DOE’s Integrated Research Infrastructure (IRI) in the AI era

The 11th Innovating the Network for Data-Intensive Science (INDIS 2024)
November 18, 2024 | Atlanta, Georgia

Benjamin Brown, Ph.D.
Director, Facilities Division
Advanced Scientific Computing Research

Outline

- The DOE Office of Science’s major **research infrastructure**
- DOE’s **Integrated Research Infrastructure (IRI)** program
- ESnet’s pivotal role
- High performance networking and R&E networking in the AI era

Acknowledgement:

Thank you to Inder Monga and Chin Guok, ESnet, for collaboration on slides

The dawn of the AI era.
The dawn of the nuclear era.

understand risk

harness potential

U.S. Department of ENERGY | **Office of Science**

Our Mission:

Deliver scientific discoveries and major scientific tools to transform our understanding of nature and advance the energy, economic, and national security of the United States.

More than **34,000 r**esearchers supported at more than **300** institutions and **17** DOE national laboratories

Steward **10** of the 17 DOE national laboratories

More than **39,500** users of **28** Office of Science scientific user facilities

$8.24B
FY 2024 enacted

DOE National Laboratories

Office of Science Laboratories

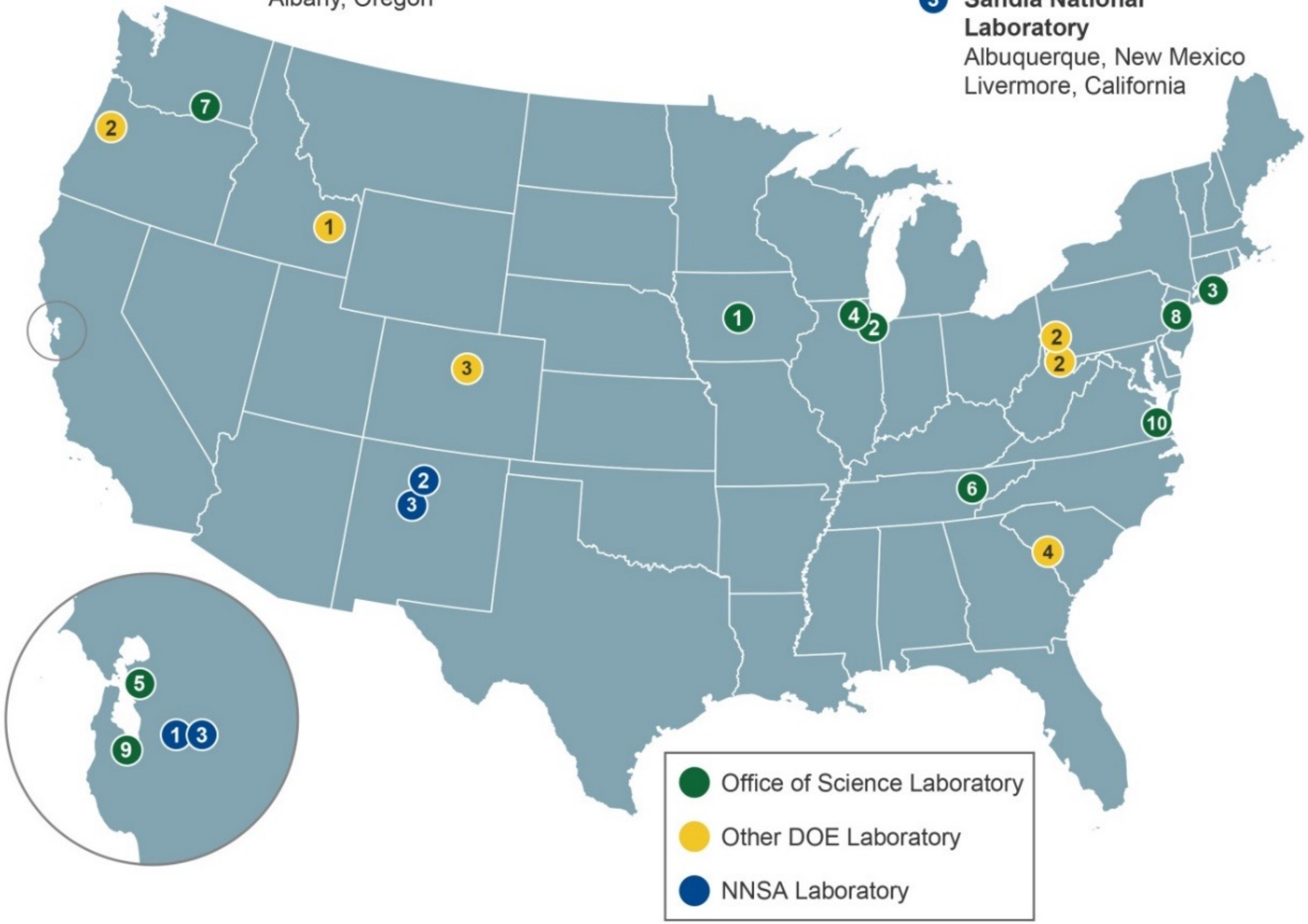
1. **Ames Laboratory**
   Ames, Iowa
2. **Argonne National Laboratory**
   Argonne, Illinois
3. **Brookhaven National Laboratory**
   Upton, New York
4. **Fermi National Accelerator Laboratory**
   Batavia, Illinois
5. **Lawrence Berkeley National Laboratory**
   Berkeley, California
6. **Oak Ridge National Laboratory**
   Oak Ridge, Tennessee
7. **Pacific Northwest National Laboratory**
   Richland, Washington
8. **Princeton Plasma Physics Laboratory**
   Princeton, New Jersey
9. **SLAC National Accelerator Laboratory**
   Menlo Park, California
10. **Thomas Jefferson National Accelerator Facility**
    Newport News, Virginia

Other DOE Laboratories

1. **Idaho National Laboratory**
   Idaho Falls, Idaho
2. **National Energy Technology Laboratory**
   Morgantown, West Virginia
   Pittsburgh, Pennsylvania
   Albany, Oregon
3. **National Renewable Energy Laboratory**
   Golden, Colorado
4. **Savannah River National Laboratory**
   Aiken, South Carolina

NNSA Laboratories

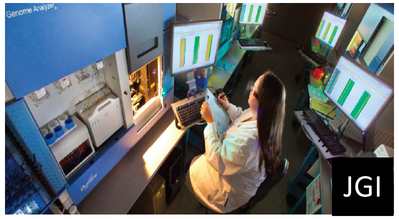
1. **Lawrence Livermore National Laboratory**
   Livermore, California
2. **Los Alamos National Laboratory**
   Los Alamos, New Mexico
3. **Sandia National Laboratory**
   Albuquerque, New Mexico
   Livermore, California

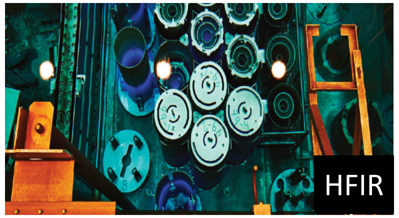


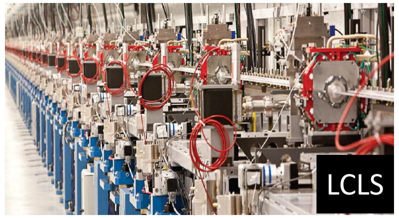
U.S. DEPARTMENT OF
ENERGY

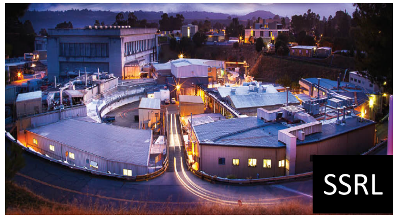
The Office of Science User Facilities

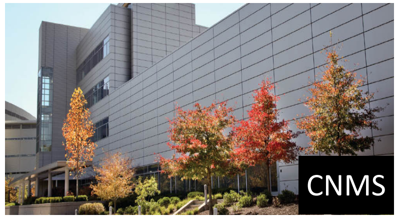
FY 2024
28 scientific
user facilities
>39,500 users

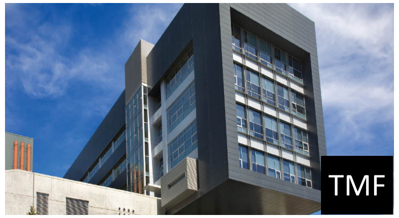


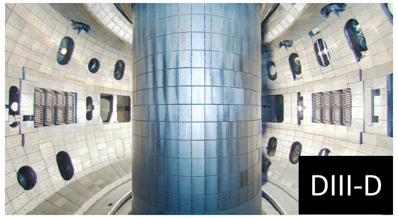


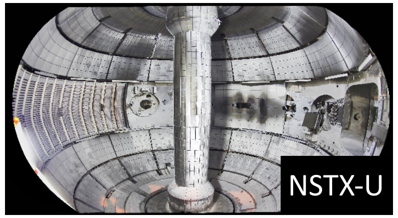


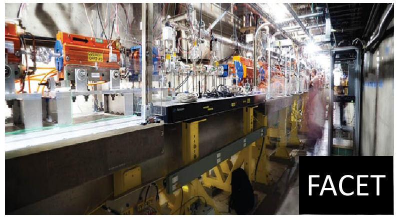


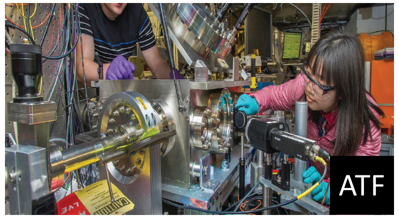


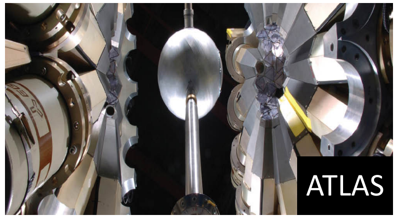






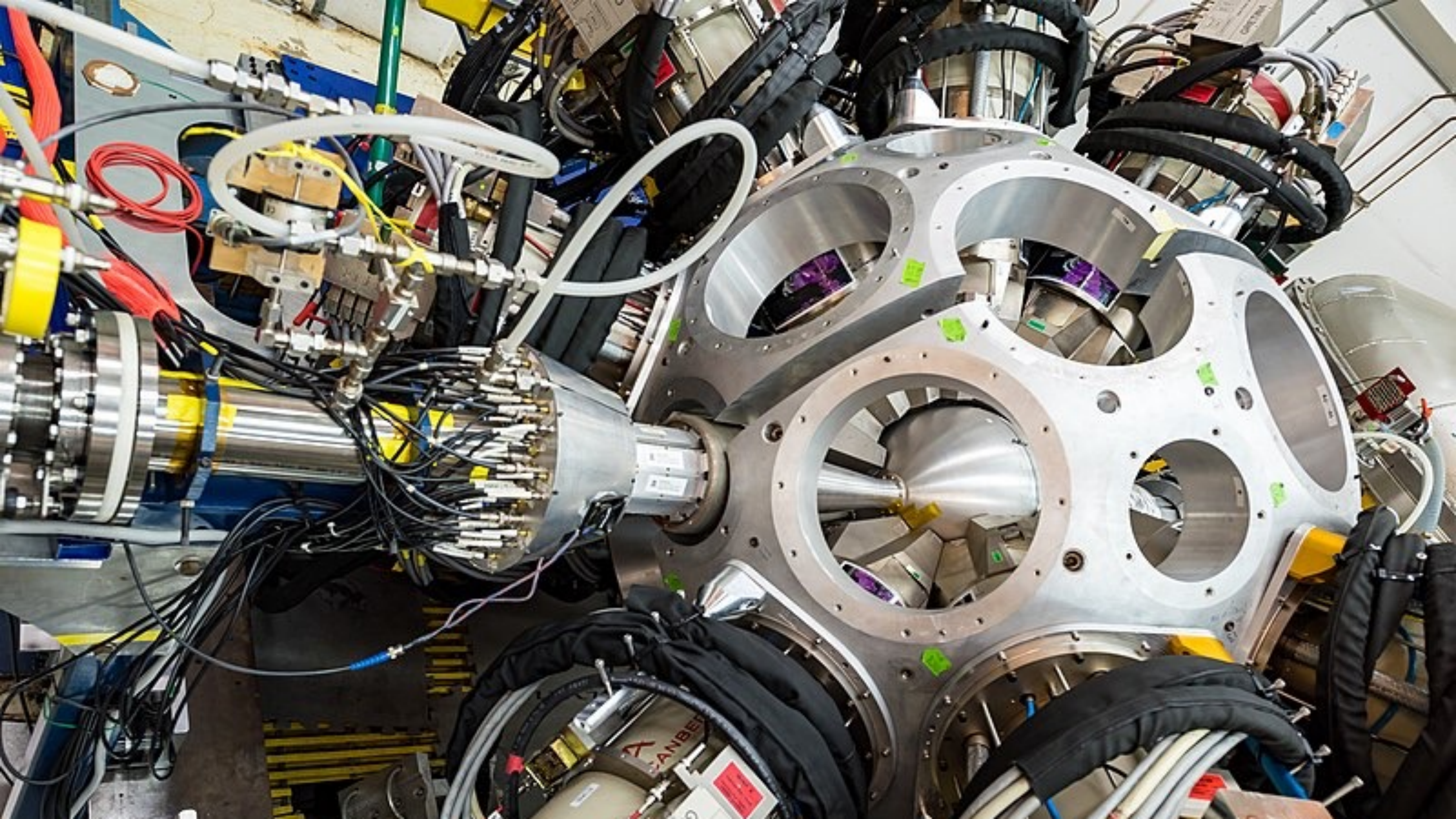




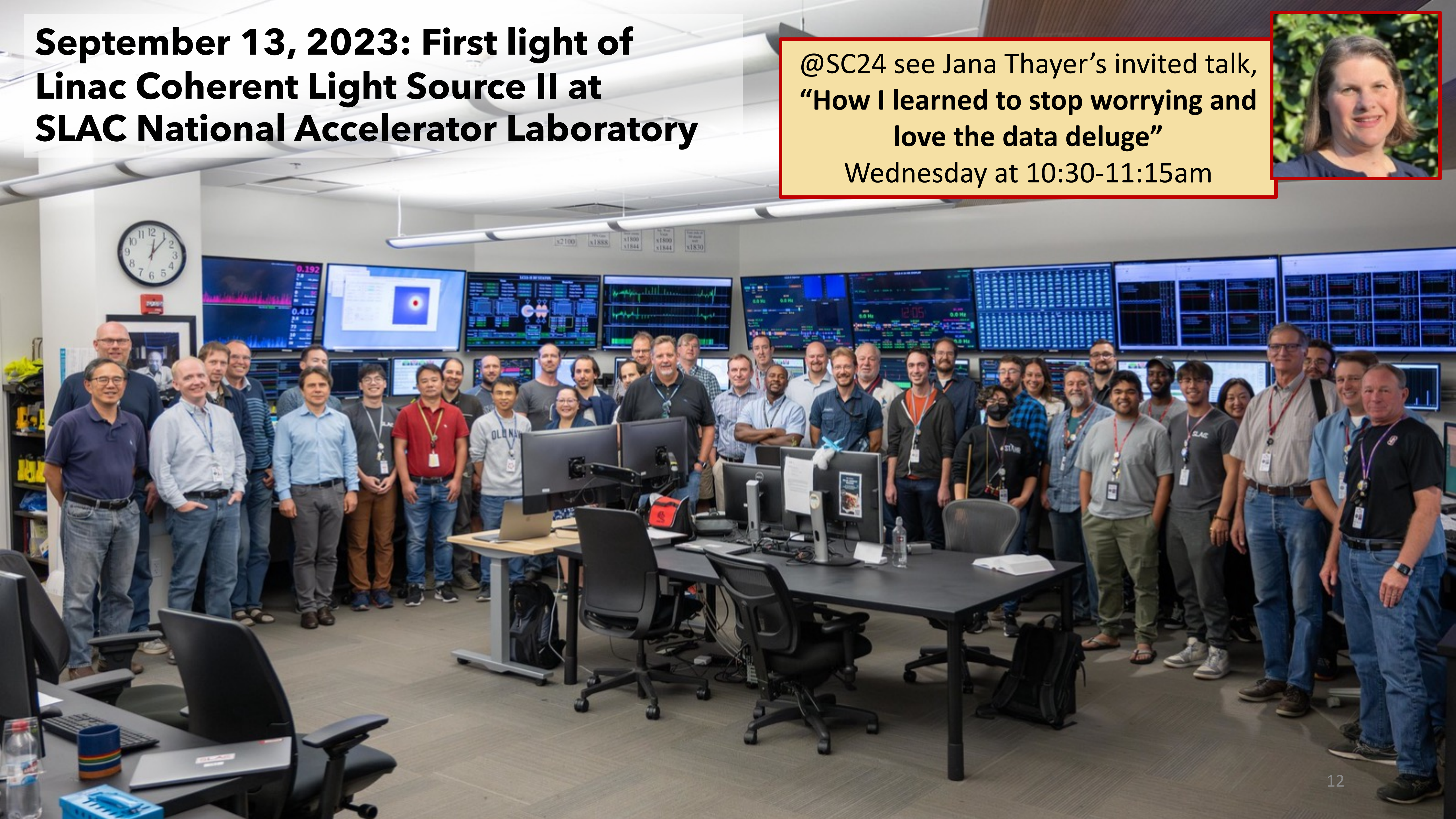


U.S. DEPARTMENT OF
ENERGY
Office of Science


T2a
PS6
T1b
EWS
PS5



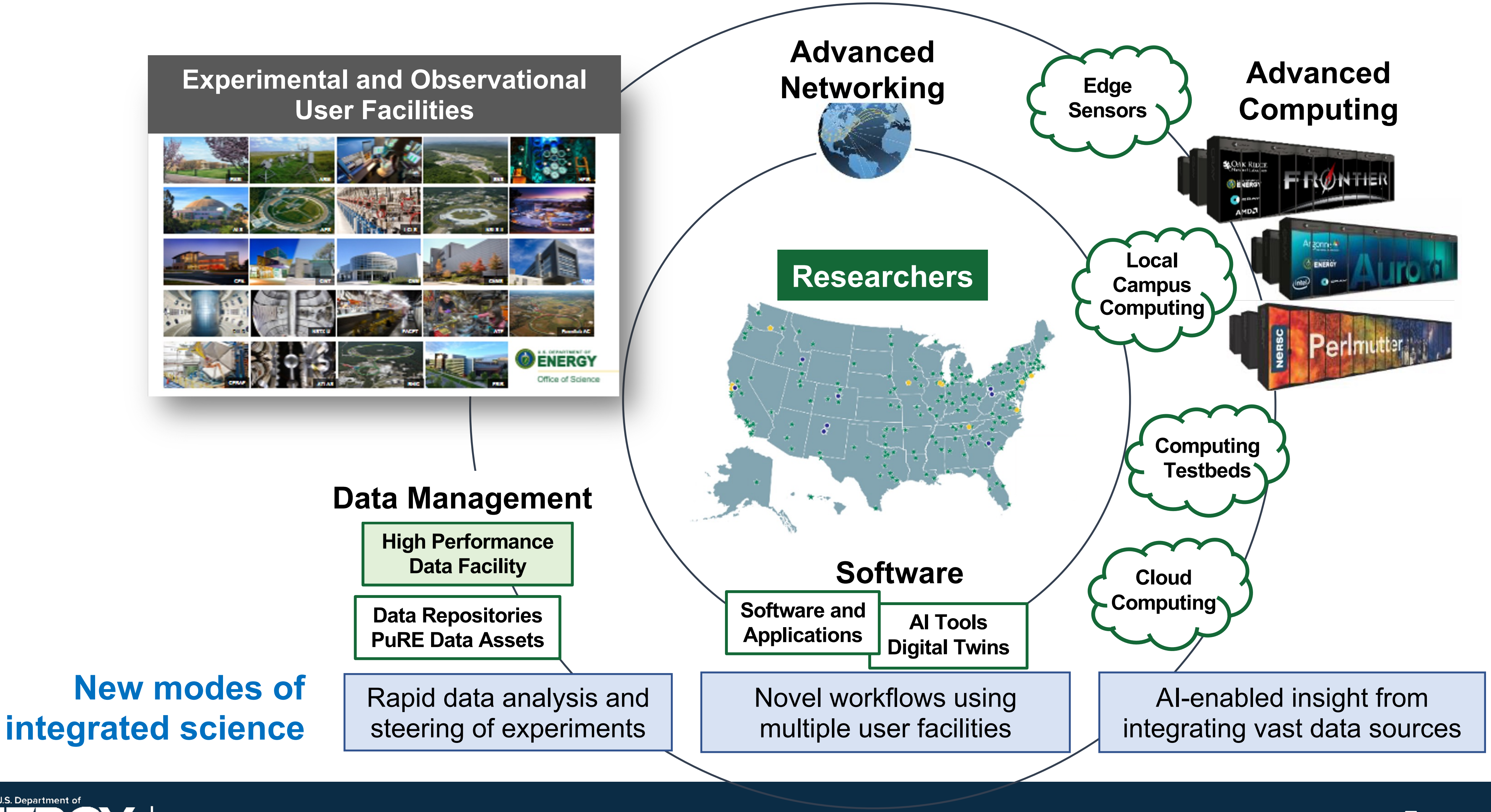
September 13, 2023: First light of Linac Coherent Light Source II at SLAC National Accelerator Laboratory



@SC24 see Jana Thayer's invited talk,
"How I learned to stop worrying and love the data deluge"
Wednesday at 10:30-11:15am

DOE's Integrated Research Infrastructure (IRI) Vision:

To empower researchers to meld DOE's world-class research tools, infrastructure, and user facilities seamlessly and securely in novel ways to radically accelerate discovery and innovation



DOE's Integrated Research Infrastructure (IRI) Vision:

To empower researchers to meld DOE's world-class research tools, infrastructure, and user facilities seamlessly and securely in novel ways to radically accelerate discovery and innovation

The IRI Vision:

It's about empowering <u>people</u>.

It's about <u>data</u>.

Saswata Hier-Majumder, ALCF PM

Argonne Leadership Computing Facility

Jordan Thomas, NERSC PM

NERSC National Energy Research Scientific Computing Center

Christine Chalk, OLCF PM

Oak Ridge Leadership Computing Facility

Carol Hawk, ESnet PM

ESnet Energy Sciences Network

Standup of the IRI Program is a DOE FY24-25 Agency Priority Goal

FY 2021 President's Budget Request includes **Integrated** Computation and Data Infrastructure Initiative

ASCR IRI Task Force launch

ASCR IRI Task Force report

SC IRI Blueprint Activity launch

IRI Blueprint Activity results

FY 2024 PBR advances IRI and the High Performance Data Facility

HPDF Selection

GO

IRI Program Development

Jan **2020** | Jan **2021** | Jan **2022** | Jan **2023** | Jan **2024**

Vision → **Strategy** → **Implement**

The IRI Architecture Blueprint Activity established a framework for serious planning

Download link

The IRI Blueprint Activity created a framework for IRI implementation

IRI Science Patterns (3)

Time-sensitive pattern has *urgency*, requiring real-time or end-to-end performance with high reliability, e.g., for timely decision-making, experiment steering, and virtual proximity.

Data integration-intensive pattern requires combining and analyzing data from multiple sources, e.g., sites, experiments, and/or computational runs.

Long-term campaign pattern requires sustained access to resources over a long period to accomplish a well-defined objective.

Convened over **150 DOE national laboratory experts** from **all 28 SC user facilities** across **13 national laboratories** to consider the **technological, policy, and sociological challenges** to implementing IRI.

IRI Practice Areas (6)

User experience practice will ensure relentless attention to user perspectives and needs through requirements gathering, user-centric (co)-design, continuous feedback, and other means.

Resource co-operations practice is focused on creating new modes of cooperation, collaboration, co-scheduling, and joint planning across facilities and DOE programs.

Cybersecurity and federated access practice is focused on creating novel solutions that enable seamless scientific collaboration within a secure and trusted IRI ecosystem.

Workflows, interfaces, and automation practice is focused on creating novel solutions that facilitate the dynamic assembly of components across facilities into end-to-end IRI pipelines.

Scientific data life cycle practice is focused on ensuring that users can manage their data and metadata across facilities from inception to curation, archiving, dissemination, and publication.

Portable/scalable solutions practice is focused on ensuring that transitions can be made across heterogeneous facilities (portability) and from smaller to larger resources (scalability).

IRI Program value propositions *(authored by the SC IRI Coordination Group)*

For the taxpayer, for all of us:

Achieve greater productivity and avoid duplication of effort.

For the researcher:

Achieve transformational reduction in *time to insight* and *complexity*.

For program/RI/institutional leaders:

- **Achieve greater effectiveness and efficiency in coordinating efforts;**
- **Achieve more nimble solutions than would be possible alone;**
- **Gain leverage with partners who possess like requirements;**
- **Avoid single points of failure; and**
- **Gain access to expertise and shared experience.**

IRI Program launch is a DOE FY24-25 Agency Priority Goal. ASCR is implementing IRI through these four major elements.

1. **Invest in IRI foundational infrastructure**
2. **Stand up the IRI Program governance and FY24 workstreams**
3. **Bring IRI projects into formal coordination**
4. **Deploy an IRI Science Testbed across the ASCR Facilities**

These are all connected.
These are each essential.

This year DOE conducted a decadal assessment of major facilities. These are overarching recommendations (direct quotations):

Recommendation 1: Ensure the continued support and development of all five ASCR computational facilities reviewed—ALCF, OLCF, NERSC, HPDF, and ESnet—as they are central and essential to all SC science programs and broader national science and engineering research programs.

Recommendation 2: Science demands integration. We advocate viewing ASCR facilities not as isolated entities, but as integral components of a single, larger integrated computational *ecosystem* (henceforth referred to as *Ecosystem*), with a single governance model. … *Further, this integrated ecosystem is required for programs of other agencies, and industry. Its critical role in bolstering national scientific and technological capabilities, as well as its status as a model internationally, cannot be overstated*.

Recommendation 3: A comprehensive, coordinated R&D program delivering multiple prototype computing systems over a five-year timescale must be mounted to inform pathways for this integrated ecosystem, operational by 2034, due to (a) rapidly evolving economic and technical landscapes of the semiconductor and computing industries and (b) changing research practices.

"Recommendation 1 is necessary but not sufficient for success."

This year DOE conducted a decadal assessment of major facilities. These are overarching recommendations (direct quotations):

Recommendation 1: **Ensure the continued support and development of all five ASCR computational facilities reviewed—ALCF, OLCF, NERSC, HPDF, and ESnet—as they are central and essential to all SC science programs and broader national science and engineering research programs.**

Recommendation 2: **Science demands integration. We advocate viewing ASCR facilities not as isolated entities but as integral components of a single, larger integrated computational *ecosystem* (henceforth referred to as *Ecosystem*) with a single governance model.** … *Further, this integrated ecosystem is required for programs of other agencies, and industry. Its critical role in bolstering national scientific and technological capabilities, as well as its status as a model internationally, cannot be overstated.*

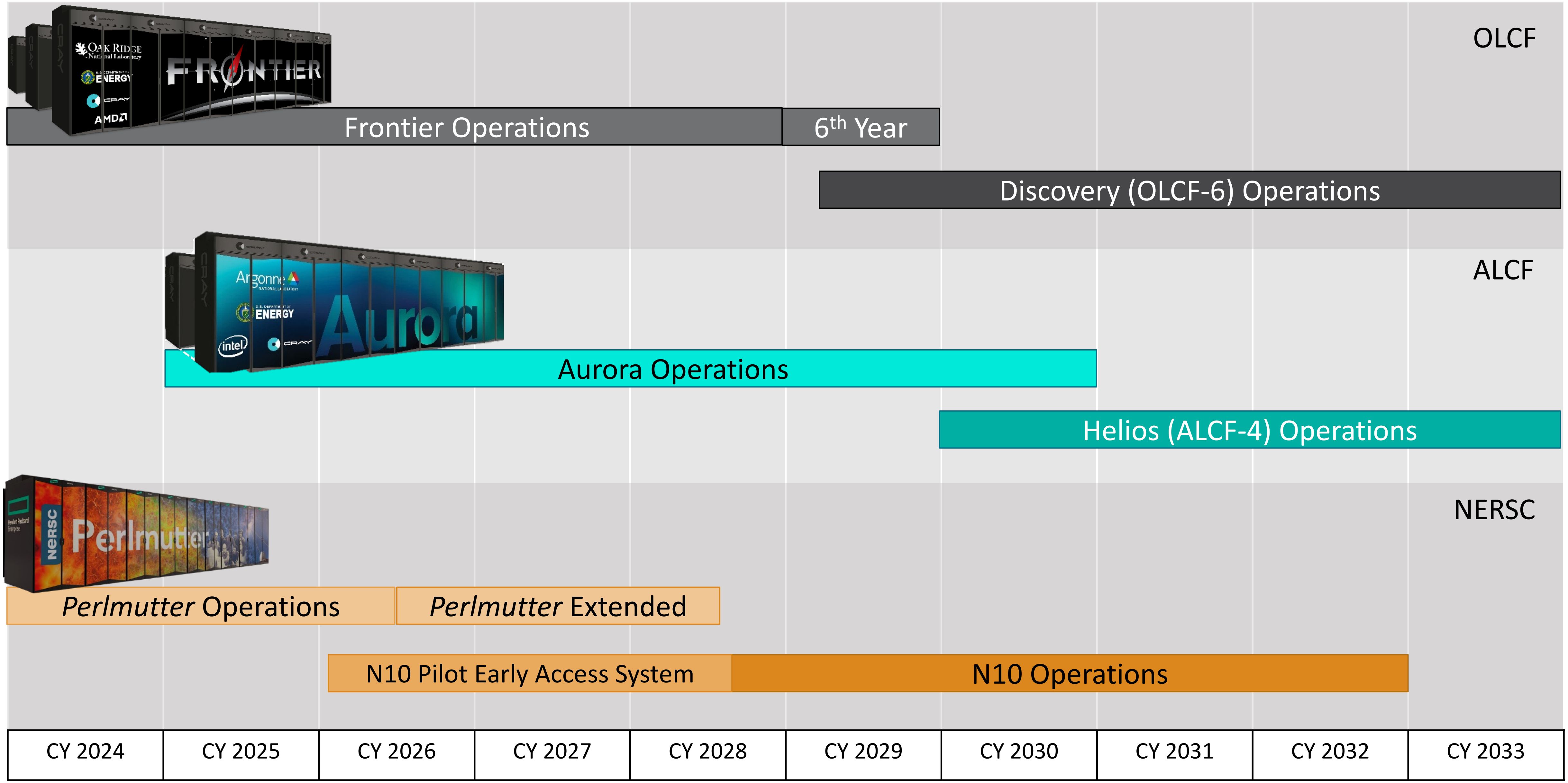
Recommendation 3: **A comprehensive, coordinated R&D program delivering multiple prototype computing systems over a five-year timescale must be mounted to inform pathways for this integrated ecosystem, operational by 2034, due to (a) rapidly evolving economic and technical landscapes of the semiconductor and computing industries and (b) changing research practices.**

Individual Facility Assessment

"Without ESnet, the entire vision collapses; none of the facilities, nor the integrated Ecosystem, nor IRI could function."

"Recommendation 1 is necessary but not sufficient for success."

ASCR High Performance Computing Upgrade Projects Timelines




Jefferson Lab
U.S. DEPARTMENT OF ENERGY
BERKELEY LAB
HPDF
High Performance Data Facility

High Performance Data Facility Project start-up

Oct. 15, 2023: Selection announcement

Nov. 13, 2023: Breaking ground meeting at SC23

Feb. 13, 2024: Project kick-off meeting

Mar. 5-6, 2024: Benchmarking meeting at LBNL

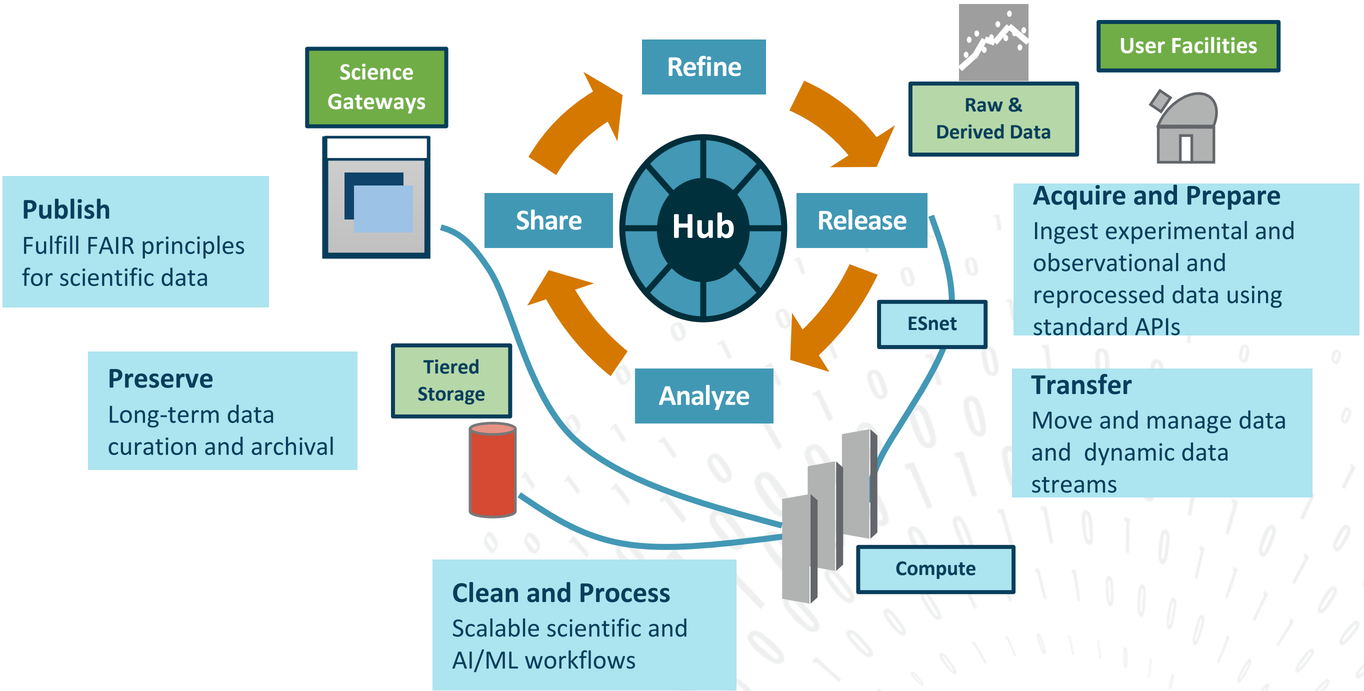
April, 2024: Integrated Project Team meetings start

May, 2024: FY 2024 appropriations: $8.0M

July, 2024: IRI/HPDF Coordination Kick-off Meeting (100+ attendees)

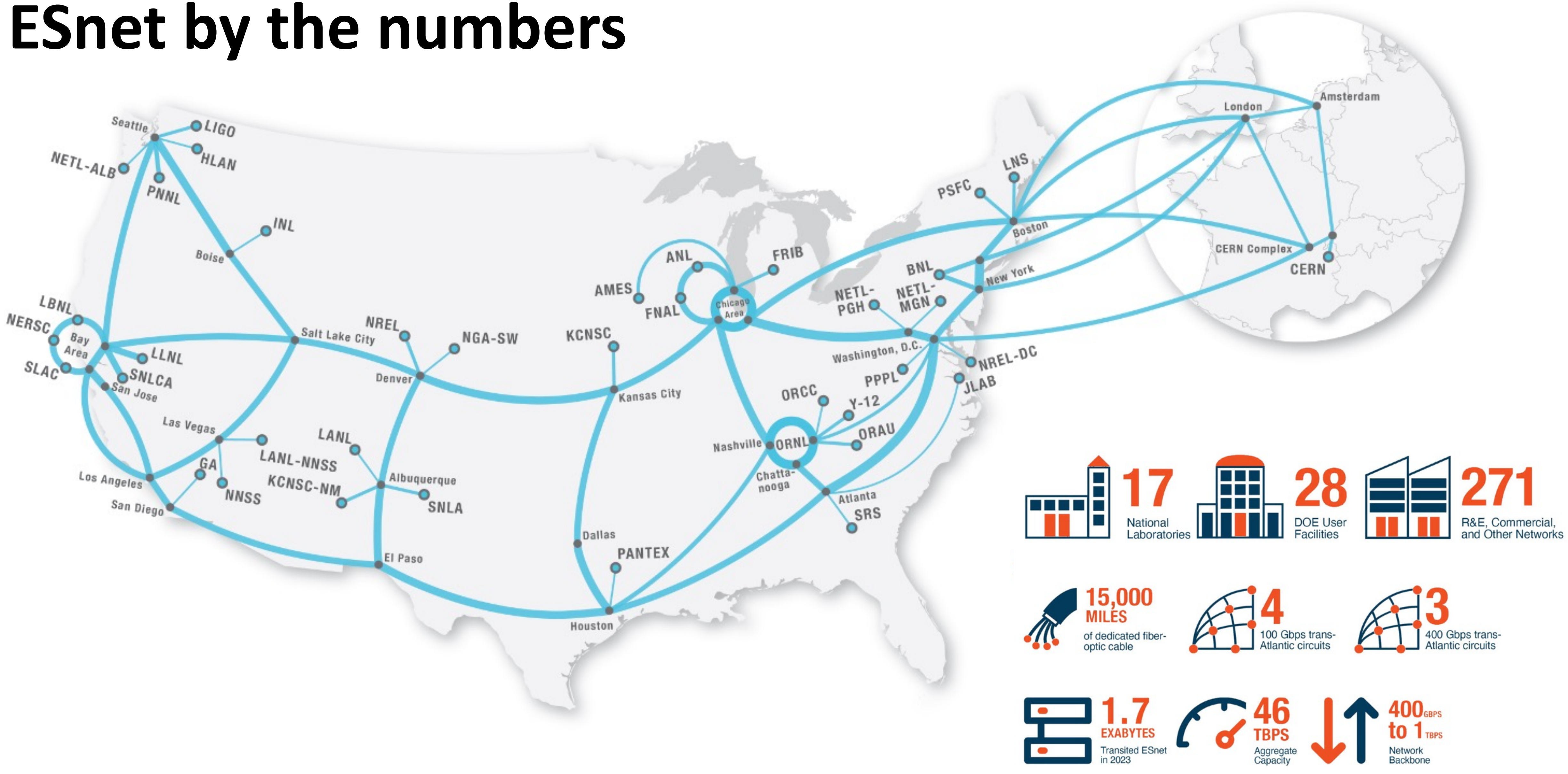
Flexible & Full Life Cycle Coverage

- **Management** – A dynamic and scalable data management infrastructure integrated with the DOE computing ecosystem
- **Capture** – Dynamically allocatable data storage and edge computing at the point of generation
- **Staging** – Dynamic placement of data in proximity to appropriate computing for reduction, analysis, and processing
- **Archiving** – Extreme-scale distributed archiving and cataloging of data with FAIR principles – findability, accessibility, interoperability, and reusability
- **Processing** – Resources for workflow and automation for processing and analyses of data at scale



Data science requires curated and annotated data that adheres to FAIR principles, and data reuse will be an HPDF metric. Office of Scientific and Technical Information services will complement HPDF to provide full life cycle coverage.

ESnet by the numbers



Evolution of the ESnet over the past 30+ years

Magnetic Fusion Energy Network (MFENET)
[1976 - 1986]

ESnet1
[1986 - 1994]
Building an open standards network
- *IPv4*
- *BGP*

ESnet2
[1994 - 2000]
Adopting emerging technologies
- *ATM*
- *IPv6*
- *Video Conf Services*

ESnet3
[2000 - 2006]
Building predictable network services
- *QoS and TE (MPLS)*
- *OSCARS*

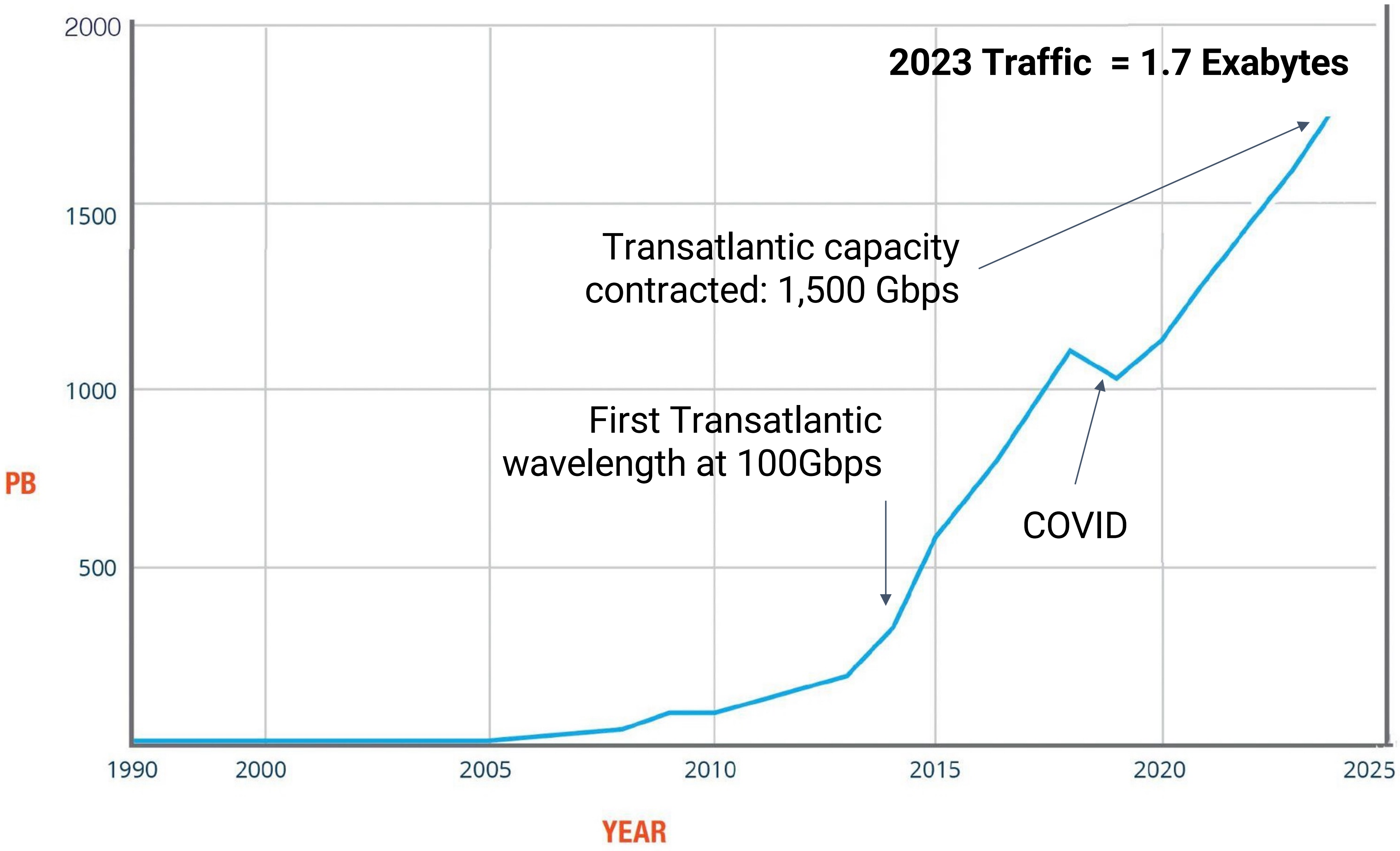
ESnet4
[2006 - 2011]
Deploying purpose built architectures
- *Science Data Network (SDN)*
- *ScienceDMZ*
- *perfSONAR*

ESnet5
[2011 - 2022]
Expanding the network breadth and depth
- *OLS operations*
- *Transatlantic connectivity*

ESnet6
[2022 - TBD]
Developing comprehensive network automation and visibility capabilities
- *Greenfield Build*
- *Orchestration/Automation*
- *High-Touch*

1970s → **1980s** → **1990s** → **2000s** → **2010s** → **Present**

Exponential traffic growth across 3 decades



ESnet traffic analysis: significance of major facilitie

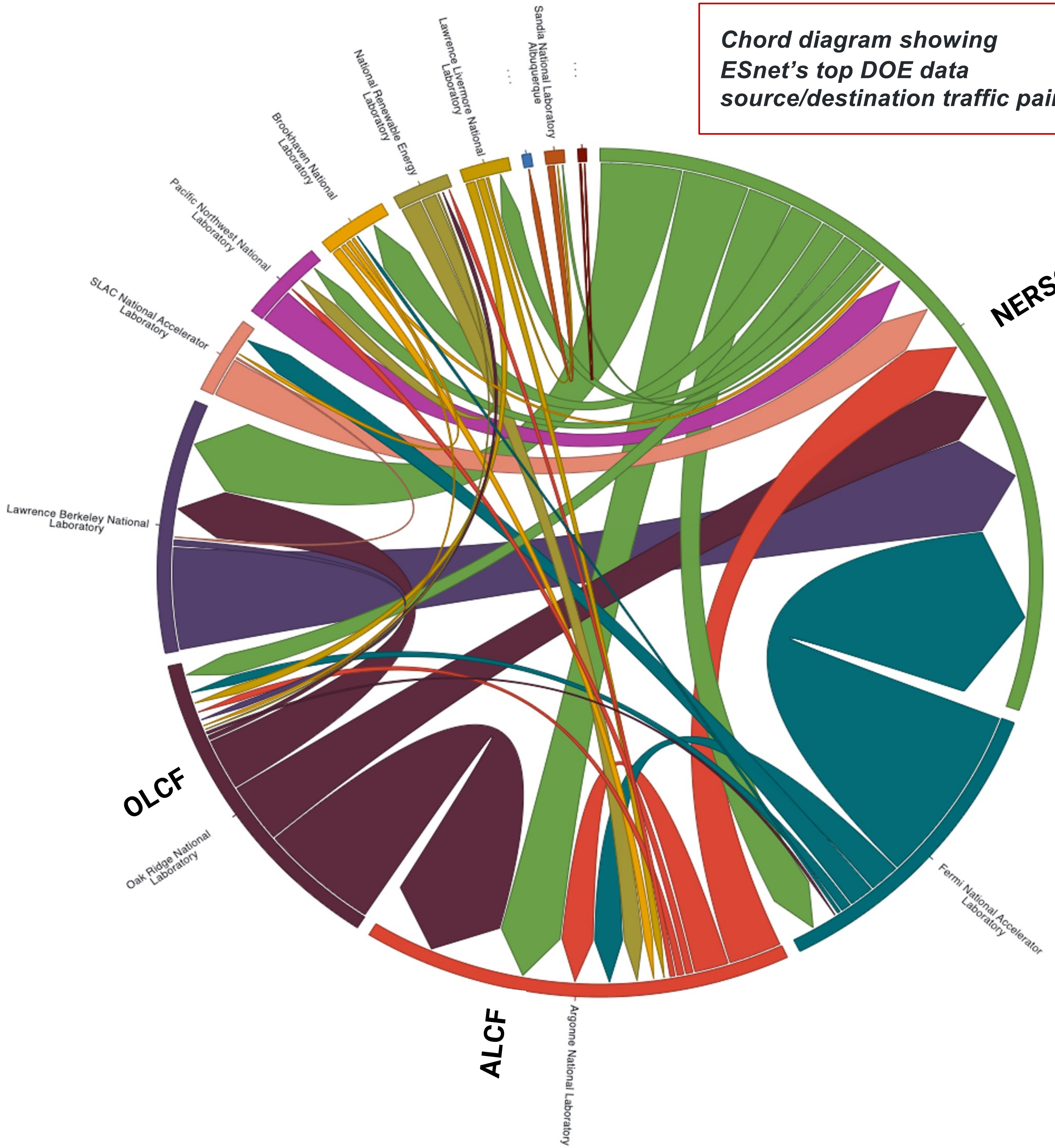
- User facilities, especially the ASCR HPC centers, are significant data providers and consumers

- Combined ASCR facilities (ESnet + Computing) are central to SC's data-intensive science

- DOE Labs and Facilities already use ESnet for connectivity and data exchange

IRI plans to make these interactions consistent, intuitive, and broad.

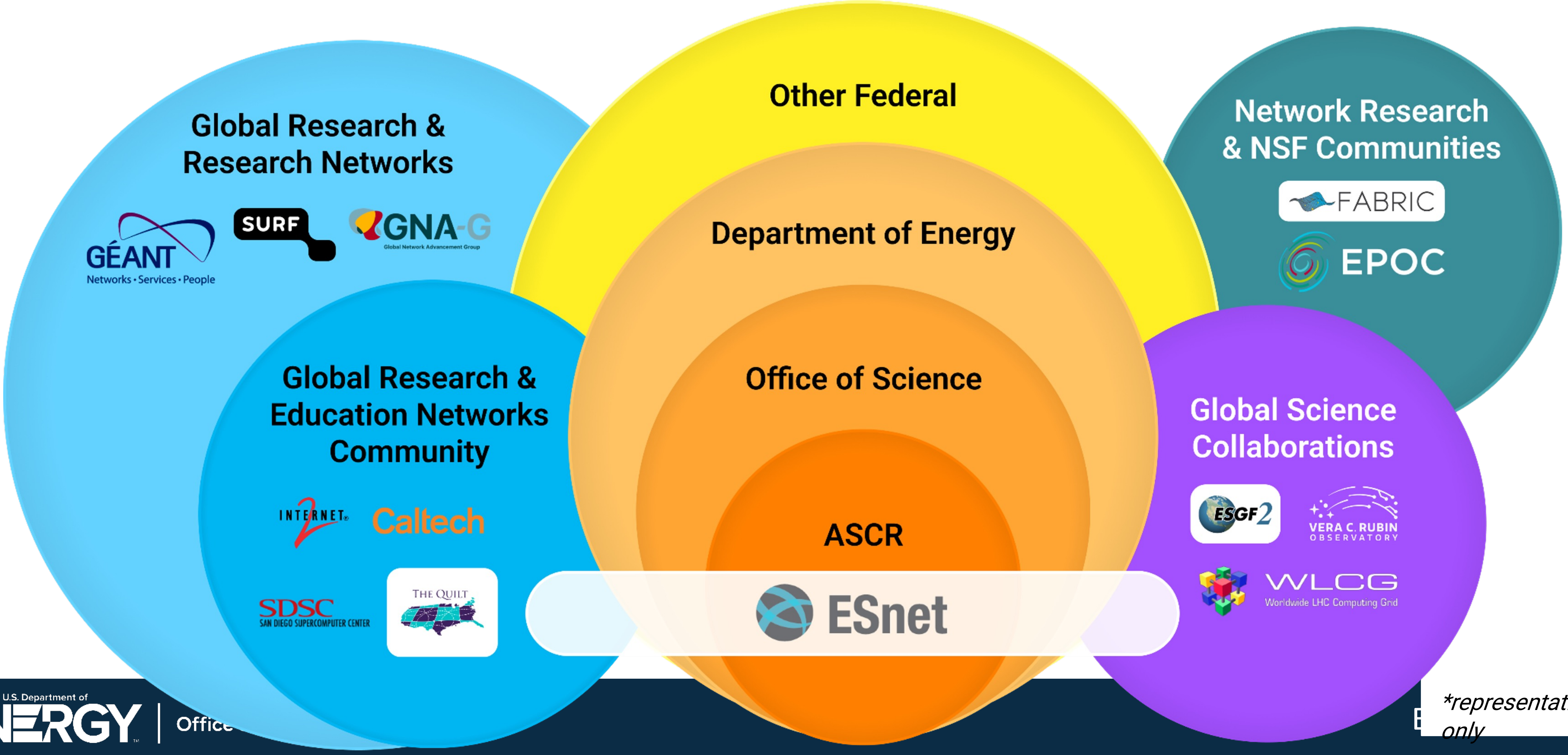
Chord diagram showing ESnet's top DOE data source/destination traffic pairs



Slide credit: Inder Monga and Chin Guok, ESnet

Networks need to be global team player

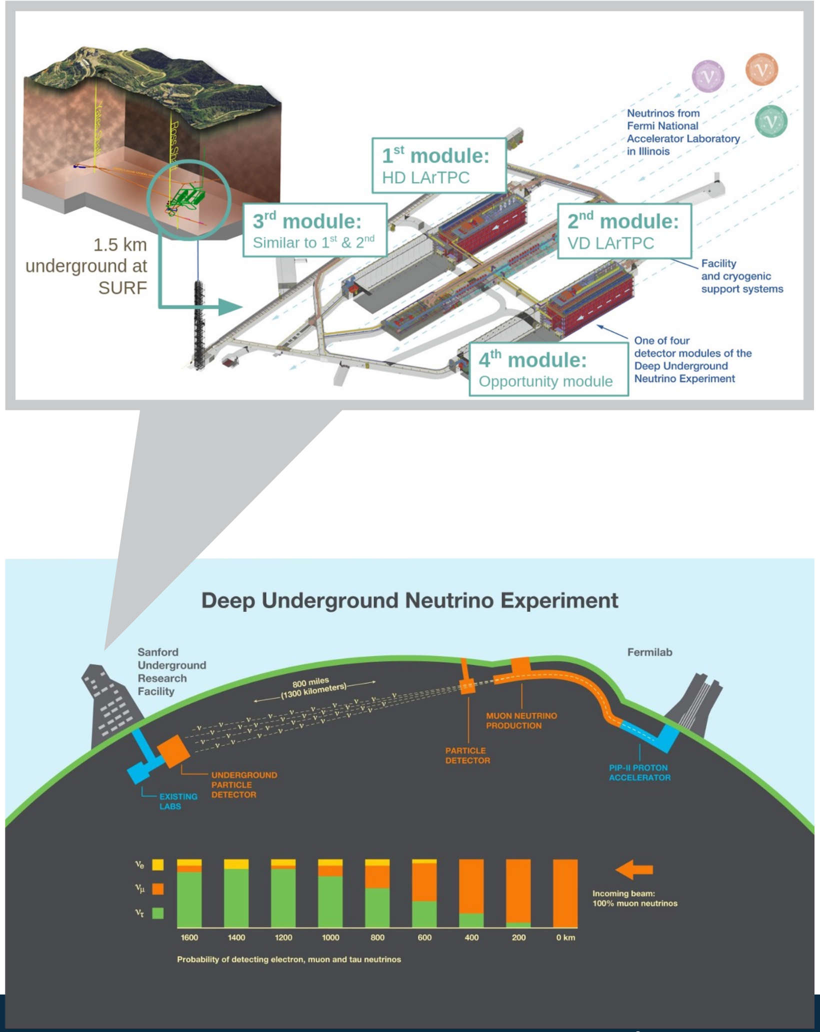
ESnet is an integral part of many ecosystems, not just DOE/federal ones.
ESnet ***contributes to, bridges, and leads*** in multiple research communities.*
Other Research and Education networks play similar roles in their communities



*representative logos only

Changing the current balance of the user facilities - Increasing need for specialized computing

Different computing resources (e.g., HPCs, Cloud, local compute, etc) can offer unique capabilities to help the scientists achieve results more efficiently.



Deep Underground Neutrino Experiment (DUNE)

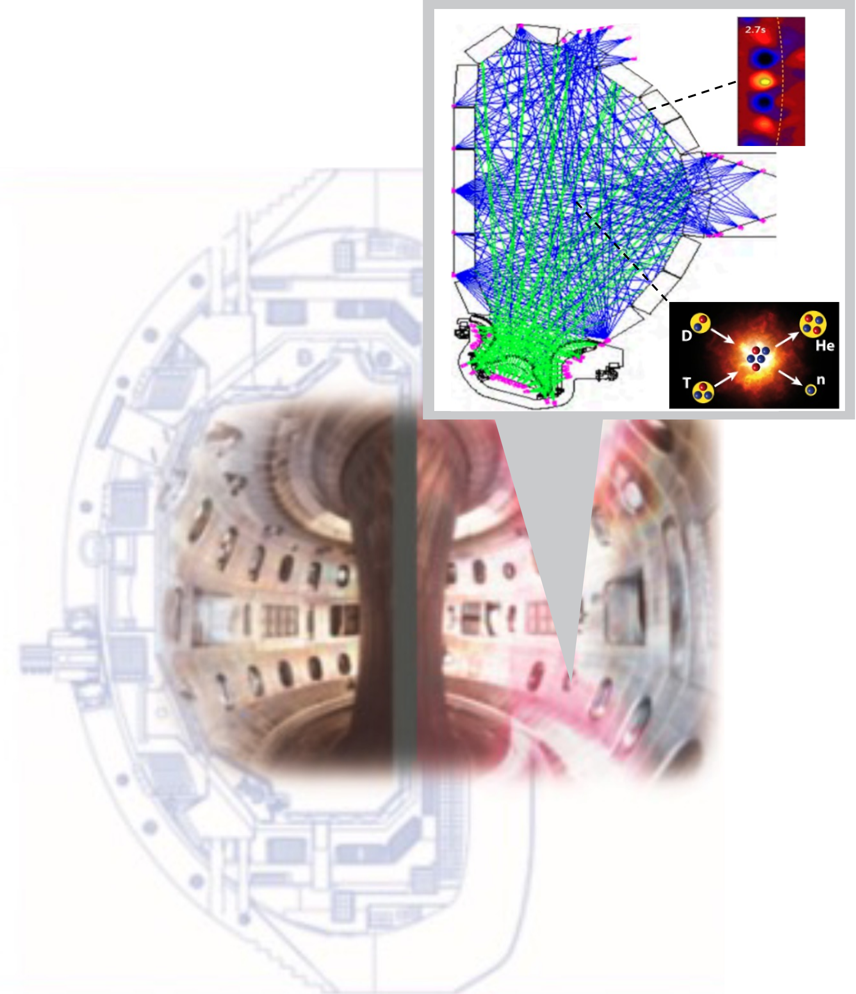
- Measure neutrino oscillations by studying neutrinos that will be sent from Fermilab to the DUNE detectors at the Sanford Underground Neutrino Facility.
- In the HL-LHC Data Challenge 2021, compute at FNAL was used for normal processing, but inference model training was done using GPUs in Google Cloud.
- ***Flexibility to select the right (combination of) compute will be essential in optimizing the efficiency and effectiveness of the workflow.***

Changing the current balance of the user facilities - Supporting complex multi-modal science workflows

The need for observational, simulation, and experimental data to be collected, transformed, and combined for scientific analysis.

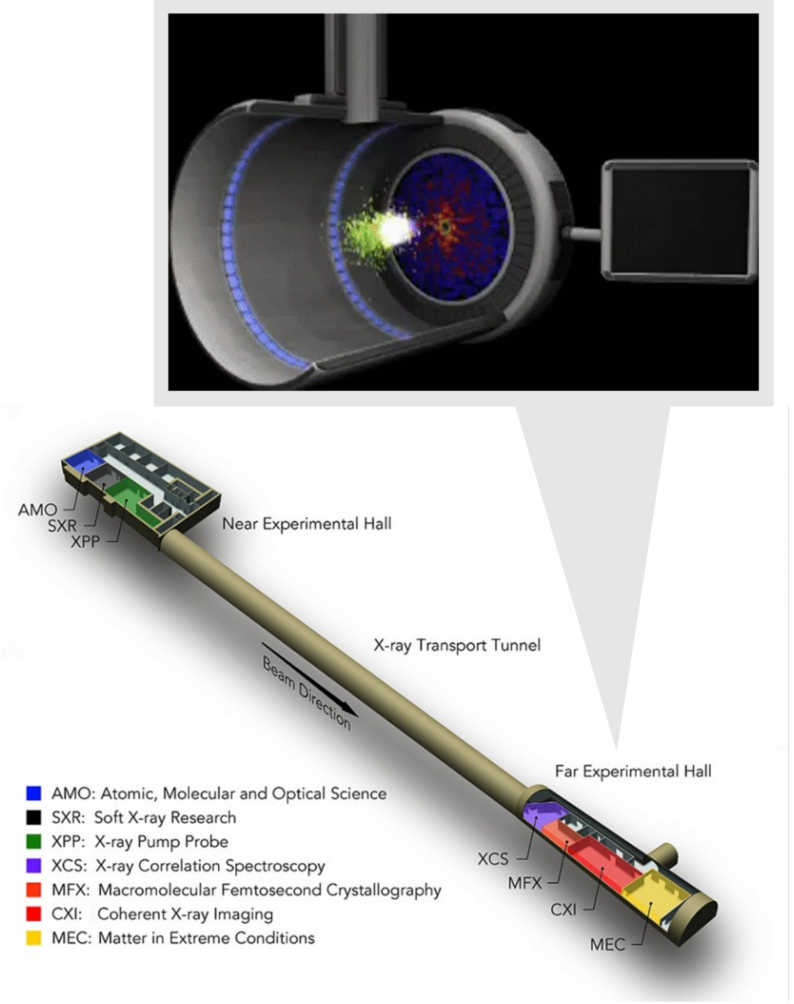
DIII-D Magnetic Fusion Tokamak

- Magnetic confinement fusion using a toroidal solenoid to confine high-temperature plasmas.
- Integrated Data Analysis (IDA) combines all diagnostics from the tokamak, yielding smaller uncertainties in the density function, resulting in better control of the magnetic confinement.
- ***IDA output can be used to build both experiment analysis and simulation databases, coupled with ML to develop models for magnetic confinement for future tokamaks (e.g., ITER).***



Changing the current balance of the user facilities - Optimizing the user facilities

Supporting real-time fast feedback while the experiment is running increases the probability of successful research and shortens the time to results.



Linac Coherent Light Source (LCLS)

- Ultrafast X-ray pulses from LCLS are used like flashes from a high-speed strobe light, producing stop-action movies of atoms and molecules.
- Both data processing and scientific interpretation demand intensive computational analysis.
- ***Leveraging HPC resources for real-time calibration to verify proper alignment is critical, misalignments would result in a wasted experiment.***

Slide credit: Inder Monga and Chin Guok, ESnet

Building capabilities to support IRI touches many ESnet areas

Predictable (end-to-end) network services

- OSCARS guaranteed b/w dynamic provisioning
- Operational measurement & performance monitoring

High bandwidth and rich connectivity

- ESnet6 capacity deployment
- Transatlantic spectrum
- ESnet Cloud Connect for Virtual Private Clouds (VPC)

Application/network interaction

- SENSE multi-domain resource orchestration
- OpenAPI & OpenTelemetry

"Friction-free" data movement

- ScienceDMZ data transfer optimization architecture
- Petascale DTN for HPC oriented data transfers

Network computational storage

- DTNaaS in-network caching
- EJFAT FPGA based real-time DSP processing for edge compute

Multi-modal network connectivity

- CBRS pilot field deployment (EESA) Mt. Crested Butte, Co

Programming constructs

- SURF's Workflow Orchestrator
- SENSE and NSI network interdomain APIs
- JANUS container management

Common access framework

- Federated Identity and Access Control Management

Resource allocation policies

- ESnet does not have an allocation policy, but this may change to support IRI (e.g., time sensitive workflows)

Development and testing environments

- ESnet Testbed Next-Gen
- FABRIC (NSF) Testbed
- ROVER network orchestration testing environment

Co-design services

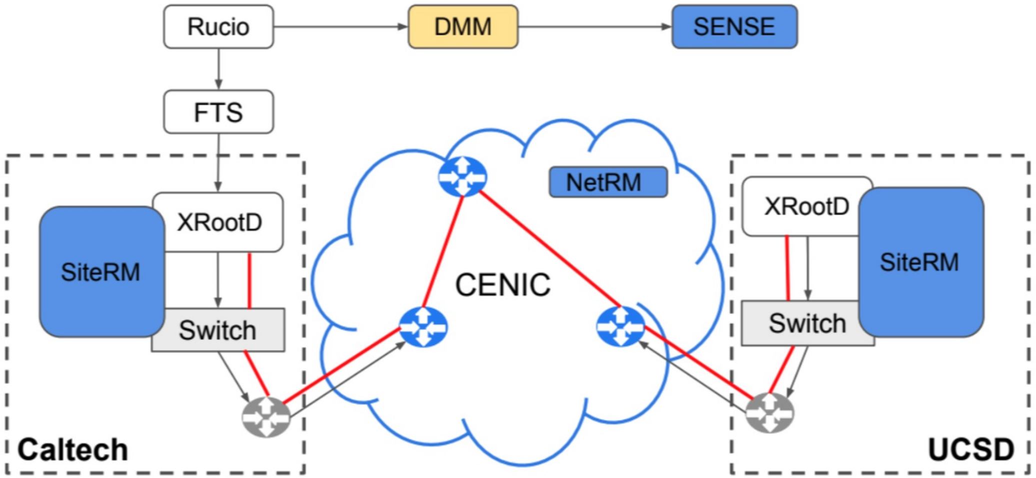
- GRETA network integration
- EJFAT ESnet JLab FPGA Accelerated Transport

Engagement and partnerships

- DOE program requirements reviews
- ESnet ConFAB

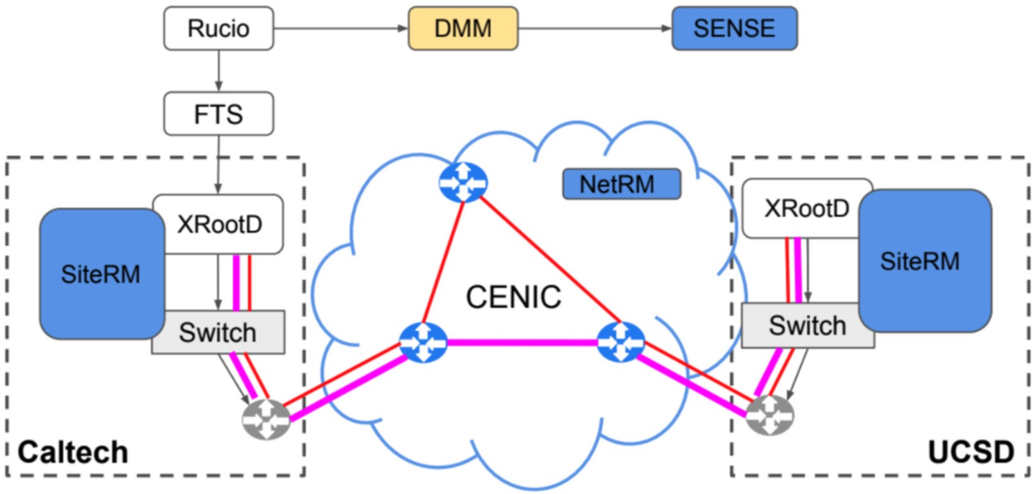
Indicators show ESnet's current experience or implementation status in the area.

Network Application Service Composability: Integration of data management software with network APIs enabling applications to request network outcomes



For **non-priority** Rucio request, Rucio will contact the Data Movement Manager (DMM) and receive endpoints that use the (red) path for **best-effort data movement**.

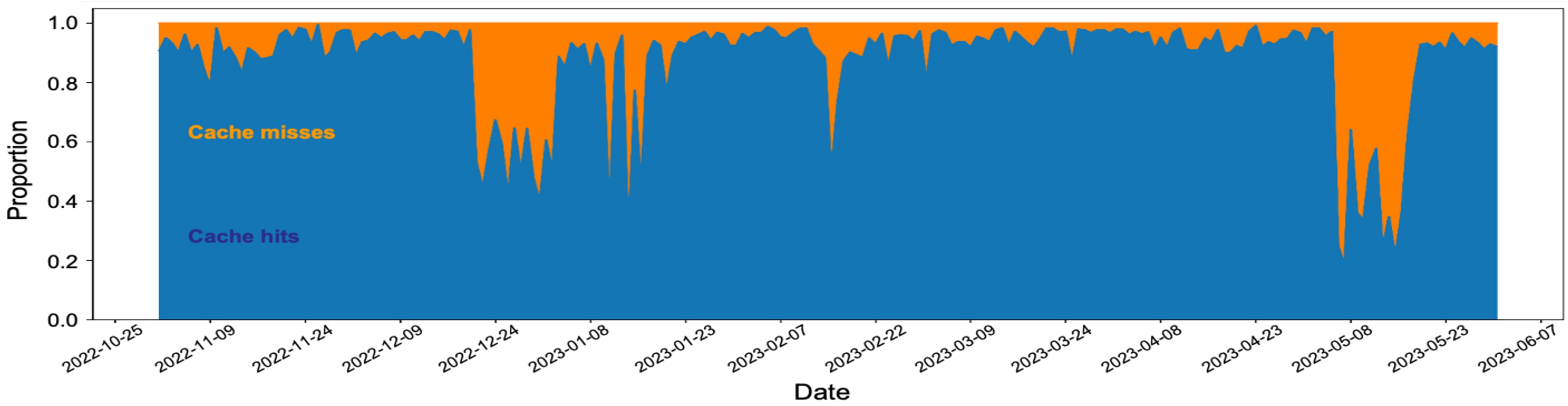
SENSE provides end-to-end (network) orchestration functions for **IRI Time-Sensitive Pattern** workflows



For **priority** Rucio request, Rucio will contact the DMM for endpoints associated with the (pink) guaranteed bandwidth path. The DMM would concurrently request a bandwidth allocation from SENSE to set up the guaranteed bandwidth path. SENSE will instruct both the SiteRm and NetRM to implement specific routing and QoS, facilitating an **end-to-end guaranteed bandwidth data movement**.

ESnet's early work with in-network storage and computing demonstrates that science data caching improves data accessibility

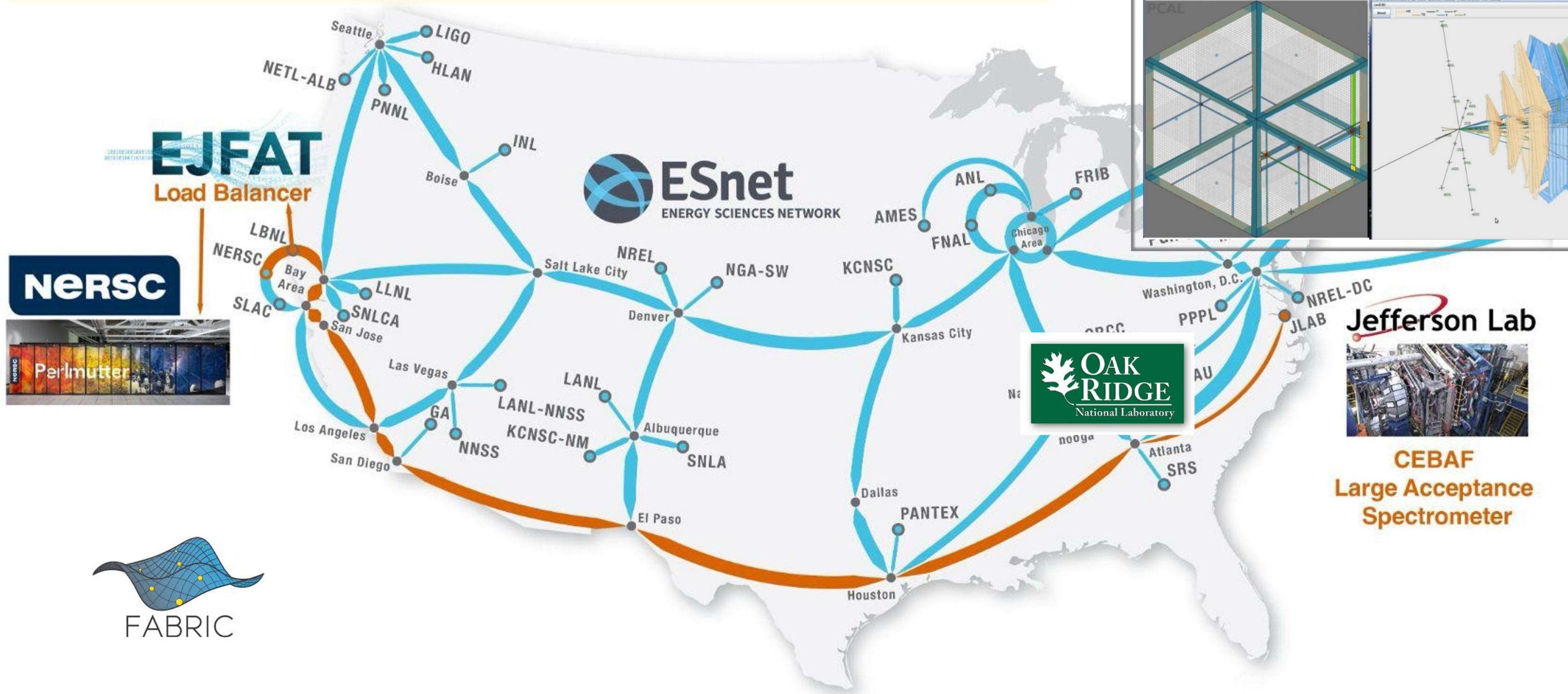
- Pilot installations in 5 locations: currently supporting LHC ATLAS/CMS, DUNE, and LIGO datasets
- Leveraging Open Science Grid caching solution with ESnet's DTN-as-a-service virtualization software stack
- Early studies show lower latency of access for scientists and the reduced traffic on network backbone — a win-win result

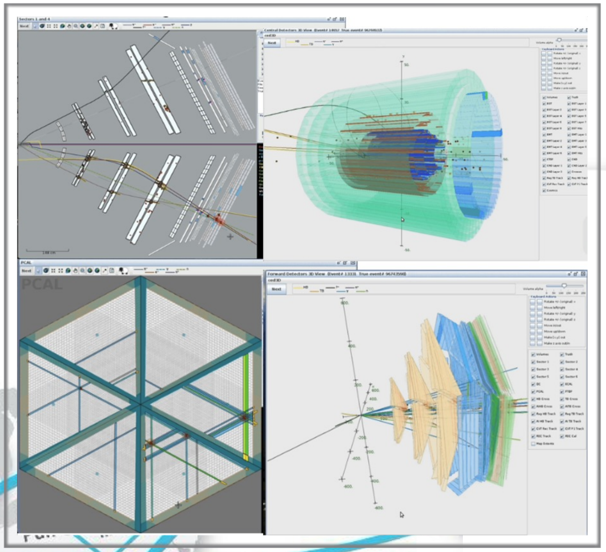


A strong spirit of co-design of the *Ecosystem* is starting to thrive under IRI

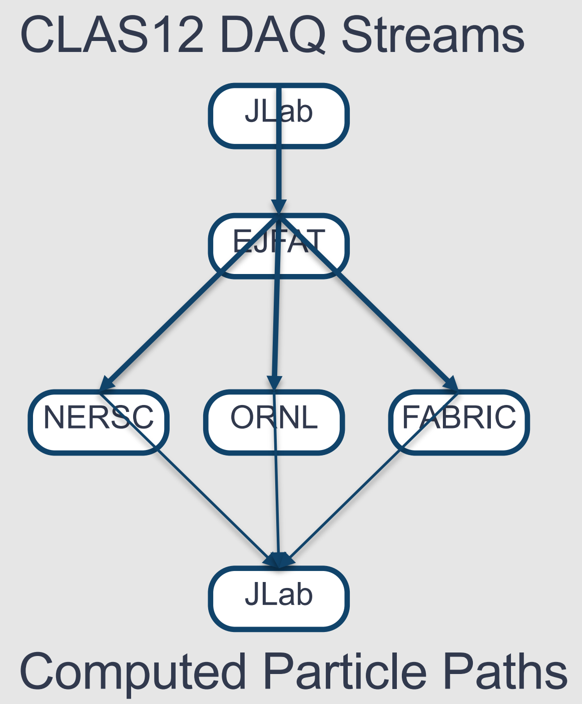
Example: Detector-to-Compute, Real-Time Streaming Prototype EJFAT

Raw physics data streamed from JLAB to NERSC at 100Gbps in real time,
with no buffering or temporal storage,
with no data loss or latency-related problems.



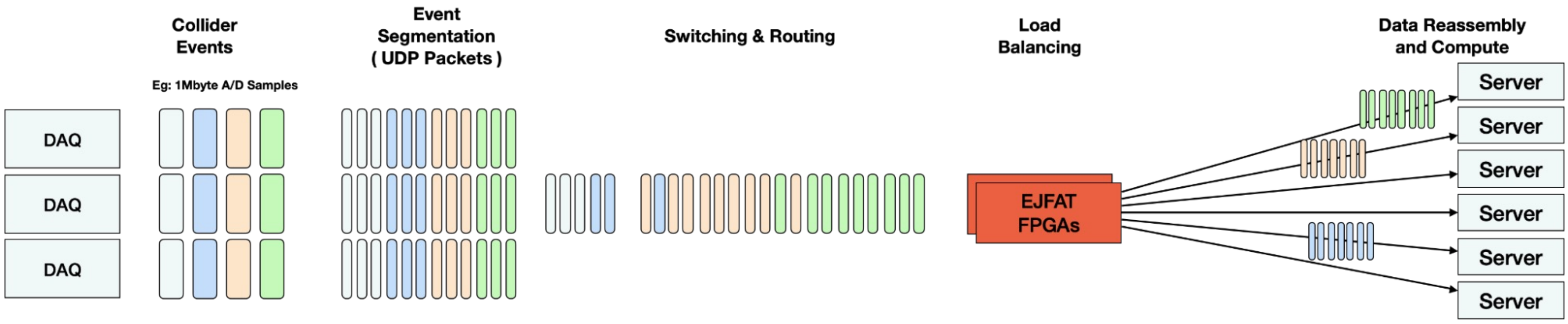


Three facilities, NERSC , ORNL and FABRIC, combine their compute resources to process a single real-time stream (IRI).



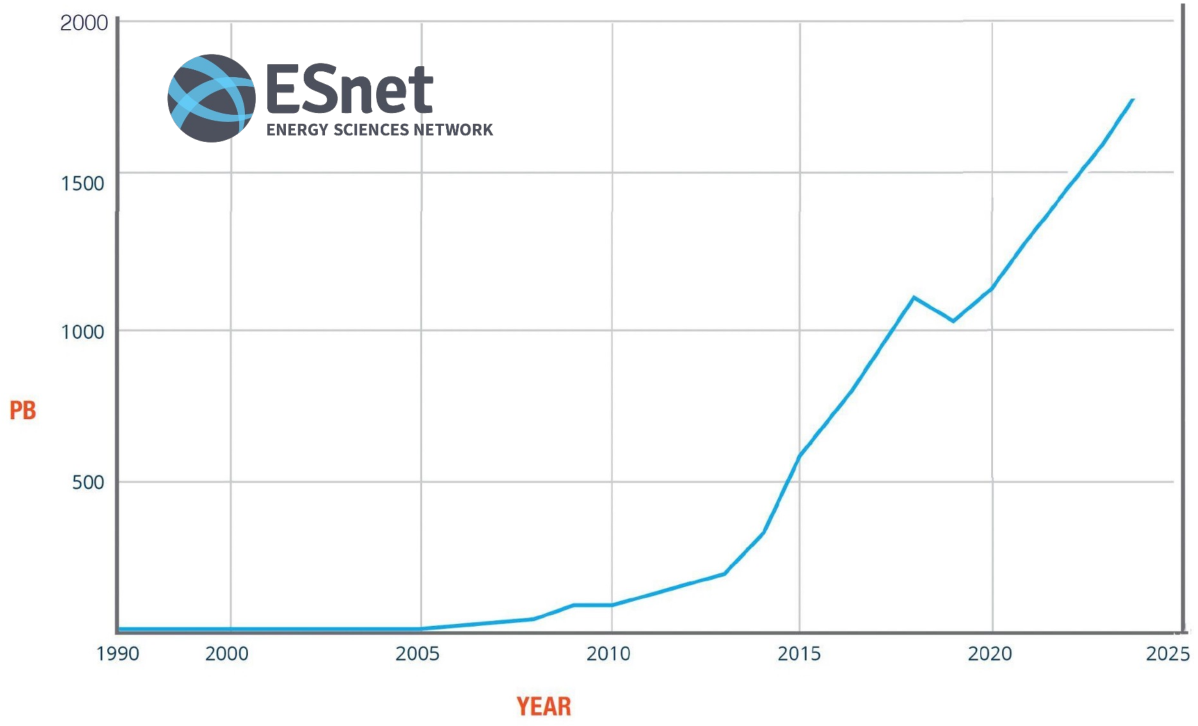
Argonne Photon Source (APS), Advanced Light Source (ALS), Facility for Rare Isotope Beams (FRIB) in active development to trial this approach

EJFAT design

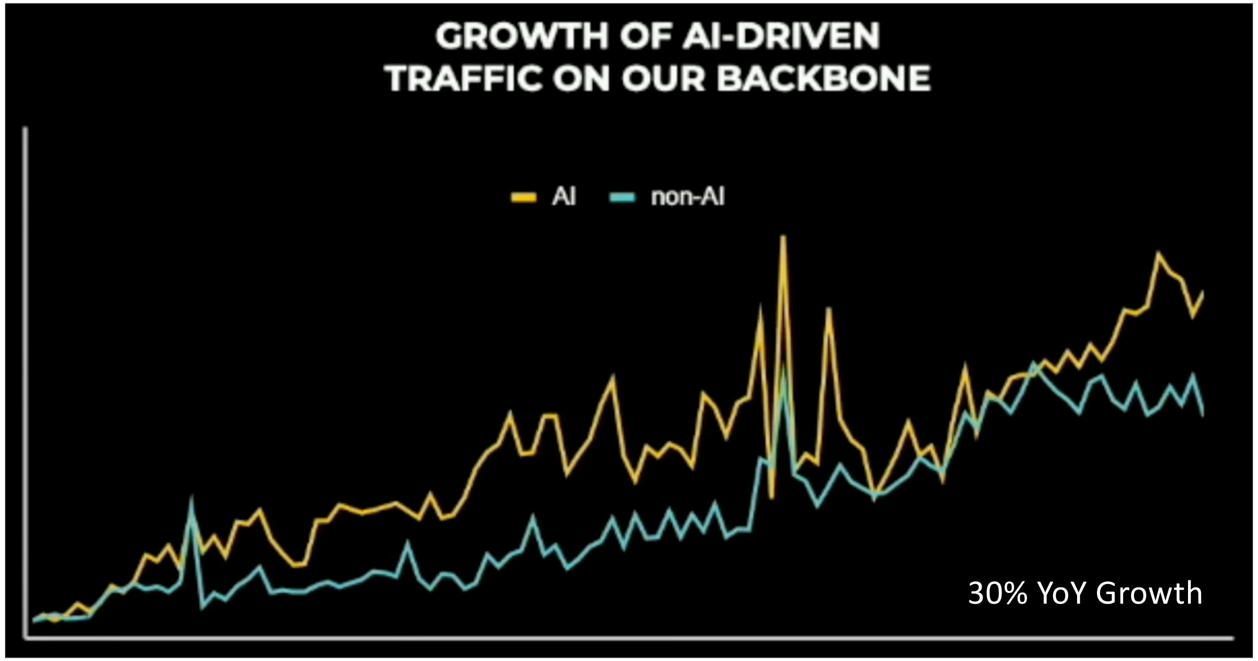


AI will be the next big driver of data

ESnet has seen exponential growth for the last 3 decades, but AI data has not been a growth factor (so far).



In recent years, some hyperscalers are seeing AI data overtake non-AI data on the backbone.



Networking @Scale 2024, AI Impact on the Backbone

Networks are more than moving data.

HPC centers are more than compute engines.

Experiments are more than data sources.

>>> We will need automation and AI to help manage the complexity.

Questions for you, the INDIS audience:

What opportunities do you see for network innovation to support research in the AI era?

What would you create?

What would you change?

Visit
https://iri.science
for SC24 demos, talks, office hours