"High Performance Cyberinfrastructure Is Required for the Era of Big Data"

Opening Workshop Presentation

"Whither Science in Mexico: an Analysis for Action from the Academic, Industry and Technology."

Held at CICESE, Ensenada, Mexico

March 14, 2013

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

A Ten Year Journey : Creating a 10Gbps Optical Fiber Link Between UCSD and CICESE

- UCSD Meeting on Joint CICESE/Calit2 Proposal Sept 2002
- SDSU's Eric Frost Talk at CUDI Meeting at CICESE April 2003
- Arzberger PRAGMA Talk-CUDI in Puebla, Mexico October 2003
- Visit by CICESE and CONACYT to Calit2 Jan 2004
- Visit by Calit2 and OptIPuter to CICESE March 2004
- Visit by CICESE and CONACYT to Calit2 AHM April 2004

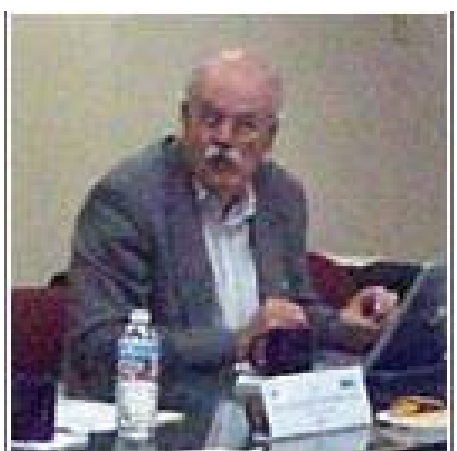




Jaime Parada, Felipe Rubio, & Carlos Duarte at the Calit2 All Hands Meeting



Maxine Brown, Larry Smarr, Javier Mendieta, Tom DeFanti



Carlos Casasús, director of CUDI, Mexico's Internet2

CUDI-CENIC Fiber Dedication at Border Governor's Conference, July 14, 2005



US Mexico

Torreon Conference---Fiber Dedication Linking Mexico and US, crossing at San Diego-Tijuana



- **Shared Security**
- **Energy**
- **Trans-National Crime**
- **Education and Research**
- **Business Development**

Culmination of Three Years of Work Between Calit2, CICESE, CENIC, and CUDI



Success in First Phase—
OptIPortal is Installed in CICESE—First in Mexico



CICESE, Mexico

September 19, 2008

The Final Push

- LS OptIPuter Talk at CUDI Fall Meeting 2006 San Luis Potosi
- Dec 2006 CONACYT Funds Calit2/Mexico Collaborations
- 2006-7 Calit2 Sets up Funding Contracts for CUDI and CICESE
- 2007-8 CICESE Gets Training in Visualization from Calit2
- 2007 Visits Between Calit2 and CICESE
- Sept 2008 CICESE Constructs OptIPortal
- 2009-11 Investigations of Networking Possibilities
- 2011 CONACYT Letter Directing Calit2 to Work with CUDI
- 2011 CUDI Negotiates Multi-Year Networking Agreement with Televisa/BESTEL
- Feb 2012 NSF IRNC Upgrade of Cross-Border from 1G to 10G
- March 2012 First Light Calit2→Tijuana→CICESE
- March 2013 CENIC 2013 Meeting CICESE/Calit2 Demo

CONACYT

AMLIGHT RESEARCH · EDUCATION COLLABORATION

CICESE 40 Aniversario

cudi

Televisa

Bestel

cenic Corporation for Education Network Initiatives in California

NSF

it²

Accepting the Award
CENIC 2012



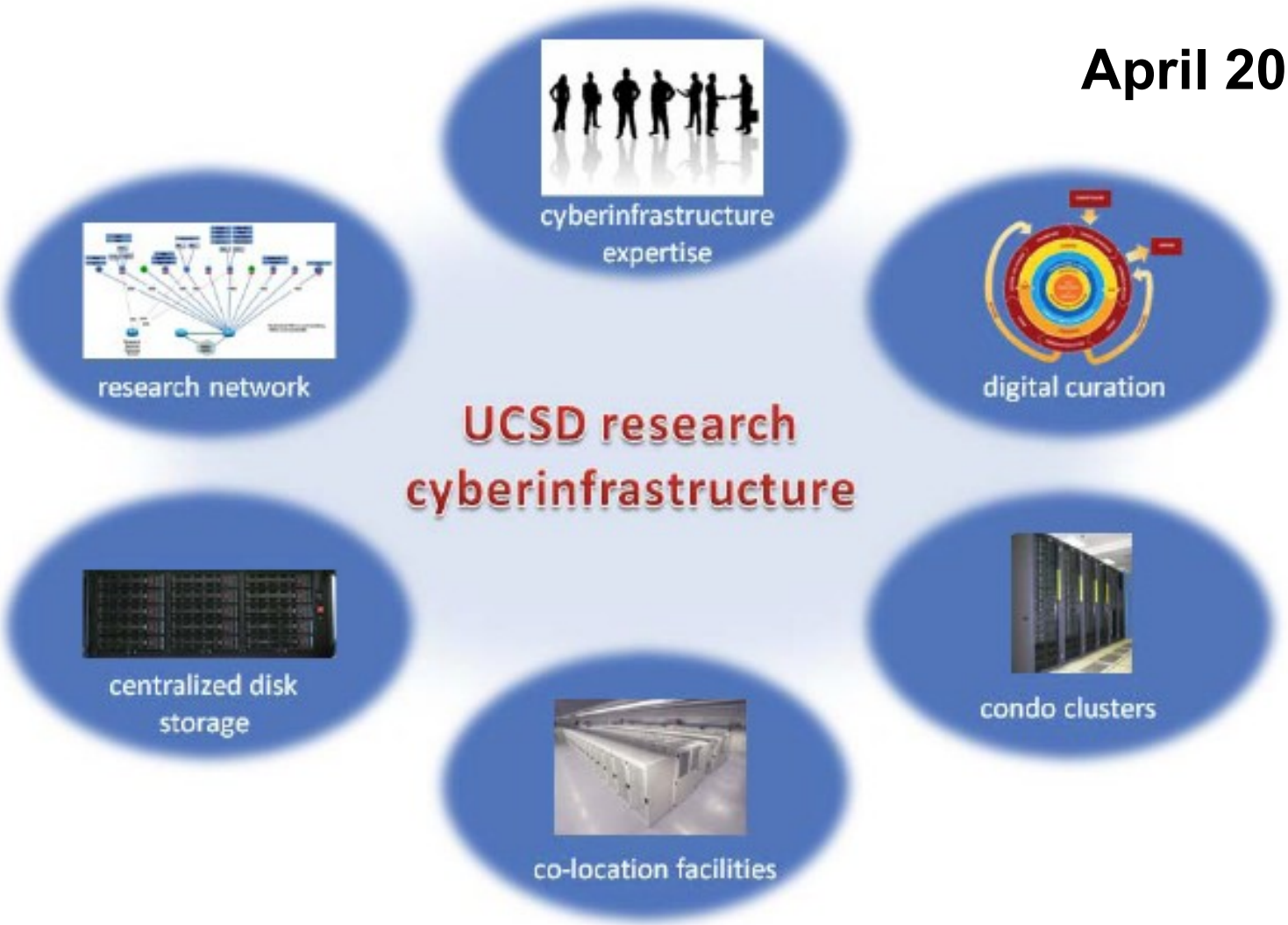
In the photo you see me holding the glass award (very cool looking!), flanked by CUDI (Mexico's R&E network) director Carlos Casasus on my right and CICESE (largest Mexican science institute funded by CONACYT) director-general Federico Graef on my left. The CENIC award was presented by Louis Fox, President of CENIC (right of Carlos) and Doug Hartline, UC Santa Cruz, CENIC Confernce Committee Chair (left of Federico). The Calit2/CUDI/CICESE technical team is on the right.

"Blueprint for the Digital University"--Report of the UCSD Research Cyberinfrastructure Design Team

- **A Five Year Process Begins Pilot Deployment This Year**

April 2009

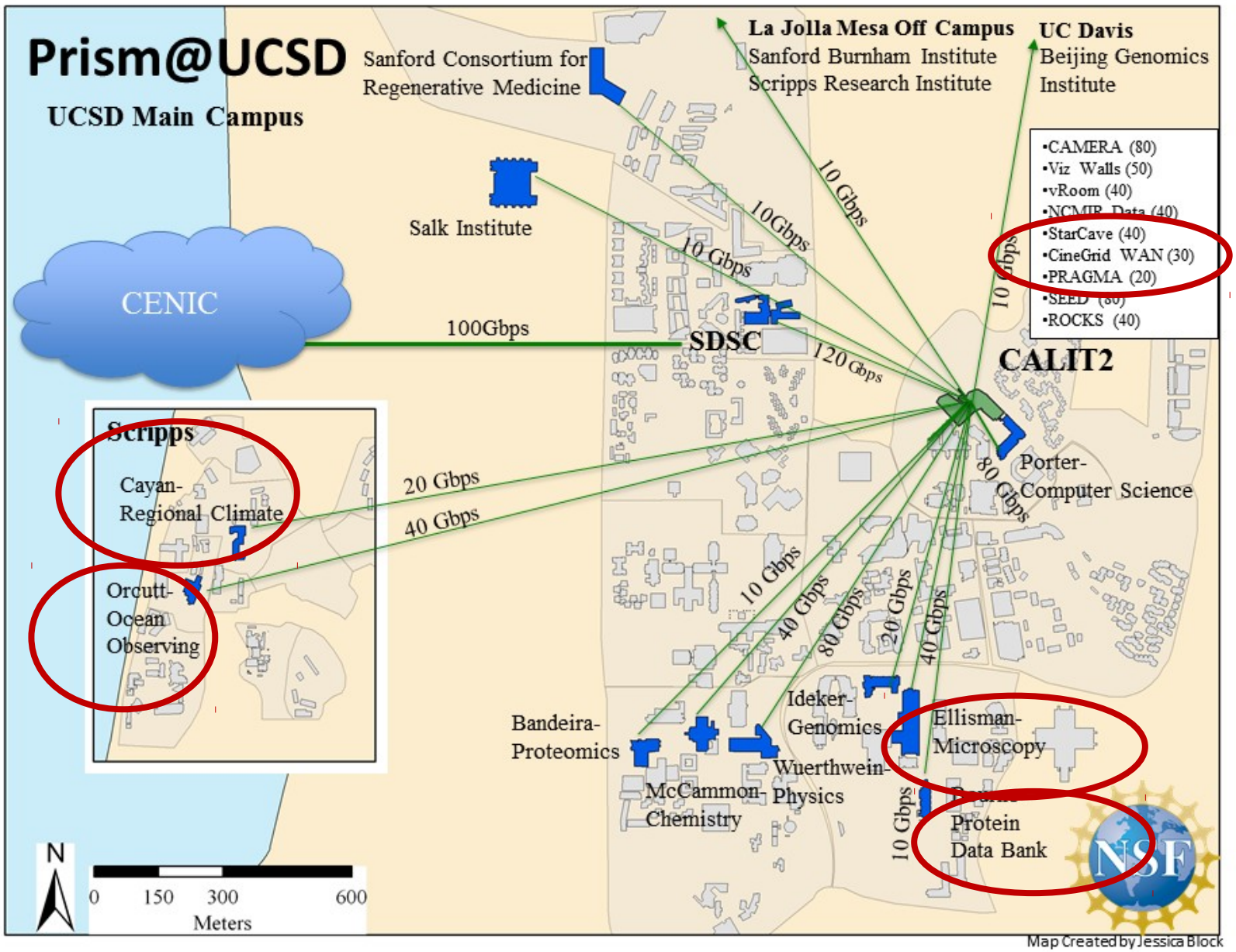
No Data Bottlenecks --Design for Gigabit/s Data Flows



SDSC research.ucsd.edu/documents/rcidt/RCIDTReportFinal2009.pdf

The Next Step: Creating a "Big Data Freeway" System Connecting Instruments, Computers, & Storage

Phil Papadopoulos, PI
Larry Smarr co-PI

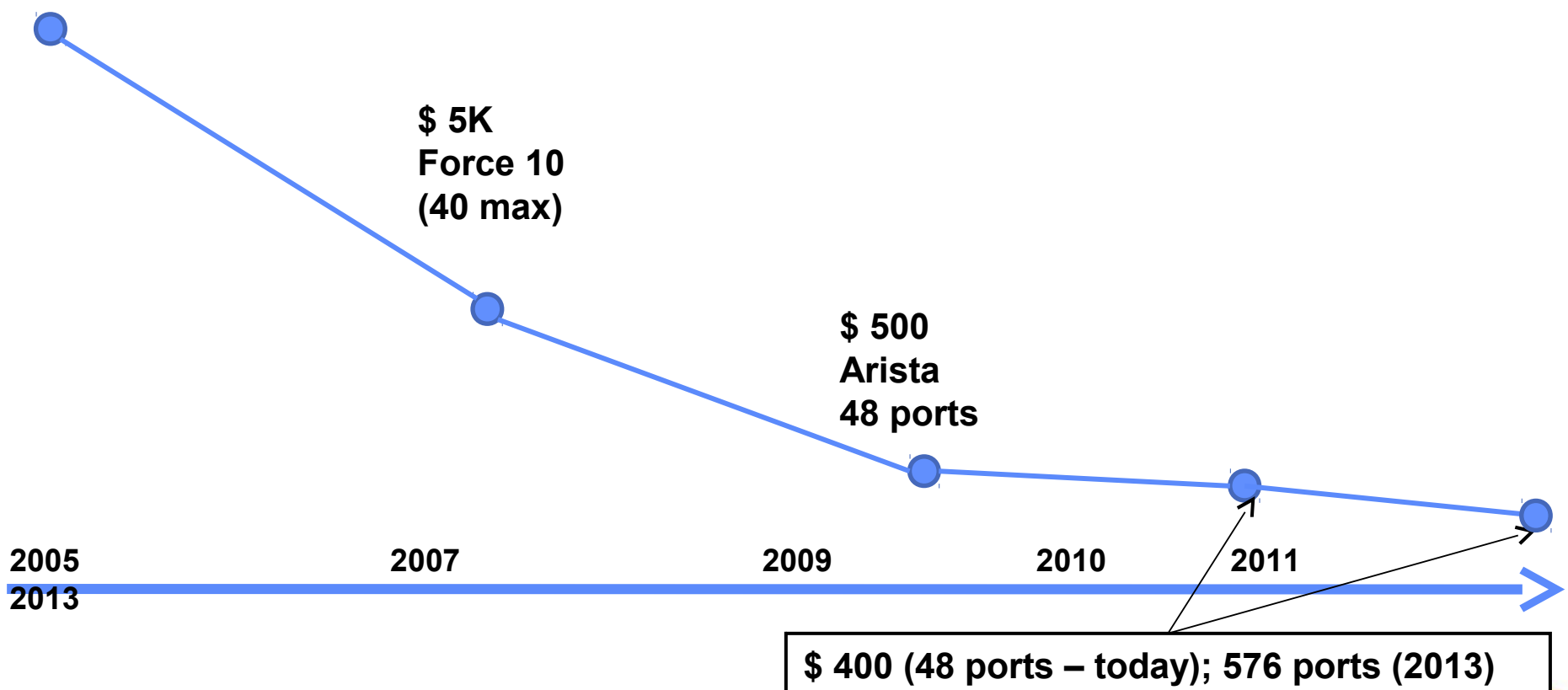


PRISM @UCSD

Start Date 1/1/13

Rapid Evolution of 10GbE Port Prices Makes Campus-Scale 10Gbps CI Affordable

- **Port Pricing is Falling**
- **Density is Rising – Dramatically**
- **Cost of 10GbE Approaching Cluster HPC Interconnects**

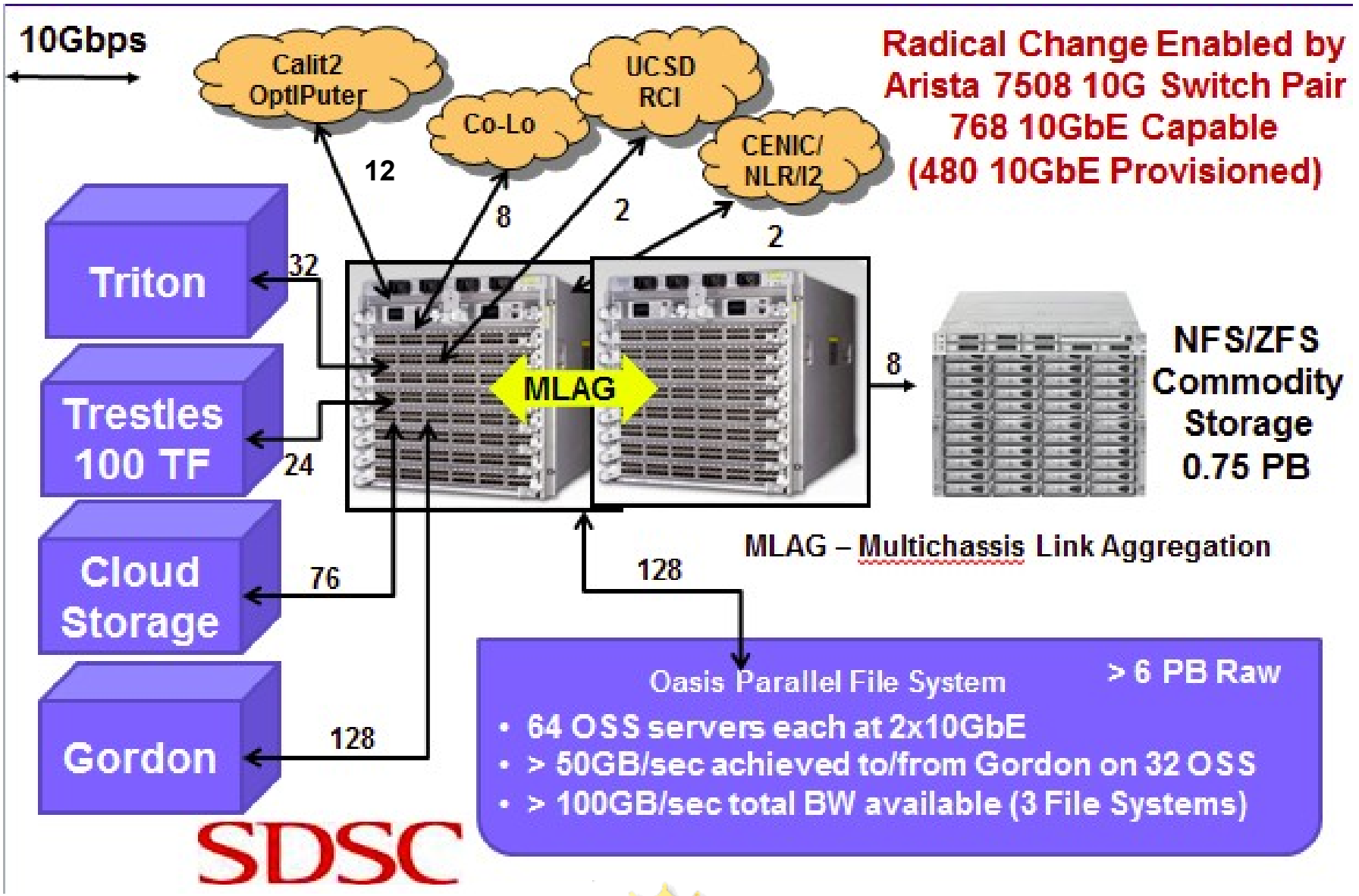


SDSC

Source: Philip Papadopoulos, SDSC/Calit2

it²

Arista Enables SDSC's Massively Parallel 10G Switched Data Analysis Resource



Many Disciplines Beginning to Need Dedicated High Bandwidth on Campus

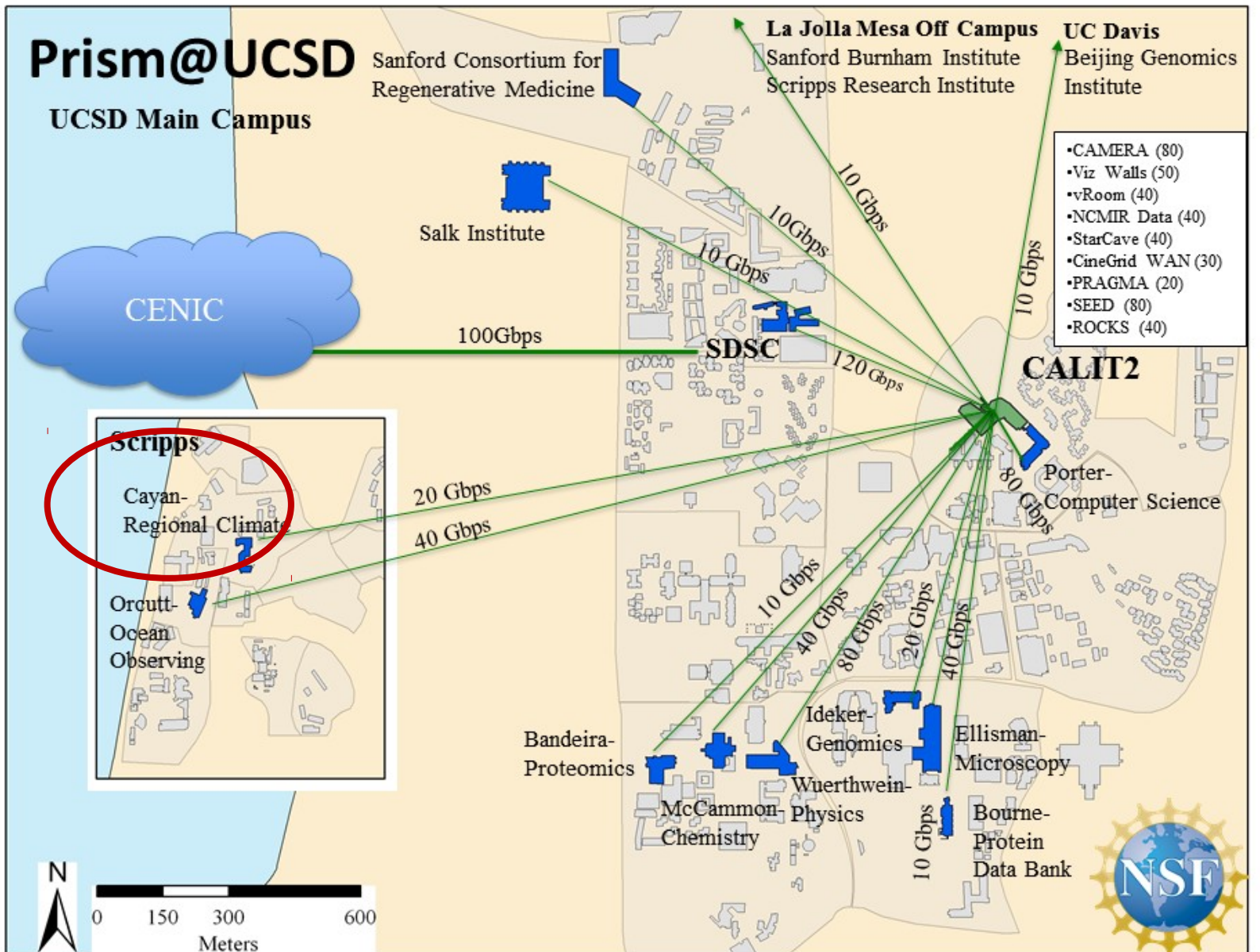
How to Terminate a CENIC 100G Campus Connection

- Remote Analysis of Large Data Sets
  - Regional Climate Change
- Connection to Remote Campus Compute & Storage Clusters
  - Ocean Observatory
  - Microscopy
- Providing Remote Access to Campus Data Repositories
  - Protein Data Bank
- Enabling Remote Collaborations
  - National and International

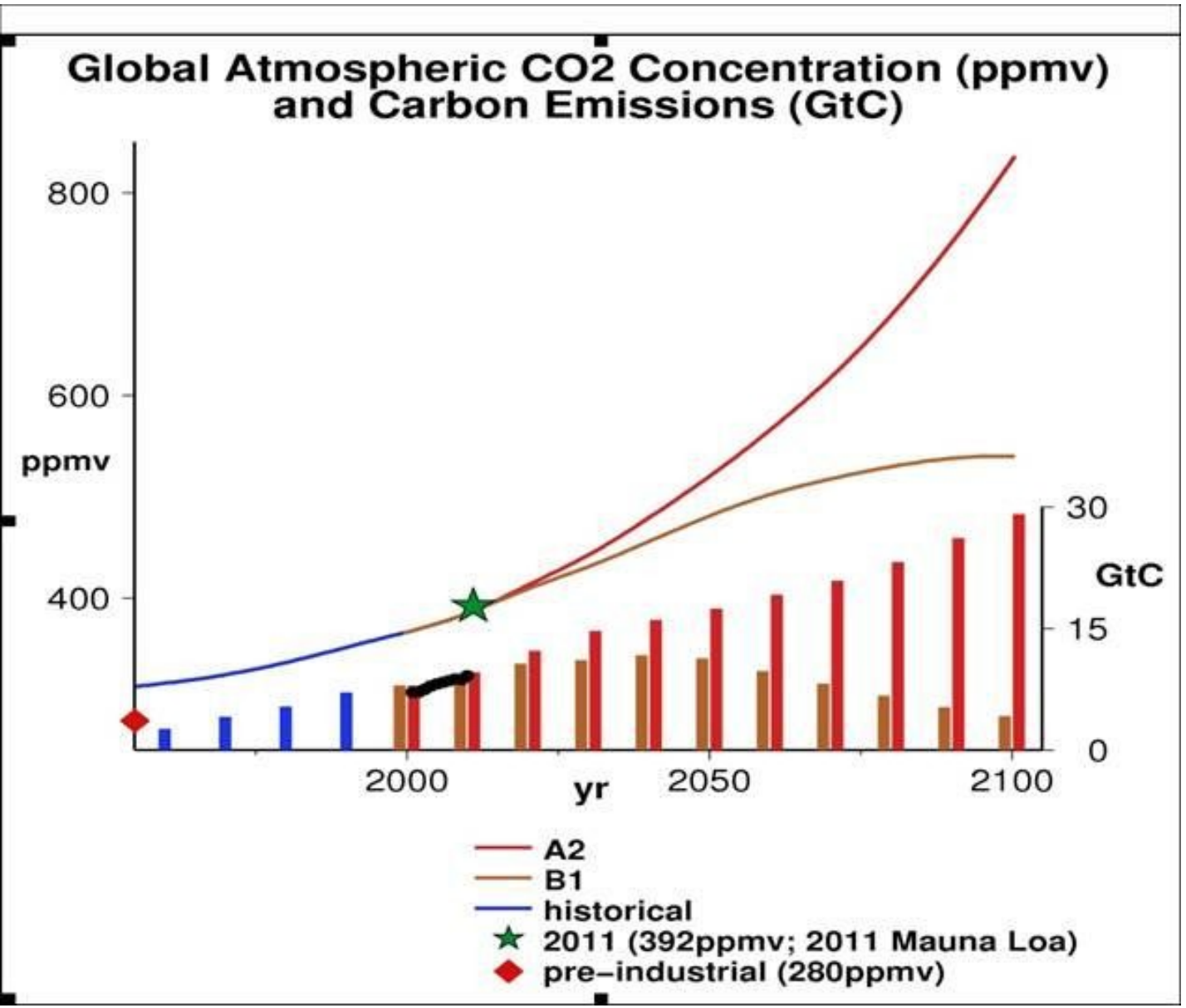
SDSC

it²

PRISM@UCSD Enables Remote Analysis of Large Data Sets



UCSD Campus Climate Researchers Need to Download Results from Remote Supercomputer Simulations



Greenhouse Gas Emissions and Concentration CMIP3 GCM's

Source: Dan Cayan, SIO UCSD

Global to Regional Downscaling

GCMs ~150km downscaled to Regional models ~ 12km

Many simulations IPCC AR4 and IPCC AR5

INCREASING VOLUME OF CLIMATE SIMULATIONS

in comparison to 4th IPCC (CMIP3) GCMs :

Latest Generation CMIP5 Models Provide:

- **More Simulations**
- **Higher Spatial Resolution**
- **More Developed Process Representation**
- **Daily Output is More Available**

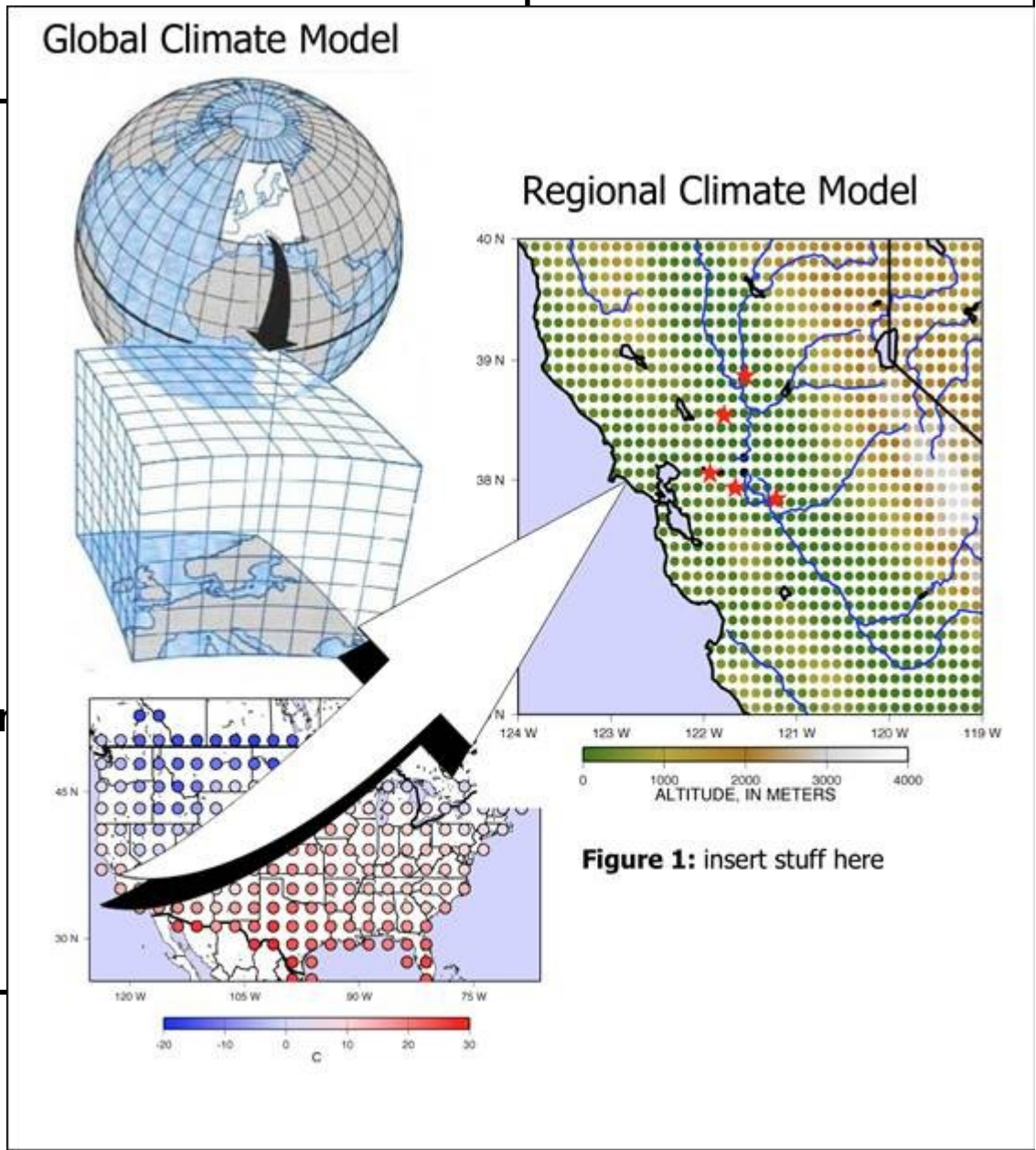
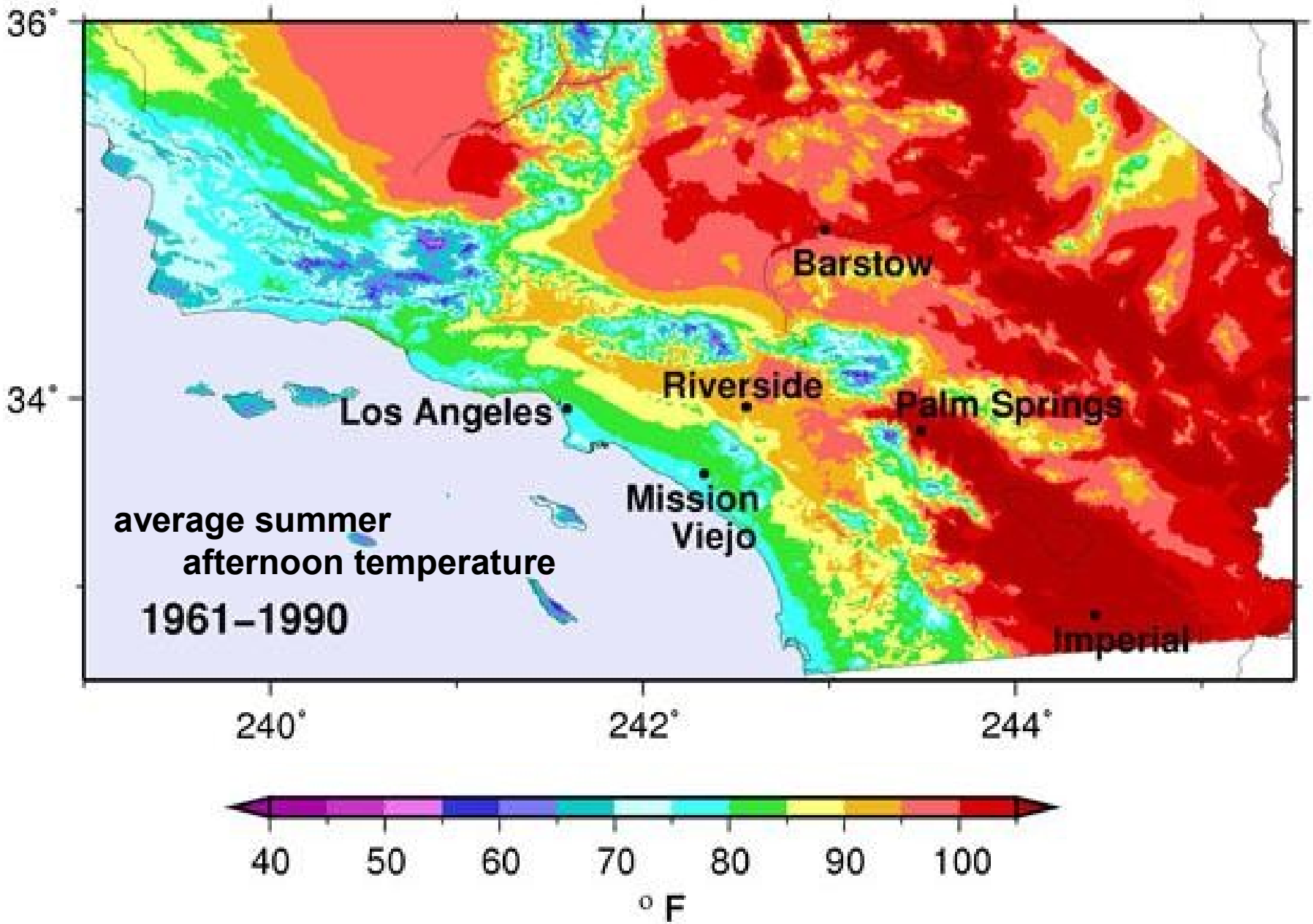


Figure 1: insert stuff here

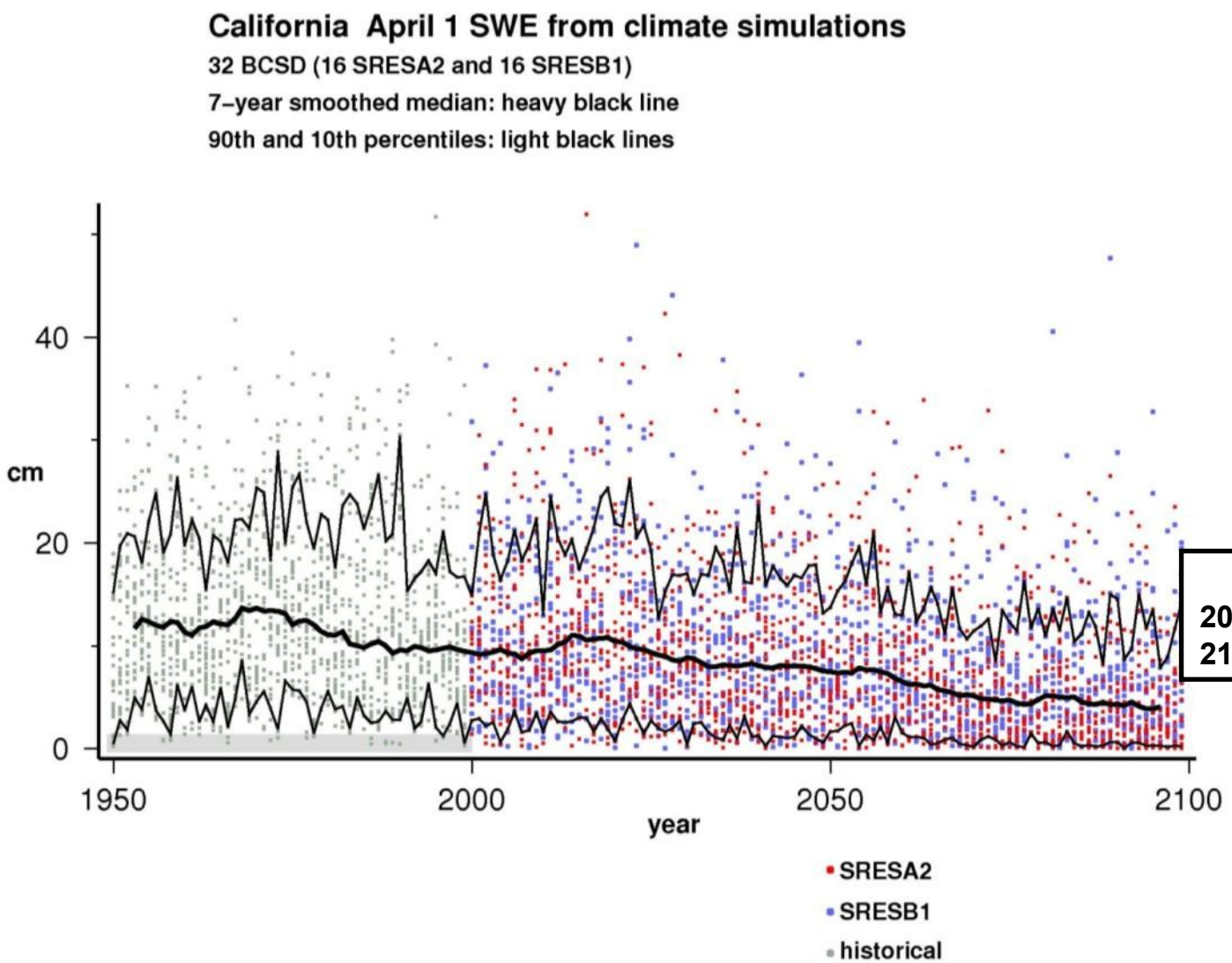
Source: Dan Cayan, SIO UCSD

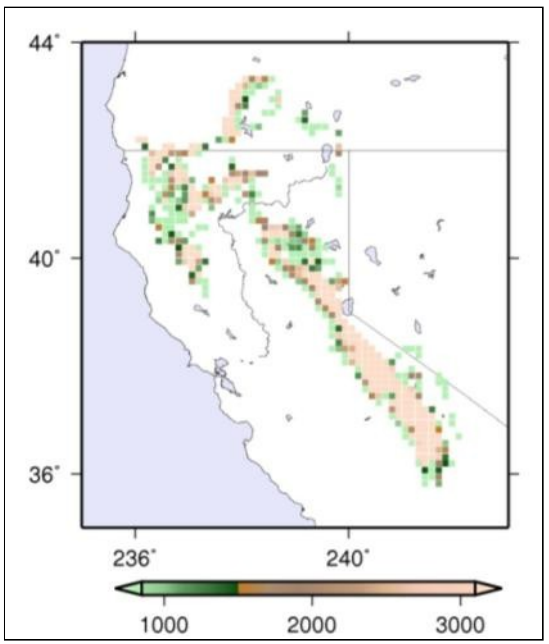


GFDL A2 1km downscaled to 1km
Hugo Hidalgo Tapash Das Mike Dettinger

HOW MUCH CALIFORNIA SNOW LOSS ?

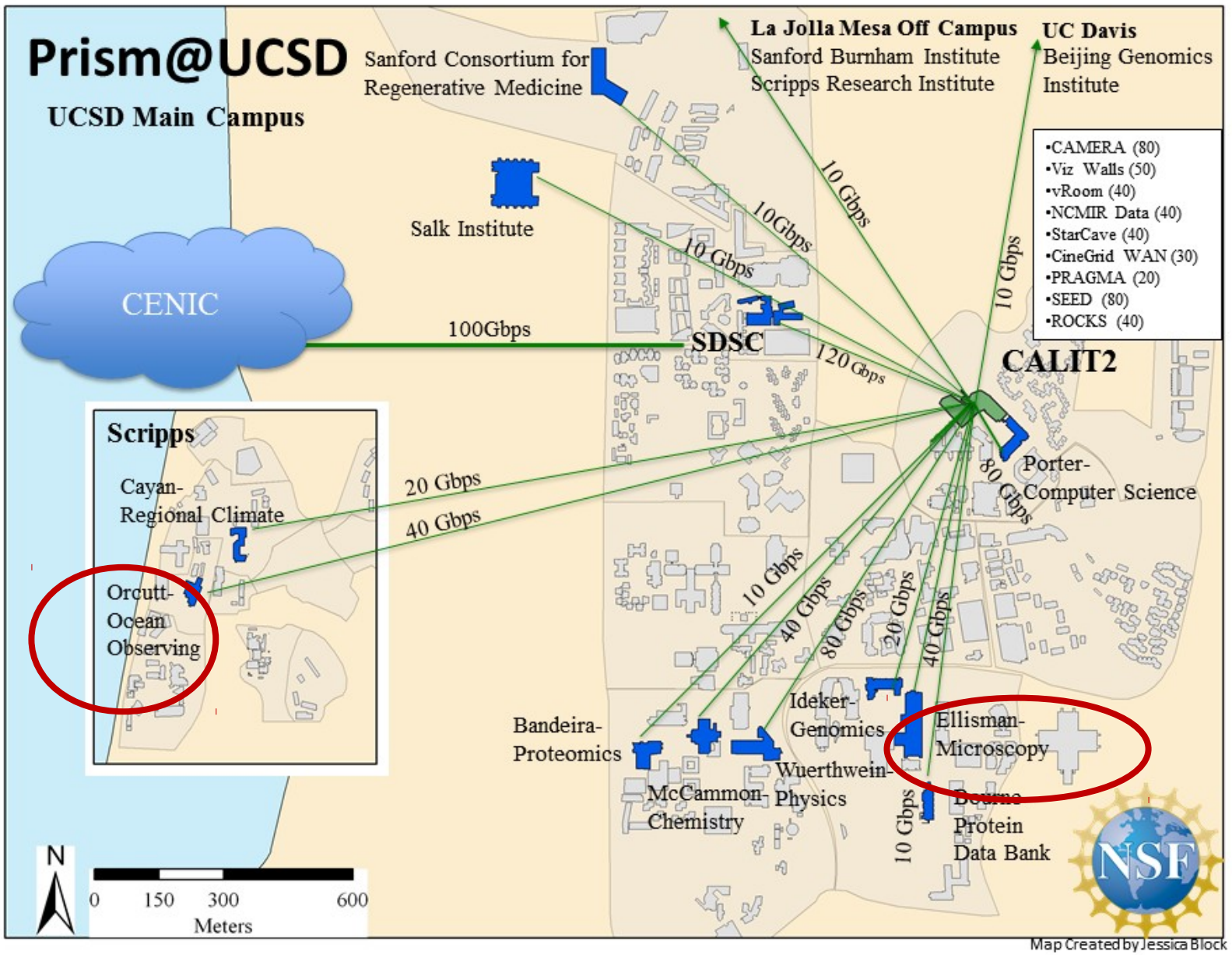
Initial projections indicate substantial reduction in snow water for Sierra Nevada+





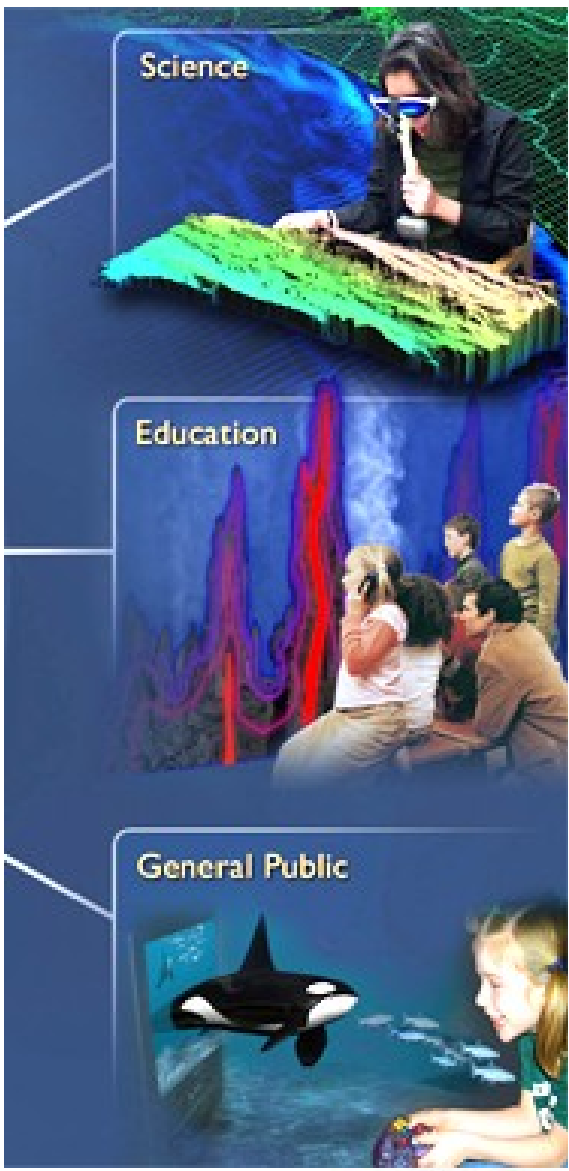
declining Apr 1 SWE:
2050 median SWE ~ 2/3 historical median
2100 median SWE ~ 1/3 historical median

PRISM@UCSD Enables
Connection to Remote Campus Compute & Storage Clusters



The OOI CI is Built on Dedicated 10GE and Serves Researchers, Education, and Public

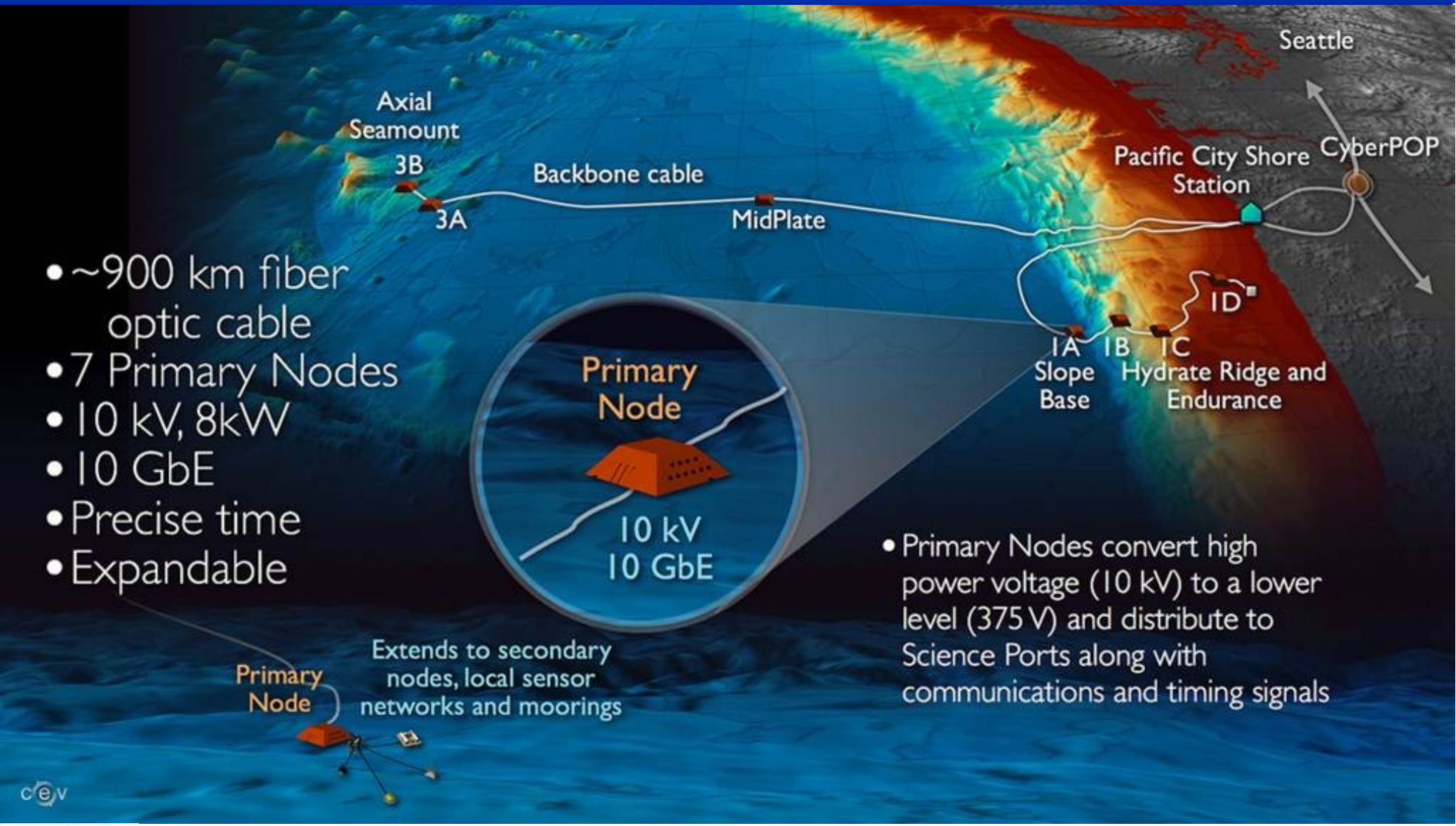




- Distribution Point
- Acquisition Point/ Engineering Center
- Global Array
- Regional Node
- Coastal Array
- Dedicated 10GE
- VPN over Internet

Source: Matthew Arrott, John Orcutt OOI CI

Reused Undersea Optical Cables Form a Part of the Ocean Observatories

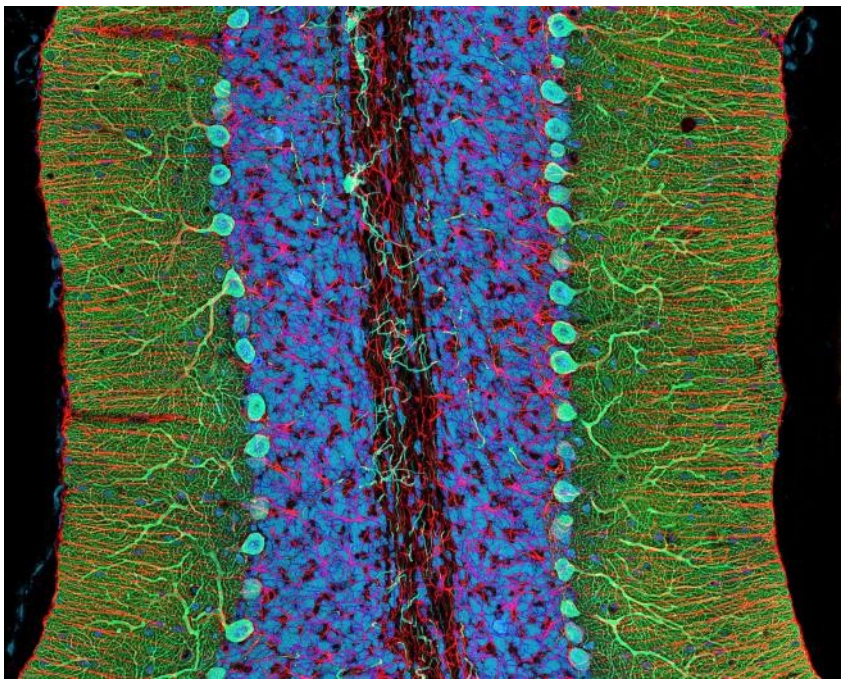


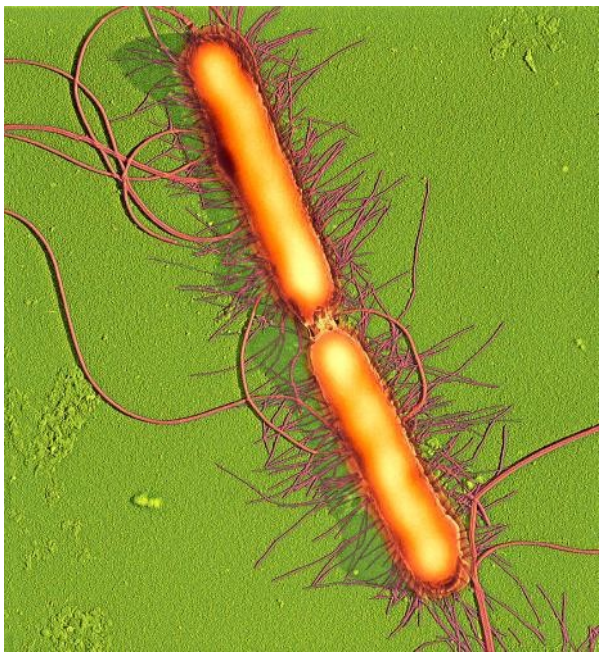
Source: John Delaney UWash OOI

OOI CI Team at Scripps Institution of Oceanography Needs Connection to Its Server Complex in Calit2

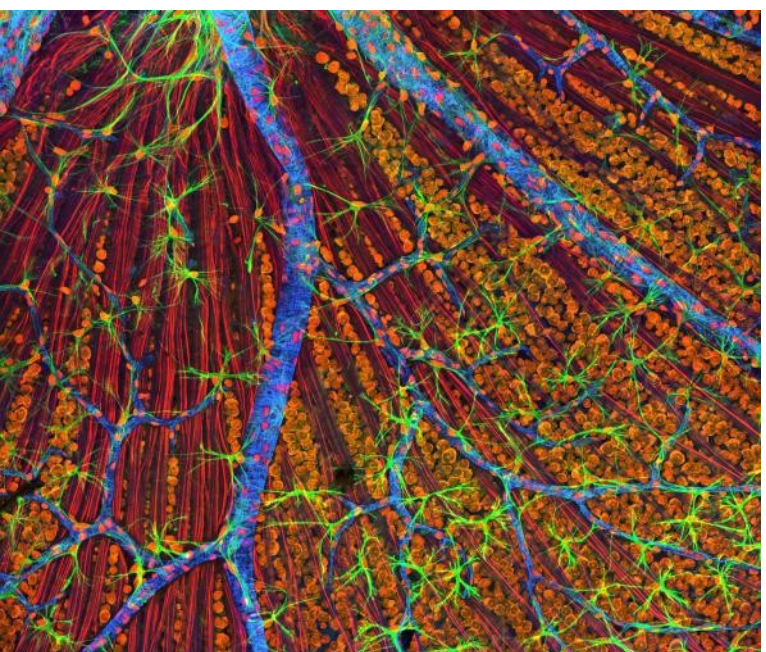


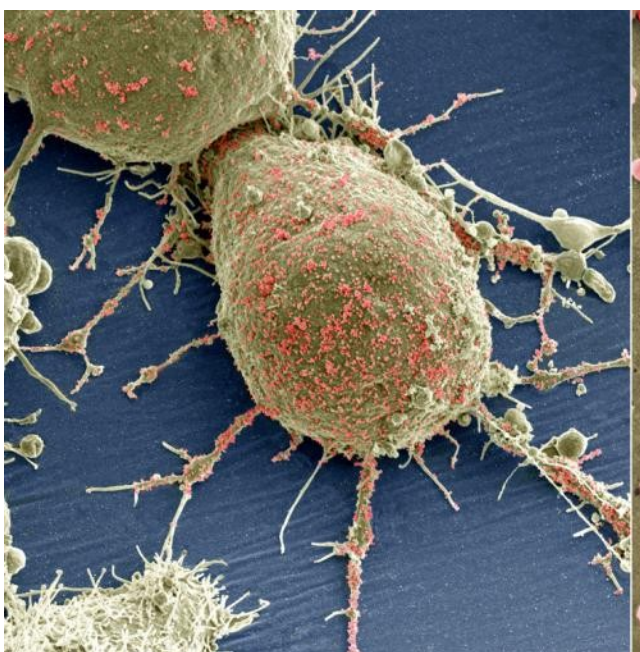
Ultra High Resolution Microscopy Images Created at the National Center for Microscopy Imaging

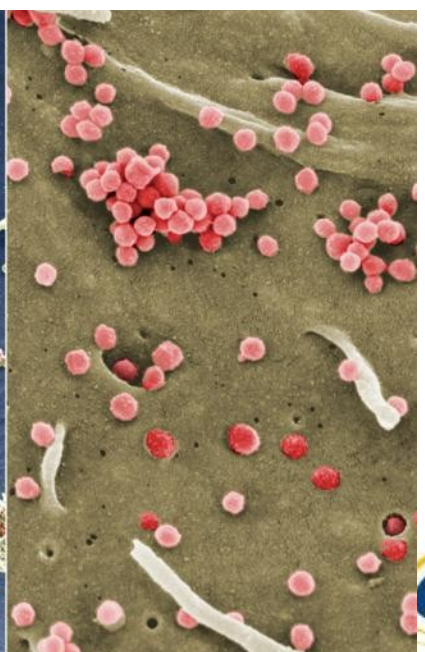




NCMIR







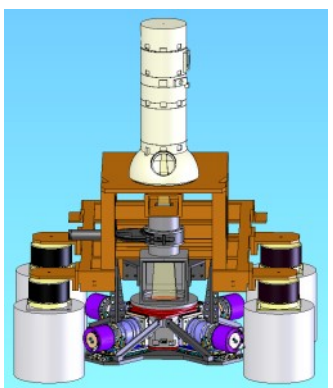
it²

Microscopes Are Big Data Generators – Driving Software & Cyberinfrastructure Development



Zeiss Merlin 3View w/ 32k x 32k Scanning and Automated Mosaicing:

Current= 1-2 TB/week → soon 12 TB/week



JEOL-4000EX w/ 8k x 8k CD, Automated Mosaicing, and Serial Tomography:

Current= 1 TB/week

NCMIR



FEI Titan w/ 4k x 4k STEM, EELS, 4k x 3.5k DDD, 4k x4k CCD, Automated Mosaicing, and Multi-tilt Tomography:

Current= 1 TB/week

200-500 TB/year Raw → >2 PB/year Aggregate

Source: Mark Ellisman, School of Medicine, UCSD

it²

NIH National Center for Microscopy & Imaging Research
Integrated Infrastructure of Shared Resources

NCMIR

Telemicroscopy Resources:
Osaka Univ., Oxford Univ.,
The Jackson Labs

Neuroinformatics Resources

Shared Infrastructure

Development Systems
Storage Resources
OptIPortal Tiled Display
California Institute of Telecommunications and Information Technology
UCSD
NBCR
SDSC
CRBS
Storage Resources
Production Systems
Triton Petascale Resource
San Diego Supercomputer Center

Scientific Instruments

Ultra-wide Field Intra-vital LM Instruments
Super-resolution PALM
Serial Blockface Scanning EM
Ultra-wide field Serial Section EM Tomography
Multi-color EM, EFTEM
Wide-field STEM/TEM

End User Workstations

Graphics Workstations
OptIPortal Tiled Display

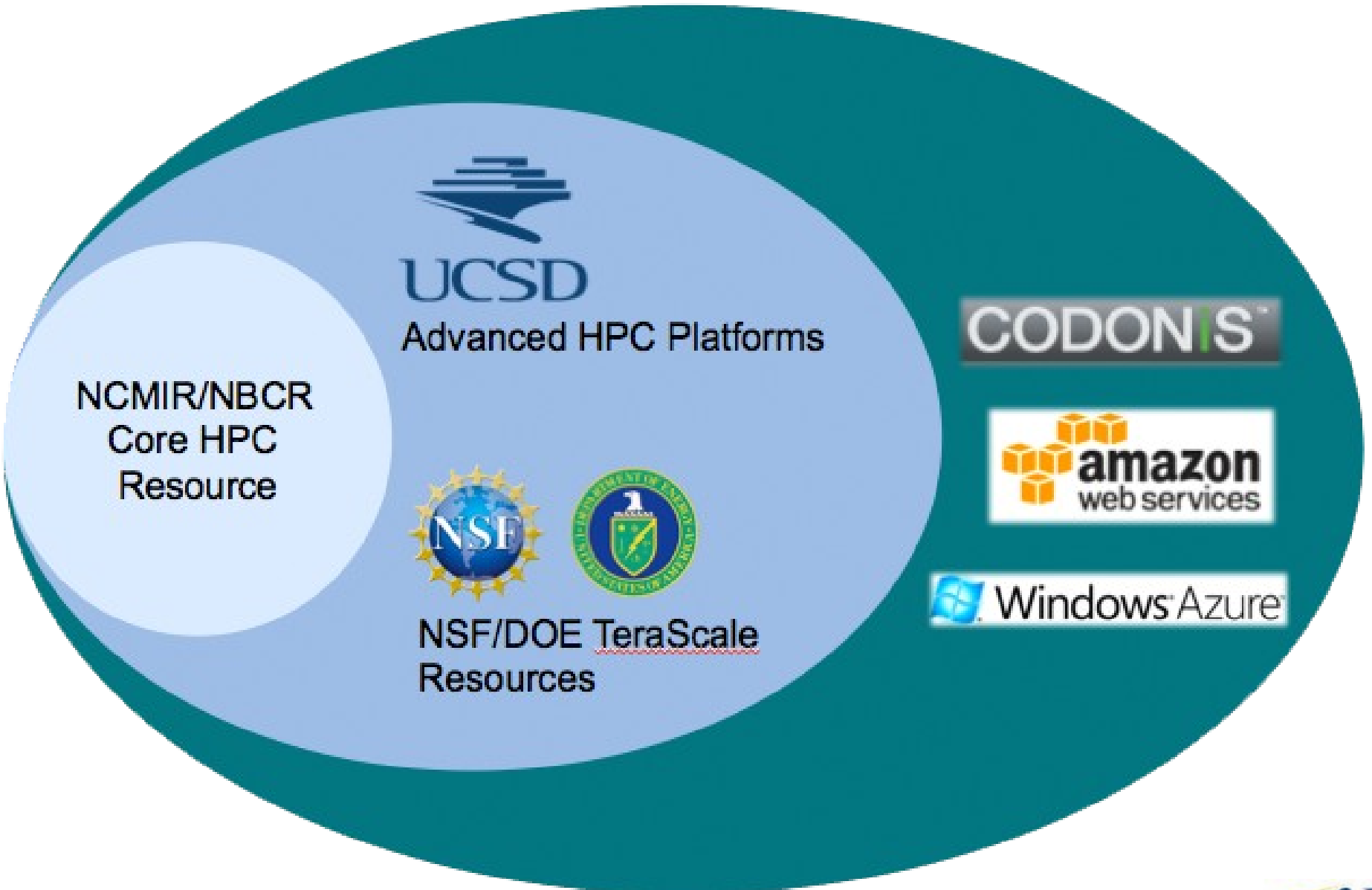
Local SOM Infrastructure

NCMIR Data Storage Servers
Critical Servers (Mail, Web, DB)
Large Memory Servers
NBCR Compute Clusters
GreenLight Project
GPU Compute Clusters

SDSC

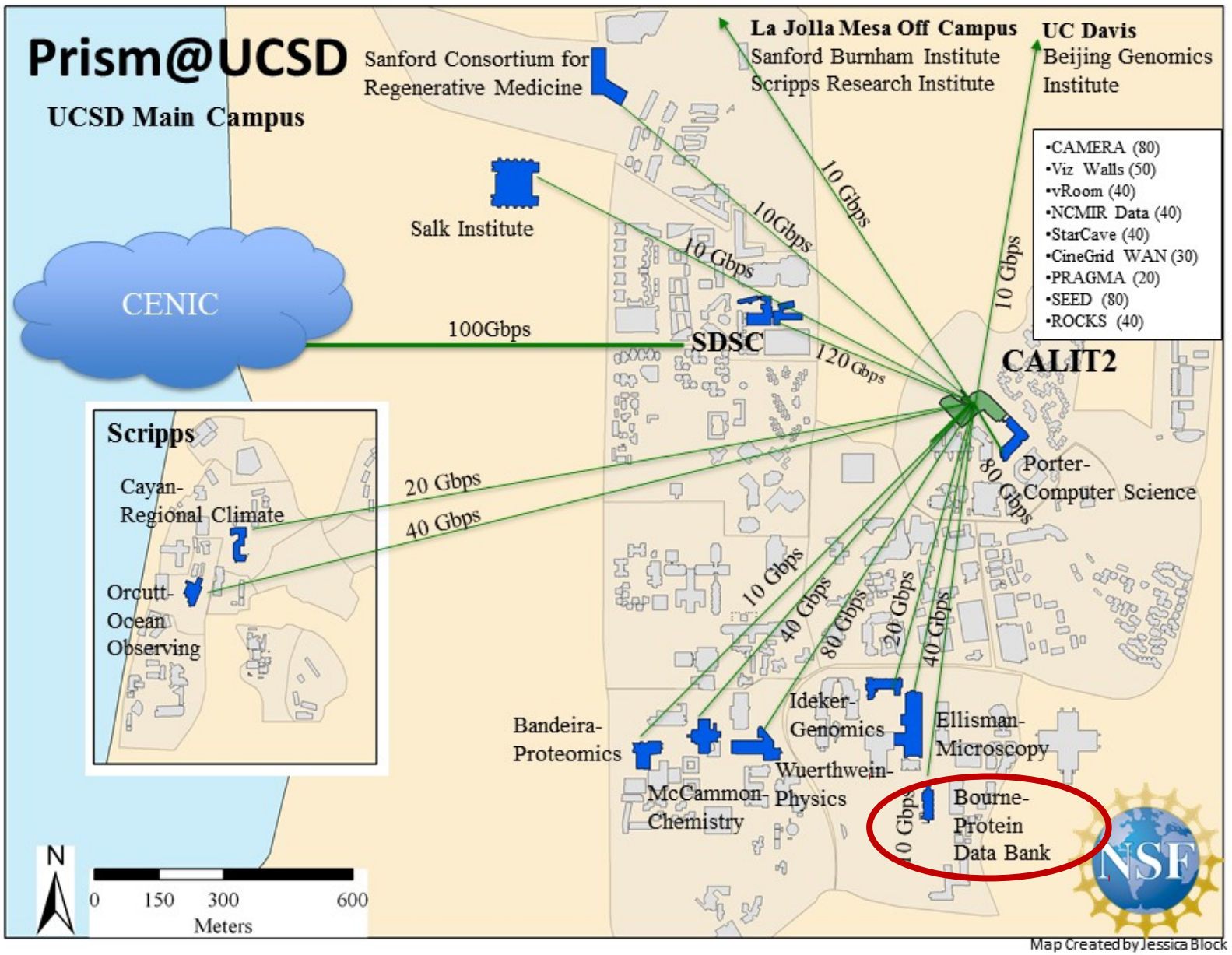
Source: Steve Peltier, Mark Ellisman, NCMIR

Agile System that Spans Resource Classes



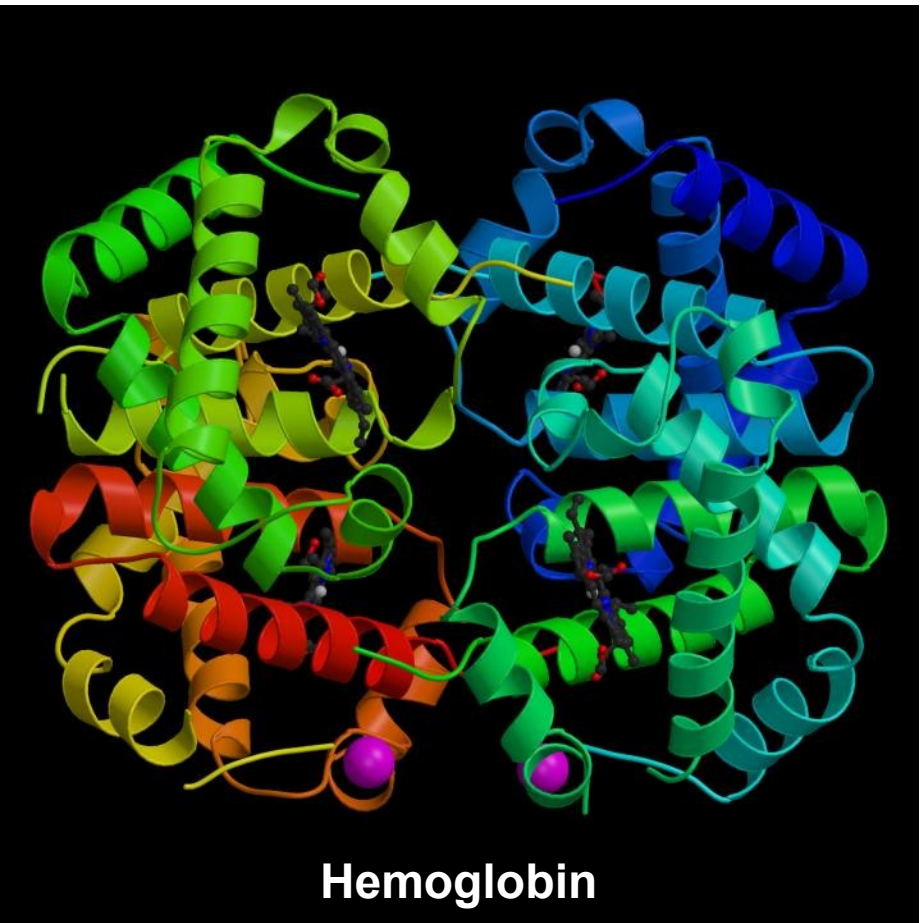
NCMIR

PRISM@UCSD Enables
Providing Remote Access to Campus Data Repositories

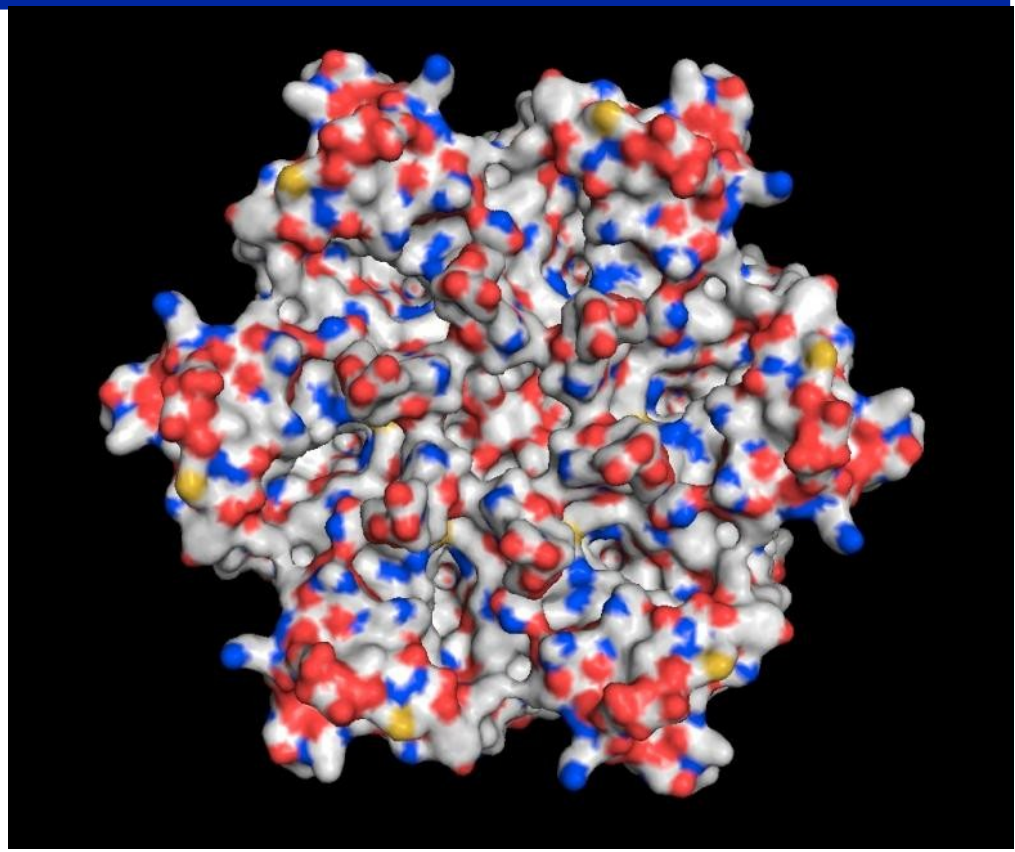


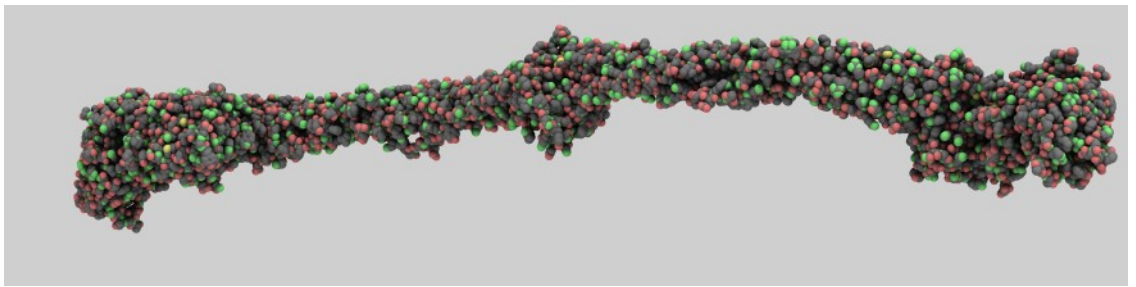
Protein Data Bank (PDB) Needs Bandwidth to Connect Resources and Users

- Archive of experimentally determined 3D structures of proteins, nucleic acids, complex assemblies
- One of the largest scientific resources in life sciences



Hemoglobin



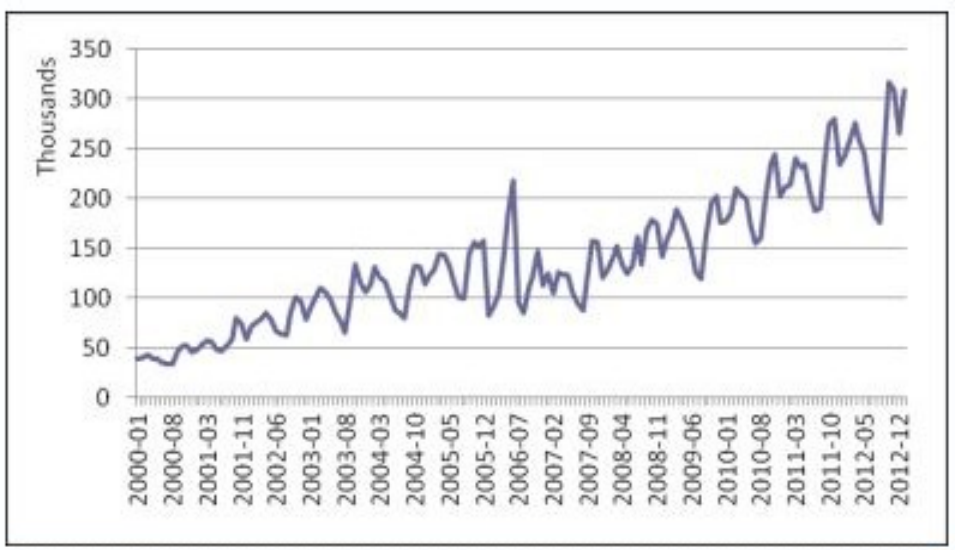


Source: Phil Bourne and Andreas Prlić, PDB

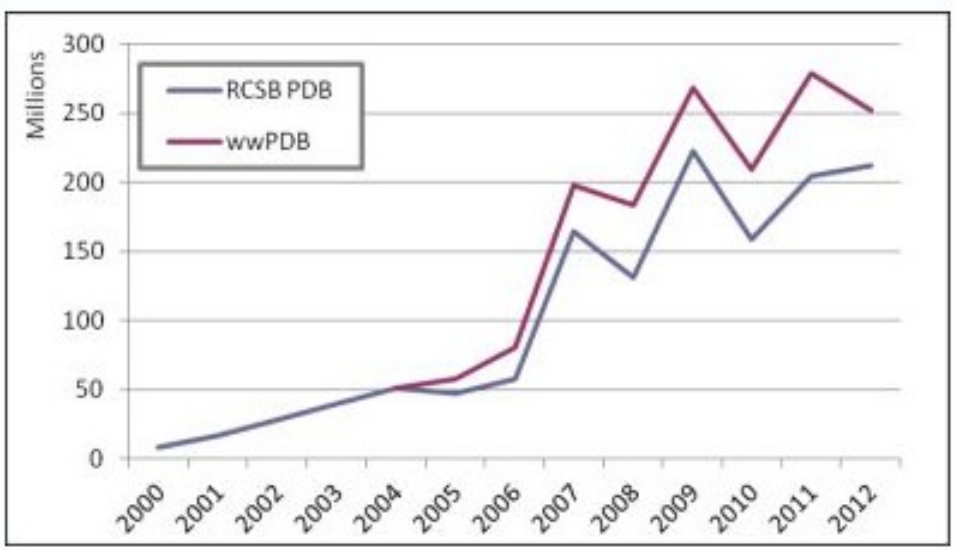
PDB Usage Is Growing Over Time

- More than 300,000 Unique Visitors per Month
- Up to 300 Concurrent Users
- ~10 Structures are Downloaded per Second 7/24/365
- Increasingly Popular Web Services Traffic

RCSB PDB Website Unique Users by Month

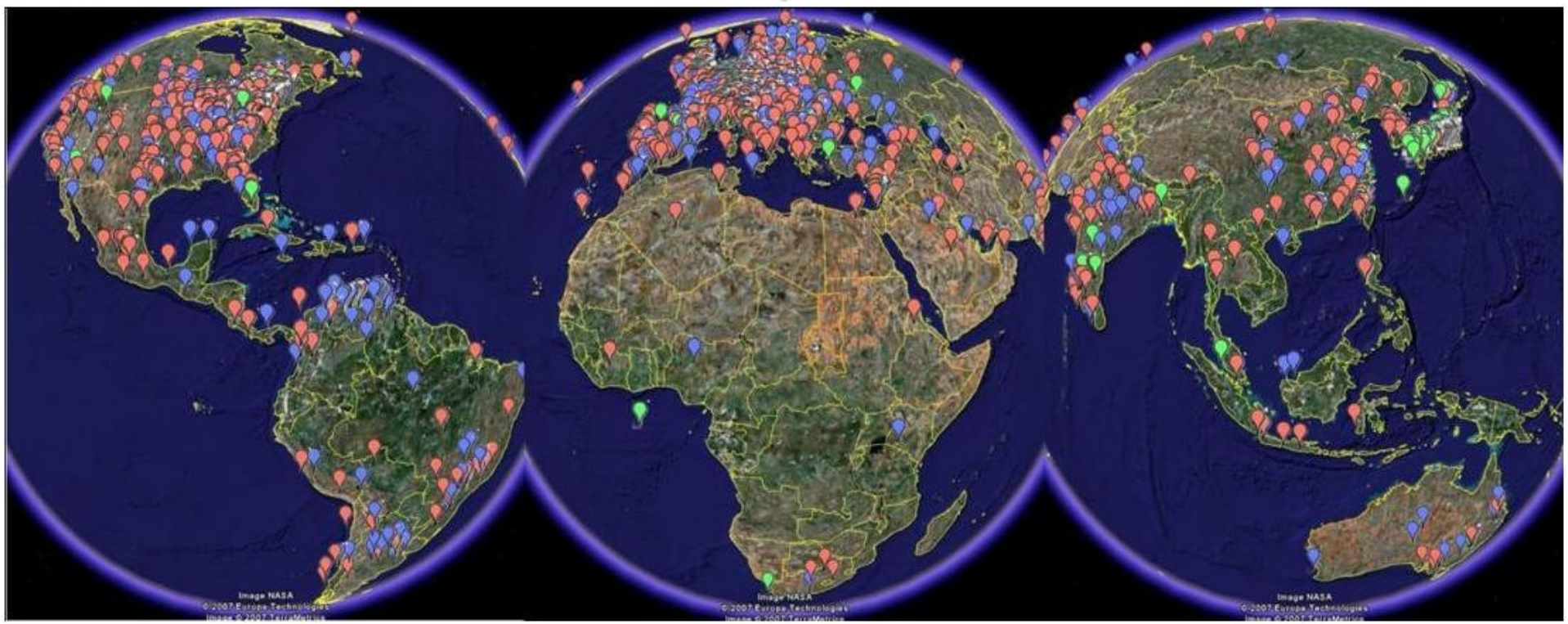


FTP Downloads by Year



Source: Phil Bourne and Andreas Prlić, PDB

The Global Users of the PDB: 2010 FTP Traffic



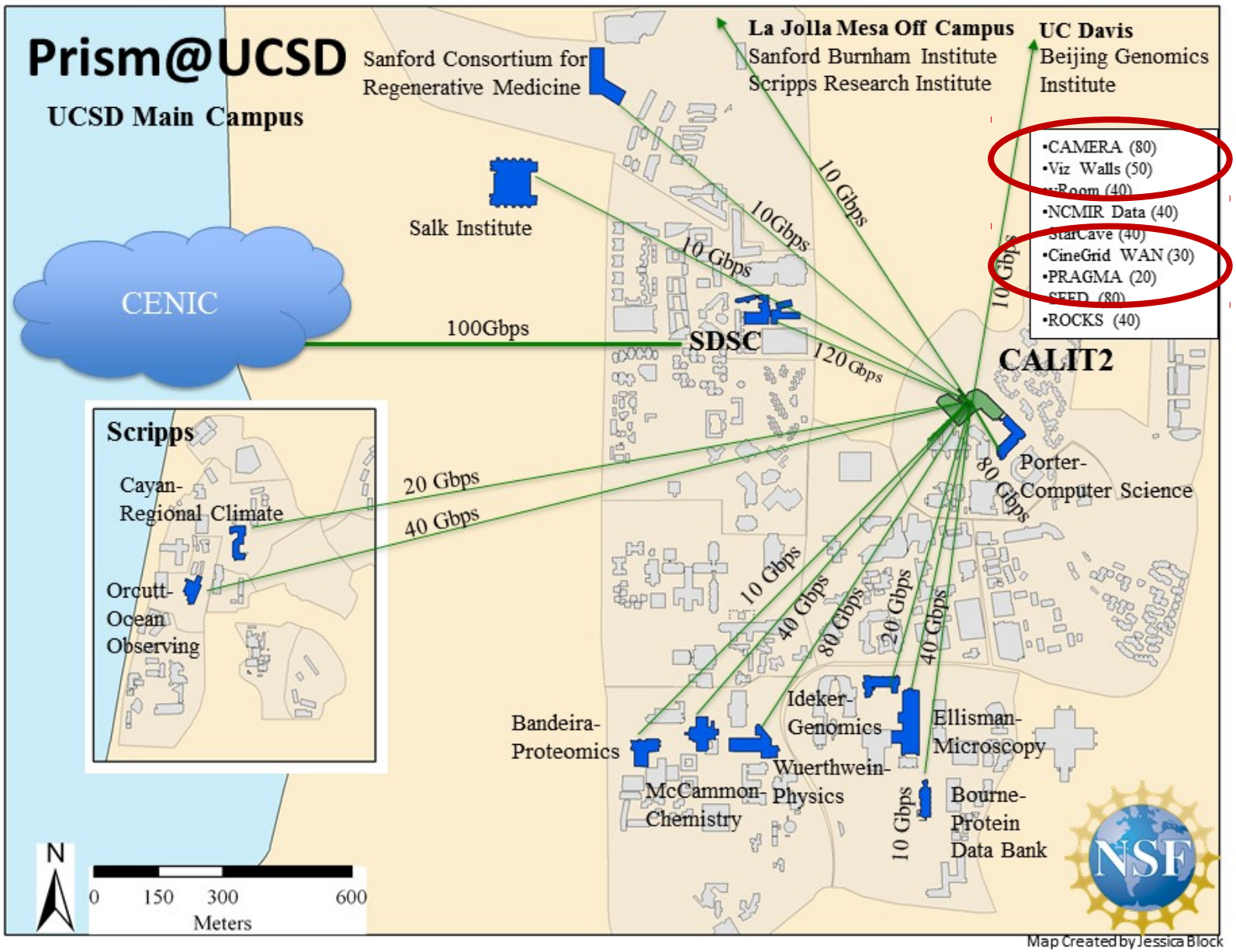
RCSB PDB
159 million
entry downloads

PDBe
34 million
entry downloads

PDBj
16 million
entry download

Source: Phil Bourne and Andreas Prlić, PDB

PRISM@UCSD Enables
Enabling Remote National and International Collaborations



Tele-Collaboration for Audio Post-Production
Realtime Picture & Sound Editing Synchronized Over IP



Skywalker Sound@Marin **Calit2@San Diego**

Collaboration Between EVL's CAVE2 and Calit2's VROOM Over 10Gb Wavelength



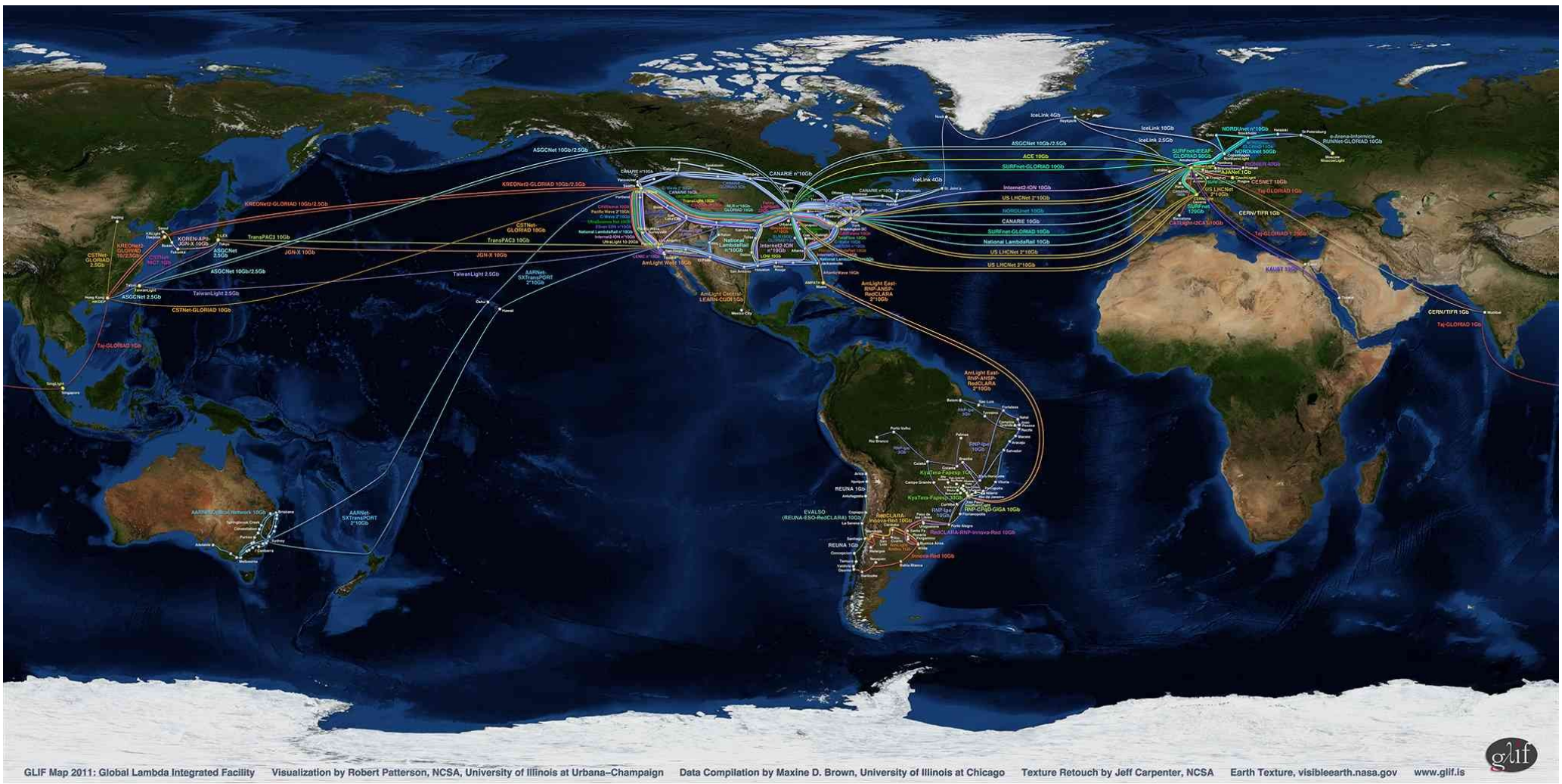
Source: NTT Sponsored ON*VECTOR Workshop at Calit2 March 6, 2013

Calit2 is Linked to CICESE at 10G Coupling OptIPortals at Each Site



The Global Lambda Integrated Facility--CICESE Becomes a Member of the Planetary-Scale High Bandwidth Collaboratory

Research Innovation Labs Linked by 10G Dedicated Lambdas



Next Step – Extend to Other Big Data Sites in Mexico

www.glif.is/publications/maps/GLIF_5-11_World_2k.jpg