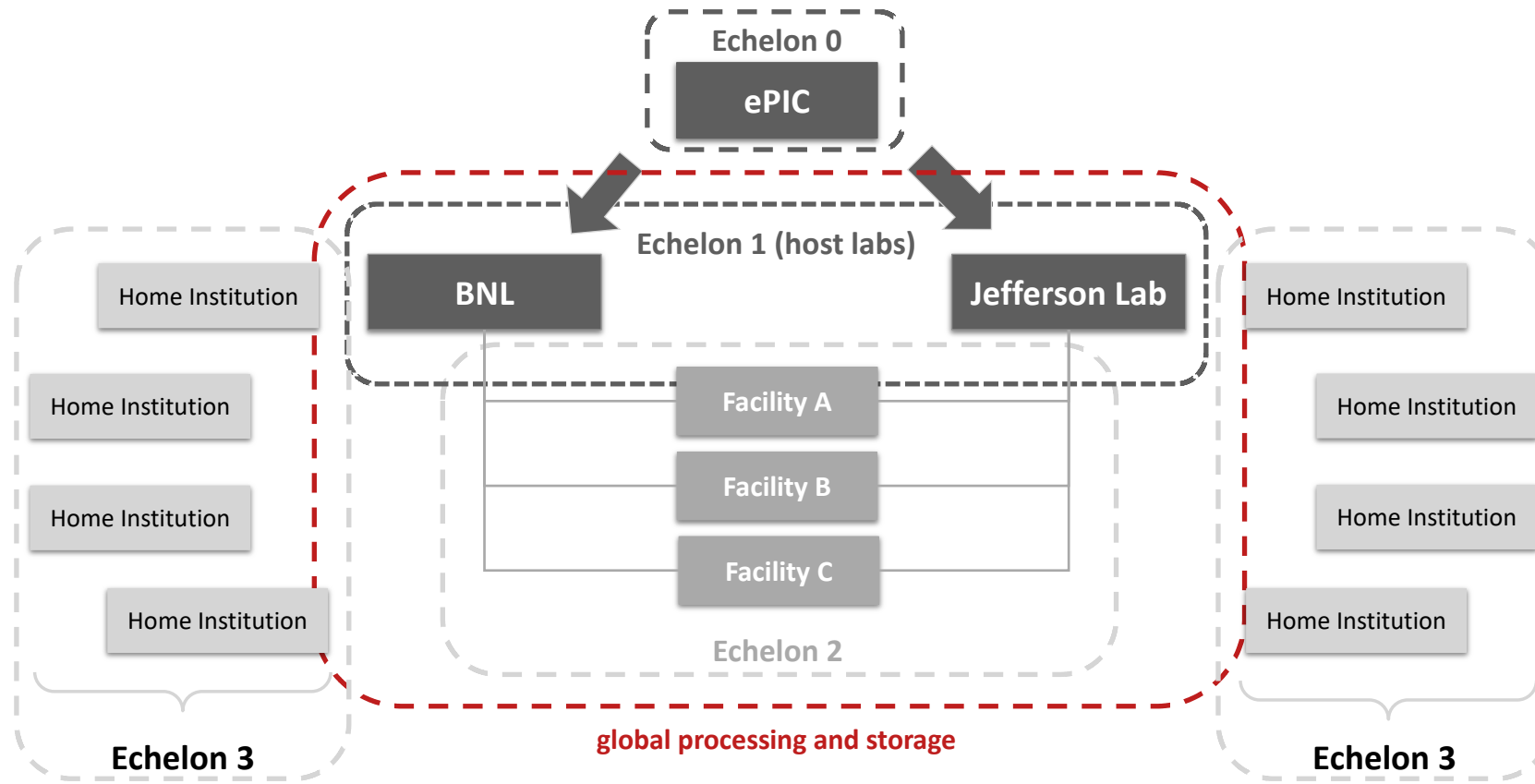


ECSJI and EICO Annual Meeting Report: EIC Networking and Echelon 2 Requirements



Markus Diefenthaler (Jefferson Lab)

Echelon 2 Roles

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, and this flexibility will influence the fraction of use cases handled at Echelon 2. We have decided to keep the currently assumed fraction.

Echelon 2 Resource Needs (2034)

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 an integral part of our computing model.

~350 PB to store data of one year.

The resource estimates are based on a luminosity of $L = 1 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1} = 1 \text{ kHz}/\mu\text{b}$. While there is ongoing discussion regarding the EIC strategy and the luminosity during the first years of running, we will not update the resource estimates at this time. The current estimates will be maintained until the EIC strategy is better defined.

Echelon 2 Participation Model: Discussion

We decided against defining fixed categories of Echelon 2 sites. Instead, we will specify minimum compute and storage requirements for each use case. Individual sites will determine which use cases they choose to support.

Illustrative example of how fixed categories could have looked

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

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).

By using capability envelopes, the computing model remains realistic and use-case driven, while allowing Echelon 2 contributions to sum to the required global compute and storage targets.

Use-Case-Specific Compute and Storage Estimates

Processing by Use Case [cores]	20%	10%
Streaming Data Storage and Monitoring	-	-
Alignment and Calibration	1200	600
Prompt Reconstruction	-	-
First Full Reconstruction	9606	4803
Reprocessing	43227	21613
Simulation	73996	36997

Storage Estimates by Use Case [PB]	20%	10%
Streaming Data Storage and Monitoring	7	3.5
Alignment and Calibration	0.4	0.2
Prompt Reconstruction	-	-
First Full Reconstruction	0.6	0.3
Reprocessing	1.8	0.9
Simulation	21.5	10.7

10% and 20% represent possible Echelon 2 contribution levels.

The goal is to identify meaningful site contributions while keeping operational load manageable.

Next Steps

Timeline

- **EIC RRB update:** Present progress on the Echelon 2 participation model.

Define capability envelopes by use case

- ePIC defines the requirements; EICO feedback will ensure they are practical for participating sites.
- **Near-term deliverable**

Develop the contribution framework with EICO

- Clarify how individual site contributions will combine to meet global resource needs.
- Site commitments will be addressed later, but the process should be understood now.