

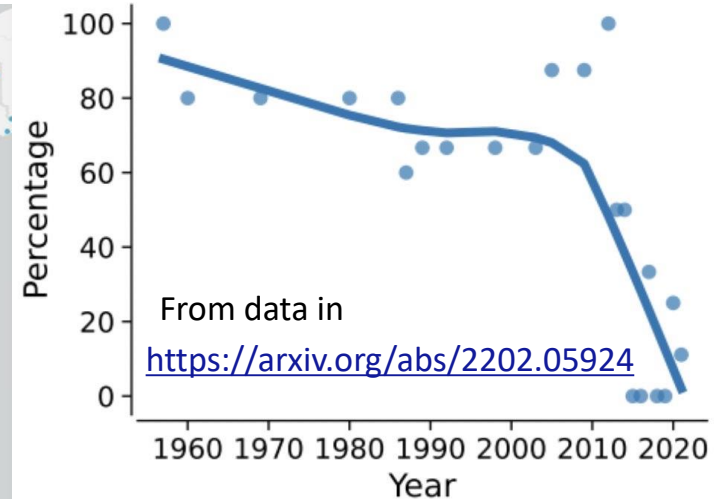
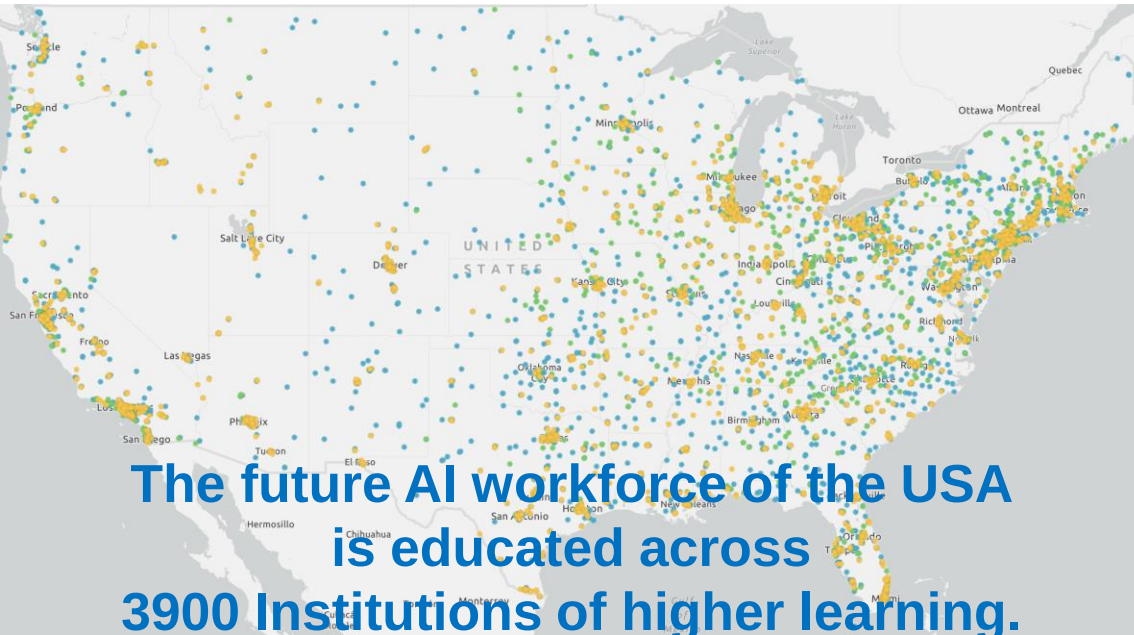
Open Cyberinfrastructure for an Open Society

Frank Würthwein
Director, San Diego Supercomputer Center

June 10th 2024



Why



% of large-scale AI
results from academia
is basically zero

**To guarantee access to AI education at these 3900 institutions
requires rethinking of how to provide cyberinfrastructure.**



The Minds We Need

- **Connect every community college, every minority serving institution, and every college and university, including all urban, rural, and tribal institutions** to a world-class and secure R&E infrastructure, with particular attention to institutions that have been chronically underserved;
- **Engage and empower every student and researcher** everywhere with the opportunity to join collaborative environments of the future, because we cannot know where the next Edison, Carver, Curie, McClintock, Einstein, or Katherine Johnson will come from; and

What

Vision of an Open Infrastructure

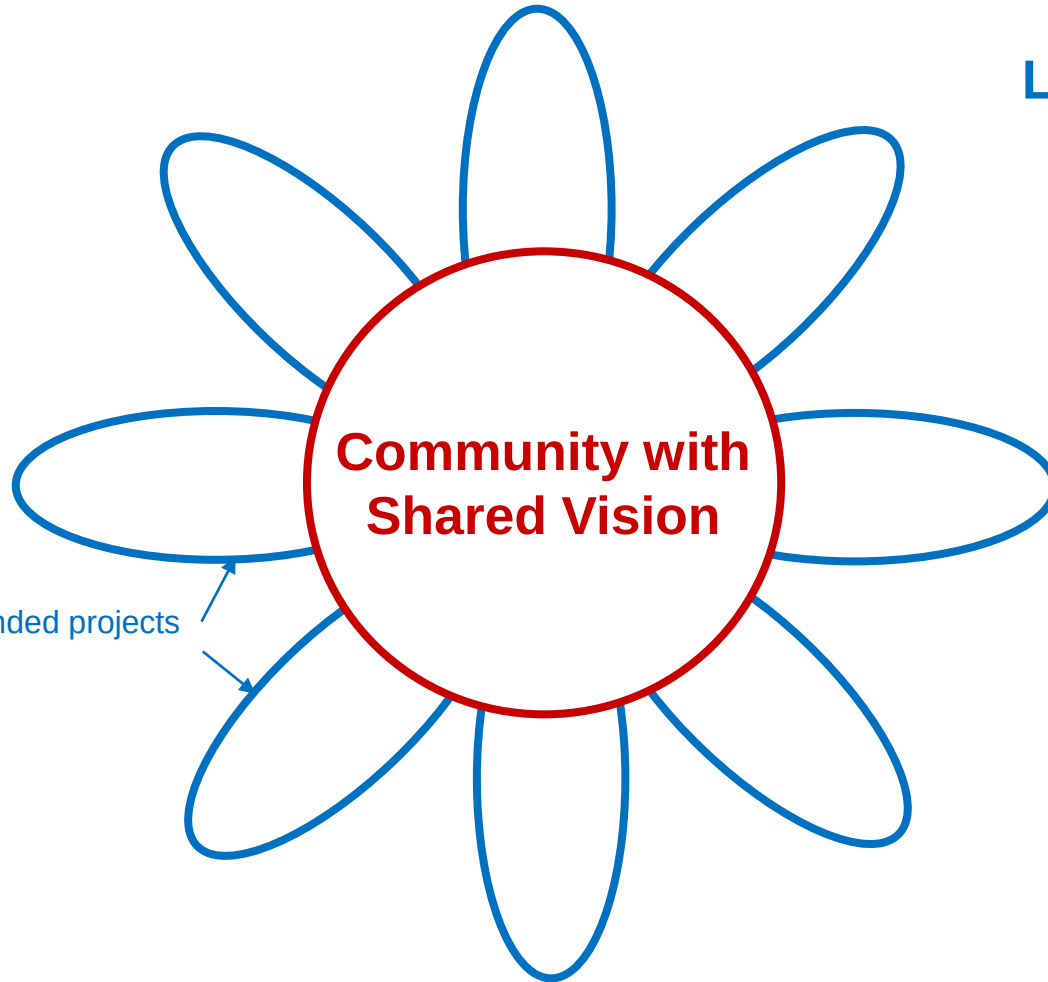
Horizontally open => institutions can integrate their resources

Vertically open => projects can build on the infrastructure

- Create an Open National Cyberinfrastructure that allows the federation of CI at all ~4,000 accredited, degree granting higher education institutions, non-profit research institutions, and national laboratories.
 - Open Science
 - Open Data
 - Open Source
 - Open Infrastructure
 - ← Open Compute
 - ← Open Storage & CDN
 - ← Open devices/instruments/IoT, ...?

Openness for an Open Society

How



Lot's of funded projects that contribute to this **shared vision** in different ways.

Hardware funded by NSF, DOD, DOE, ...

Petals of this flower include:

Prototype National Research Platform
PATH

Open Science Data Federation

Open Science Compute Federation

National Data Platform

Pelican

Fusion Data Platform for AI

R&E platforms for multiple campuses

GP-Engine

TIDE

...

Open Infrastructure is “owned” and “built” by the community for the community

NATIONAL RESEARCH PLATFORM (NRP)

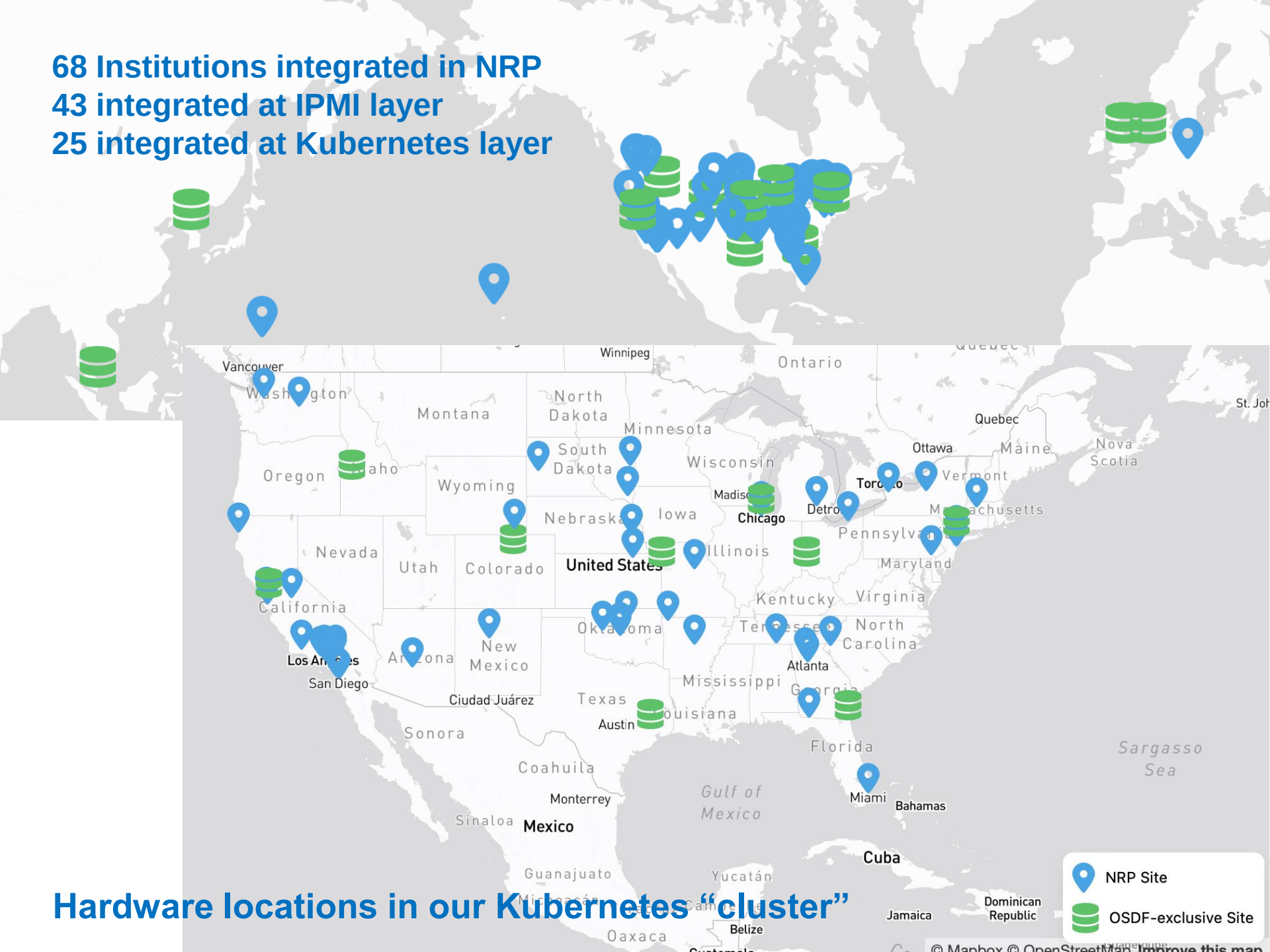
HOW WE EXECUTE ON THIS VISION

- Depending on effort available and control desired, you can build on NRP both vertically and horizontally at different layers of the stack.



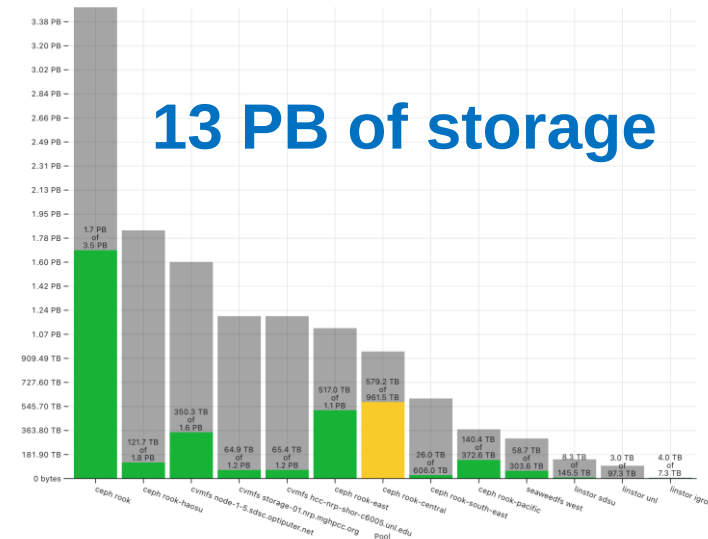
- NRP is a non-local extendable container deployment platform, thus allowing many uses unthinkable for a SLURM cluster in a data center.

68 Institutions integrated in NRP
43 integrated at IPMI layer
25 integrated at Kubernetes layer

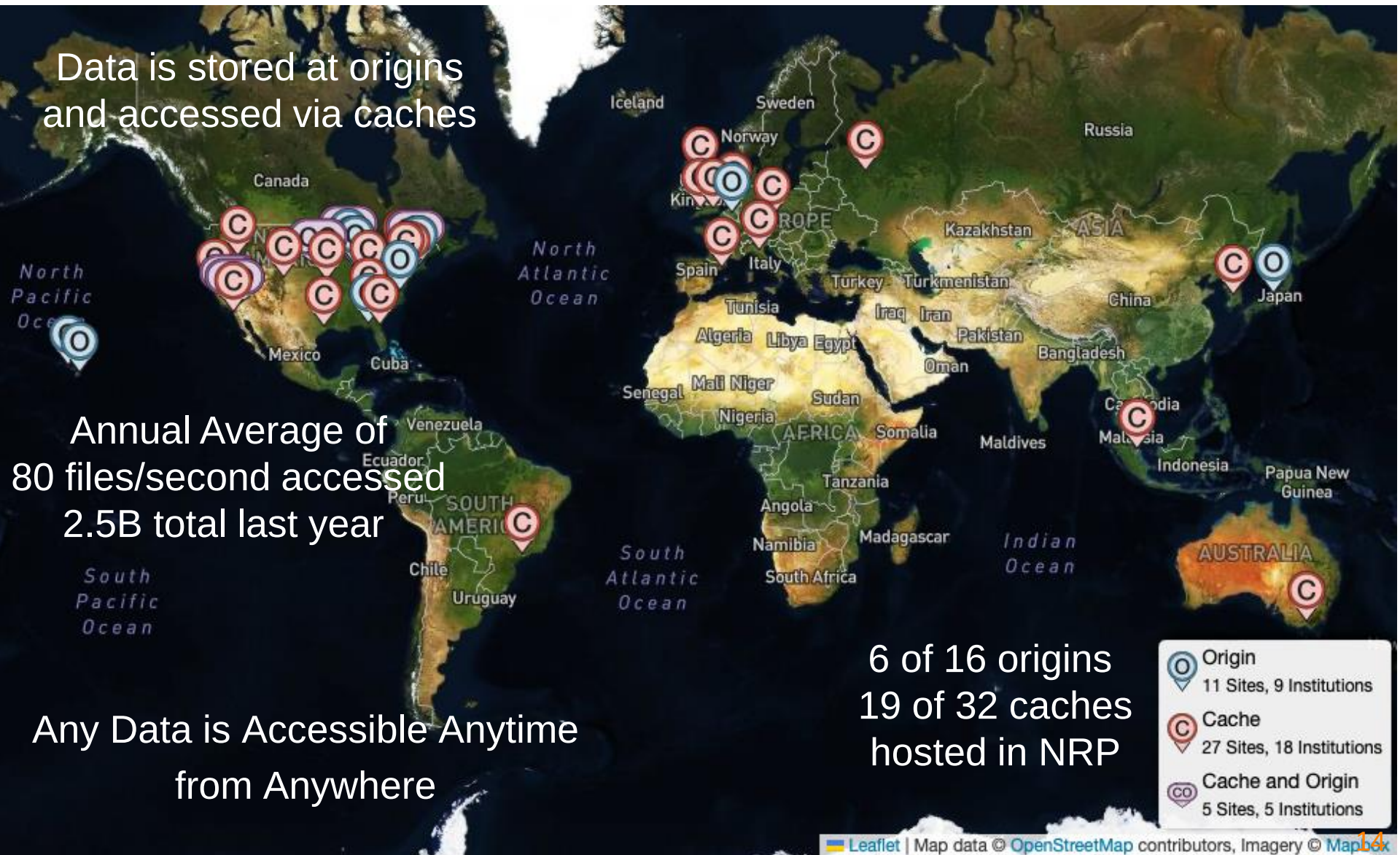


Hardware locations in our Kubernetes “cluster”

Type of Accelerator	# of that type in NRP	Type of Accelerator	# of that type in NRP
A100 80G	224	4090	10
A100 40G	2	3090	247
A10	288	2080Ti	152
A40	10	1080Ti	139
A6000	48	Xilinx U55C	32
A5000	6	Other GPUs	104
A4000	32	Total	1262



In addition, there are 21,070 x86 CPU cores, 7,406 are in nodes that have no accelerators.





143 Institutions joining at the batch or storage system layer



Combining all 3 layers, we integrate ~200 institutions across 5 continents

Bridging Education & Research by having them co-exist on one platform

A lot of the smaller campuses care more about
educational use than research use on NRP.

Building AI Education on NRP

... Two Examples ...

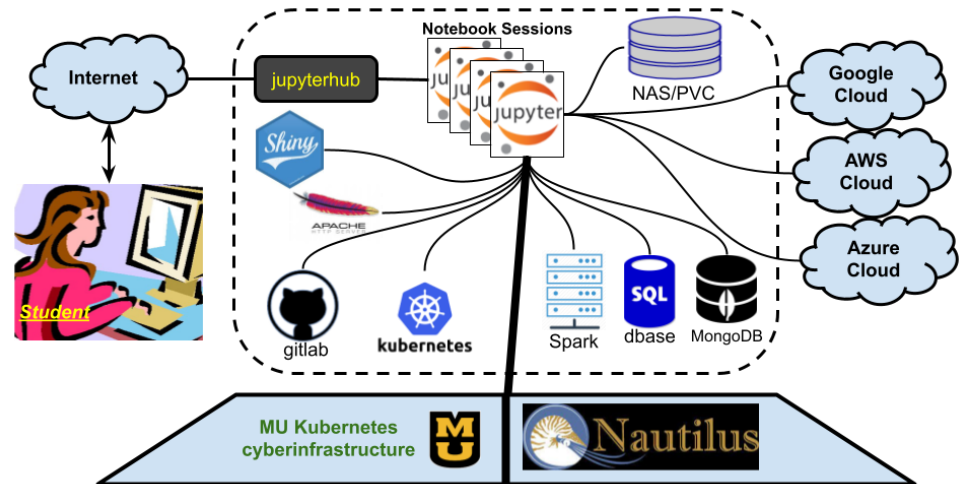
- GP-Engine (NSF OAC 2322218) builds a regional compute cluster across 7 states in the Great Planes Regional Network.
 - Hardware is mostly GPUs, with strong focus on supporting STEM education, especially AI.
- **CENIC AI Resource**
 - Collaboration between California's regional R&E network, NRP, and multiple California State Universities with strong focus on AI education
 - Includes "The California State University System Technology Infrastructure for Data Exploration (TIDE)" ([NSF OAC 2346701](#))
 - The CSU system includes 23 campuses, 21 of which are Hispanic-Serving Institutions.
 - "... most diverse higher education system in America ... provides 50% of all bachelor degrees in California ..."
 - " TIDE creates a pioneering computational core facility within the California State University (CSU) system, focused on ML and AI "

Grant Scott

<https://scottgs.mufaculty.umsystem.edu/>



NSF OAC 2322218



- Computer Science Undergraduate Classes
- Computer Science HPC Classes
- Undergraduate Data Science
- HPC Emphasis Graduate Data Science
- State & Federal Government Training Programs

- AI Filmmaking Hackathon
 - Using Generative AI to create film
 - Uses NRP to host the AI models and the inference, ie, standard diffusion

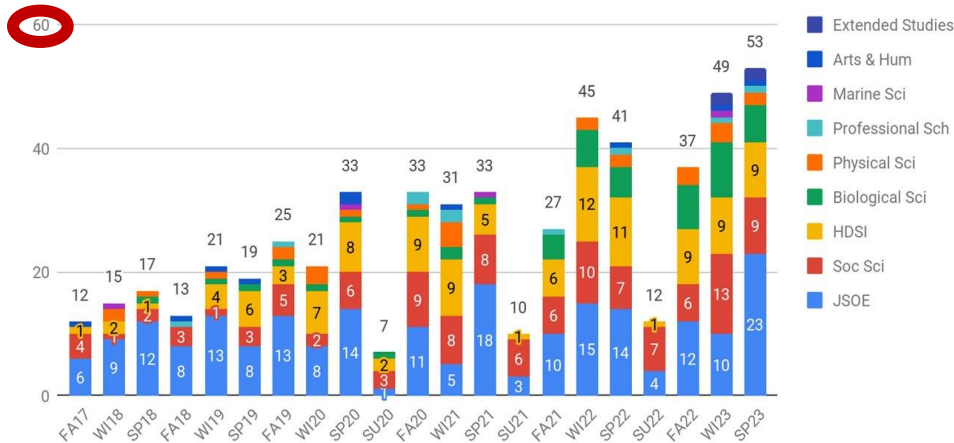


Twomey @ UNL's Johnny Carson School of Theatre & Film

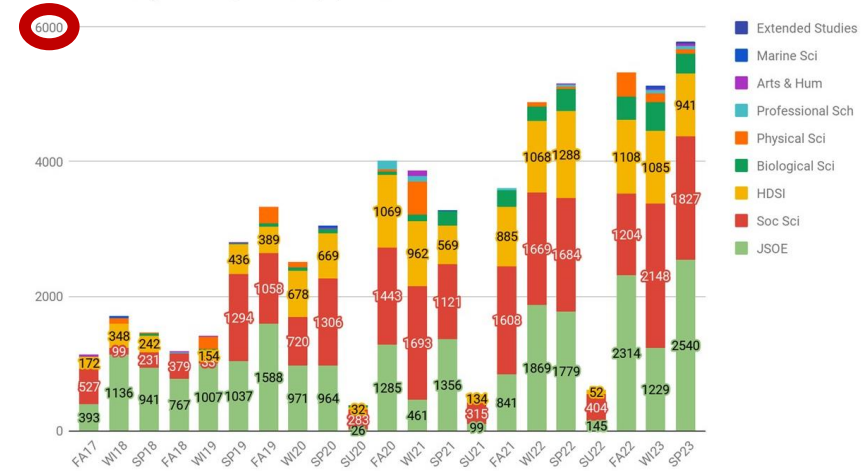


Image generated with stable diffusion
instance on the NRP

#Courses by School/Division, Quarter



#Students by School/Division, Quarter



Source: Adam Tilghman, UCSD ITS
Data Science/ML Platform - 5/2023 Update

UCSD operates a modest size cluster (~140) 32-bit GPUs for use in the classroom

During FY23, this was used by 6,502 undergraduates & 1847 graduate students

33% of 1.5M pod-hours per year are used by AI/ML containers

In Spring Quarter 2023, 53 courses across 8 schools were taught on this system

- Advanced Computer Vision
- Bioinformatics for Immunologists
- Computational Physics: Probabilistic Models/Sim.
- Data Analysis/Design for Biologists
- Data Science/Spatial Analysis
- Deep Learning and Applications
- Intro to Causal Inference
- Neural Networks/Pattern Recognition
- Numerical Analysis for Multiscale Biology
- Robot Manipulation and Control

AI is accelerating a trend that already existed prior.

- **NRP Operates a Matrix Chat infrastructure**
 - All Educators across all institutions are signed up to a channel to share experiences.
 - Each course has its own chat, moderated by the educator(s) that teach the course
- **NRP supports Jupyter**
 - One JupyterHub for all newcomers
 - Experienced Institutions tend to clone their own, and build their own custom infrastructure on NRP
 - Join their own hardware & customize user experience for their users
- **We support interactive and batch access to GPUs**
 - Jupyter, cube-ctl (Unicorn as scheduler), HTCondor

Example: CSU San Bernardino

- < 1,100 faculty members
- < 19,000 students
- < Serves 2 of CA's largest counties
- < Hispanic Serving Institution
- < 57% Pell Grant recipients
- < Many student oriented projects





CSUSB JupyterHub

Institutional logon

- Accessible to anyone with a CoyoteID
- Each user is provided with 50 GB of dedicated storage, expandable upon request
- Resources include large RAM and storage and multiple CPUs and GPUs
- Pre-installed software includes Linux desktop, Python, Matlab, R, Scipy, Pyspark, Tensorflow, a data science package, and SageMath
- Other software can be installed upon request
- Can be used for research and classes

dedicated hubs for labs

CSUSB JupyterHub for faculty research

- Dedicated to faculty research
- Contact [Youngsu Kim](#) or [Dung Vu](#) to access this server

Server Options

Advanced Options

Image

☐	Stack Minimal
☐	Stack Datascience
☐	Stack R-Studio, Vs-code for Dr. Becerra's class
☐	Stack Desktop Apps - VS Code
☐	Stack Desktop Apps - Pgadmin4
☐	Stack Desktop Apps - Blender
☐	Stack PySpark
☐	Stack PyTorch2
☐	Stack R-Studio
☐	Stack R-Studio for BIOL-5050
☐	Stack SageMath

<https://csusb-metashape.nrp-nautilus.io>: 3D modeling

<https://csusb-vasp1.nrp-nautilus.io> Viena Ab initio Simulation package (VASP)

<https://csusb-cousins-lab.nrp-nautilus.io>: VASP simulation

<https://csusb-becerra.nrp-nautilus.io> AI/ML project

<https://csusb-biol-5050.nrp-nautilus.io>: Biology course

<https://csusb-cse-salloum.nrp-nautilus.io> Summer Research

<https://csusb-drhamoudahub.nrp-nautilus.io> Data Analytics

<https://csusb-ratnasingam.nrp-nautilus.io> Data Analytics

<https://csusb-zhang.nrp-nautilus.io> AI/ML project

Wide range of dedicated hubs

5 Departments teach courses on NRP

CSE 5160 Machine Learning

IST 2510 Big Data Analytics

BIOL 5050 - Biostatistics and Experimental Design

IST 6110 - Foundations of Analytics and Big Data

IST 6620 - Business Analytics and Decision-Making

MATH 3465 - Computational Statistics

14 AI/ML Projects on NRP

- Brain Tumor classification with Deep Learning
- Public Opinion Sentiment Analysis
- Medical Notes Labelling
- Cloud Removal from Satellite Imaging
- Wearable Technology for Assessing Autism
- GPT empowered personalized eLearning for programming
-

- **NRP has a very ambitious vision**
 - Horizontally open
 - Hardware across 68 institutions has joined NRP
 - Vertically Open
 - We have built the “Open Science Data Federation” on top of NRP,
 - ... and are starting to build “Fusion Data Platform for AI” on top of NRP
 - ... and are starting to build elements of the National Discovery Cloud for Climate on top of NRP (Pelican, National Data Platform, NCAR integration, ...)
 - “Playground” for CS R&D on the same platform as “Production” system for Domain Scientists
- **Education increasingly requires significant CI**
- **NRP is owned & built by the community for the community**
 - We are always looking for new partners!

- This work was partially supported by the NSF grants OAC-1541349, OAC-1826967, OAC-2030508, OAC-1841530, OAC-2005369, OAC-21121167, CISE-1713149, CISE-2100237, CISE-2120019, OAC-2112167

