Perspective of Some Trends

September 30th , 2024

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego

Data is becoming a first class citizen in software & computing

Two big Trends driving Data

Artificial Intelligence

- National AI Research Resource (**NAIRR**)
  - Expected funding of $2.6B over 6 years
  - Across more than a dozen federal agencies with NSF as lead
- Frontiers in AI for Science, Security and Technology (**FASST**)
  - AI Centers at DOE labs for $30M/year for 5-7 years

Instrument integration

- Instrument labs at Universities that can produce TB/hour data volumes
  - High Throughput instrumentations that produces curated data
- DOE Integrated Research Infrastructure program (**IRI**)
  - High Performance Data Facility

NAIRR

Currently in Pilot Stage.
See https://nairrpilot.org for details.

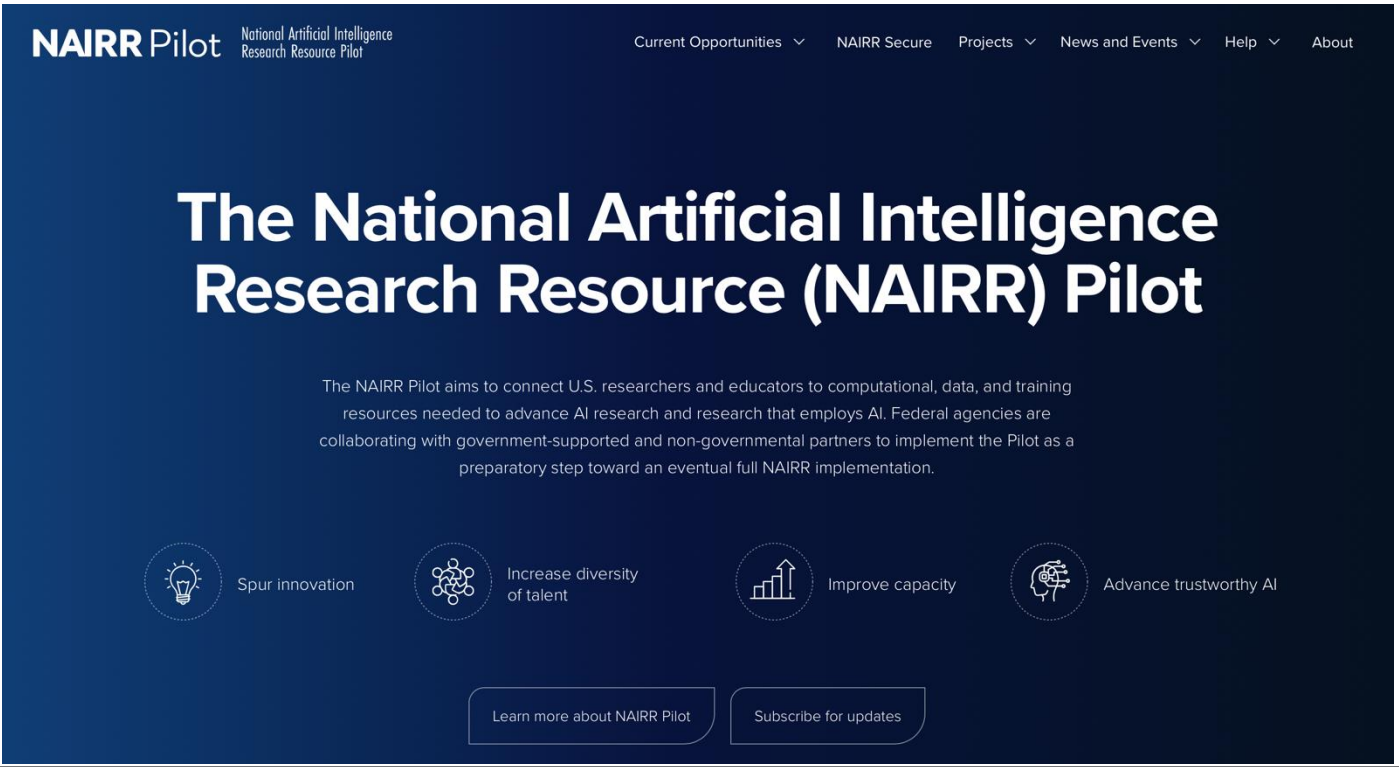


Figure 2. NAIRR Strategic Objective and Goals

[NAIRR TF report](#) suggested funding at the level of $2.6B spread over 6 years.

FASST

https://www.energy.gov/fasst

Currently a draft bill

NEW: Request for Information

DOE is seeking public input to inform how DOE can leverage its existing assets at its 17 national labs and partner with external organizations to support building a national AI capability. Responses are due by Monday, November 11, 2024.

FASST Overview

FASST will build the world's most powerful integrated scientific AI systems through four key interconnected pillars:

- AI-Ready Data +
- Frontier-Scale AI Computing Infrastructure and Platforms +
- Safe, Secure, and Trustworthy AI Models and Systems +
- AI Applications +

Download the factsheet on Frontiers in Artificial Intelligence for Science, Security and Technology (FASST)

The draft bill as written emphasizes creation of at least
8 multidisciplinary AI Research and Development Centers at DOE National Labs
Each center shall receive an annual budget of no less than $30M
for a duration between 5-7 years, with a possible renewal for another 5 years.

Integrated Research Infrastructure

Integrated Research Infrastructure Architecture Blueprint Activity (Final Report 2023)

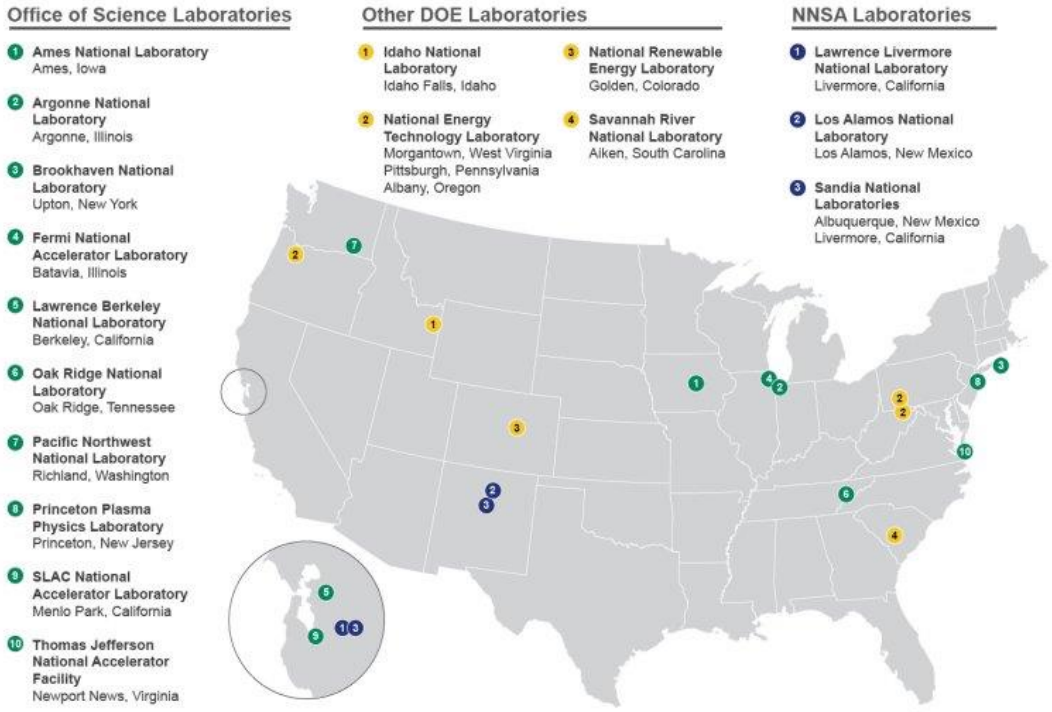
TECHNICAL REPORT · 03 July 2023

DOI: https://doi.org/10.2172/1984466 · OSTI ID: 1984466

Miller, William L. [1]; Bard, Deborah [2]; Boehnlein, Amber [3]; Fagnan, Kjiersten [4]; Guok, Chin [5]; Lançon, Eric [6]; Ramprakash, Sreeranjani (Jini) [7]; Shankar, Mallikarjun [8]; Schwarz, Nicholas [9]; Brown, Benjamin L. [10]

+ Show Author Affiliations

The complexity of scientific pursuits is increasing rapidly with aspects that require dynamic integration of experiment, observation, theory, modeling, simulation, visualization, machine learning (ML), artificial intelligence (AI), and analysis. Research projects across the Department of Energy (DOE) are increasingly data and compute intensive. Innovative research teams are accelerating the pace of discovery by using high-performance computational and...



There are 10 DOE Office of Science Laboratories and 28 DOE Scientific User Facilities, incl. supercomputers (NERSC, ANLCF, ORLCF, …), accelerators, light sources, neutron scattering sources, Tokamaks, … and ESnet

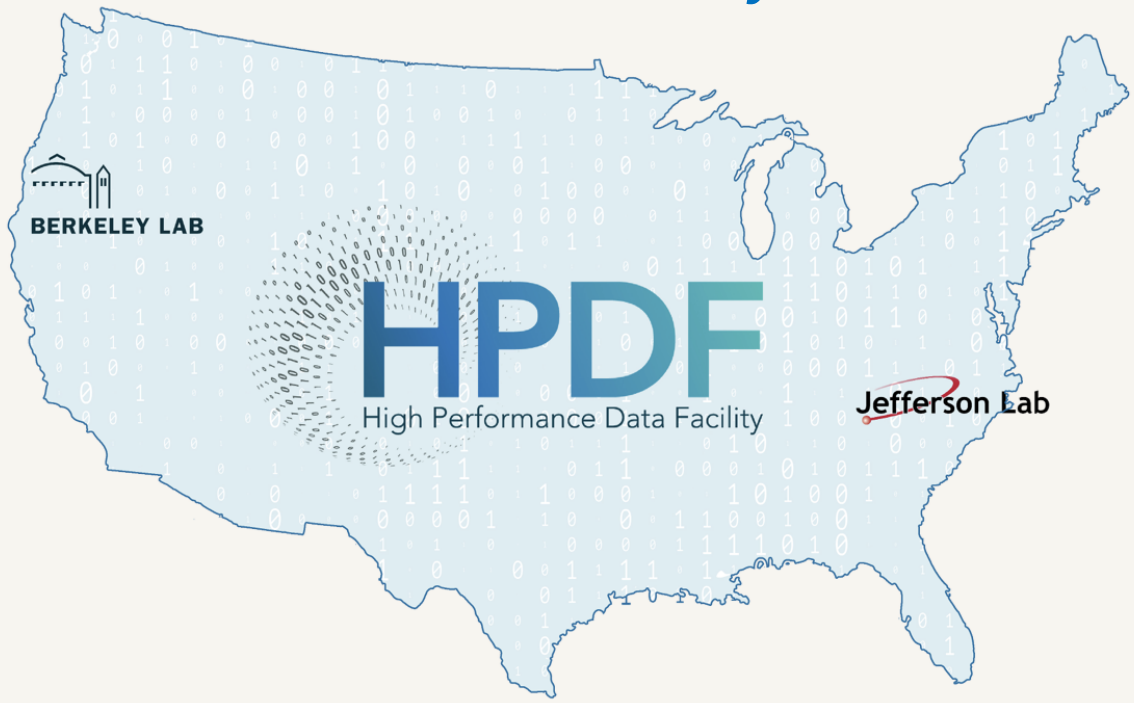
DOE scientific facilities are to be integrated for cross facility science workflows.

HPDF: A New Element in the ASCR Ecosystem

Leading-edge science demands integration, and the U.S. Department of Energy's science programs increasingly require advanced computation, data solutions, and networking. The High Performance Data Facility, a first-of-its-kind project, is envisioned as a state-of-the-art resource for data science and research. HPDF will join other Advanced Scientific Computing Research facilities engaged in implementing the DOE's Integrated Research Initiative.

As the cornerstone of the IRI, HPDF is committed to supporting scientific research patterns and data stewardship within the nation's research communities. Along with the other ASCR facilities – Energy Sciences Network, Argonne Leadership Computing Facility, Oak Ridge Leadership Computing Facility, and the National Energy Research Scientific Computing Center – HPDF will help forge the IRI's foundational infrastructure.

$300-500M "construction project" for the next 10 years



HPDF will be the 5th DOE ASCR facility

The project's hub-and-spoke model will maximize availability, resilience, and accessibility. The Hub core infrastructure, physically located at Thomas Jefferson National Accelerator Facility and Lawrence Berkeley National Laboratory, will support centralized resources. Multiple mission-application Spokes will connect to the Hub via ESnet.

Immersive Virtual Environments for Science

We make data and computationally intensive capabilities accessible to millions of researchers, educators, and students across disciplines through science gateways, secure cloud enclaves, research software development, and immersive visualizations.

From Science Gateways to Immersive Virtual Environments

Future UI/UX will have O(1M) audiences, providing immersive experiences with content and data

Initially web portals lowered the bar for access to HPC. Today, we are building immersive virtual environment enabling access to compute & sharing of data & content.

10,000 Unix Accounts over 5 years

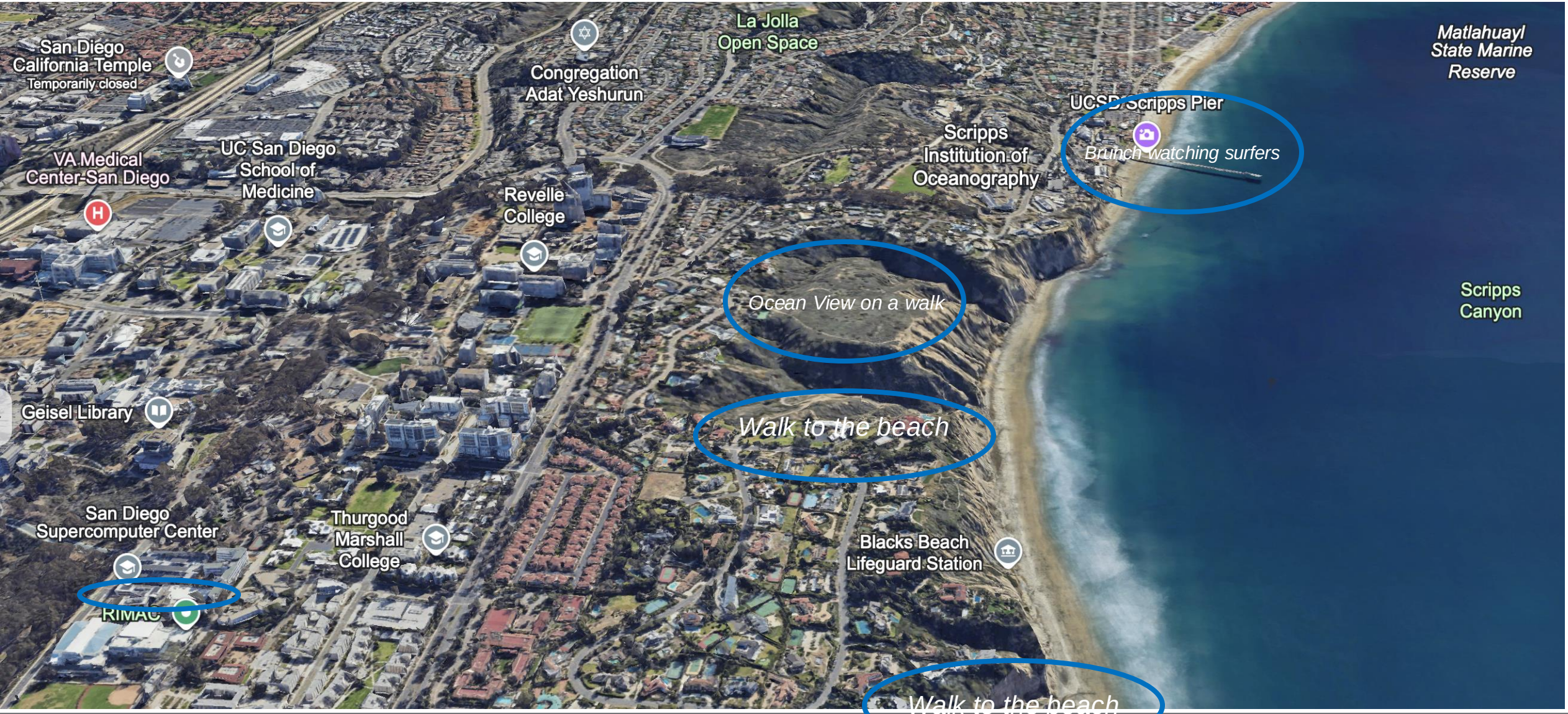
100,000 user/year on gateways

1M users/year for immersive environments

~10x users/year by enabling web access to HPC

~10x users/year by enabling content & data sharing

Welcome to UC San Diego



Welcome to UC San Diego