only analysis activity

Today STAR analysis activity represents ~20% of total STAR CPU allocation

In 5 years (2030+) analysis need per experiment *may be* 30% of today's consumption: 2 X 30% X STAR analysis share that is



Data volume and preservation levels

[PB]	PHENIX	sPHENIX	STAR
RAW	20	160-300	130
Analysis Objects	5	50-100 (one processing)	45
Other archive	10	50-100 (prev. processing)	?

Other data: historical data, previous processing, etc..

Need to be added for planning purpose

What are the other archive?

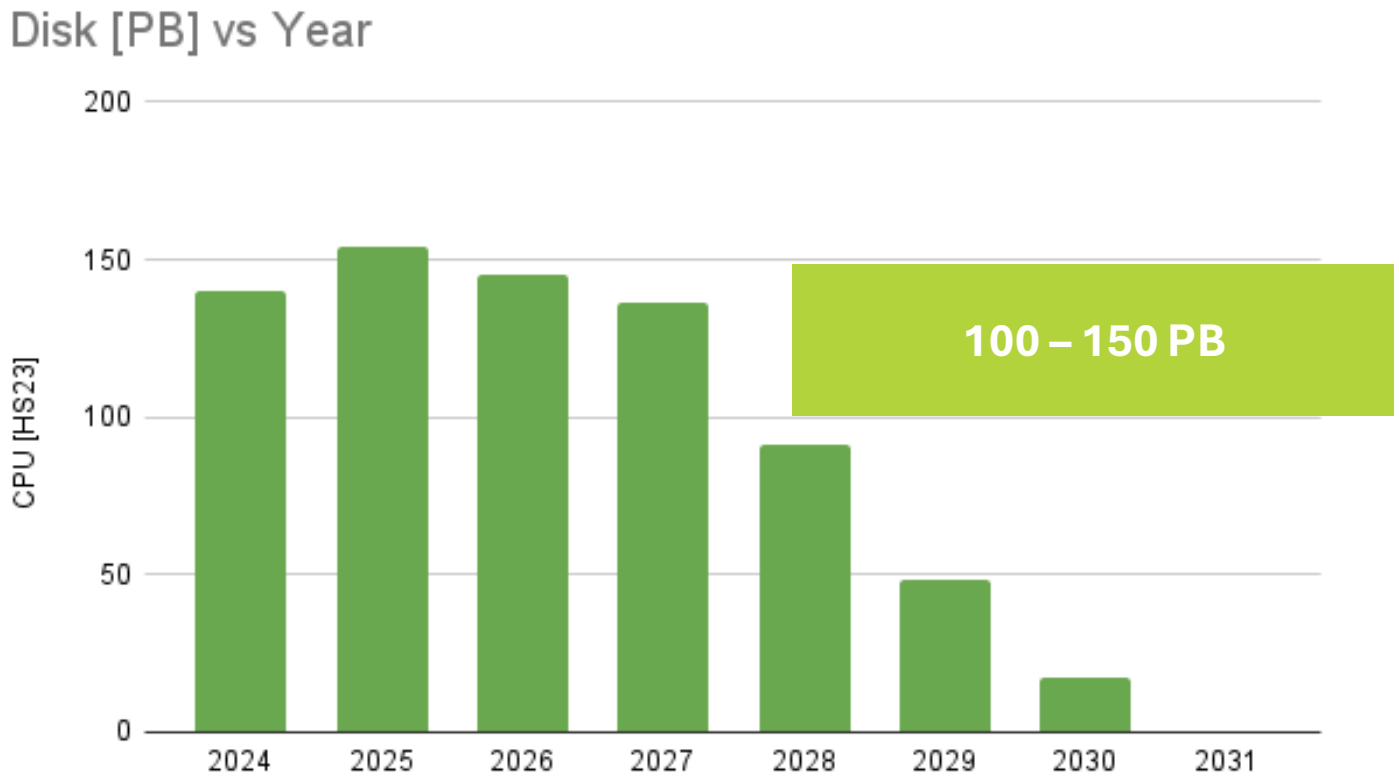
Data volume and preservation levels

[PB]	PHENIX	sPHENIX	STAR	Total
RAW	20	160-300	130	310-450
Analysis Objects	**5**	**50-100 (one processing)**	**45**	**100-150**
Other archive	10	50-100 (prev. processing)	?	?

Analysis Objects: Preservation level 3
Ideally on disk
Needs : 100-150 PB

Evolution of available RHIC Disk Storage

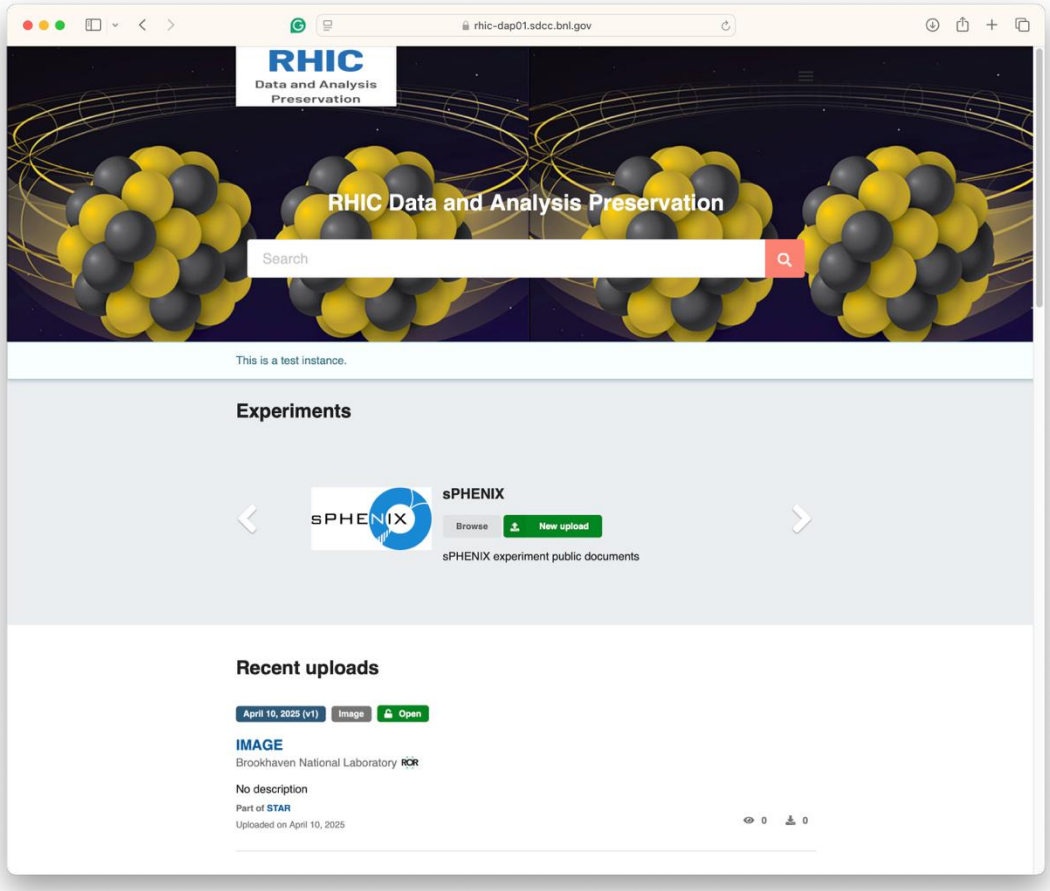
Deploying and maintaining 100–150 PB of disk storage for Analysis Objects presents significant challenges

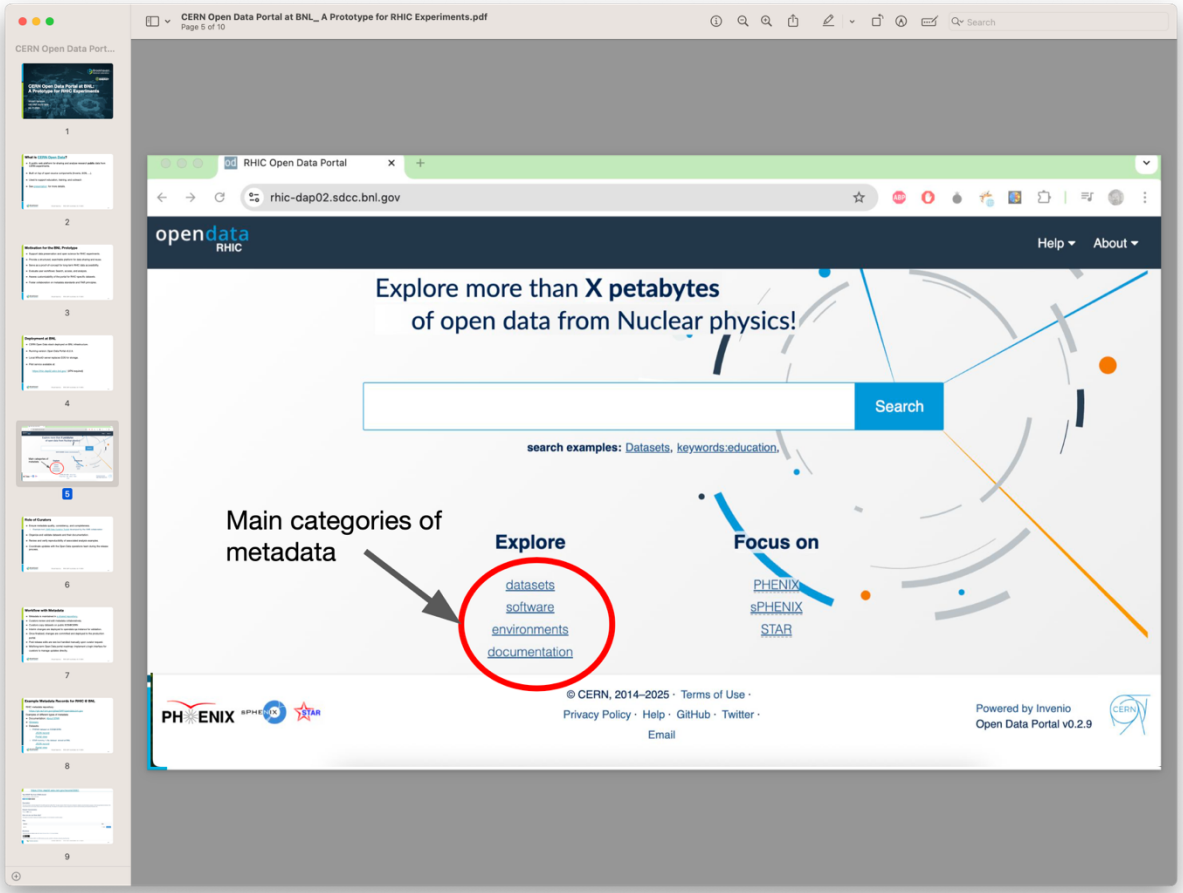


Give it a try!

https://rhic-dap01.sdcc.bnl.gov/ — BNL network access required

https://rhic-dap02.sdcc.bnl.gov/ (VPN required)





Today

1. Hardware projections - Shigeki

- Next meeting: **Thursday, 06/04**