

ePIC Software & Computing Report



Markus Diefenthaler (Jefferson Lab) on behalf of ePIC Software & Computing

ePIC Software & Computing Organization



Coordinators and WG Conveners



Development



Dmitry Kalinkin



Simulation WG: Simon Gardner, Sakib Rahman. **Reconstruction WG:** Derek Anderson, Chandra Chatterjee, Shujie Li.

Infrastructure



Torre Wenaus

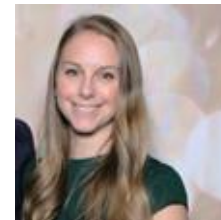
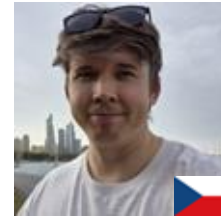


Streaming Computing WG: Marco Battaglieri, Taku Gunji, Jeff Landgraf.

Operations

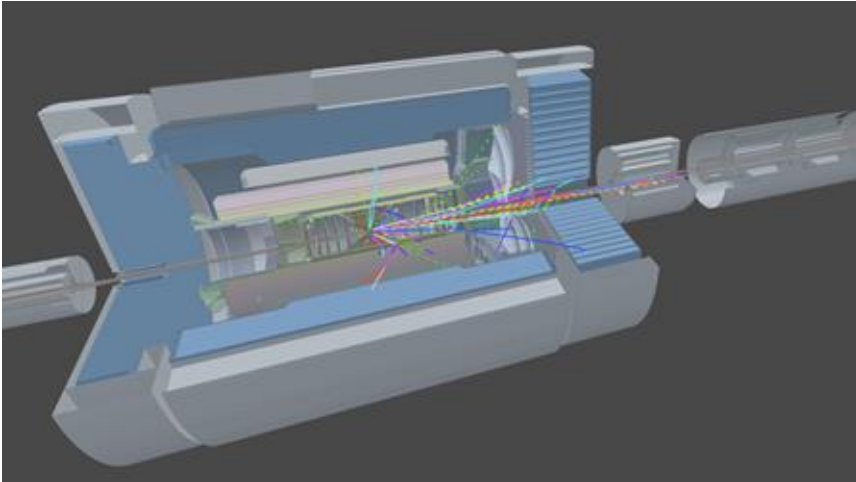


Holly Szumila-Vance



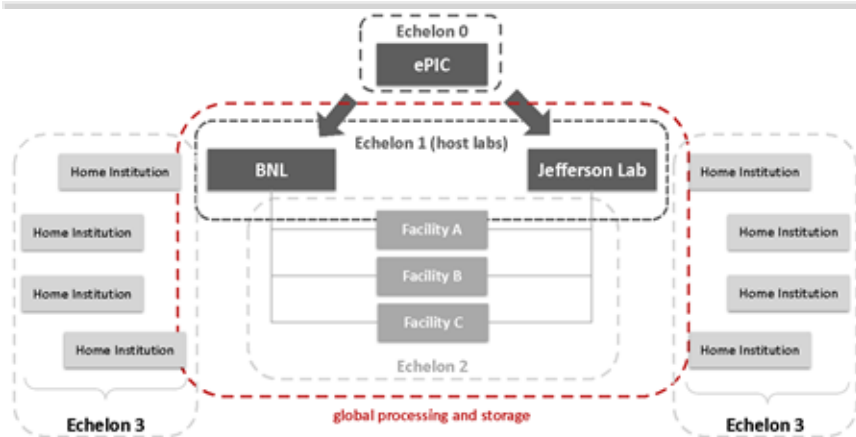
Production WG: Thomas Britton, Sakib Rahman. **User Learning WG:** Stephen Kay, Sasha Prozorov. **Validation WG:** Torri Jeske.

Roles



Enable Science

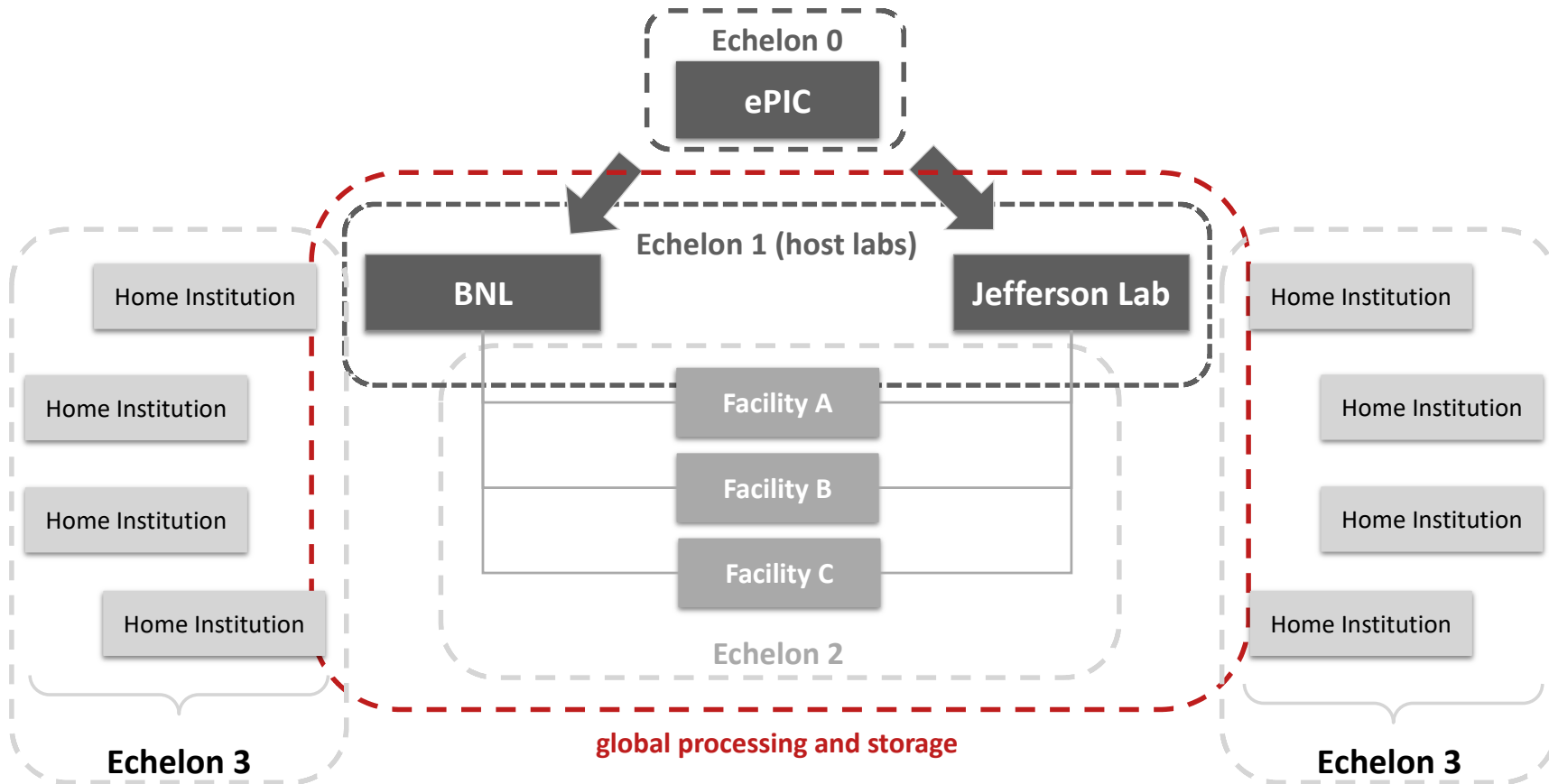
- Development, operations, and infrastructure for preTDR and future milestones.
- **Activity:** Monthly software releases and simulation campaigns tailored to collaboration needs.



Compute–Detector Integration

- Maximize and accelerate science with the ePIC computing model
- **Activity:** Define and test DAQ–computing interfaces to mitigate integration risks. Develop and deliver a functional testbed to validate workflow management for autonomous detector calibration

The ePIC Streaming Computing Model



Successfully reviewed in
2023–2025 and documented

The ePIC Streaming Computing Model
Version 2, Fall 2024

Marco Battaglieri¹, Wouter Deconinck², Markus Diefenthaler³, Jin Huang⁴, Sylvester Joosten⁵, Dmitry Kalinkin⁶, Jeffery Landgraf⁶, David Lawrence³ and Torre Wenaus⁴
for the ePIC Collaboration

¹Istituto Nazionale di Fisica Nucleare - Sezione di Genova, Genova, Liguria, Italy.
²University of Manitoba, Winnipeg, Manitoba, Canada.
³Jefferson Lab, Newport News, VA, USA.
⁴Brookhaven National Laboratory, Upton, NY, USA.
⁵Argonne National Laboratory, Lemont, IL, USA.
⁶University of Kentucky, Lexington, KY, USA.

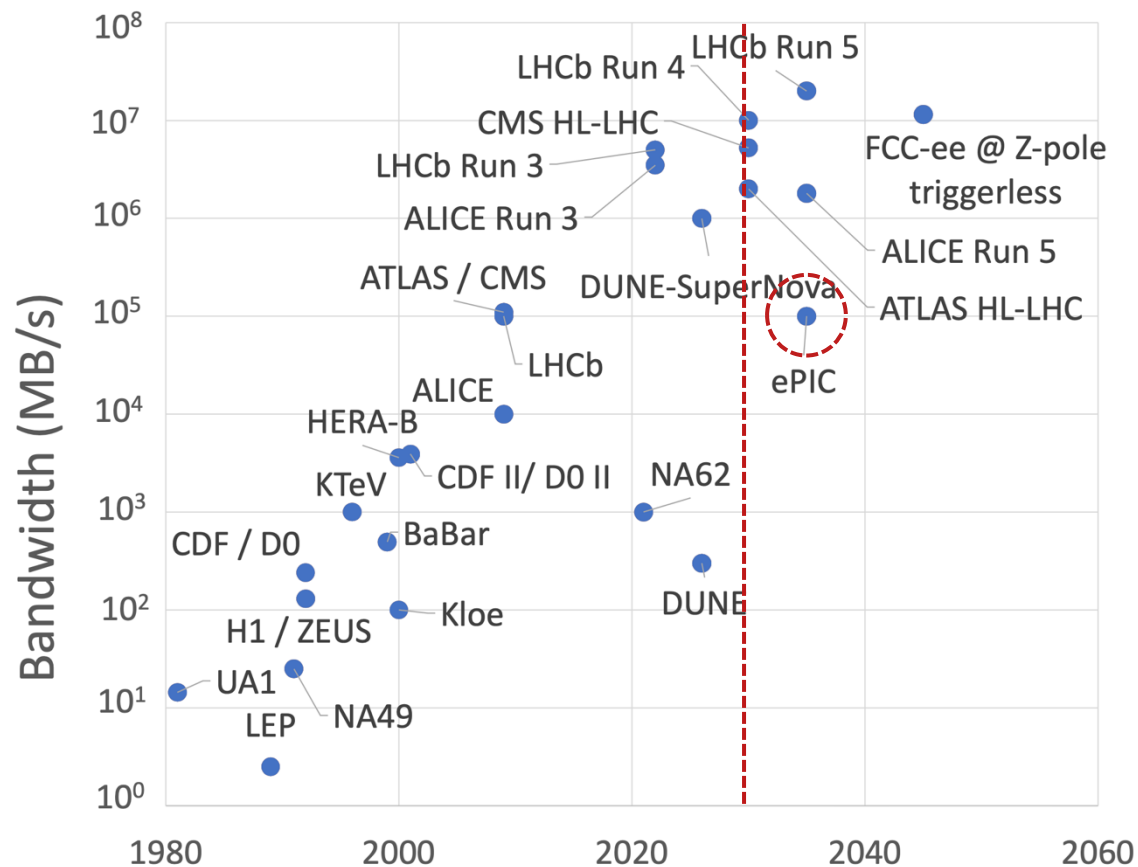
Abstract

This second version of the ePIC Streaming Computing Model Report provides a 2024 view of the computing model, updating the October 2023 report with new material including an early estimate of computing resource requirements; software developments supporting detector and physics studies, the integration of ML, and a robust production activity; the evolving plan for infrastructure, dataflows, and workflows from Echelon 0 to Echelon 1; and a more developed timeline of high-level milestones. This regularly updated report provides a common understanding within the ePIC Collaboration on the streaming computing model, and serves as input to ePIC Software & Computing reviews and to the EIC Resource Review Board. A later version will be submitted for publication to share our work and plans with the community. New and substantially rewritten material in Version 2 is dark green. The present draft is preliminary and incomplete and is yet to be circulated in ePIC for review.

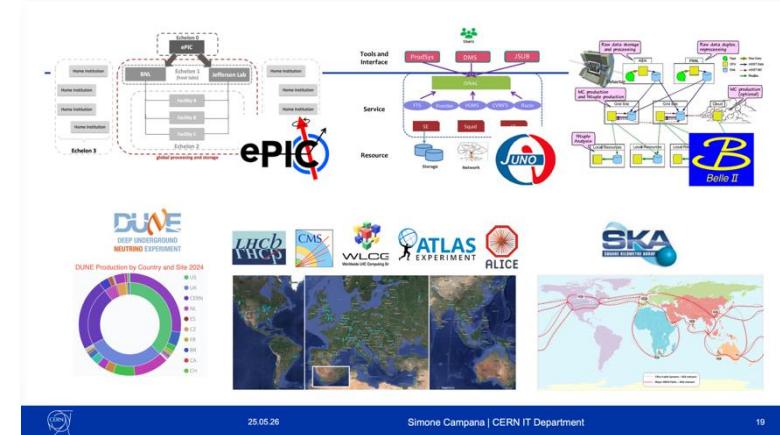
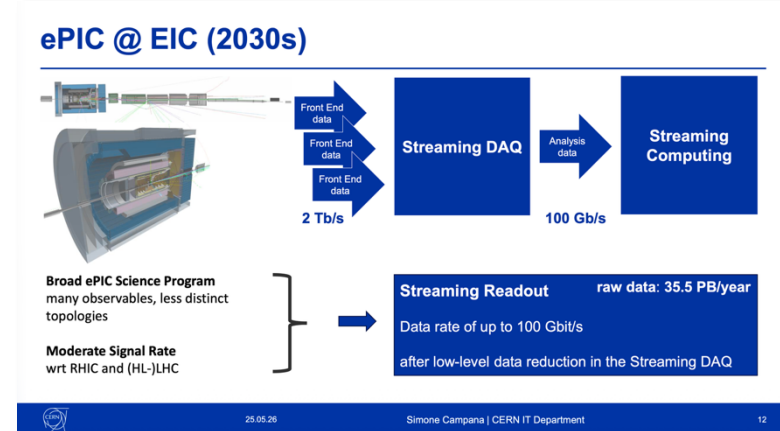
<https://doi.org/10.5281/zenodo.14675920>

The ePIC Streaming Computing Model is **physics-driven**. It has been developed to **accelerate the pace of discovery** and **enhance scientific precision** through streaming readout, AI, and distributed computing.

ePIC in the Global Particle-Physics Computing Landscape



From Open Symposium on the European Strategy for Particle Physics



From CHEP 2026 keynote

The computing **model** is generating **strong interest** within the global particle physics community, with **leadership recognized** through advisory roles in the HSF, HL-LHC Computing Review Panel, and the LHC Experiments Committee.



Distributed Computing for ePIC

- **Echelon 1** sites uniquely perform the **low-latency streaming workflows**.
- Apart from low-latency, **Echelon 2** sites fully participate in use cases and **accelerate** them.

Strong international interest and engagement:



Use Case	Echelon 0	Echelon 1	Echelon 2	Echelon 3
Streaming Data Storage and Monitoring	✓	✓		
Alignment and Calibration		✓	✓	
Prompt Reconstruction		✓		
First Full Reconstruction		✓	✓	
Reprocessing		✓	✓	
Simulation		✓	✓	
Physics Analysis		✓	✓	✓
AI Modeling and Digital Twin		✓	✓	

Substantial role for Echelon 2 in preliminary resource requirements model.

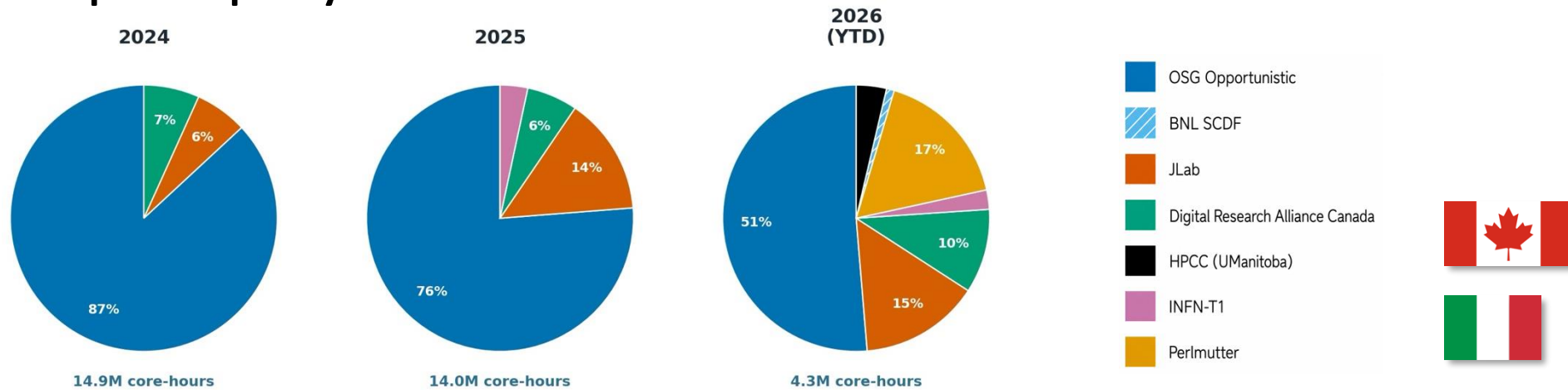
Assumed Fraction of Use Case Done Outside Echelon 1	
Alignment and Calibration	50%
First Full Reconstruction	40%
Reprocessing	60%
Simulation	75%

Distributed computing provides **flexibility**, influencing the fraction of use cases handled outside Echelon 1.

Distributed Computing and Storage: Status

Monthly simulation campaigns for preTDR and CD-2 preparations:

Compute Capacity



Storage Capacity



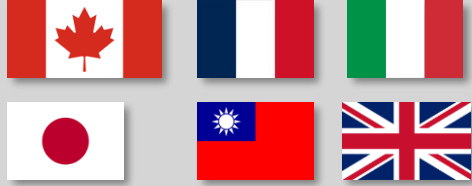
ePIC is building its distributed compute and storage fabric now, with ongoing discussions about additional international contributions.



EIC International Computing Organization

Discussion Focus

- International partner engagement:



- ePIC Streaming Computing Model:

- Coordination of data challenges with compute and data centers.
 - Echelon 2 roles and participation.
 - **Networking:** ESnet and LHCONE
- **WLCG:** Technical exchange and associated membership path.



ECSJI and EICO Annual Meeting (May 18–19, 2026, CERN)

ePIC is working closely with ECSJI and EICO to define international participation in the ePIC Streaming Computing Model and advance its implementation.

Use-Case-Driven Echelon 2 Participation

Participation being defined by use-case-specific capabilities, not by fixed site categories:

Capability Envelope

- Describes the **smallest set of resources and operational capabilities required** for a site to participate for a given use case.
- It is not a single number, but a combination of **requirements across multiple dimensions**:
 - compute · storage · network throughput · I/O performance · operational reliability
- A site **must satisfy all dimensions simultaneously**; otherwise, it cannot effectively execute the workload.
- Envelope will be defined pragmatically as an appropriate balance between added value and operational load.

Approach separates two distinct questions:

- **Qualification**: whether a site has sufficient capabilities to participate in a given use case (defined by the envelope).
- **Contribution**: what fraction of the global workload the site takes on (expressed as a percentage and coordinated across sites to match total resource needs).

Capability envelopes keep the computing model realistic and use-case driven, while allowing Echelon 2 contributions to meet global compute and storage targets.

Echelon 2 Resource Needs

Based on a luminosity of $L = 1 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1} = 1 \text{ kHz}/\mu\text{b}$

Processing by Use Case [cores]	Echelon 1	Echelon 2
Streaming Data Storage and Monitoring	-	-
Alignment and Calibration	6,004	6,004
Prompt Reconstruction	60,037	-
First Full Reconstruction	72,045	48,030
Reprocessing	144,089	216,134
Simulation	123,326	369,979
Total estimate processing	405,501	640,147

Storage Estimates by Use Case [PB]	Echelon 1	Echelon 2
Streaming Data Storage and Monitoring	71	35
Alignment and Calibration	1.8	1.8
Prompt Reconstruction	4.4	-
First Full Reconstruction	8.9	3.0
Reprocessing	9	9
Simulation	107	107
Total estimate storage	201	156

O(1M) core-years to process a year of data:

- Even with performance gains over the years, the required processing scale remains substantial.
- Highlights the need to leverage distributed and opportunistic resources from the outset.
- Distributed computing is an integral part of our computing model.

~350 PB to store data of one year.

Echelon 2 qualification may require a **defined fraction of the estimated resource needs**, e.g., 10% to 20% of the values shown, with the appropriate threshold under discussion within ePIC and EICO.

Streaming DAQ and Computing Milestones



FY25	FY26	FY27	FY28	FY29	FY30	FY31
PicoDAQ	MicroDAQ	MiniDAQ	Full DAQ-v-1	Production DAQ		DAQ
Streaming Orchestration			Streaming Challenges			
AI-Empowered Streaming Data Processing			Analysis Challenges		Computing	
				Distributed Data Challenges		
AI-Driven Autonomous Calibration			AI-Driven Autonomous Alignment, Calibration, and Control			AI

- **Compute-Detector Integration:**

- Joint deliverables between **DAQ** and **computing** to develop an integrated system.
- We define and test DAQ–computing interfaces to reduce integration risk and enable early system testing.
- **FY28Q1 priorities:** Testbeds and prototypes for streaming orchestration and autonomous detector calibration for one detector system

- **Integrating AI with ePIC Detector:**

- AI has been built into the ePIC computing model from the outset.
- We are developing AI-driven workflows for alignment, calibration, validation, control, and diagnostics.
- This enables end-to-end data processing from detector readout to physics analysis.

Streaming Orchestration: BNL Testbed

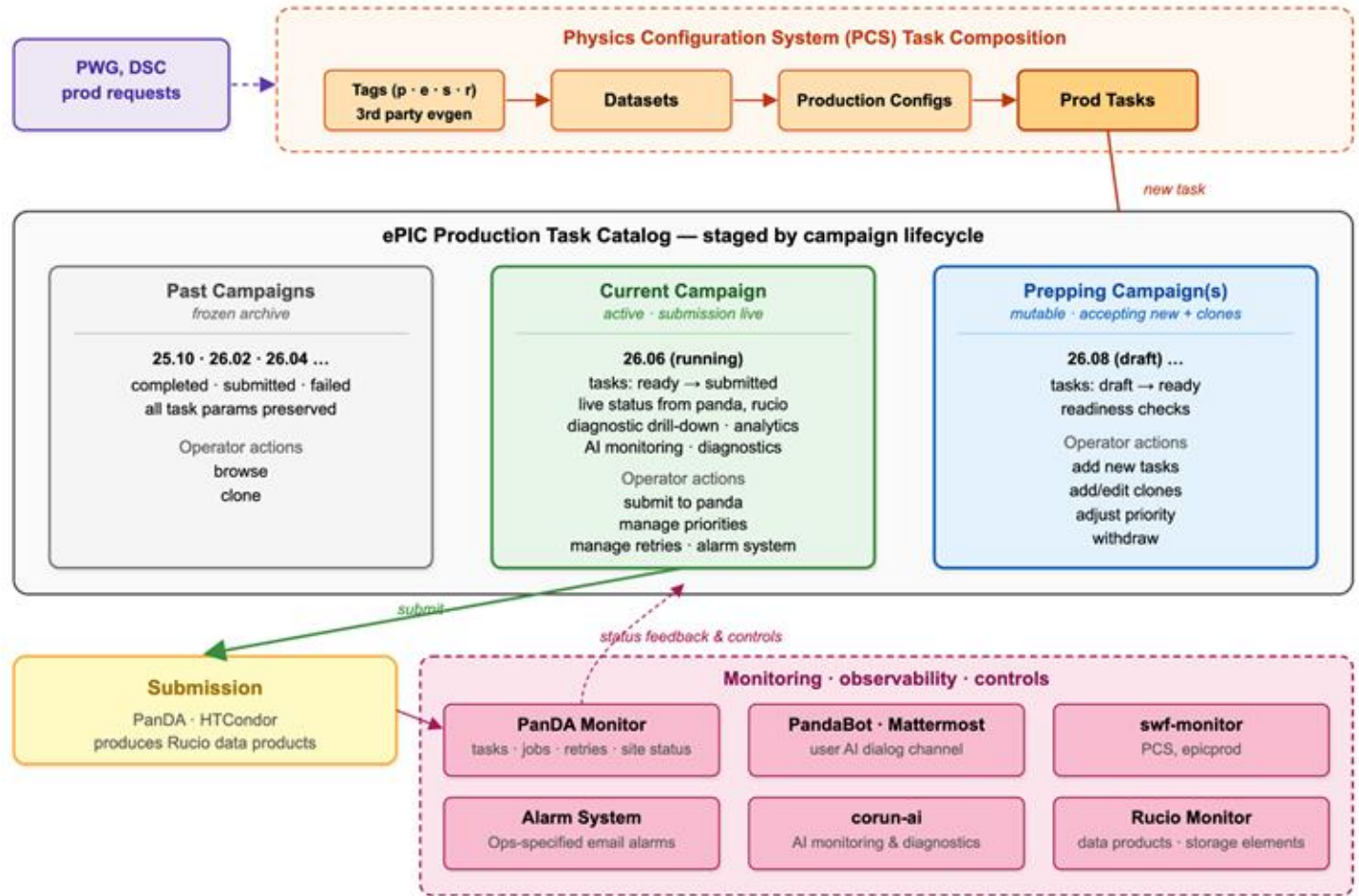


Rapid Progress

- Build streaming orchestration with **PanDA** and **Rucio**
- Enhance Panda with **MCP Tools** and **AI Services** for autonomous workflows

Prototype workflow

- Automated simulation campaigns
- Addresses urgent need
- Rollout at July collaboration meeting



Near-term milestone of automated simulation campaigns will **exercise streaming orchestration with PanDA and Rucio** and clarify where ePIC streaming workflows differ from traditional offline processing.

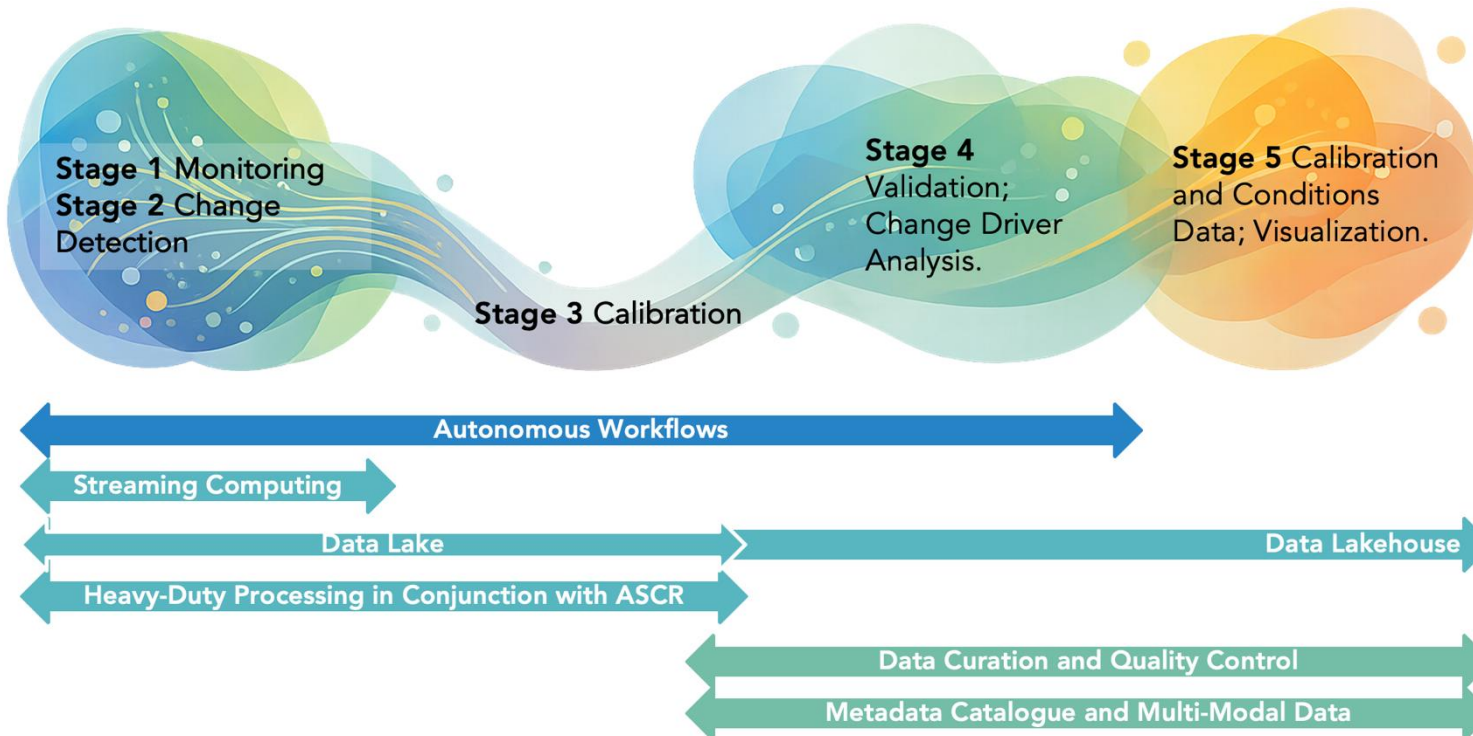


Autonomous Calibration as a Data Lifecycle Demonstrator



HPDF will be a cutting-edge high-performance data platform that provides advanced storage, processing, and analytical services, empowering data-intensive research and AI-driven discoveries.

In collaboration with the **HPDF Project**, we are demonstrating how **AI-ready data lifecycle management** and FAIR principles can enhance **autonomous calibration workflows** and their **data pipelines**. This is key in the era of **AI and data**, and complementary to the BNL testbed.

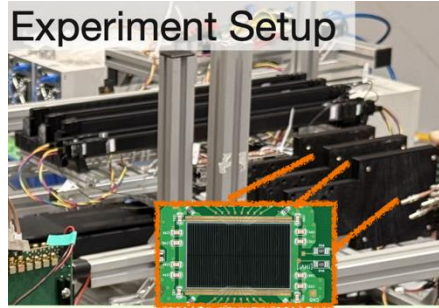


Embed FAIR principles across the data lifecycle

1. Data Capture
2. Initial Processing
3. Central Processing and Enrichment
4. Storage, Curation, and Archiving
5. Dissemination and Visualization



Integrated DAQ-Computing Tests in Japan



- Beam test by the barrel TOF DSC at RARiS, Tohoku, Japan, March 3-7, 2026.
- Mini ePIC streaming DAQ with SPADI was demonstrated in a beam-test environment.
- Two EICrecon integration paths partially established:

File-based path

AMANEQ → NestDAQ → EDM4eic file sink → XRootD → EICrecon

Achieved EDM4eic output, EICrecon file reading, and XRootD-based file transfer.

Message-queue path

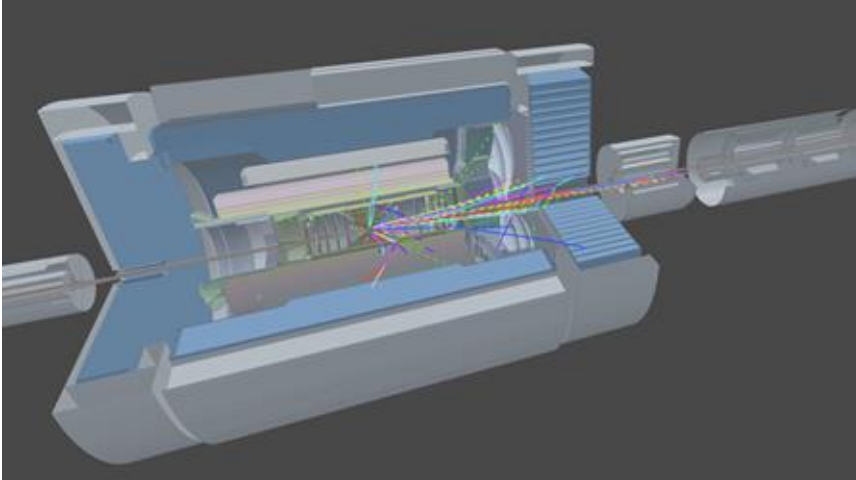
AMANEQ → NestDAQ → ZeroMQ → EICrecon

Initial EICrecon ZeroMQ receiver code was prepared, and the basic receiver appears to work.

- **Key outcomes**
 - Actual beam-test data were written as **edm4eic::RawTrackerHit**.
 - **EICrecon** produced calibrated hits and clusters from EDM4eic data.
 - **XRootD**-based transfer from RARiS to RCNP/University of Osaka was demonstrated.

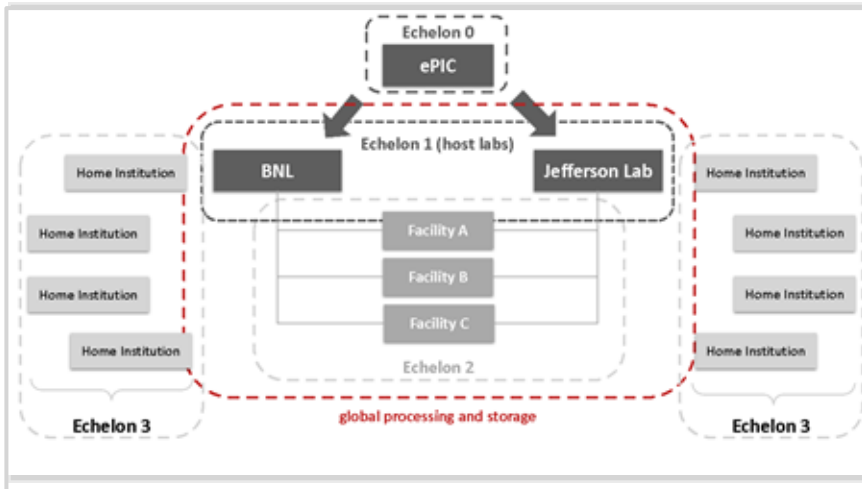
ePIC streaming DAQ with SPADI: Beam-test data were integrated into EICrecon through both file-based and message-queue paths, demonstrating progress toward end-to-end DAQ-computing integration

Roles



Enable Science

- Development, operations, and infrastructure for preTDR, Early Science Report, and future milestones.
- **Activity:** Monthly software releases and simulation campaigns tailored to collaboration needs.



Compute–Detector Integration

- Maximize and accelerate science with the ePIC computing model
- **Activity:** Define and test DAQ–computing interfaces to mitigate integration risks. Develop and deliver a functional testbed to validate workflow management for autonomous detector calibration

Updated Landing Page



The screenshot shows the EIC Software Portal landing page. At the top, there is a navigation bar with 'Get Started', 'Tutorials', and 'Software' menus, and a search bar containing 'Ask anything about EIC...'. Below the navigation is a large header section with a globe icon and the text 'EIC Software Portal' and 'This is the main portal to the EIC software, repositories, documentation and resources. It is developed and maintained by the EICUG and the ePIC Collaboration.' Below this are three main buttons: 'Get Started', 'ePIC Tutorials', and 'HEP Software Training Center'. A grid of service icons follows, including EPIC, Zenodo, Mattermost, Indico, GitHub, Wiki, Phonebook, and EICUG. At the bottom, there is a mailing list subscription link and a footer with copyright information and a build timestamp.

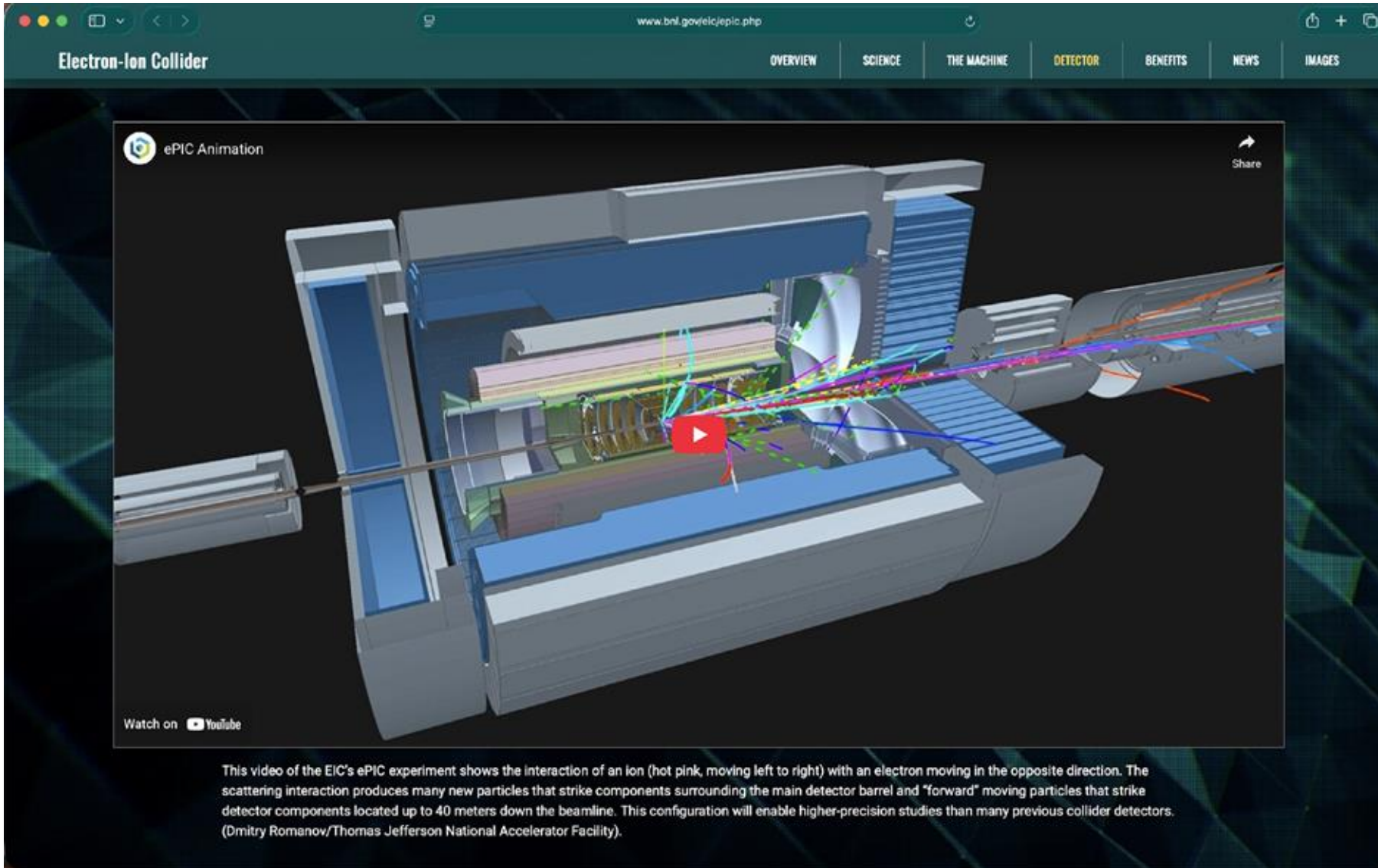
This screenshot shows a search result for 'how do I install eic-shell'. The search bar at the top contains the query. Below it, a snippet of the answer is visible, including a terminal command: `cd ~/eic && curl -s -L https://get.epic-esc.org | bash`. The snippet also mentions creating a script and running it. Below the snippet are two buttons: 'Get Started' and 'ePIC'.

The screenshot shows the HEP Software Training Center. At the top, it says 'TRAINING CENTER' and 'Browse the current ePIC tutorial collection.' Below this is a grid of four tutorial cards. Each card has a title, a brief description, and an 'Open tutorial' button. The cards are: 'Setting up an environment', 'Analysis and simulation output', 'File Access with Rucio', and 'Simulating detectors'.

Reorganization and new features, including mobile-device support and a search bar, ensure that the landing page remains an excellent resource for onboarding new collaboration members and supporting training.



Firebird Event Display



Electron-Ion Collider

OVERVIEW SCIENCE THE MACHINE **DETECTOR** BENEFITS NEWS IMAGES

ePIC Animation

Share

Watch on YouTube

This video of the EIC's ePIC experiment shows the interaction of an ion (hot pink, moving left to right) with an electron moving in the opposite direction. The scattering interaction produces many new particles that strike components surrounding the main detector barrel and "forward" moving particles that strike detector components located up to 40 meters down the beamline. This configuration will enable higher-precision studies than many previous collider detectors. (Dmitry Romanov/Thomas Jefferson National Accelerator Facility).

Firebird supports debugging, quality control, and outreach through intuitive visualizations of detector responses, particle interactions, and reconstruction outputs.

A recording of the ePIC event display is available on the [EIC webpage](#).

Three-Year Software & Computing Plan

Bottom-up • Collaborative



1

Work with WGs

Build on existing three-year outlooks.



2

Define 3-Year Goals

Define 3-year goals for each WG.



3

WG Plans

Each WG develops its plan using an example plan as guidance



4

Draft Plan

Include infrastructure plans and resource estimates.



5

Discuss & Refine

Present draft at collaboration meeting and refine through discussion.



6

Final Plan

Prepare final plan for October review.



Summary

Computing turns detector data into physics insight through calibration, reconstruction, validation, and analysis.

The **ePIC Computing Model** defines a global fabric required for data processing and analysis. The model is physics-driven, built on streaming readout, AI-enabled automation, and heterogeneous computing.

ePIC is building its **distributed compute and storage fabric** now. A substantial Echelon 2 role is foreseen.

Priorities

- Compute-Detector Integration:** Define DAQ-computing interfaces and deliver a testbed for streaming orchestration, with autonomous calibration workflows as test cases.
- Enable Science:** Monthly software releases and simulation campaigns tailored to collaboration needs.

