“Towards a High-Performance National Research Platform Enabling Digital Research”

Closing Keynote
CNI Spring 2018
San Diego, CA
April 13, 2018

Dr. Larry Smarr
Director, California Institute for Telecommunications and Information Technology
Harry E. Gruber Professor,
Dept. of Computer Science and Engineering
Jacobs School of Engineering, UCSD
http://lsmarr.calit2.net

PACIFIC RESEARCH PLATFORM

Abstract

Research in data-intensive fields is increasingly multi-investigator and multi-institutional, depending on ever more rapid access to ultra-large heterogeneous and widely distributed datasets, which in turn is demanding new technological solutions in visualization, machine learning, and high-performance cyberinfrastructure. I will describe how my NSF-funded Pacific Research Platform (PRP), which provides an Internet platform with 100-1000 times the bandwidth of today's commodity Internet to all the research universities on the West Coast, is being designed from the application needs of researchers. The disciplines which are engaged in partnering with the PRP range from particle physics to climate to human health, as well as archaeology, digital libraries, and social media analysis. The next stage, well underway, is understanding how to scale this prototype cyberinfrastructure to a National and Global Research Platform.

PRP
PACIFIC RESEARCH PLATFORM

it²

30 Years Ago NSF Brought to University Researchers a DOE HPC Center Model

1985/6







NCSA Was Modeled on LLNL

SDSC Was Modeled on MFEnet

PRP PACIFIC RESEARCH PLATFORM

NSF

it²

Thirty Years After NSF Adopts DOE Supercomputer Center Model
NSF Adopts DOE ESnet's Science DMZ for High Performance Applications

- **A Science DMZ integrates 4 key concepts into a unified whole:**
  - **A network architecture designed for high-performance applications, with the science network distinct from the general-purpose network**
  - **The use of dedicated systems as data transfer nodes (DTNs)**
  - **Performance measurement and network testing systems that are regularly used to characterize and troubleshoot the network**
  - **Security policies and enforcement mechanisms that are tailored for high performance science environments**

Science DMZ Coined 2010

The DOE ESnet Science DMZ and the NSF "Campus Bridging" Taskforce Report Formed the Basis for the NSF Campus Cyberinfrastructure Network Infrastructure and Engineering (CC-NIE) Program



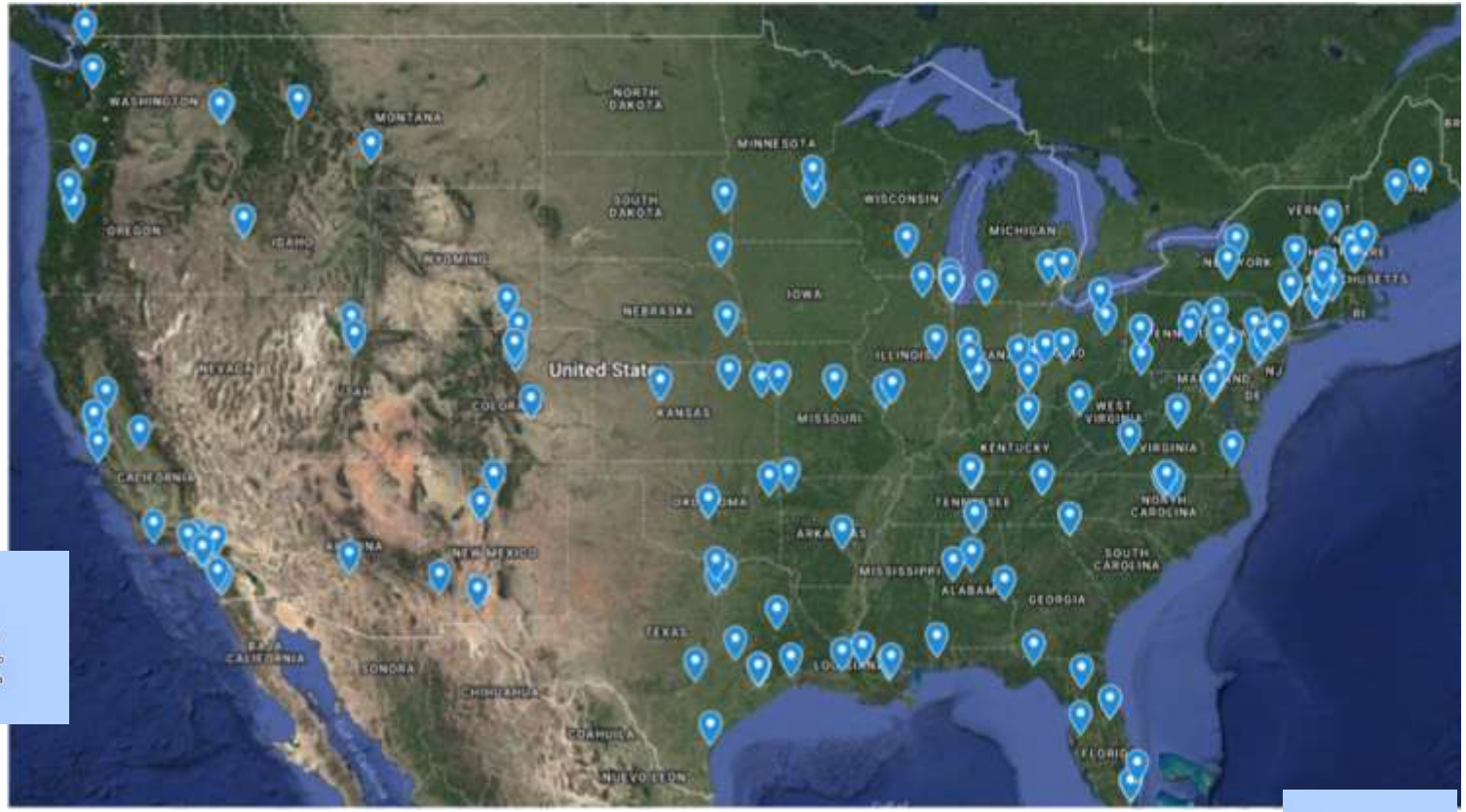
ESnet
FASTERDATA KNOWLEDGEBASE

PRP
PACIFIC RESEARCH PLATFORM

http://fasterdata.es.net/science-dmz/



Based on Community Input and on ESnet's Science DMZ Concept, NSF Has Made Over 200 Campus-Level Awards in 44 States







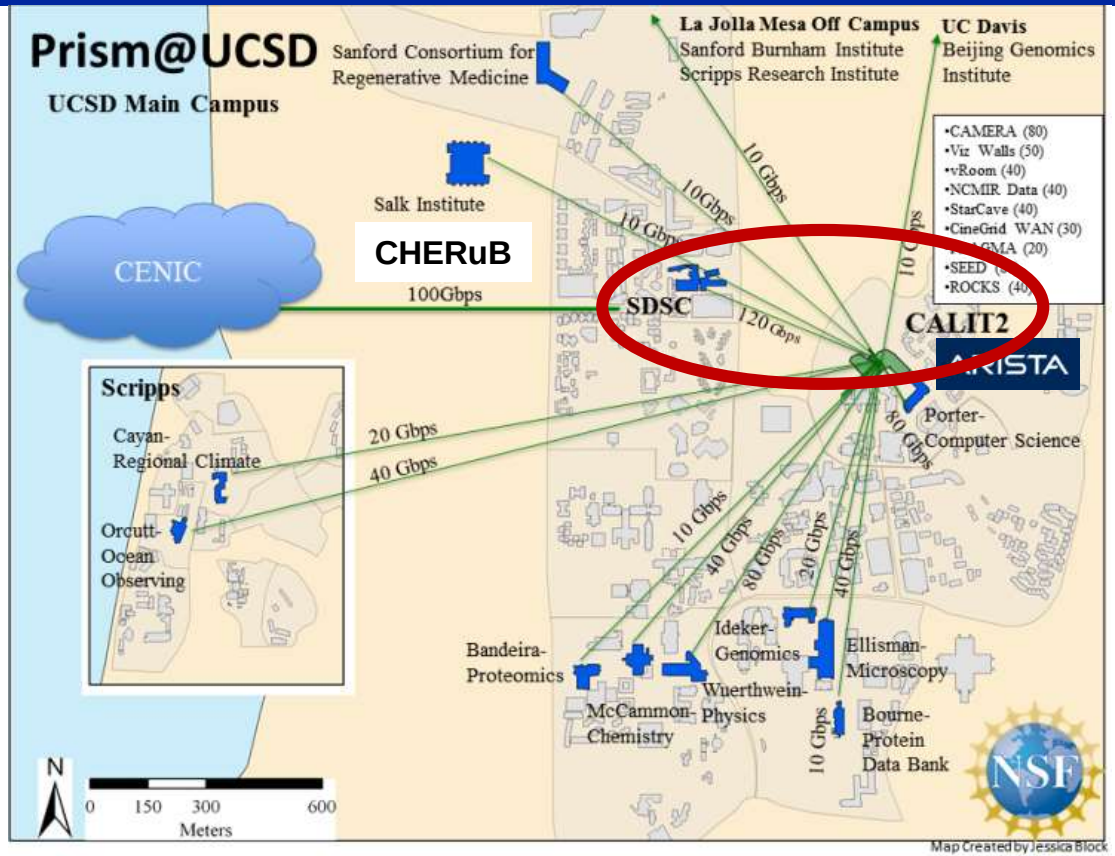
2012-2017 CC-NIE / CC*IIE / CC*DNI PROGRAMS

Source: Kevin Thompson, NSF

PRP PACIFIC RESEARCH PLATFORM

it2

Creating a "Big Data" Freeway on Campus: NSF-Funded CC-NIE Grants Prism@UCSD and CHeruB



Prism@UCSD, Phil Papadopoulos, SDSC, Calit2, PI (2013-15)
CHERuB, Mike Norman, SDSC PI

PRP PACIFIC RESEARCH PLATFORM

SDSC it²

How UCSD DMZ Network Transforms Big Data Microbiome Science: Preparing for Knight/Smarr 1 Million Core-Hour Analysis

NIH

NCBI
National Center for Biotechnology Information

INTERNET2

CENIC

Center for Microbiome Innovation

Knight Lab
FIONA





Gordon

Knight 1024 Cluster
In SDSC Co-Lo





1.3Tbps

CHERuB
100Gbps

Data Oasis
7.5PB,
200GB/s

120Gbps

SDSC

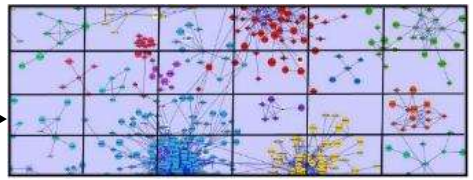
ARISTA

10Gbps

40Gbps



Prism@UCSD



Emperor & Other Vis Tools

64Mpixel Data Analysis Wall



QUALCOMM INSTITUTE

IBM Cloud

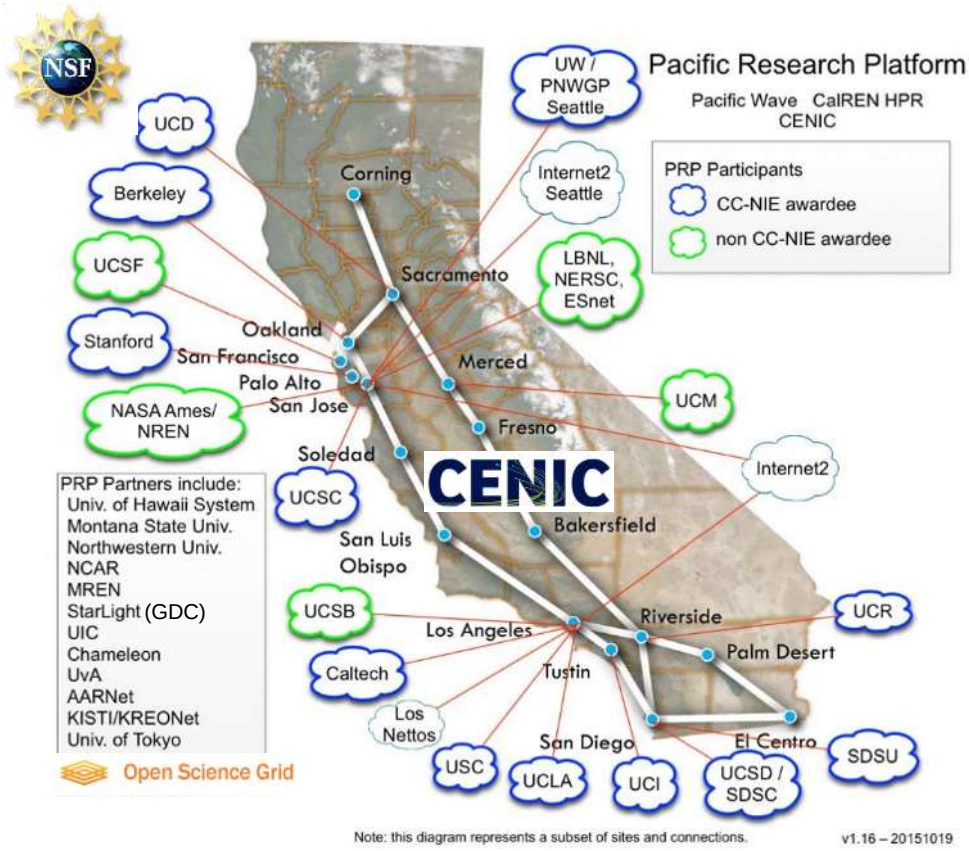
IBM Watson

PRP
PACIFIC RESEARCH PLATFORM

IGM
Institute for Genomic Medicine

it2

Logical Next Step: The Pacific Research Platform Networks Campus DMZs to Create a Regional End-to-End Science-Driven "Big Data Superhighway" System



Source: John Hess, CENIC

NSF CC*DNI Grant
$5M 10/2015-10/2020

PI: Larry Smarr, UC San Diego Calit2

Co-PIs:

- **Camille Crittenden, UC Berkeley CITRIS,**
- **Tom DeFanti, UC San Diego Calit2/QI,**
- **Philip Papadopoulos, UCSD SDSC,**
- **Frank Wuerthwein, UCSD Physics and SDSC**

Letters of Commitment from:

- 50 Researchers from 15 Campuses
- 32 IT/Network Organization Leaders

NSF Program Officer: Amy Walton

PACIFIC RESEARCH PLATFORM

SDSC CITRIS BANATAO INSTITUTE it2

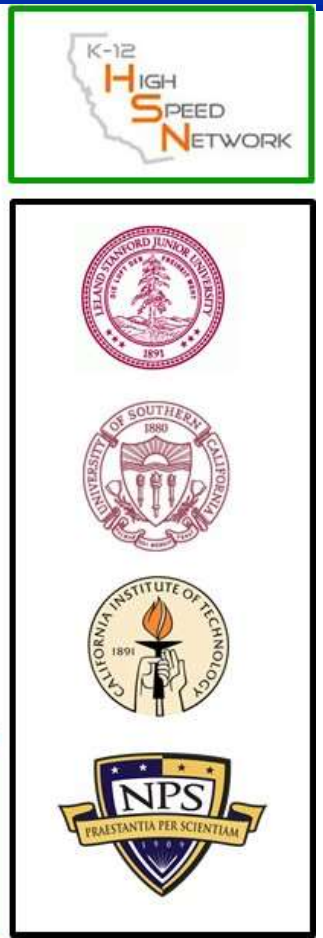
California's Research and Education Network (CENIC) Provides A World-Class Network Driving Innovation, Collaboration, & Economic Growth

- **Charter Associates:**
  - California K-12 System (~10,000)
  - California Community Colleges (114)
  - California State University System (23)
  - Stanford, Caltech, USC
  - University of California (10)
  - California Public Libraries (1150)
  - Naval Postgraduate School

CENIC

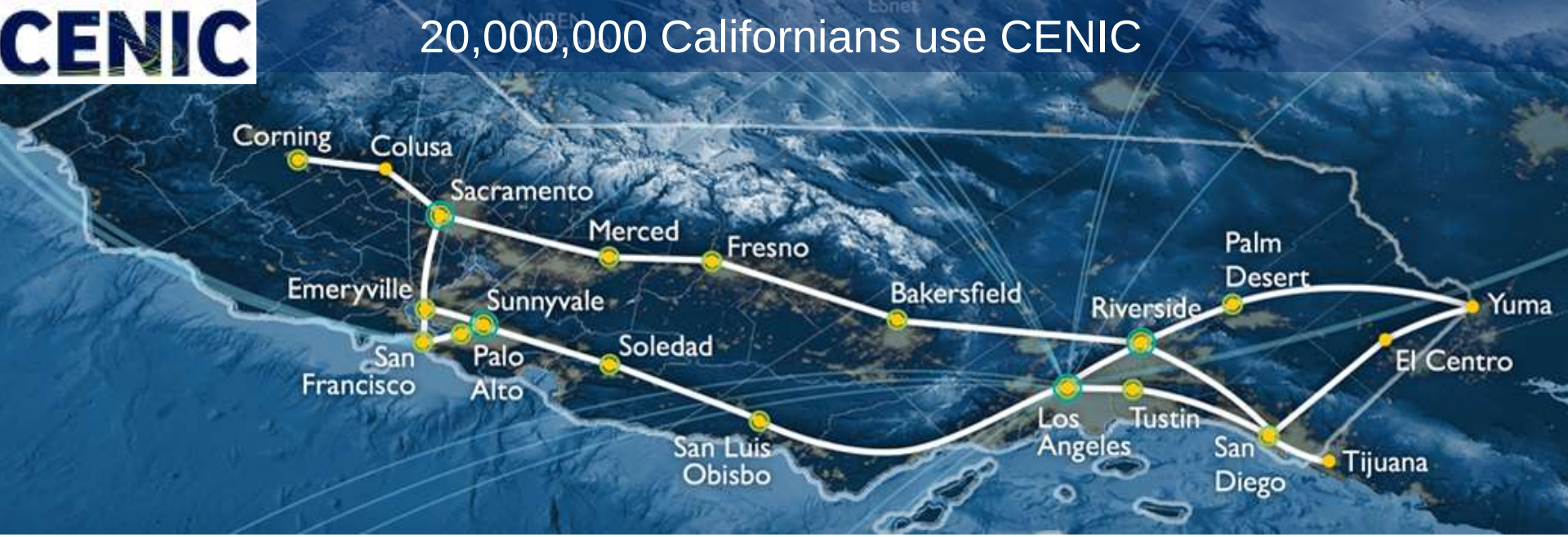
PRP PACIFIC RESEARCH PLATFORM





it²

20,000,000 Californians use CENIC



- **8,000+** miles of optical fiber
- **Members in all 58 counties** connect via fiber-optic cable or leased circuits from telecom carriers
- **Over 12,000 sites** connect to CENIC
- **A non-profit** governed by it's members
- Collaborates with over **750 private sector partners and contributes > $100,000,000** to the CA Economy
- **20 years** of connecting California

Key Innovation: UCSD Designed FIONAs To Solve the Disk-to-Disk Data Transfer Problem at Full Speed on 10/40/100G Networks

UCSD Designed FIONAs
To Solve the Disk-to-Disk
Data Transfer Problem
For Big Data
at Full Speed
on 10G, 40G and 100G Networks



FIONette—1G, $250



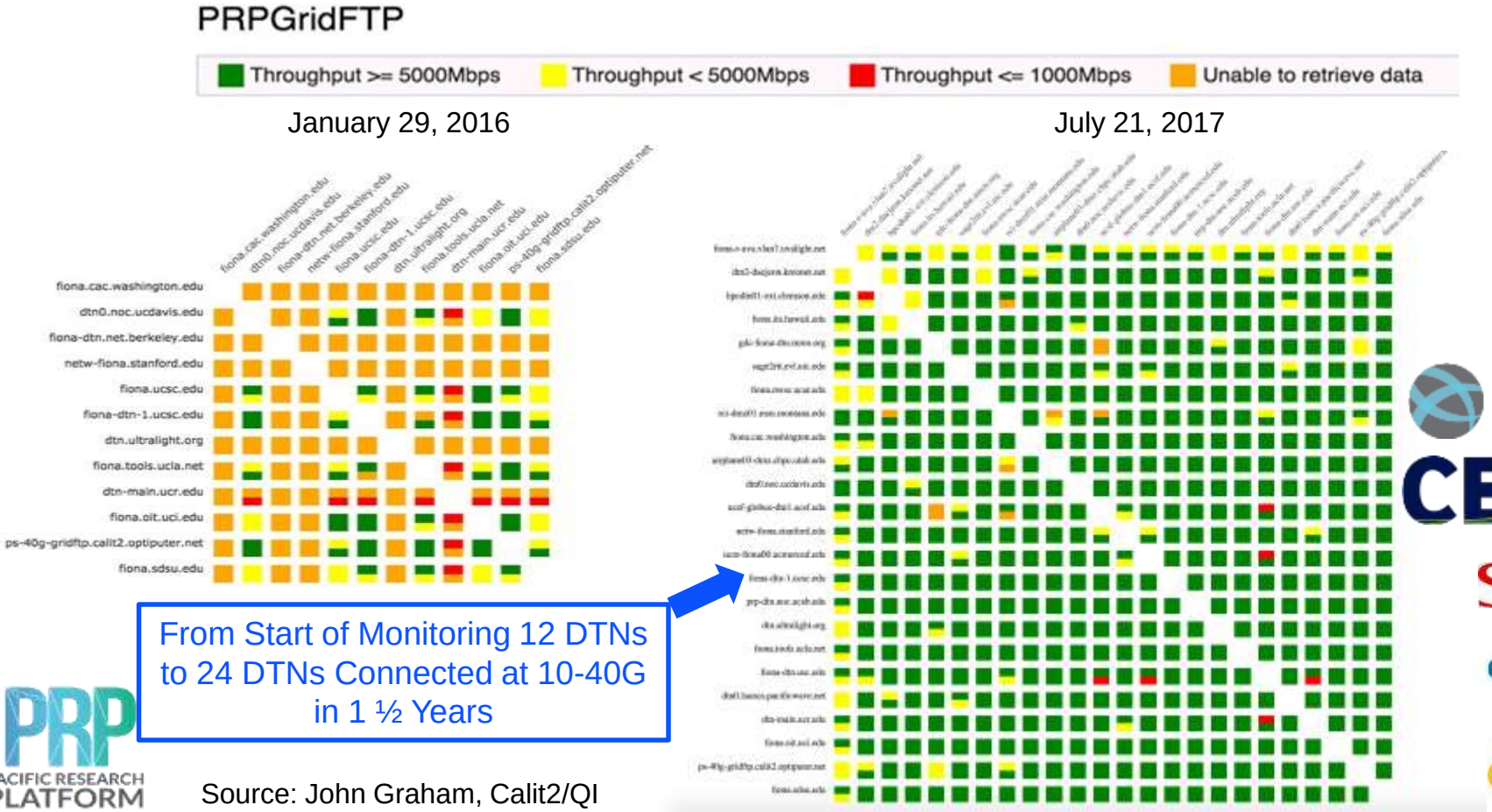
PRP PACIFIC RESEARCH PLATFORM

SDSC

Phil Papadopoulos, SDSC &
Tom DeFanti, Joe Keefe & John Graham, Calit2

it²

We Measure Disk-to-Disk Throughput with 10GB File Transfer 4 Times Per Day in Both Directions for All PRP Sites



From Start of Monitoring 12 DTNs to 24 DTNs Connected at 10-40G in 1 ½ Years

Source: John Graham, Calit2/QI

We Aggressively Use Kubernetes to Manage Containers Across the PRP

CADE METZ BUSINESS 06.10.14 01:15 PM

GOOGLE OPEN SOURCES ITS SECRET WEAPON IN CLOUD COMPUTING

WIRED

"Kubernetes is a way of stitching together
a collection of machines into, basically, a big computer,"
--Craig Mcluckie, Google
and now CEO and Founder of Heptio

"Everything at Google runs in a container."
--Joe Beda,Google

PRP
PACIFIC RESEARCH PLATFORM

CLOUD NATIVE COMPUTING FOUNDATION

How Kubernetes Conquered 2017
(and is Positioned for 2018)





Dan Meyer
December 25, 2017

sdx central

"Kubernetes has emerged as
the container orchestration engine of choice
for many cloud providers including
Google, AWS, Rackspace, and Microsoft,
and is now being used in HPC and Science DMZs.
--John Graham, Calit2/QI UC San Diego

Rook is Ceph Cloud-Native Object Storage ‘Inside’ Kubernetes



Battle-tested, production storage

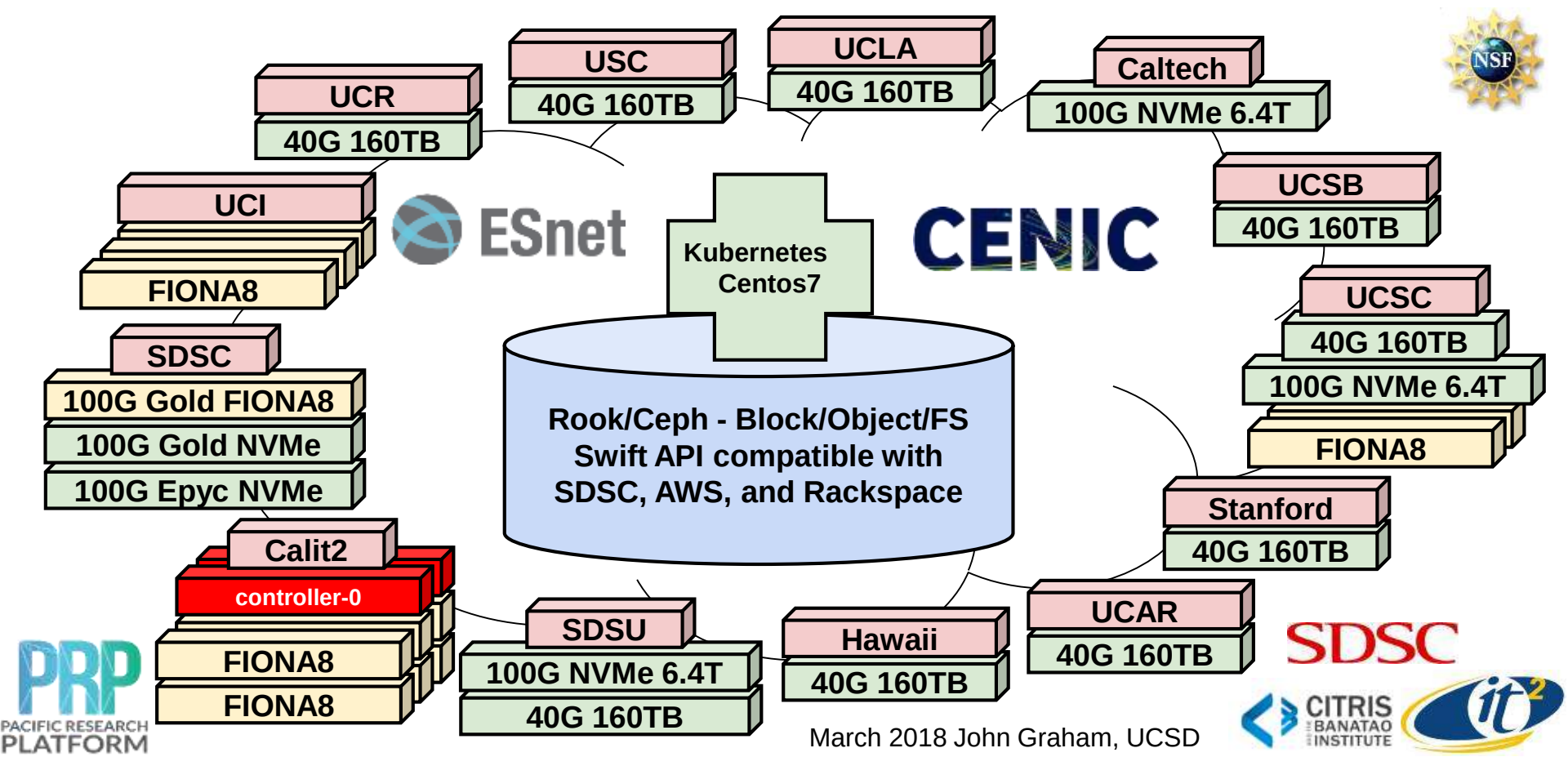
Rook is based on an embedded version of Ceph, which has 10+ years of production deployments and runs some of the worlds largest clusters.

Cloud-native environment integration

Rook runs as a cloud-native service for optimal integration with applications in need of block, object, or file storage.

https://rook.io/

Running Kubernetes/Rook/Ceph On PRP
Allows Us to Deploy a Distributed PB+ of Storage for Posting Science Data



March 2018 John Graham, UCSD

Increasing Participation Through PRP Science Engagement Workshops









Source: Camille Crittenden, UC Berkeley

PRP's First 2 Years:
Connecting Multi-Campus Application Teams and Devices



CMS

Particle Physics

Biomedical 'omics

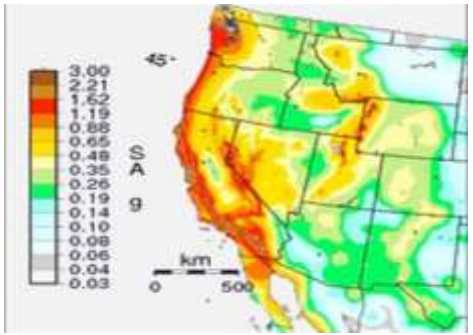




INTERMEDIATE PALOMAR TRANSIENT FACTORY

Telescope Surveys

Earth Sciences

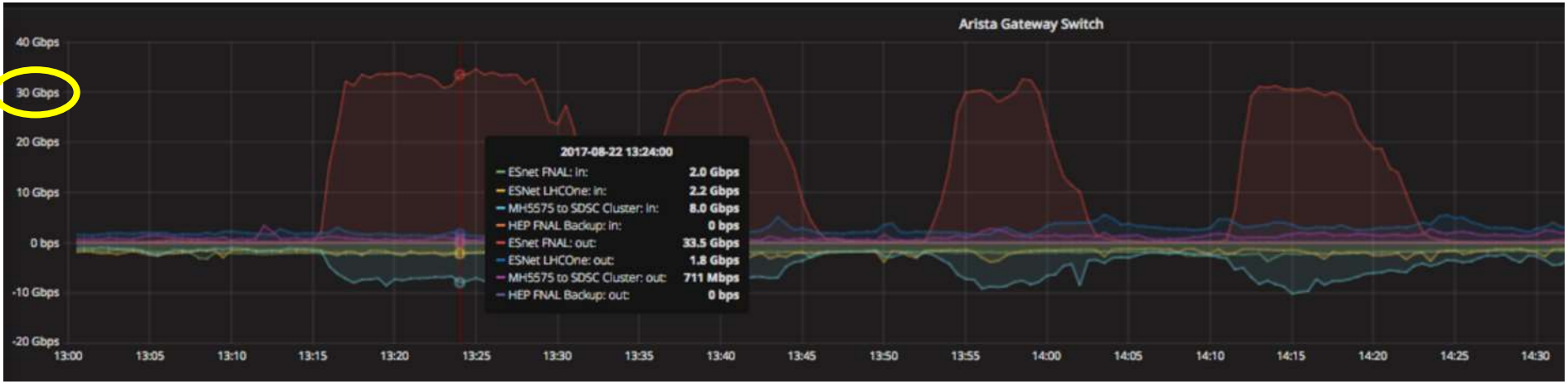


Visualization, Virtual Reality, Collaboration



Data Transfer Rates From 40 Gbps DTN in UCSD Physics Building, Across Campus on PRISM DMZ, Then to Chicago's Fermilab Over CENIC/ESnet

Based on This Success,
Will Upgrade 40G DTN to 100G
For Bandwidth Tests & Kubernetes
to OSG, Caltech, and UCSC

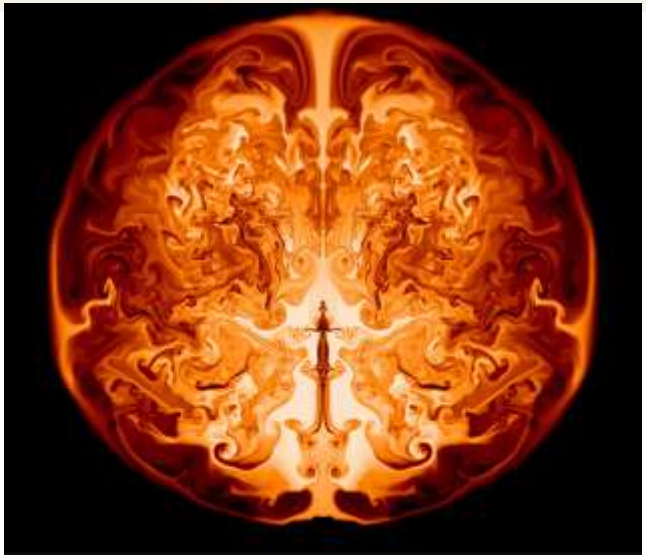


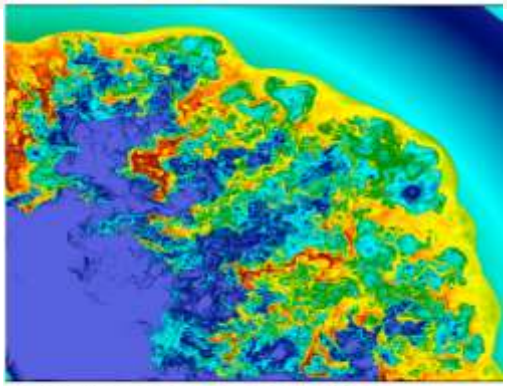
Source: Frank Wuerthwein, UCSD, SDSC

PRP PACIFIC RESEARCH PLATFORM NSF CENIC ESnet SDSC CITRIS BANATAO INSTITUTE it²

PRP Over CENIC
Couples UC Santa Cruz Astrophysics Cluster to LBNL NERSC Supercomputer

CENIC recognizes UCSC's Hyades supercomputer cluster connection





A 2D superluminous supernova simulation generated with the CASTRO code. (Image: Ken Chen, National Astronomical Observatory of Japan)

K. Chen, S. Woolsey, T. Sukhbold, The Astrophysical Journal, 832, 1, Nov. 2016

CENIC 2018 Innovations in Networking Award for Research Applications

PRP PACIFIC RESEARCH PLATFORM

BERKELEY LAB Bringing Science Solutions to the World

NERSC

NERSC Project PI:
S. Woolsey, UC Santa Cruz

SDSC

CITRIS BANATAO INSTITUTE

it²

100 Gbps FIONA at UCSC Allows for Downloads to the UCSC Hyades Cluster from the LBNL NERSC Supercomputer for DESI Science Analysis



INTERMEDIATE PALOMAR TRANSIENT FACTORY

300 images per night.
100MB per raw image

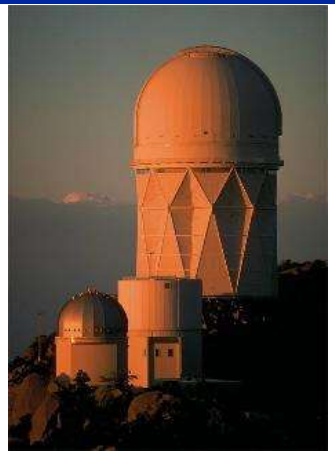
120GB per night

Source: Peter Nugent, LBNL
Professor of Astronomy, UC Berkeley

Precursors to LSST and NCSA



NSF-Funded Cyberengineer
Shaw Dong @UCSC
Receiving FIONA
Feb 7, 2017



Dark Energy Spectroscopic Instrument

250 images per night.
530MB per raw image

800GB per night

BERKELEY LAB
Bringing Science Solutions to the World

NERSC

PRP PACIFIC RESEARCH PLATFORM

NSF

CENIC

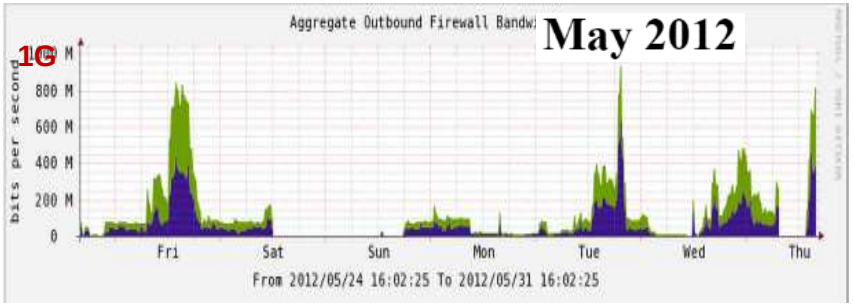
ESnet

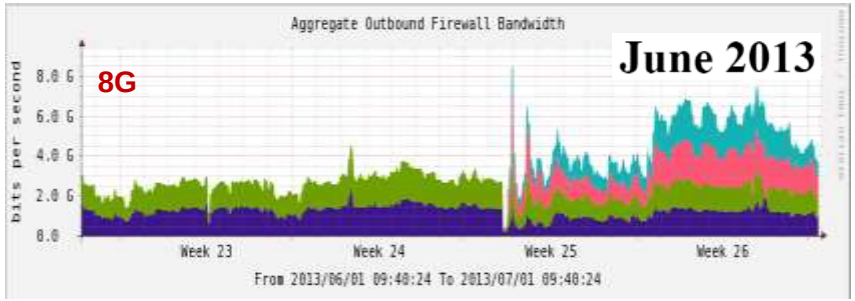
SDSC

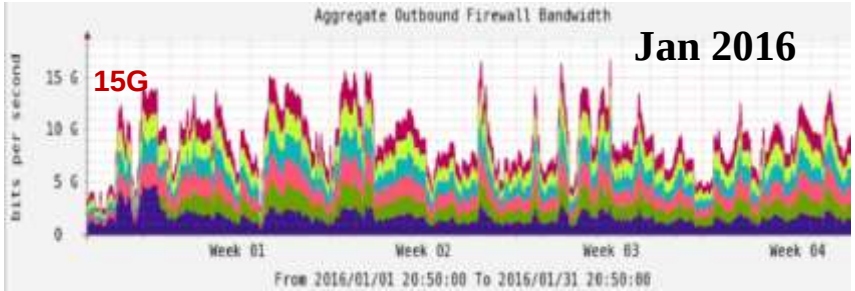
CITRIS BANATAO INSTITUTE

it²

Cancer Genomics Hub (UCSC) Was Housed in SDSC, But NIH Moved Dataset From SDSC to Uchicago - So the PRP Deployed a FIONA to Chicago's MREN









Genomic Data Commons heralds new era of data sharing for cancer research

June 6, 2016

UChicagoNews

Data Source: David Haussler, Brad Smith, UCSC

PRP PACIFIC RESEARCH PLATFORM

it2

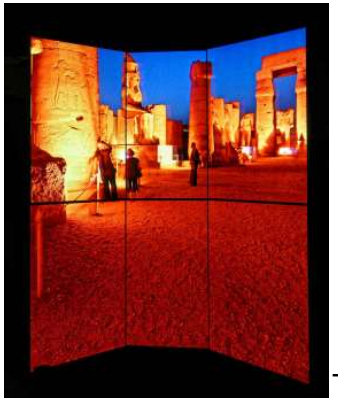
The Prototype PRP Has Attracted New Application Drivers



Frank Vernon, Graham Kent, & Ilkay Altintas, Wildfires

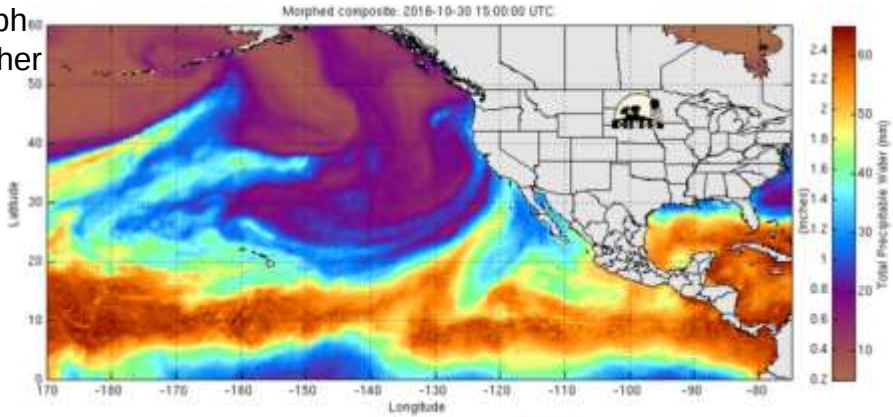


Jules Jaffe – Undersea Microscope



Tom Levy At-Risk Cultural Heritage

Scott Sellars, Marty Ralph
Center for Western Weather and Water Extremes



jupyterhub

PRP PACIFIC RESEARCH PLATFORM

NSF

CENIC

ESnet

SDSC

CITRIS BANATAO INSTITUTE

it²

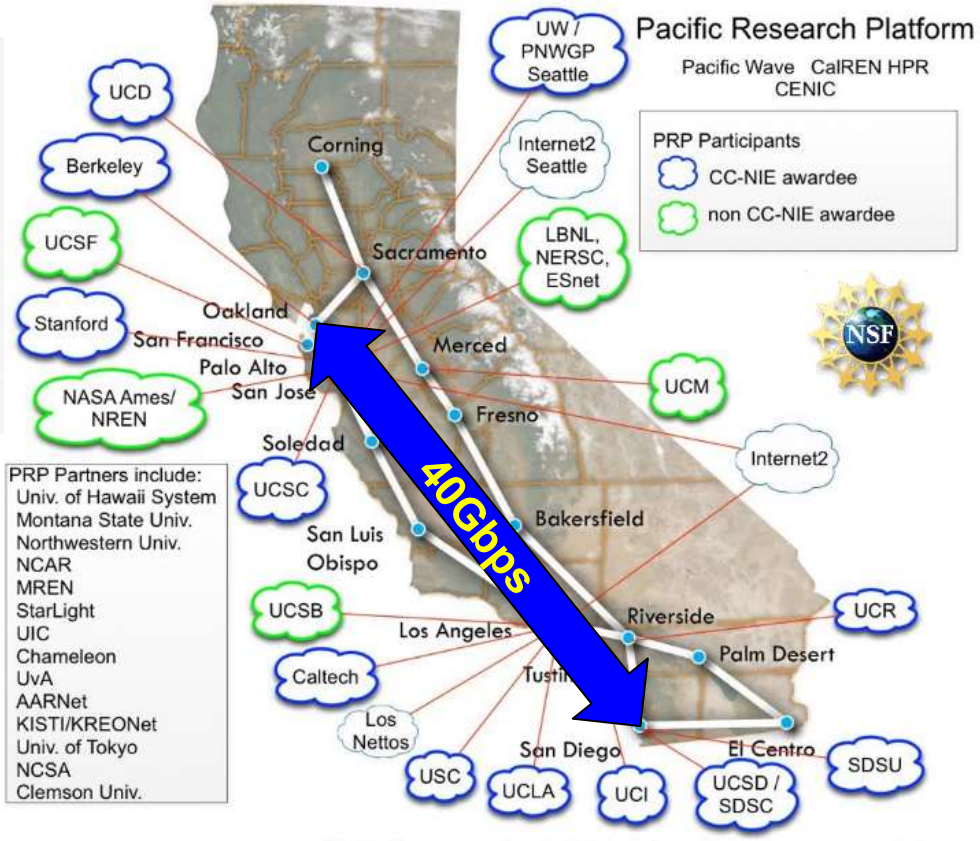
Jupyter Has Become the Digital Fabric for Data Sciences
PRP Creates UC-JupyterHub Backbone

Source: John Graham, Calit2

jupyterhub

A multi-user version of the notebook designed for companies, classrooms and research labs

Goal: Jupyter Everywhere



PRP PACIFIC RESEARCH PLATFORM

ESnet

CENIC

CITRIS BANATAO INSTITUTE

it²

SDSC

PRP Links At-Risk Cultural Heritage and Archaeology Datasets at UCB, UCLA, UCM and UCSD with CAVEkiosks

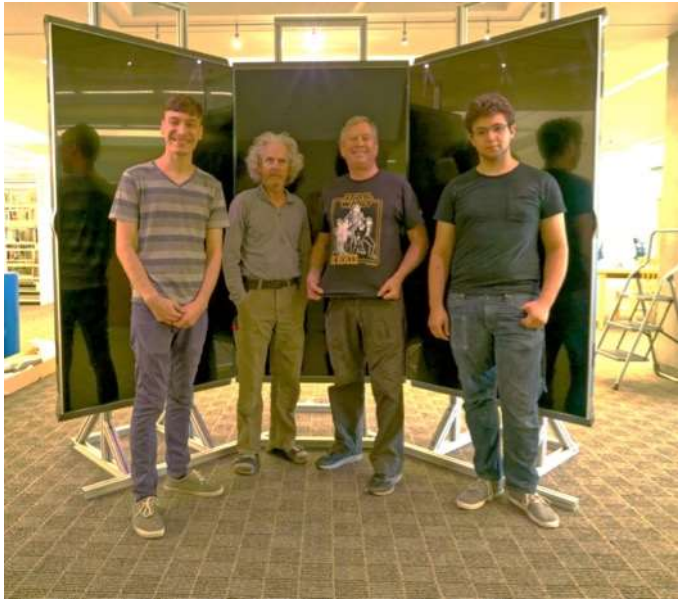
UC President Napolitano's Research Catalyst Award to UC San Diego (Tom Levy), UC Berkeley (Benjamin Porter), UC Merced (Nicola Lercari) and UCLA (Willeke Wendrich)



48 Megapixel CAVEkiosk
UCSD Library



48 Megapixel CAVEkiosk
UCB Library



24 Megapixel CAVEkiosk
UCM Library

PRP PACIFIC RESEARCH PLATFORM

NSF

CENIC

SDSC

CITRIS BANATAO INSTITUTE

it²

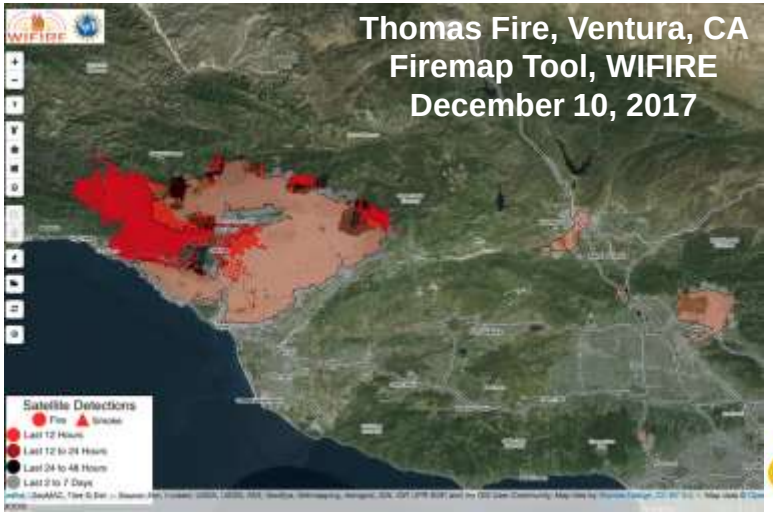
New PRP Application:
Coupling Wireless Wildfire Sensors to Computing

Technology Projects to Combat California Wildfires Are Recognized with CENIC Innovation Award

WIFIRE HPWREN Nevada Seismological Laboratory
Mackay School of Earth Sciences & Engineering
UNIVERSITY OF NEVADA, RENO

CENIC 2018
Innovations in Networking Award for Experimental Applications





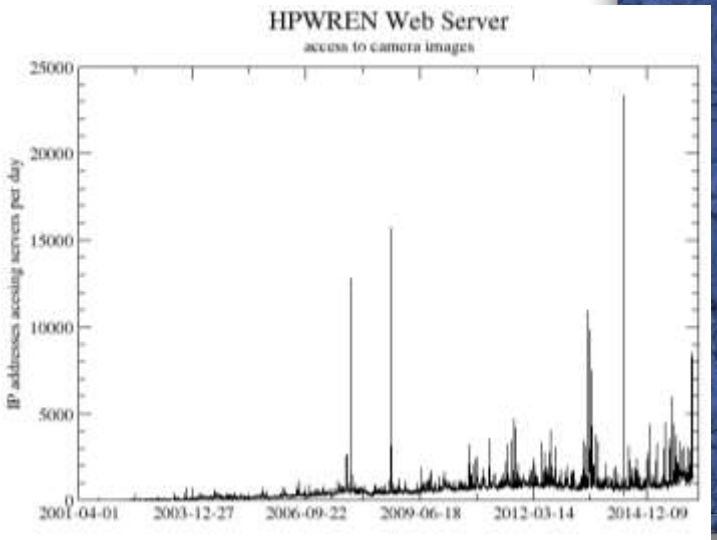
PRP PACIFIC RESEARCH PLATFORM

it²

CENIC/PRP Backbone Sets Stage for 2017 Wireless Expansion of HPWREN into Orange and Possibly Riverside Counties

- **CENIC/PRP Will Connect UCSD and SDSU**
  - **Data Redundancy**
  - **Disaster Recovery**
  - **High Availability**
- **CENIC Extension to UCI & UCR**
  - **Data Replication Sites**





SDSC

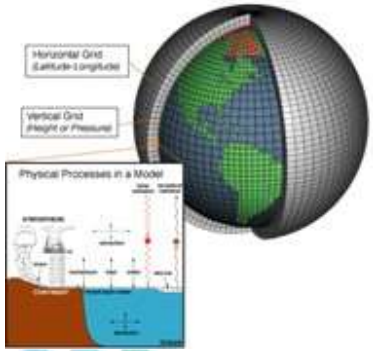
it²

Once a Wildfire is Spotted, PRP Brings High-Resolution Weather Data to Fire Modeling Workflows in WIFIRE



Real-Time Meteorological Sensors

Weather Forecast



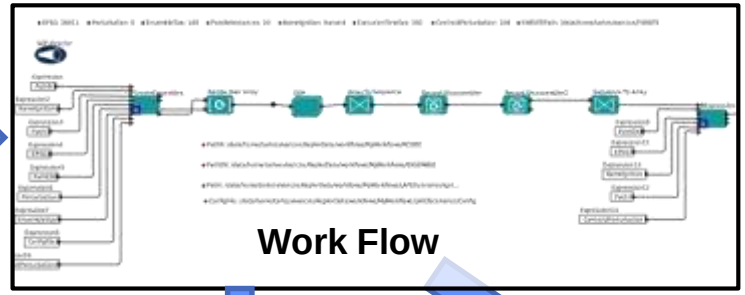
PRP

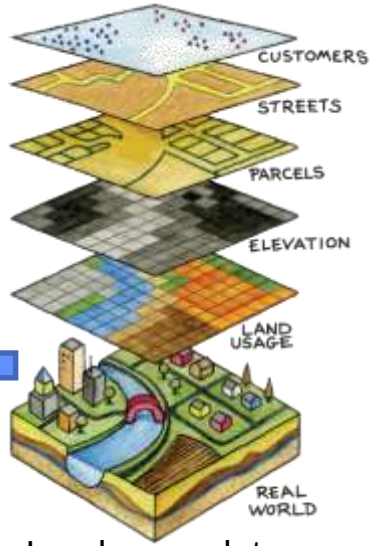


Source: Ilkay Altintas, SDSC

WIFIRE







Landscape data

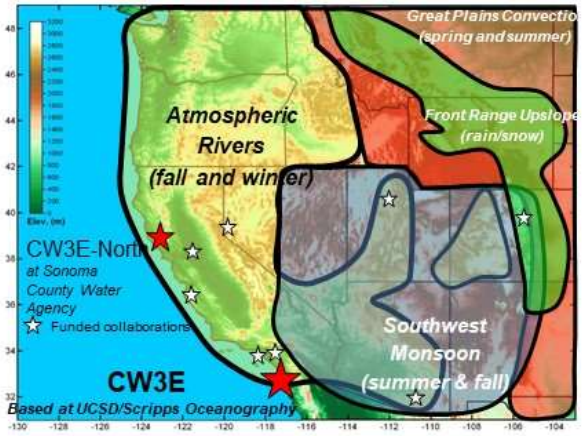




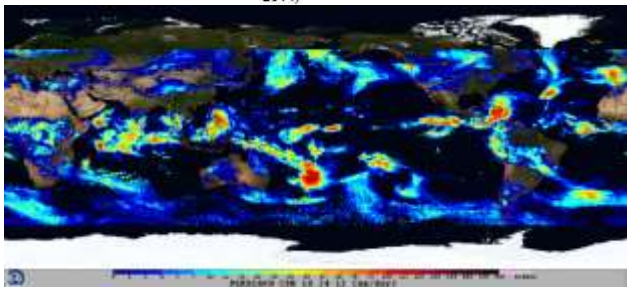


PRP
PACIFIC RESEARCH PLATFORM

Collaboration on Atmospheric Water in the West Between UC San Diego and UC Irvine



Key Phenomena Causing Extreme Precipitation in the Western U.S. (Ralph et al. 2014)



Center for Western Weather and Water Extremes
SCRIPPS INSTITUTION OF OCEANOGRAPHY AT UC SAN DIEGO

Director: F. Martin Ralph Website: cw3e.ucsd.edu

Big Data Collaboration with:

Center for Hydrometeorology & Remote Sensing
University of California, Irvine

Director, Soroosh Sorooshian, UCSD Website http://chrs.web.uci.edu

PACIFIC RESEARCH PLATFORM

Source: Scott Sellers, CW3E

Major Speedup in Scientific Work Flow Using the PRP

Pacific Research Platform (10-100 Gb/s)

Complete workflow time: 20 days→20 hrs→20 Minutes!

UC, Irvine

GPUs



Calit2's FIONA

SDSC's COMET



Source: Scott Sellers, CW3E

UC, San Diego

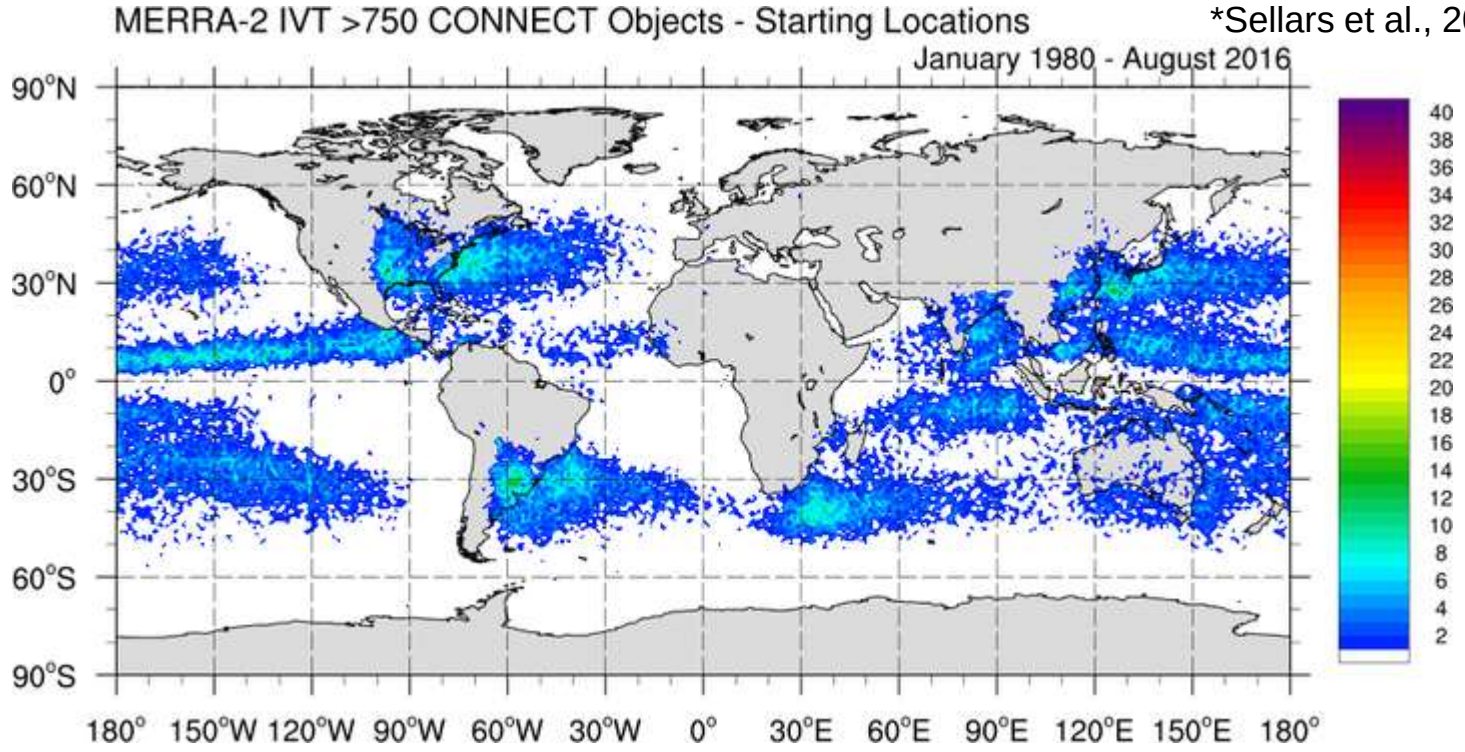
GPUs



Calit2's FIONA

Using Machine Learning to Determine the Precipitation Object Starting Locations

*Sellars et al., 2017 (in prep)



PRP PACIFIC RESEARCH PLATFORM

NSF

CENIC

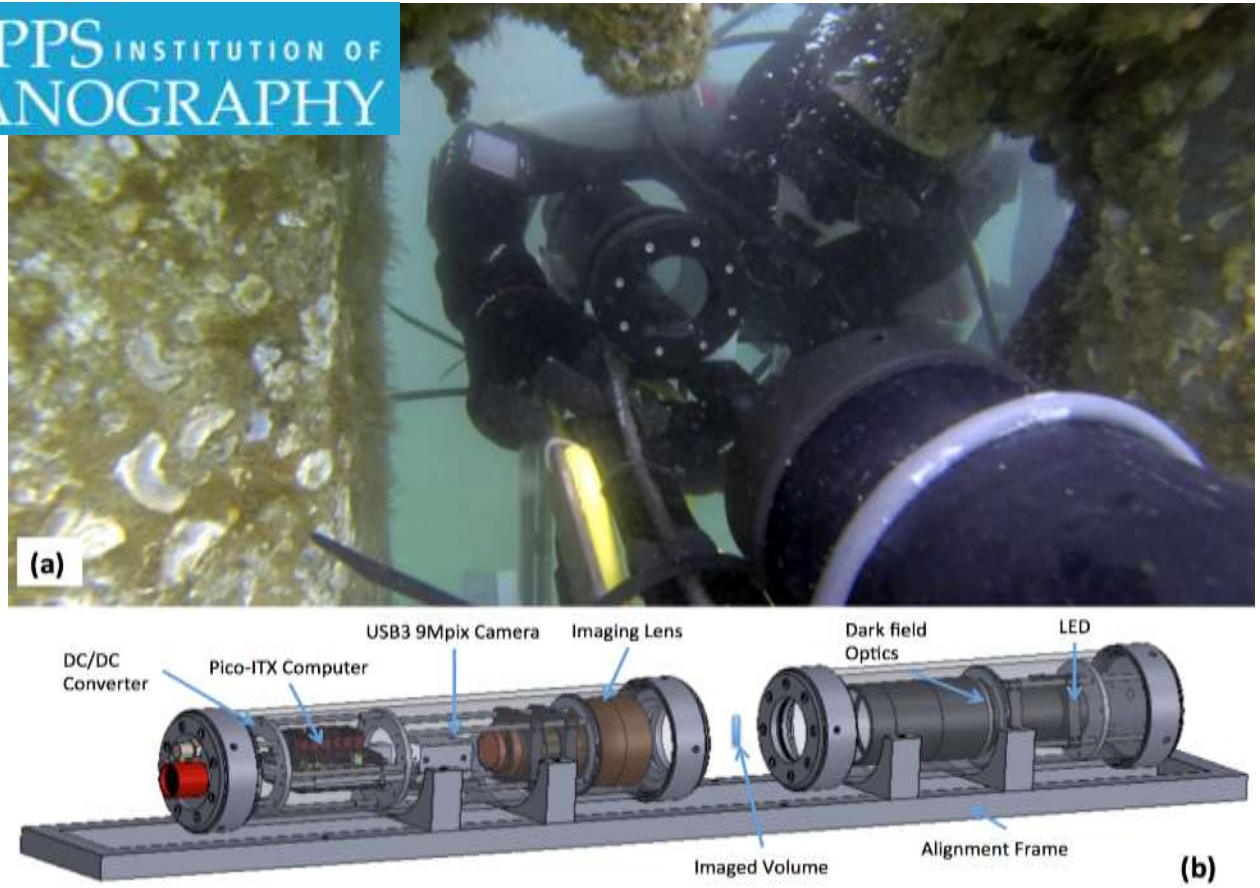
ESnet

CITRIS THE BANATAO INSTITUTE

it²

UC San Diego Jaffe Lab (SIO) Scripps Plankton Camera Off the SIO Pier with Fiber Optic Network

SCRIPPS INSTITUTION OF OCEANOGRAPHY

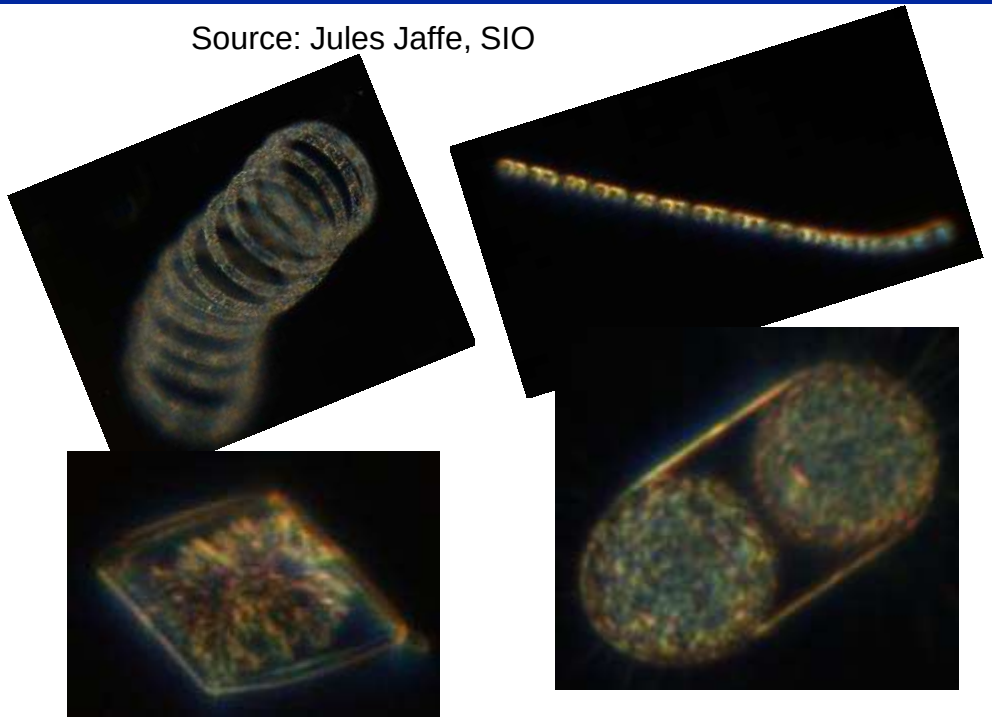


PRP PACIFIC RESEARCH PLATFORM

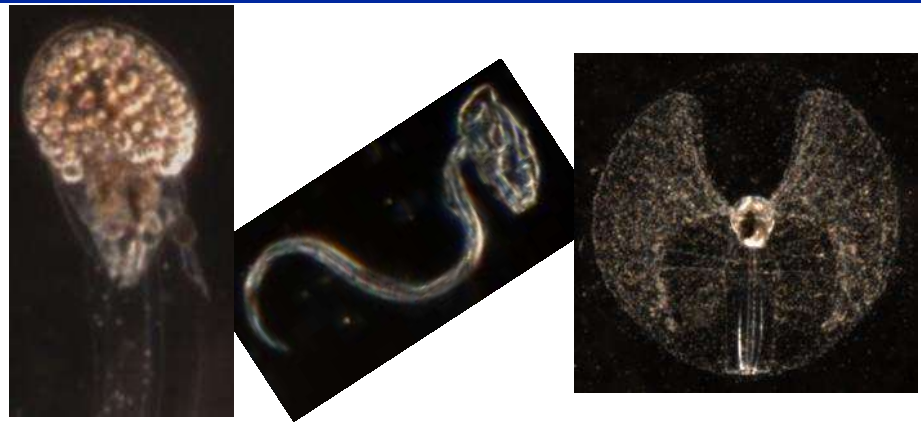
it²

Over 300 Million Images So Far!
Requires Machine Learning for Automated Image Analysis and Classification

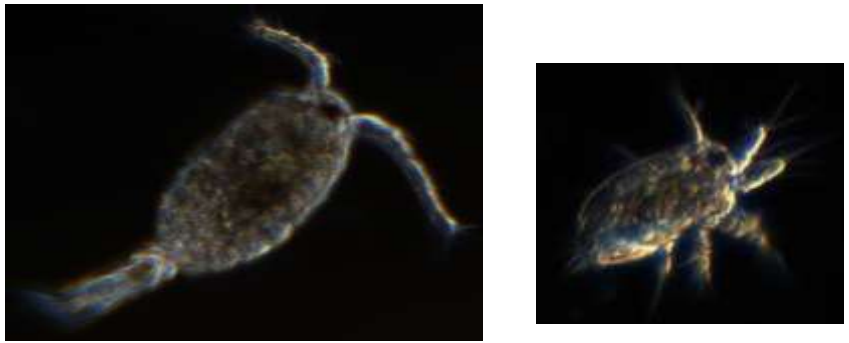
Source: Jules Jaffe, SIO



Phytoplankton: Diatoms



Zooplankton: Larvaceans

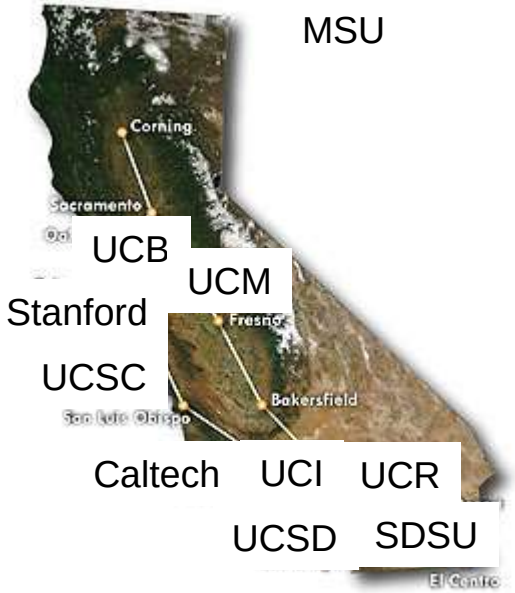


Zooplankton: Copepods

”We are using the FIONAs for image processing... this includes doing Particle Tracking Velocimetry that is very computationally intense.”-Jules Jaffe

PACIFIC RESEARCH PLATFORM

New NSF CHASE-CI Grant Creates a Community Cyberinfrastructure: Adding a Machine Learning Layer Built on Top of the Pacific Research Platform



CI-New: Cognitive Hardware and Software Ecosystem Community Infrastructure (CHASE-CI)

For the Period September 1, 2017 – August 31, 2020

SUBMITTED – January 18, 2017

PI: Larry Smarr, Professor of Computer Science and Engineering, Director Calit2, UCSD
Co-PI: Tajana Rosing, Professor of Computer Science and Engineering, UCSD
Co-PI: Ken Kreutz-Delgado, Professor of Electrical and Computer Engineering, UCSD
Co-PI: Ilkay Altintas, Chief Data Science Officer, San Diego Supercomputer Center, UCSD
Co-PI: Tom DeFanti, Research Scientist, Calit2, UCSD

NSF Program Officer: Mimi McClure

NSF Grant for High Speed "Cloud" of 256 GPUs
For 30 ML Faculty & Their Students at 10 Campuses
for Training AI Algorithms on Big Data

PRP PACIFIC RESEARCH PLATFORM

JACOBS SCHOOL OF ENGINEERING

SDSC UC San Diego it²

FIONA8: Adding GPUs to FIONAs
Supports Data Science Machine Learning



Multi-Tenant Containerized GPU JupyterHub
Running Kubernetes / CoreOS
Eight Nvidia GTX-1080 Ti GPUs

~$13K

32GB RAM, 3TB SSD, 40G & Dual 10G ports

Source: John Graham, Calit2

PACIFIC RESEARCH PLATFORM

Brain-Inspired Processors
Are Accelerating the Non-von Neumann Architecture Era





Mar. 29, 2016

Lawrence Livermore and IBM collaborate to build new brain-inspired supercomputer

Lawrence Livermore National Laboratory

Air Force Enlists IBM Research to Build Neuromorphic Supercomputer

Michael Feldman | June 26, 2017 14:01 CEST

"On the drawing board are collections of 64, 256, 1024, and 4096 chips.
'It's only limited by money, not imagination,' Modha says."

Source: Dr. Dharmendra Modha
IBM Chief Scientist for Brain-inspired Computing

PRP
PACIFIC RESEARCH PLATFORM

it²

Calit2's Qualcomm Institute Has Established a Pattern Recognition Lab For Machine Learning on GPUs and von Neumann and NvN Processors



UCSD ECE Professor Ken Kreutz-Delgado Brings the IBM TrueNorth Chip to Start Calit2's Qualcomm Institute Pattern Recognition Laboratory September 16, 2015

Source: Dr. Dharmendra Modha
Founding Director, IBM Cognitive Computing Group

PRP PACIFIC RESEARCH PLATFORM

it²

Our Pattern Recognition Lab is Exploring Mapping Machine Learning Algorithm Families Onto Novel Architectures

- Deep & Recurrent Neural Networks (DNN, RNN)
- Graph Theoretic
- Reinforcement Learning (RL)
- Clustering and other neighborhood-based
- Support Vector Machine (SVM)
- Sparse Signal Processing and Source Localization
- Dimensionality Reduction & Manifold Learning
- Latent Variable Analysis (PCA, ICA)
- Stochastic Sampling, Variational Approximation
- Decision Tree Learning

PACIFIC RESEARCH PLATFORM

Qualcomm Institute

UCSD Adding >350 Game GPUs to Data Sciences Cyberinfrastructure - Devoted to Data Analytics and Machine Learning

UC San Diego

IT SERVICES

88 GPUs
for Students

UCSD Cognitive Science

JACOBS SCHOOL OF ENGINEERING

Open Science Grid

48 GPUs for
OSG Applications

SDSC



SunCAVE 70 GPUs
WAVE + Vroom 48 GPUs

QUALCOMM INSTITUTE

CHASE-CI Grant Provides
96 GPUs at UCSD
for Training AI Algorithms on Big Data

PRP
PACIFIC RESEARCH PLATFORM

Next Step: Surrounding the PRP Machine Learning Platform With Clouds of GPUs and Non-Von Neumann Processors

IBM



64-TrueNorth Cluster

IBM Cloud

IBM Watson

64-bit GPUs

XSEDE

Extreme Science and Engineering Discovery Environment

NSF





Microsoft Installs Altera FPGAs into Bing Servers & 384 into TACC for Academic Access

Project Catapult

Established: February 2, 2015

Microsoft

KNL EDGE

HIGHLY SCALABLE NEURAL COMPUTING



CHASE-CI

PRP

PACIFIC RESEARCH PLATFORM

Google's First Machine Learning Chip (TPU) Is 30x Faster Than CPUs And GPUs

They didn't have to make a lot of new data centers.

BUILDING AN AI CHIP SAVED GOOGLE FROM BUILDING A DOZEN NEW DATA CENTERS

WIRED



SEP 30, 2016 @ 07:39 PM

Marco Chiappetta,

Forbes / Tech

Amazon Boosts Cloud-Computing Performance With New, GPU-Accelerated AWS Instances

amazon web services

Japan plans to build the fastest deep learning supercomputer

AIST

ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

4352x NVIDIA Tesla V100 GPUs

it2

PRP Hosted
The First National Research Platform Workshop on August 7-8, 2017



150 Attendees

Co-Chairs:
Larry Smarr, Calit2
& Jim Bottum, Internet2

INTERNET2 2017 global SUMMIT April 23-26 Washington DC

Announced in I2 Closing Keynote:
Larry Smarr "Toward a National Big Data Superhighway"
on Wednesday, April 26, 2017

PRP PACIFIC RESEARCH PLATFORM NSF CENIC ESnet CITRIS BANATAO INSTITUTE it2

The Second National Research Platform Workshop
Bozeman, MT August 6-7, 2018

Local Host: Jerry Sheehan, MSU



Co-Chairs:
Larry Smarr, Calit2
Inder Monga, ESnet
Ana Hunsinger, Internet2

A follow-up FIONA workshop will be held as a lead into the 2nd NRP workshop in Bozeman, starting August 2nd.

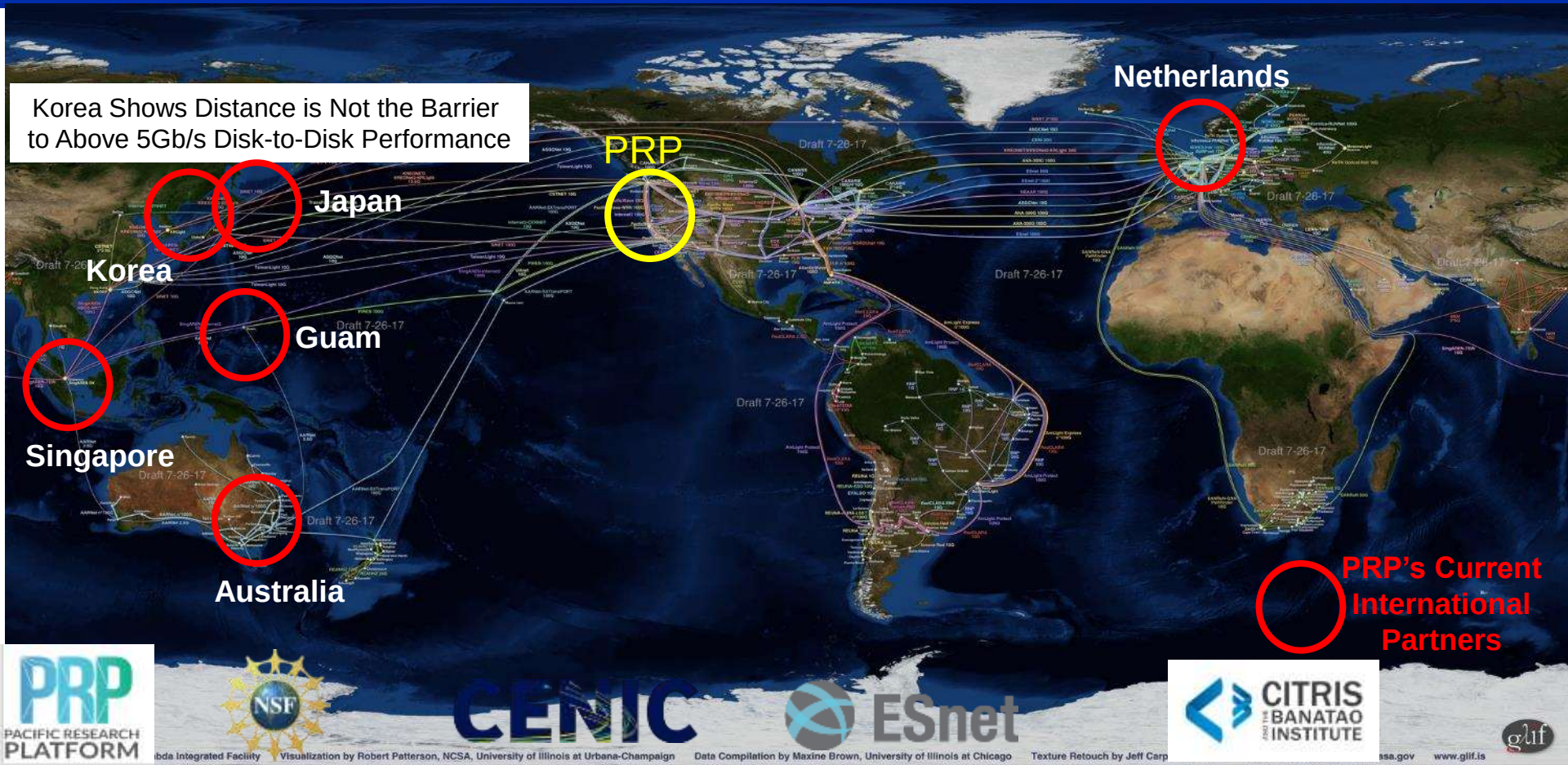
The program is being developed by Jerry Sheehan, in coordination with Richard Alo (JSU) and will focus on networking engineers and faculty interested in expanding the breadth of the NRP network.

While the workshop will be open to the community, there is a specific focus on EPSCoR affiliated and minority serving institutions.

PRP PACIFIC RESEARCH PLATFORM

SDSC CITRIS BANATAO INSTITUTE it2

Expanding to the Global Research Platform
Via CENIC/Pacific Wave, Internet2, and International Links



Many Open Research Questions for This Tightly Coupled Distributed "Computer" for Big Data Analysis

How To:

- Enable Data Discovery, Annotation, Curation
- Provide Both Working Data Storage and Archiving
- Encourage Application Teams to Adopt It?
- Strengthen Cybersecurity
- Tightly Integrate Cloud Providers
- Scale Both Technically and Socially?
- Plus Many More…

PRP PACIFIC RESEARCH PLATFORM

it²

Our Support:

- US National Science Foundation (NSF) awards
  - CNS 0821155, CNS-1338192, CNS-1456638, CNS-1730158, ACI-1540112, & ACI-1541349
- University of California Office of the President CIO
- UCSD Chancellor's Integrated Digital Infrastructure Program
- UCSD Next Generation Networking initiative
- Calit2 and Calit2 Qualcomm Institute
- CENIC, PacificWave and StarLight
- DOE ESnet

PRP PACIFIC RESEARCH PLATFORM

it²