NRP NATIONAL RESEARCH PLATFORM

UCSD

Open Infrastructure for an Open Society

Frank Würthwein
Director, San Diego Supercomputer Center

May 3rd 2023

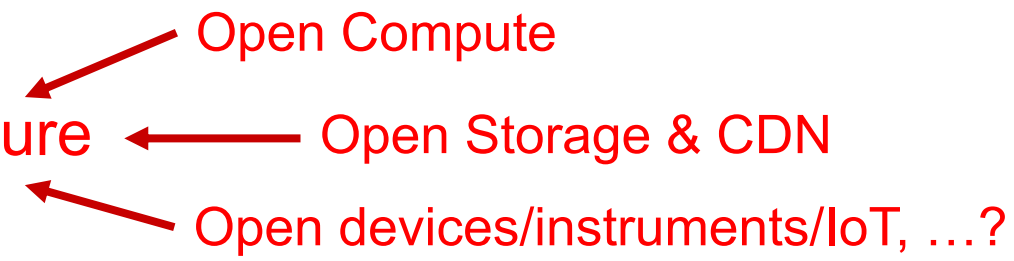
NATIONAL RESEARCH PLATFORM

UCSD

Vision

Long Term Vision

- 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



Openness for an Open Society

NATIONAL RESEARCH PLATFORM

UCSD

National Research Platform (NRP)

Community vs Funded Projects

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

We want to …
… extend NRP horizontally
… build on NRP vertically
… via many independently funded projects.

NRP is "owned" and "built" by the community for the community

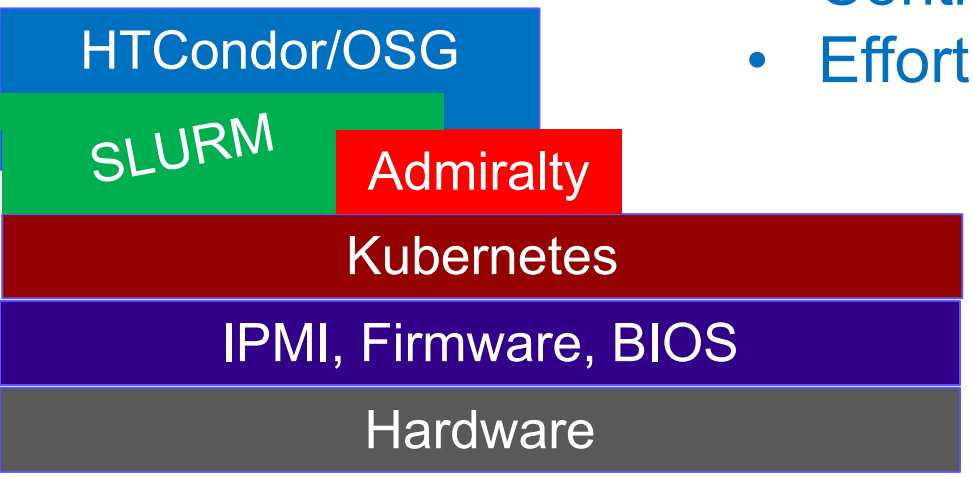
Cyberinfrastructure Stack

NRP operates at all layers of the stack, from IPMI up

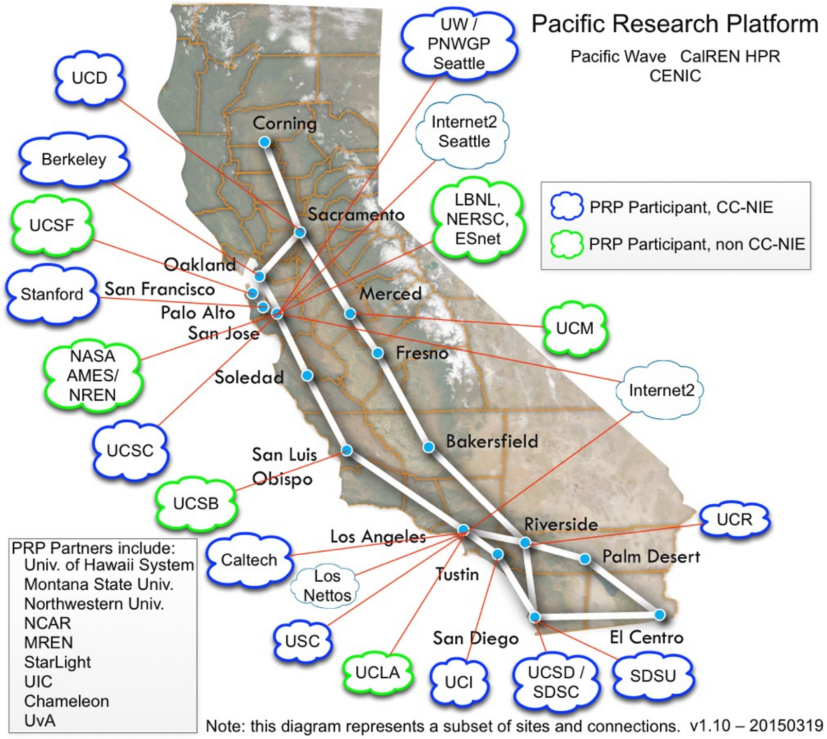
- IPMI reduces TCO and lower threshold to entry
- Kubernetes allows service deployments
  - Also the natural layer for application container deployment
- Admiralty allows K8S federation with folks who want control
  - Including cloud integration to access TPUs & other cloud only architectures
- HTCondor allows NRP to show up as a “site” in OSG

The layer you integrate at depends on

- Control you want
- Effort you can afford



NRP builds on 2 predecessors



Pacific Research Platform

A distributed Kubernetes cluster (2015 – 2023)

- **When the PRP Grant Was Funded in 2015, It Started With:**
  - 6 States ➡ Now 43 States
  - 19 Campuses ➡ Now 110 Campuses
  - 9 Minority Serving Institutions ➡ Now 23 MSIs
  - 2 NSF EPCoR States ➡ Now 19 EPSCoR States, 2 Territories, and Wash DC

NRP builds on 2 Predecessors

Open Science Grid

A Federation of Compute and Storage Clusters (founded in 2005)

NRP NATIONAL RESEARCH PLATFORM

UCSD

Complementarity in Implementation of "Bring Your Own Resource" model

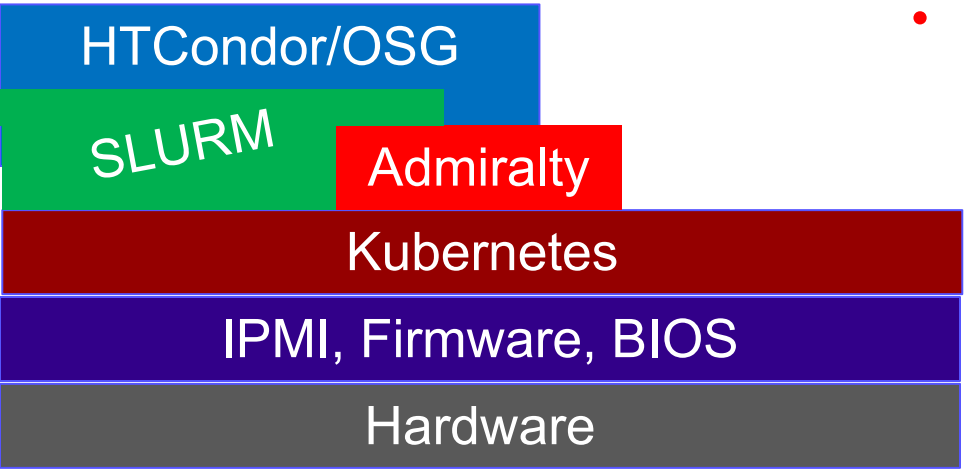
OSG focuses on campus cluster integration.

NRP focuses on individual node integration instead of clusters.

Cyberinfrastructure Stack

NRP operates at all layers of the stack, from IPMI up

- **IPMI reduces TCO and lowers threshold to entry**
- Kubernetes allows service deployments
  - Also the natural layer for application container deployment
- Admiralty allows K8S federation with folks who want control
  - Including cloud integration to access TPUs & other cloud only architectures
- HTCondor allows NRP to show up as site in OSG

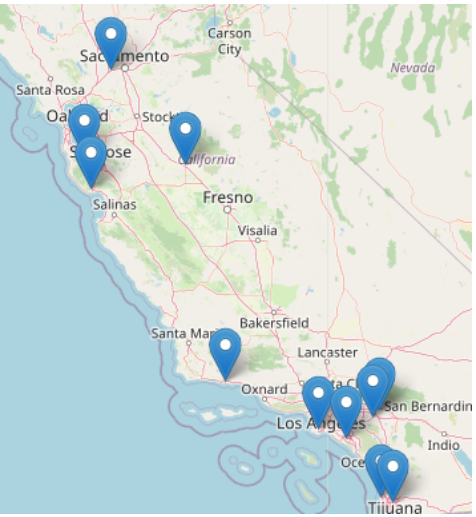


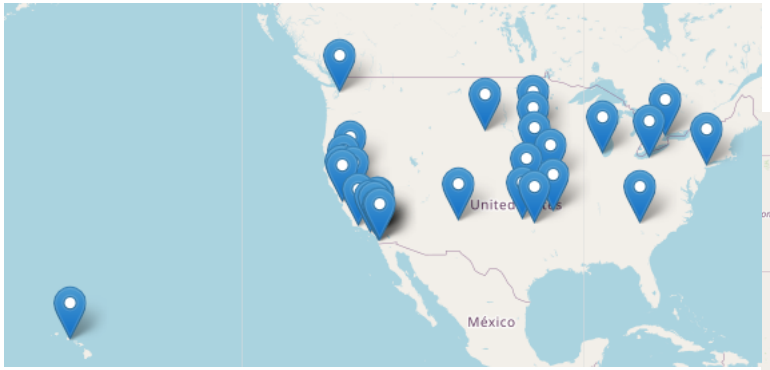
- **Under-resourced institutions**
- **Network providers and their POPs**
- **CS & ECE faculty specialized in:**
  - **AI/ML => gaming GPUs**
  - **systems R&D**

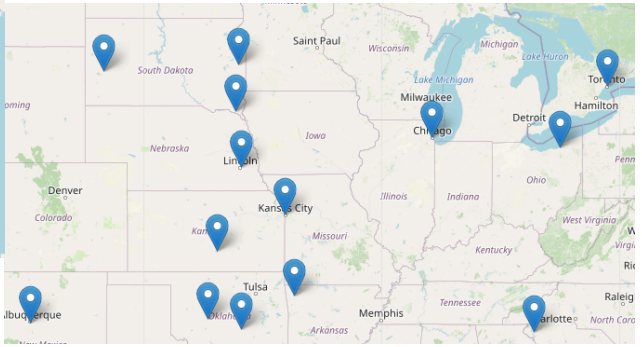
All of these find it difficult to justify staff to support all layers

Hardware on NRP is Global

NRP integrates hardware in USA, EU, and Asia







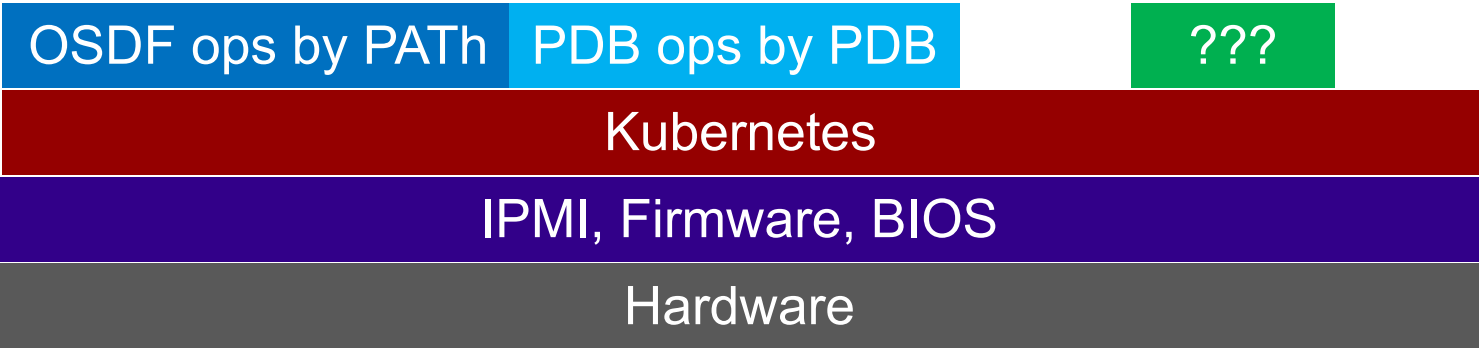
Cyberinfrastructure Stack

NRP operates at all layers of the stack, from IPMI up

- IPMI reduces TCO and lowers threshold to entry
- **Kubernetes allows service deployments**
  - Also the natural layer for application container deployment
- Admiralty allows K8S federation with folks who want control
  - Including cloud integration to access TPUs & other cloud only architectures
- HTCondor allows NRP to show up as site in OSG

NRP is unique in its support of global service deployments

- **Open Science Data Federation**
  - Origins & Caches in US, EU, Asia
- **Protein Data Bank**
  - (Future) Replicas in EU & Asia



NRP NATIONAL RESEARCH PLATFORM

UCSD

Supporting Nautilus for the next decade

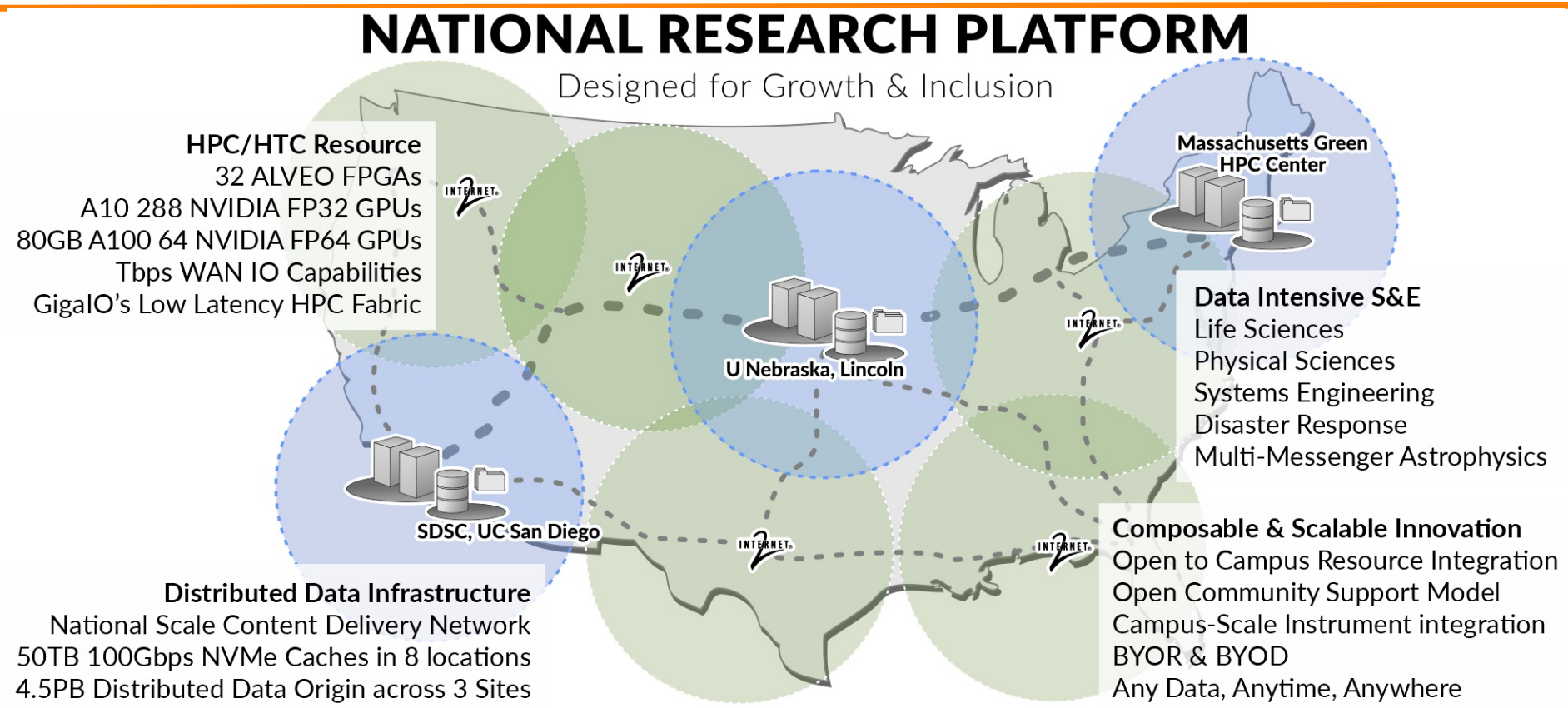
Nautilus = K8S infrastructure of PRP for the last 5+ years

Nautilus = K8S of NRP for the next 10 years

The NSF Cat-II Program

- NSF supports via the Cat-II program novel systems ideas.
  - 3 year “testbed” phase
    - The PI owns the resource, and has (some) freedom regarding who uses it.
      - No requirements for making it available via any specific allocation mechanism.
    - It is expected that not all features work on day 1.
      - 3 years of experimentation & development of features
  - 2 year “allocation” phase
    - The resource is made available via an NSF supported allocation mechanism.
  - **The solicitation mentions the possibility of an additional 5 year renewal without re-competition if system is successful.**
- **We decided that this is an ideal program to try and secure NRP core operations funding for the next 10 years**
  - And thus provide the stability necessary for growth of NRP.

Cat-II: Prototype National Research Platform (PNRP)



5 year project with $5M hardware & $6.45M people

Supports Nautilus, and thus the core NRP infrastructure

NSF Acceptance Review of System passed on March 9th 2023

PI = Wuerthwein; Co-PIs: DeFanti, Rosing, Tatineni, Weitzel

Funded as NSF 2112167

Innovations

- I1: Innovative network fabric that allows “rack” of hardware to behave like a single “node” connected via PCIe.
- I2: Innovative application libraries to expose FPGAs hardware to science apps at language constructs scientists understand (C, C++ rather than firmware)
- I3: A “Bring Your Own Resource” model that allows campuses nationwide to join their resources to the system.
- I4: Innovative scheduling to support urgent computing, including interactive via Jupyter.
- I5: Innovative Data Infrastructure, including national scale Content Delivery System like YouTube for science.

I3 & I4 & I5 turn ”PRP” into “NRP” and sustains it into the future.

I1 & I2 are totally new.

Data Infrastructure Model of NRP

- Support regional Ceph storage systems across the USA.
  - Campuses can join individual storage hosts to the Ceph system in their region.
  - All regional storage systems are Data Origins in OSG Data Federation (OSDF)
  - Deploy replication system such that researchers can decide what part of their namespace should be in which regional storage.
- Deploy caches in Internet2 backbone such that no campus nationwide is more than 500 miles from a cache.

NRP data infrastructure model combines best of PRP & OSG

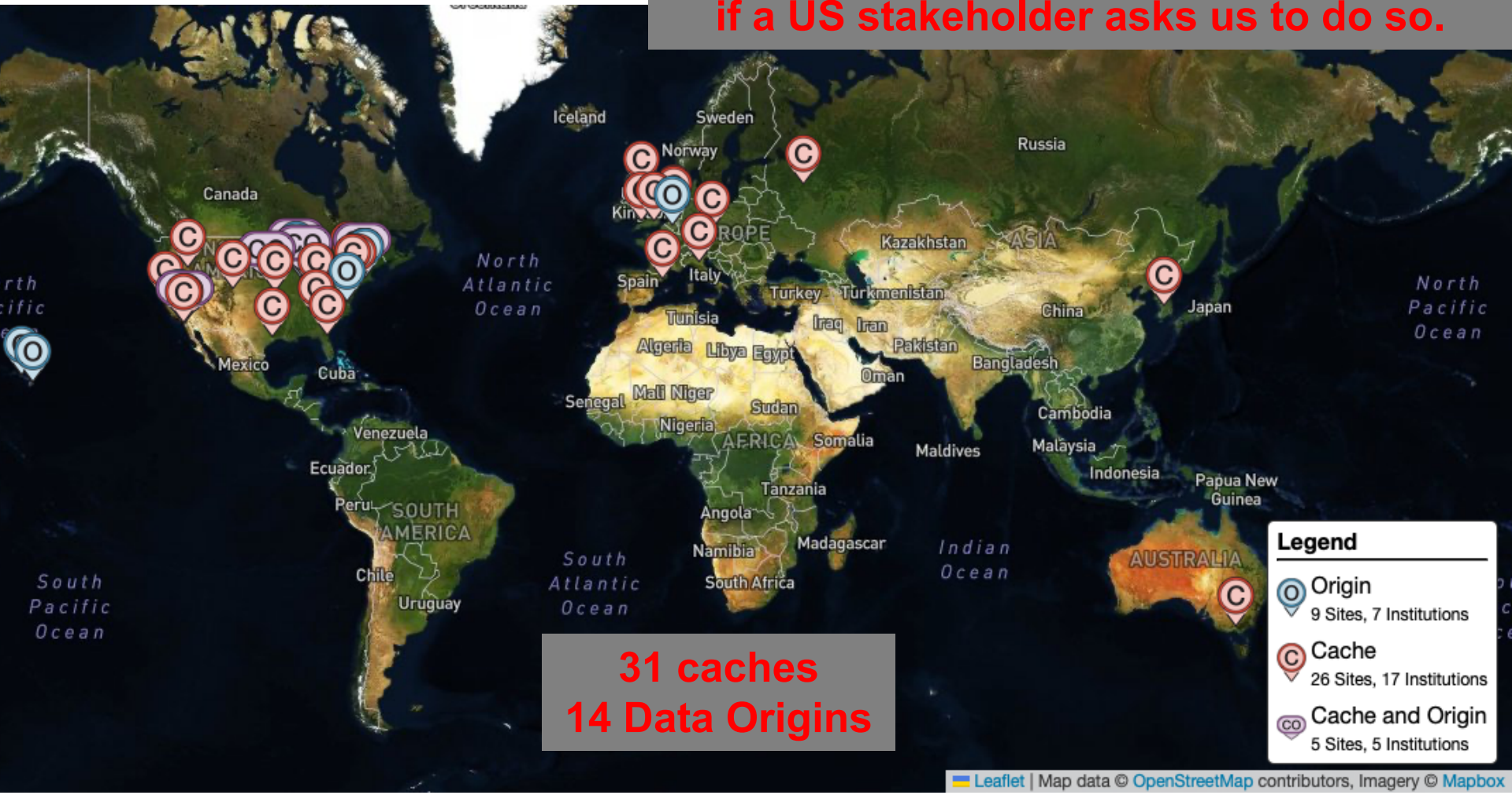
From PRP we take the regional Ceph storage concept
From OSG/PATh we take the data origin & caching concepts

And then we add as a totally new feature:
User controlled replication of partial namespaces across regions.
(We will develop this during 3 year “testbed” phase)

Want Others to build higher level data services on top

Open Science Data Federation

We integrate institutions outside the USA if a US stakeholder asks us to do so.



Global data distribution driven by LIGO, Virgo, DUNE, …

Matrix of Science x Innovations

Table 3.1 Representative Science and Engineering Use Cases

Application domain	Lead researcher & Institution	Science Driver Themes	NRP Innovations
LIGO	Peter Couvares, LIGO Lab; Erik Katsavounidis, MIT	BGS, UC, AI	I2, I3, I4, I5
IceCube	Benedikt Riedel, UW Madison	BGS, UC, AI	I3, I4
Astronomy (DKIST & Sky Surveys)	Curt Dodds, U. Hawai'i	BGS, AI	I3, I5,
Campus Scale Instrument Facilities	Mark Ellisman, NCMIR; Samara Reck-Peterson, Nicon Imaging Center; Johannes Schoeneberg, Adaptive Optics Lightsheet Microscopy; Kristen Jepsen, Institute for Genomic Medicine; Tami Brown-Brandl, Precision Animal Management	SD, UC, H	I1, I2, I3, I4, I5
Molecular Dynamics	Rommie Amaro, UCSD; Andreas Goetz, SDSC; Jonathan Allen, LLNL	MD, AI, H	I1, I2, I3
Human microbiome	Rob Knight, UCSD	G, AI, H	I1, I2, I3
Genomics & Bioinformatics	Alex Feltus, Clemson	G, AI, H	I3, I4, I5
Fluid Dynamics	Rose Yu, UCSD	AI	I1, I2, I3
Experimental Particle Physics, IAIFI	Phil Harris, MIT	AI, BGS, SD	I1, I2
Computer Vision	Nuno Vasconcelos, UCSD	AI, CV	I3
Computer Graphics	Robert Twomey, UNL	CV, AI	I3
Programmable Storage	Carlos Malzahn, UCSC	SD	I1, I2, I5
AI systems software stack for FPGAs	Hadi Esmaeilzadeh , UCSD	SD	I1, I2
WildFire Analysis & Prediction	Ilkay Altintas, UCSD	UC, AI, CV	I3, I4

Key: The NRP innovations column lists those innovations among I1 through I5 listed in Section 2.1 that a given science driver most benefits from.

Lot's of AI …
but so much more …

NSF MREFCs

Incl. 4 campus scale instrument facilities

Incl. a very diverse set of sciences and engineering

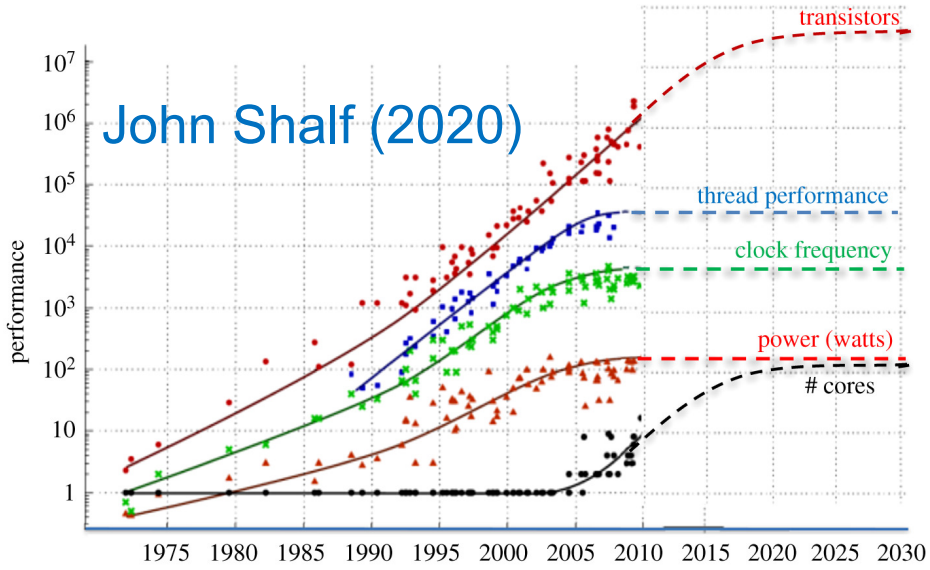
FKW's Wishlist for NRP

- **Growth of NRP infrastructure**
  - 1,000++ GPUs end of 2022 => 1,300 GPUs as of April 2023
  - 50 PB storage end of 2024
  - Growth in diversity of community
    - # and types of campuses and their researchers
- **Introduce new capabilities to NRP**
  - Machine learning at 100TB scale
  - Support Domain Specific Architecture R&D on NRP
  - Expand NRP into Wireless, Edge, IoT
  - Towards "FAIR" on OSDF
- **New Directions initiated by the Community**

"Domain Specific Architectures"

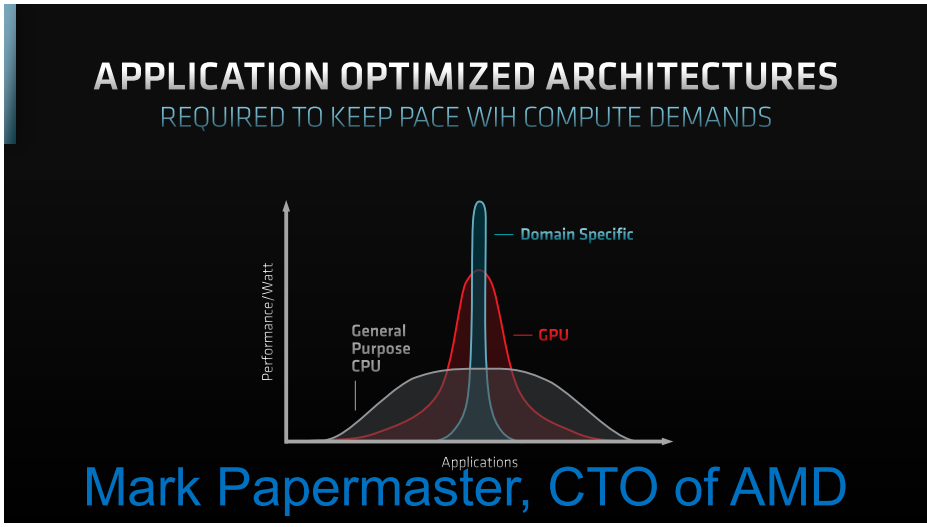
- I1: Innovative network fabric allowing "composable hardware".
- I2: Innovative application libraries allowing "domain optimized architectures" on FPGAs

"end of Moore's law" motivates new architectures



John Shalf (2020)

https://doi.org/10.1098/rsta.2019.0061



Mark Papermaster, CTO of AMD

PI, Tajana Rosing

PRISM, a Jump 2.0 project funded by SRC is early user of FPGAs@PNRP

New Data Origins

- The NSF CC* 2022 program awarded 9 campuses with $500k storage awards each.
  - We guess this pays for 5PB of storage each.
- Some of these campuses may decide to integrate their CC* storage into the OSDF.
- Some of these campuses have storage from other projects that they may integrate with the OSDF in addition.
- **NSF 23-523 includes $500k storage solicitation again, Spring & Fall 2023.**

NATIONAL RESEARCH PLATFORM

UCSD

New Infrastructure Coming up in USA

High Performance Data Facility (HPDF)

- DOE is competing (LAB 23-3020) HPDF
  - $300-500M program that is a Hub & Spokes model.
  - Deadline for proposals was May 5th 2023
- HPDF will serve as foundational element of the "Integrated Research Infrastructure" (IRI)
- IRI = "federated" infrastructure across DOE labs
- IRI Science Patterns:
  - Time-sensitive, e.g. streaming data, instrument steering, …
  - Data-integration, e.g. analysis of combined data from multiple independent sources, …
  - Long-term campaigns, e.g. large scale archiving and processing for collaborative use and reuse, …

National AI Research Resource

Task force Report from Jan. 2023 suggests:

$2.6B investment over 6 years

Cross Agency Initiative

Figure 2. NAIRR Strategic Objective and Goals

Resources suggested:

Operating Entity to run the program.

Compute Resource to train
1 Trillion parameter model

Smaller scale specialized resources:

- Data resources
- Compute resources
- Computing continuum resources

Remains to be seen if and when NAIRR gets off the ground.

Summary & Conclusions

- **10 year program to support Federation at multiple locations within the vertical stack**
  - From IPMI through K8s to clusters
  - Supporting IPMI & K8s implies a distributed open devOps infrastructure => much more than just federating compute & storage
- DOE sees a need for federated cyberinfrastructure
- USA sees a need to invest into AI Research Resources as an initiative that is coordinated across federal agencies.

Acknowledgements

- 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