ASSOCIATION OF UNIVERSITY RESEARCH PARKS







High Performance Cyberinfrastructure to Support Data-Intensive Biomedical Research Instruments

Invited Talk
Association of University Research Parks BioParks 2008
"From Discovery to Innovation"
Salk Institute
La Jolla, CA
June 16, 2008

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

Abstract

Calit2 is using 10 gigabit/s optical paths to connect people and devices on local, regional, national, and global scales. On campus this cyberinfrastructure connects a variety of data-intensive biomedical instruments (DNA arrays, genome sequencers, mass spectrographs) to distributed computing/storage.

UCSD
UCIrvine

Calit2 Continues to Pursue Its Initial Mission:

Envisioning How the Extension of Innovative Telecommunications and Information Technologies Throughout the Physical World will Transform Critical Applications Important to the California Economy and its Citizens' Quality Of Life.

Two New Calit2 Buildings Provide New Laboratories for "Living in the Future"



- "Convergence" Laboratory Facilities
  - Nanotech, BioMEMS, Chips, Radio, Photonics
  - Virtual Reality, Digital Cinema, HDTV, Gaming
- Over 1000 Researchers in Two Buildings
  - Linked via Dedicated Optical Networks



Preparing for a World in Which Distance is Eliminated…

www.calit2.net

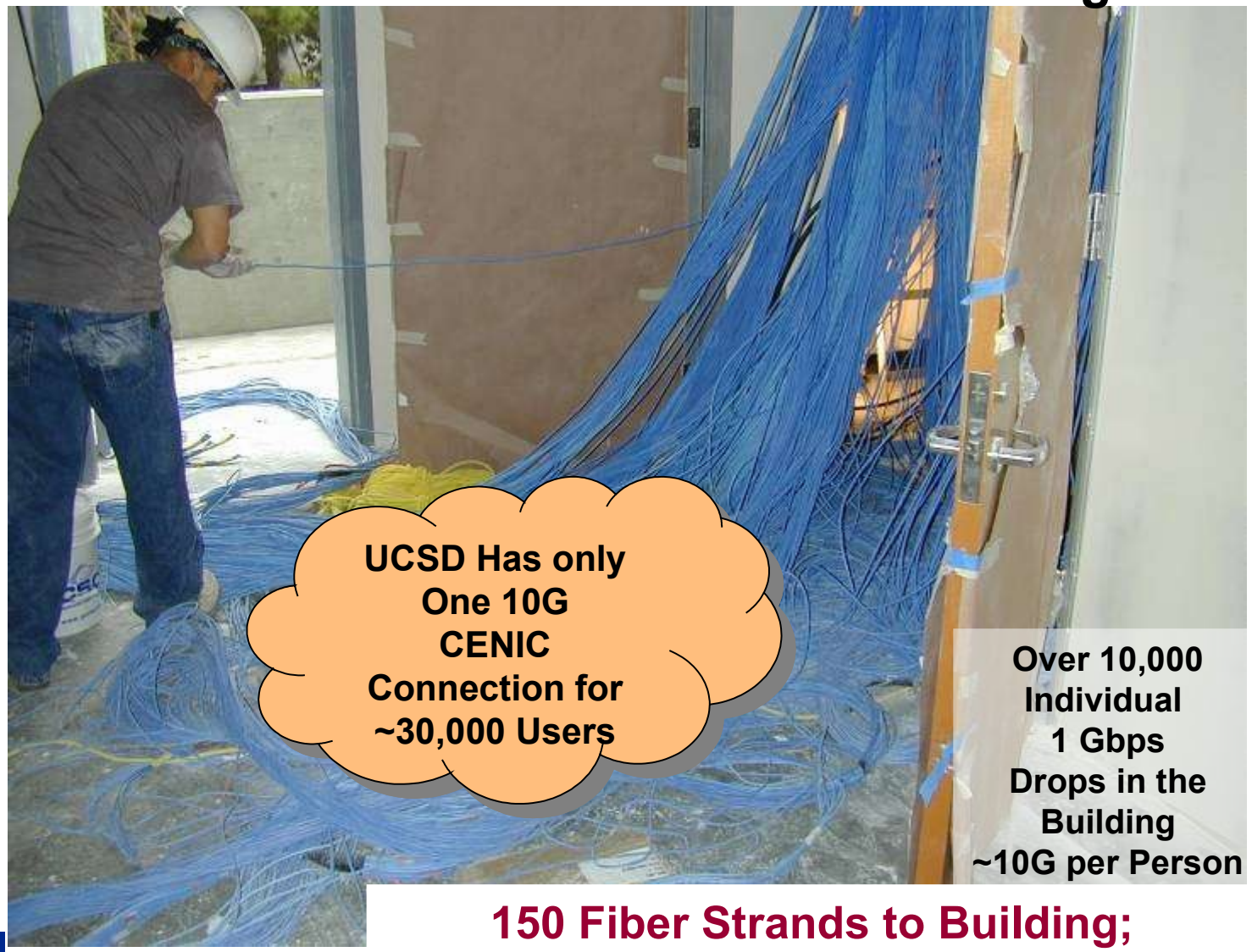
UCSD
UCIrvine

$100M From State for New Facilities

it²

The Calit2@UCSD Building is Designed for Prototyping Extremely High Bandwidth Applications

1.8 Million Feet of Cat6 Ethernet Cabling



24 Fiber Pairs to Each Lab

150 Fiber Strands to Building;
Experimental Roof Radio Antenna Farm
Ubiquitous WiFi

Photo: Tim Beach, Calit2

UCSD
UCIrvine

it²

Calit2--A Systems Approach to the Future of the Internet and its Transformation of Our Society

Calit2 Has Assembled a Complex Social Network
of Over 350 UC San Diego & UC Irvine Faculty
From Two Dozen Departments
Working in Multidisciplinary Teams
With Staff, Students, Industry, and the Community



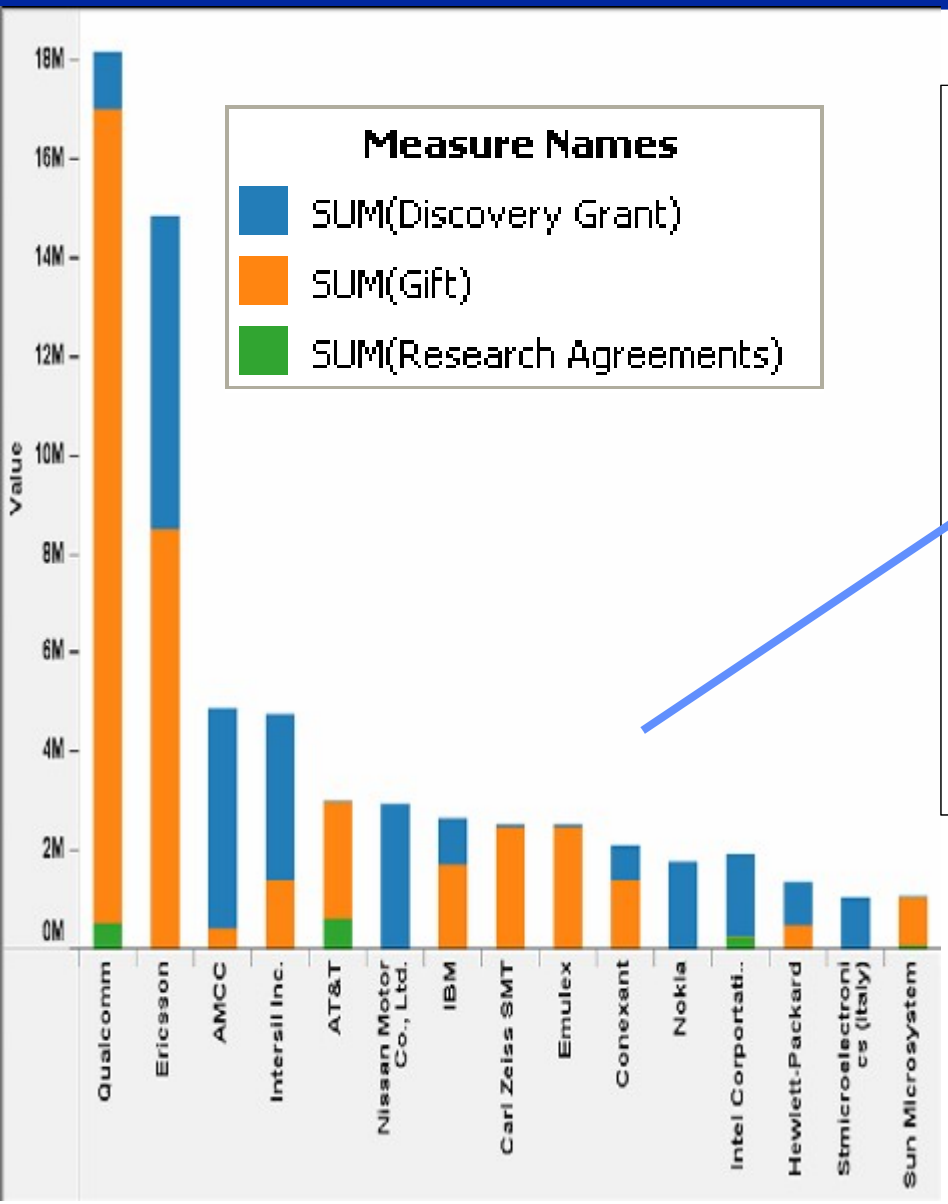
Integrating Technology Consumers and Producers
Into "Living Laboratories"

UCSD
UCIrvine

www.calit2.net

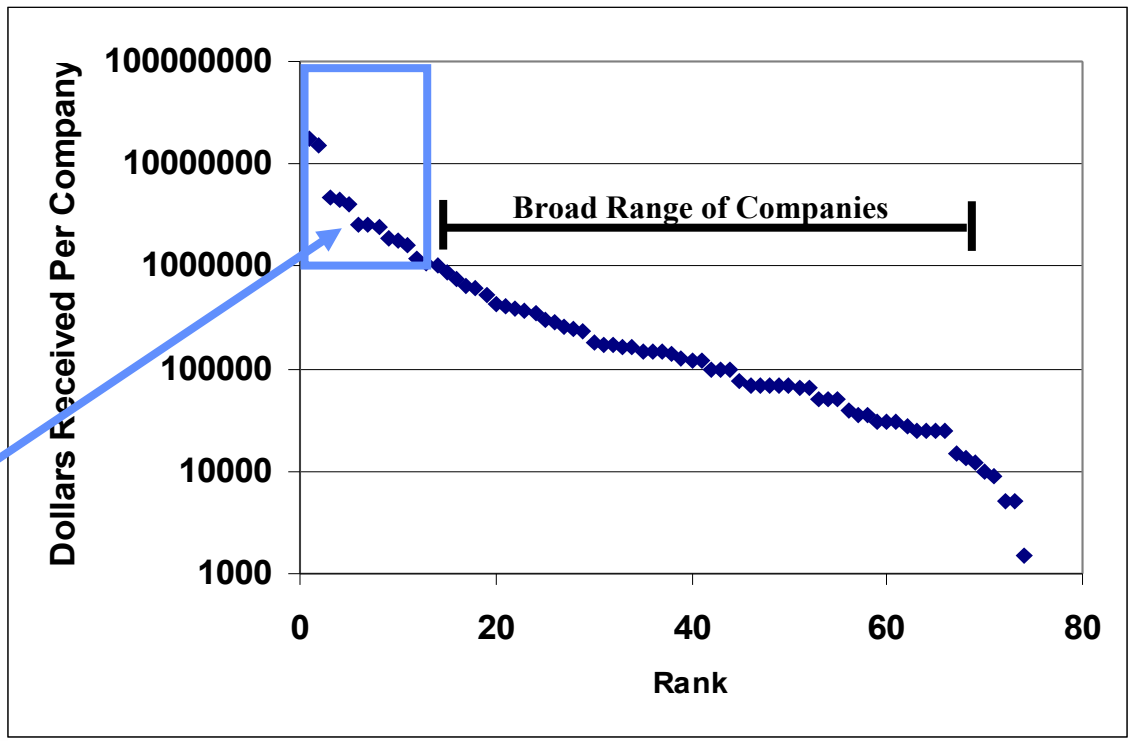
it²

In Spite of the Bubble Bursting, Calit2 Has Partnered with over 130 Companies



Industrial Partners > $1 Million

More Than 80 Have Provided Funds or In-kind

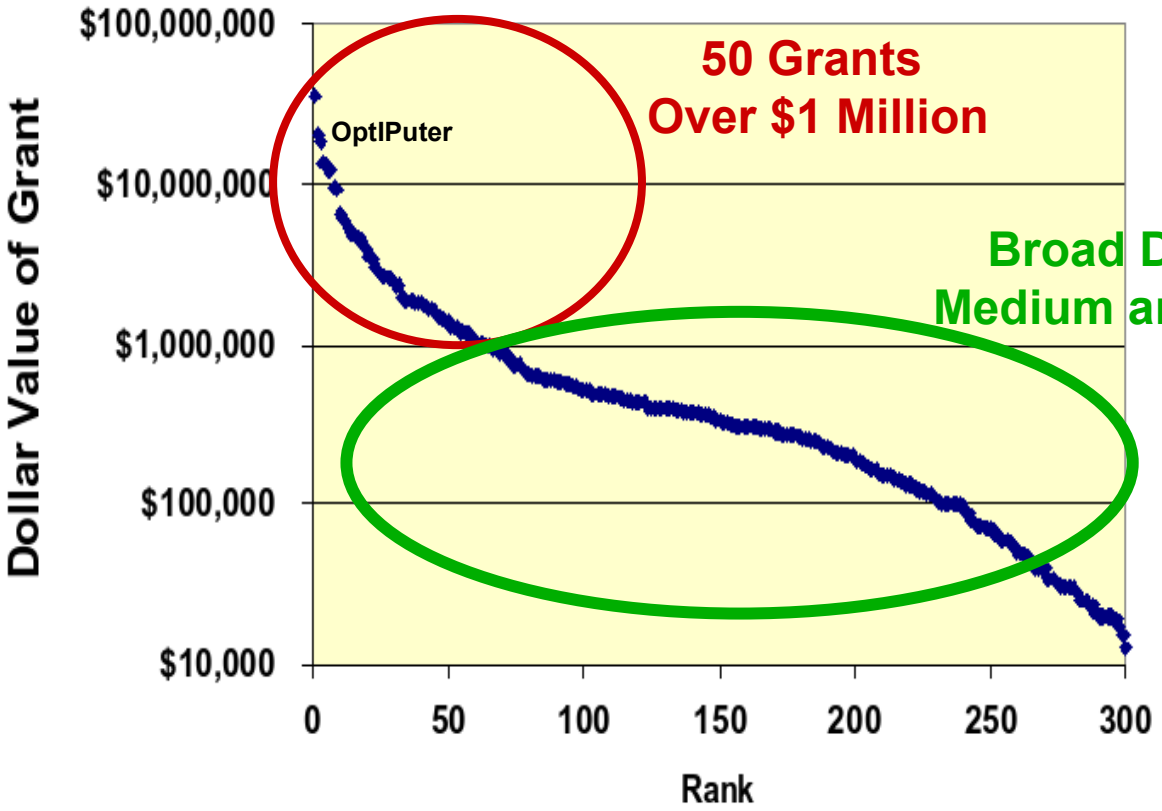


$85 Million from Industrial Partners in Matching Funds

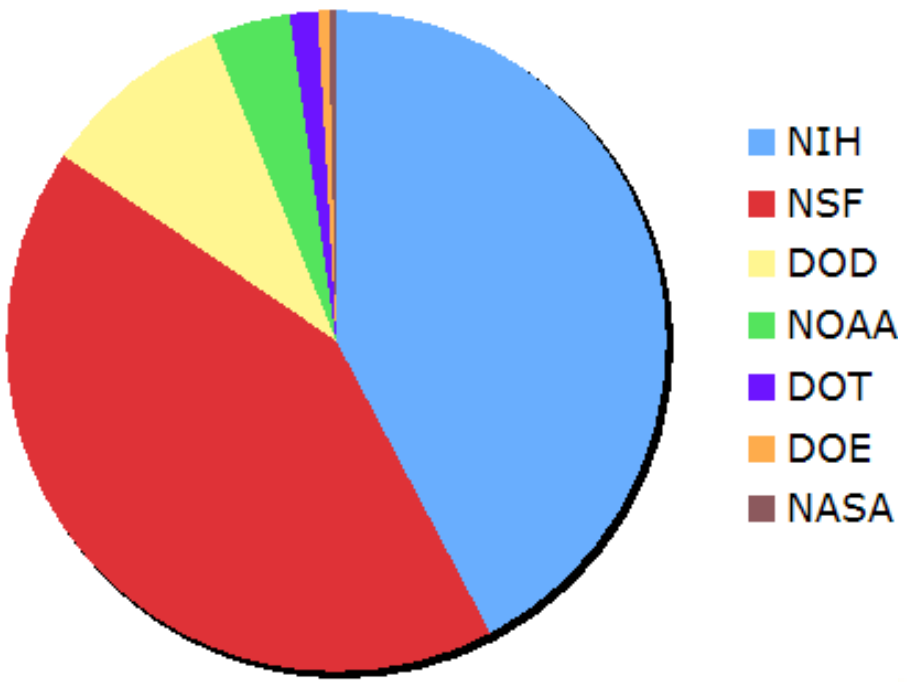
UCSD
UCIrvine

it²

Federal Agencies Have Funded $350 Million to Over 300 Calit2 Affiliated Grants



Creating a Rich Ecology of Basic Research



Federal Agency Source of Funds

UCSD
UCIrvine

Calit2 Review Report p.4,21

Calit2 Brings Computer Scientists and Engineers Together with Biomedical Researchers

- **Some Areas of Concentration:**
  - Algorithmic and System Biology
  - Bioinformatics
  - Metagenomics
  - Cancer Genomics
  - Human Genomic Variation and Disease
  - Proteomics
  - Mitochondrial Evolution
  - Biomedical Instruments
  - Multi-Scale Cellular Imaging
  - Information Theory and Biological Systems
  - Telemedicine

CAMERA Marine Microbial Ecology

NCMIR NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH at San Diego, an NIH supported resource center

National Biomedical Computation Resource

an NIH supported resource center

BIOMEDICAL INFORMATICS RESEARCH NETWORK

Information Theory and Applications

UCSD UCIrvine

Southern California Telemedicine Learning Center (TLC)

Tri-Conference at the Center for Algorithmic and Systems Biology

CASB http://casb.ucsd.edu

UCSD, November 30-December 3, 2006, San Diego, CA, USA

algorithmic biology
systems biology
computational proteomics

UC San Diego and the California Institute for Telecommunications and Information Technology (Calit2) will host three conferences devoted to the disciplines of algorithmic biology, systems biology and computational proteomics. In partnership with the International Society for Computational Biology, the RECOMB Satellite Conferences on Systems Biology and Computational Proteomics will provide a forum for new methods that broadly address either or both of these areas. Algorithmic Biology will feature leaders in the field of computational biology who will discuss the emerging challenges at the interface of biology and computer science.

ALGORITHMIC BIOLOGY
NOVEMBER 30, 2006

- PAVEL PEVZNER, UCSD (CHAIR)
- SERAFIM BATZOGLOU, STANFORD
- DAVID HAUSSLER, UC SANTA CRUZ
- MANOLIS KELLIS, MIT
- RON SHAMIR, TEL AVIV UNIV
- MICHAEL WATERMAN, USC



CALIT2 AUDITORIUM
ATKINSON HALL, UCSD
LA JOLLA, CALIFORNIA

SYSTEMS BIOLOGY
DECEMBER 1-2, 2006

- TREY IDEKER, UCSD (CHAIR)
- MARTHA BULYK, HARVARD MEDICAL
- ERIC DAVIDSON, CALTECH
- DAVID EISENBERG, UCLA
- ANNE-CLAUDE GAVIN, EMBL HEIDELBERG
- MARK GERSTEIN, YALE
- DAPHNE KOLLER, STANFORD
- BING REN, UCSD
- RON SHAMIR, TEL AVIV UNIV

Calit2 | the UC Discovery Grant

COMPUTATIONAL PROTEOMICS
DECEMBER 2-3, 2006

- VINEET BAFNA, UCSD (CHAIR)
- MARSHALL BERN, XEROX PARC
- STEVE BRIGGS, UCSD
- NEIL KELLEHER, UNIV OF ILLINOIS
- NEVAN KROGAN, UCSF
- FOREST WHITE, MIT
- RICHARD SMITH, PACIFIC NORTHWEST NATIONAL LAB

UCSD SPONSORS

ISCB | molecular systems biology

Calit2 Facilitated Formation of the Center for Algorithmic and Systems Biology

CASB

CASB Brings Together Researchers from Scripps, Burnham, GNF and Five UCSD Departments







http://casb.ucsd.edu/

it2

Challenge: What is the Appropriate Data Infrastructure for a 21st Century Data-Intensive BioMedical Campus?

- **Needed: a High Performance Biological Data Storage, Analysis, and Dissemination Cyberinfrastructure that Connects:**
  - **Genomic and Metagenomic Sequences**
  - **MicroArrays**
  - **Proteomics**
  - **Cellular Pathways**
  - **Federated Repositories of Multi-Scale Images**
    - **Full Body to Microscopy**
- **With Interactive Remote Control of Scientific Instruments**
- **Multi-level Storage and Scalable Computing**
- **Scalable Laboratory Visualization and Analysis Facilities**
- **High Definition Collaboration Facilities**

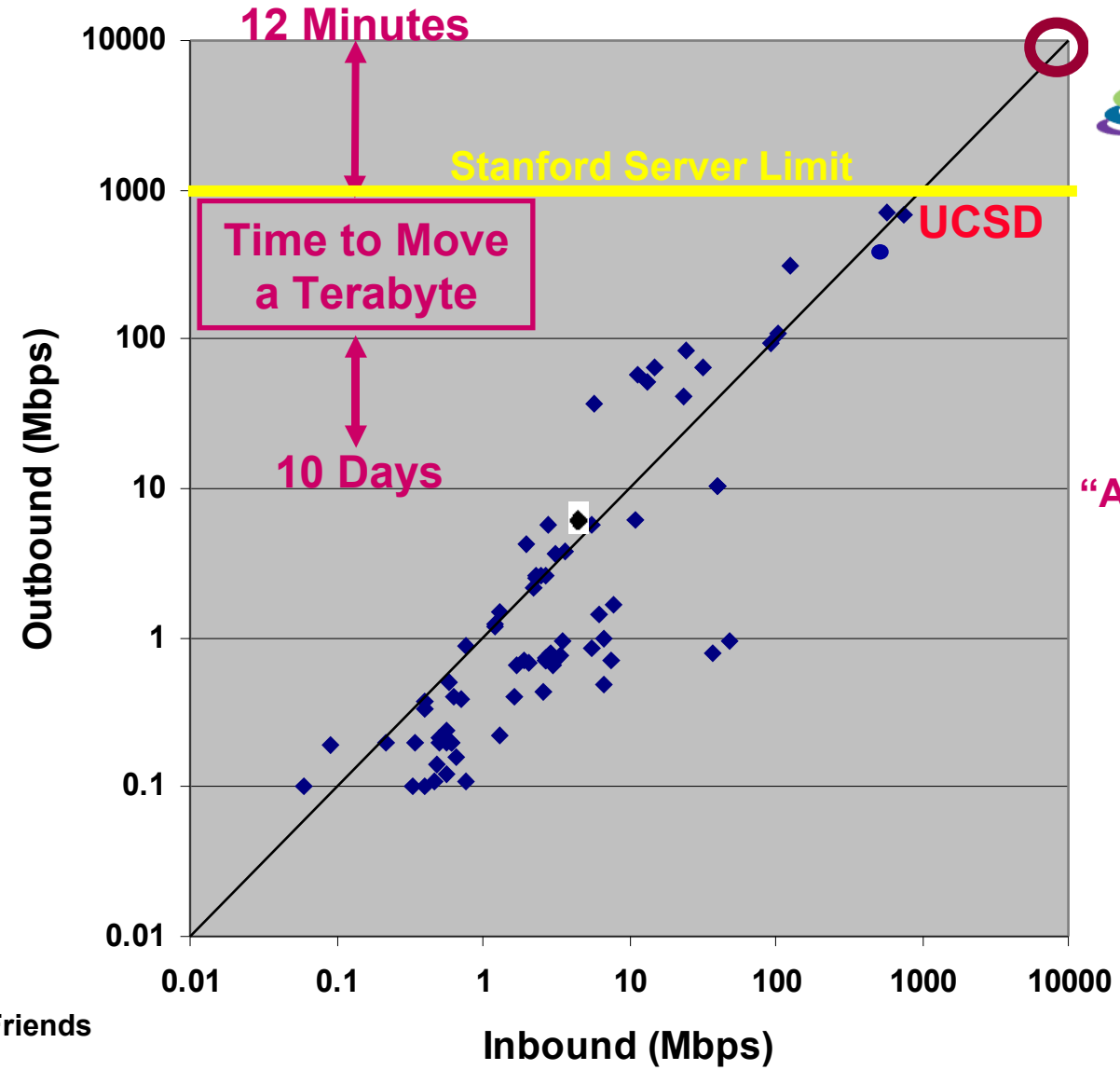
UCSD
UCIrvine

it²

Shared Internet Bandwidth: Unpredictable, Widely Varying, Jitter, Asymmetric

Computers In:

Australia
Canada
Czech Rep.
India
Japan
Korea
Mexico
Moorea
Netherlands
Poland
Taiwan
United States



OptIPuter

1000x Normal Internet!

"Average" Bandwidth

Data Intensive Sciences Require Fast Predictable Bandwidth

Source: Larry Smarr and Friends

UCSD
UCIrvine

Measured Bandwidth from User Computer to Stanford Gigabit Server in Megabits/sec
http://netspeed.stanford.edu/

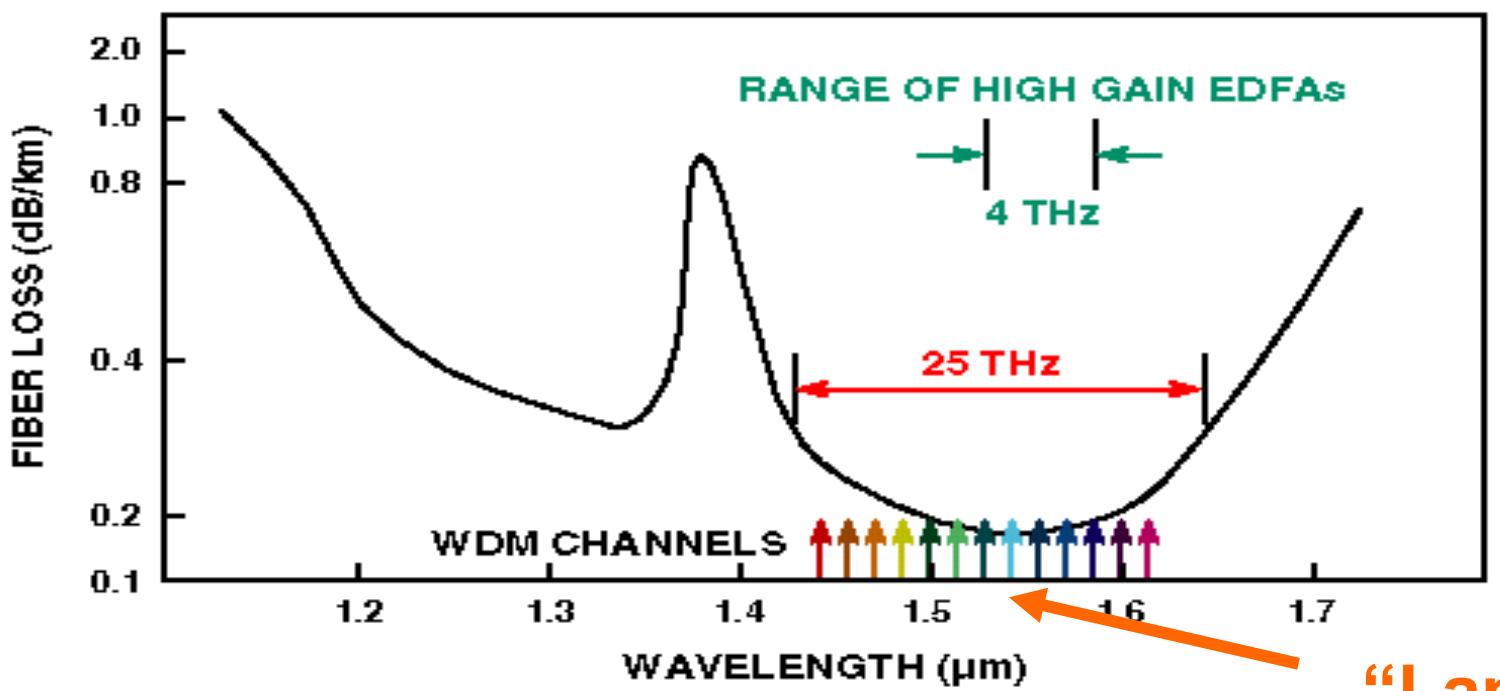
it2

Dedicated Optical Fiber Channels Makes High Performance Cyberinfrastructure Possible

WAVELENGTH DIVISION MULTIPLEXING (WDM)

10 Gbps per User ~ 500x Shared Internet Throughput

- EXPLOITS
  - ENORMOUS BANDWIDTH OF SILICA FIBER
  - HIGH-GAIN WIDEBAND OPTICAL AMPLIFIERS



"Lambdas"

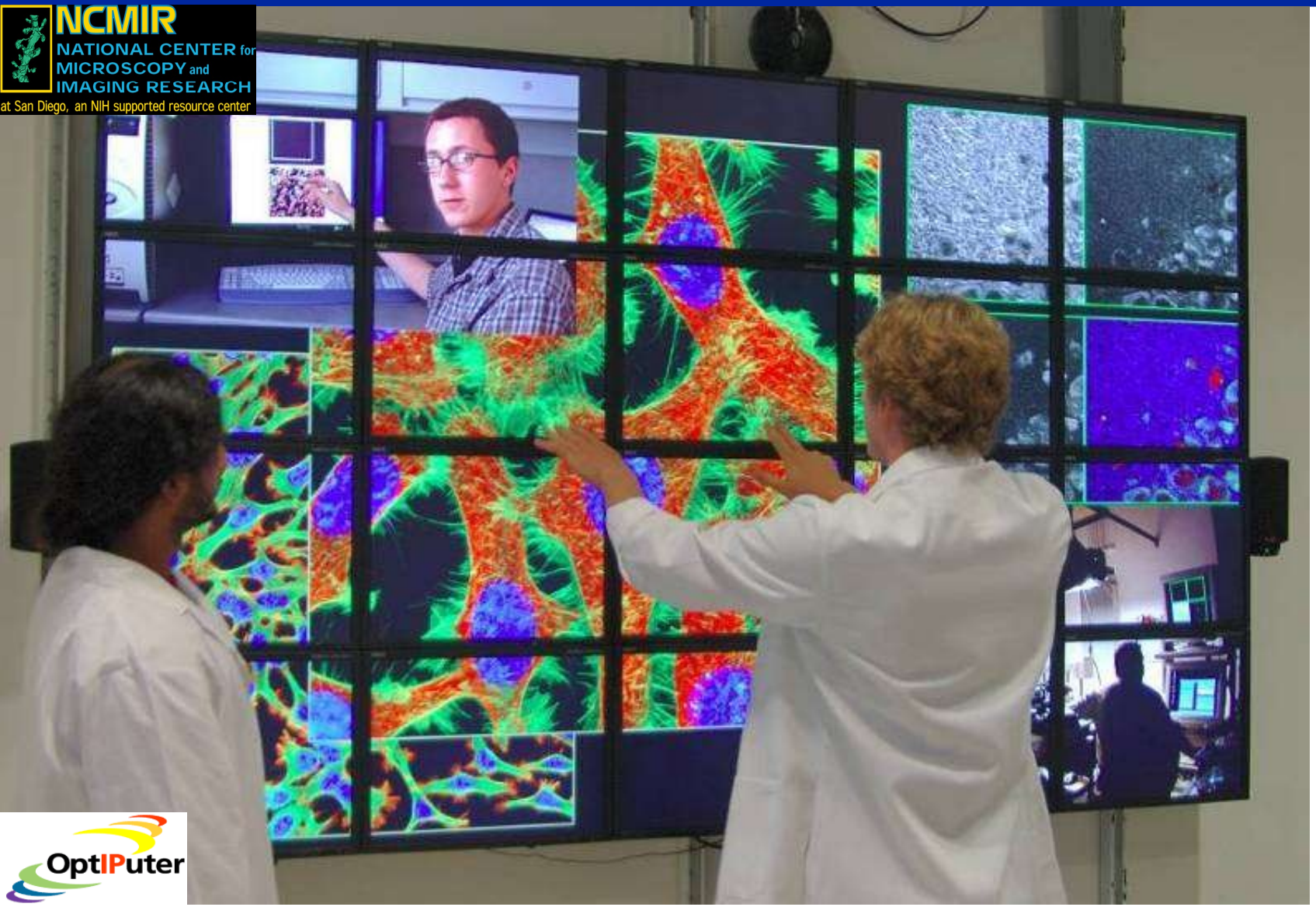
Parallel Lambdas are Driving Optical Networking The Way Parallel Processors Drove 1990s Computing

UCSD UCIrvine

it²

The OptIPuter Project: Creating High Resolution Portals Over Dedicated Optical Channels to Global Science Data

NCMIR
NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH
at San Diego, an NIH supported resource center



Scalable Adaptive Graphics Environment (SAGE)

evl electronic visualization laboratory

NSF

$13.5M Over Five Years

Picture Source: Mark Ellisman, David Lee, Jason Leigh

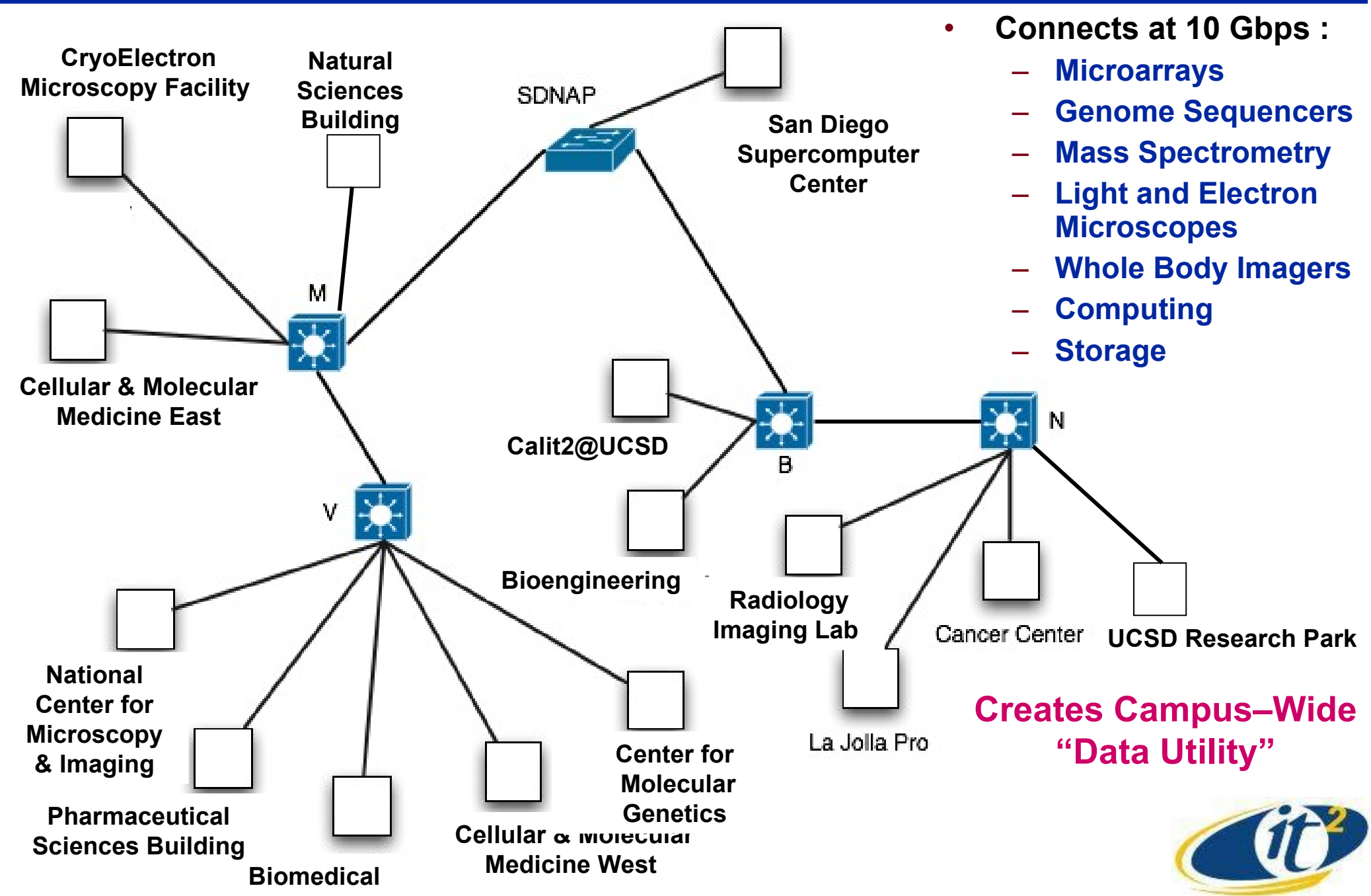
OptIPuter

UCSD
UCIrvine

Calit2 (UCSD, UCI) and UIC Lead Campuses—Larry Smarr PI
Univ. Partners: SDSC, USC, SDSU, NW, TA&M, UvA, SARA, KISTI, AIST
Industry: IBM, Sun, Telcordia, Chiaro, Calient, Glimmerglass, Lucent

it2

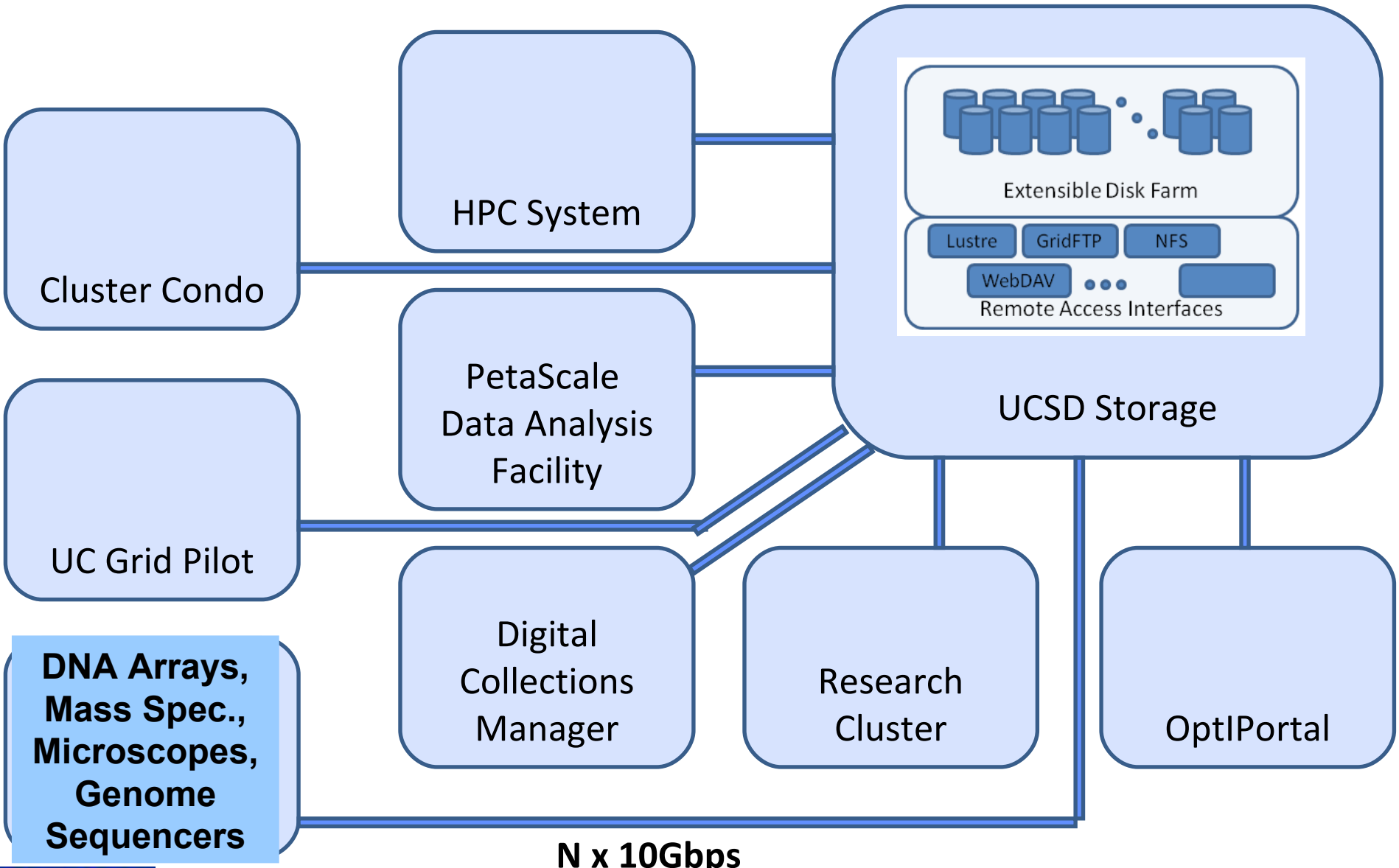
UCSD Planned Optical Networked Biomedical Researchers and Instruments



- **Connects at 10 Gbps :**
  - **Microarrays**
  - **Genome Sequencers**
  - **Mass Spectrometry**
  - **Light and Electron Microscopes**
  - **Whole Body Imagers**
  - **Computing**
  - **Storage**

Creates Campus–Wide "Data Utility"

Conceptual Architecture to Physically Connect Campus Resources Using Fiber Optic Networks



Source:Phil Papadopoulos, SDSC/Calit2

New Compute/Storage Solution for Research Parks: Optically Connected "Green" Modular Datacenters

UCSD Structural Engineering Dept.

Conducted Tests

Sun microsystems

UCSD Jacobs School of Engineering



7 Racks plus Network



Takes up 2 Parking Spaces



Data Power Cooling

- **Measure and Control Energy Usage:**
  - **Sun Has Shown up to 40% Reduction in Energy**
  - **Active Management of Disks, CPUs, etc.**
  - **Measures Temperature at 40 Points (5 Spots in 8 Racks)**
  - **Power Utilization in Each of the 8 Racks**

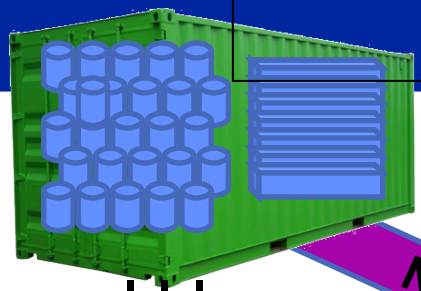
UCSD (Calit2 & School of Medicine) Bought Two Sun Boxes May 2008



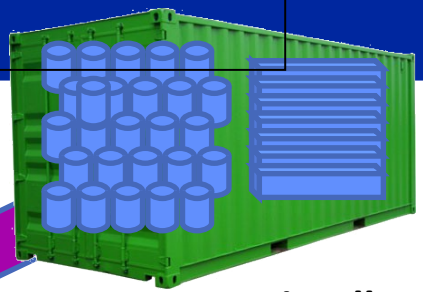
UCSD UCIrvine

it²

Planned UCSD Energy Instrumented Cyberinfrastructure



Active Data Replication



N x 10 Gbit

N x 10 Gbit

Eco-Friendly
Storage and Compute

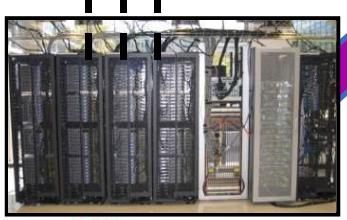
"Network in a box"
- > 200 Connections
- DWDM or Gray Optics



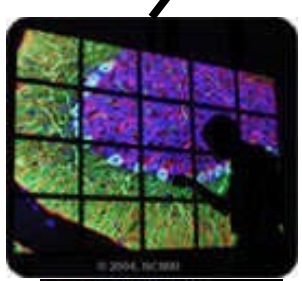
10 Gigabit L2/
L3 Switch

Wide-Area 10G
- Cenic/HPR
- NLR Cavewave
- Cinegrid
- …

On-Demand Physical
Connections



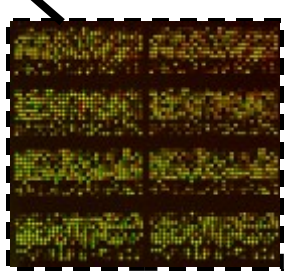
camera
Marine Microbial Ecology



NCMIR
NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH



NBCR



Microarray

OptIPuter

Your
Lab
Here

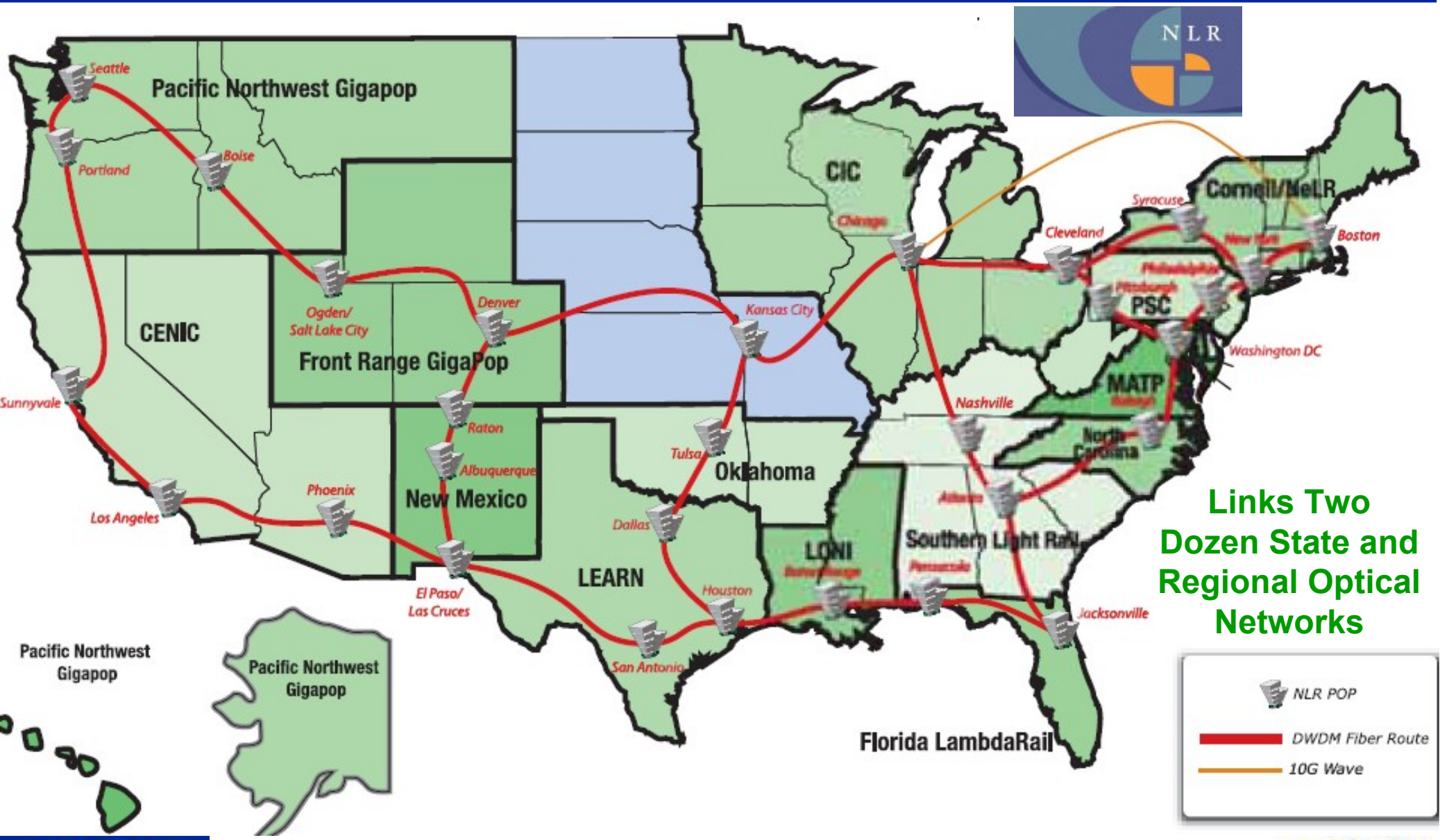
UCSD
UCIrvine

ROCKS

Source:Phil Papadopoulos, SDSC/Calit2

it2

National Lambda Rail (NLR) Provides Cyberinfrastructure Backbone for U.S. Researchers



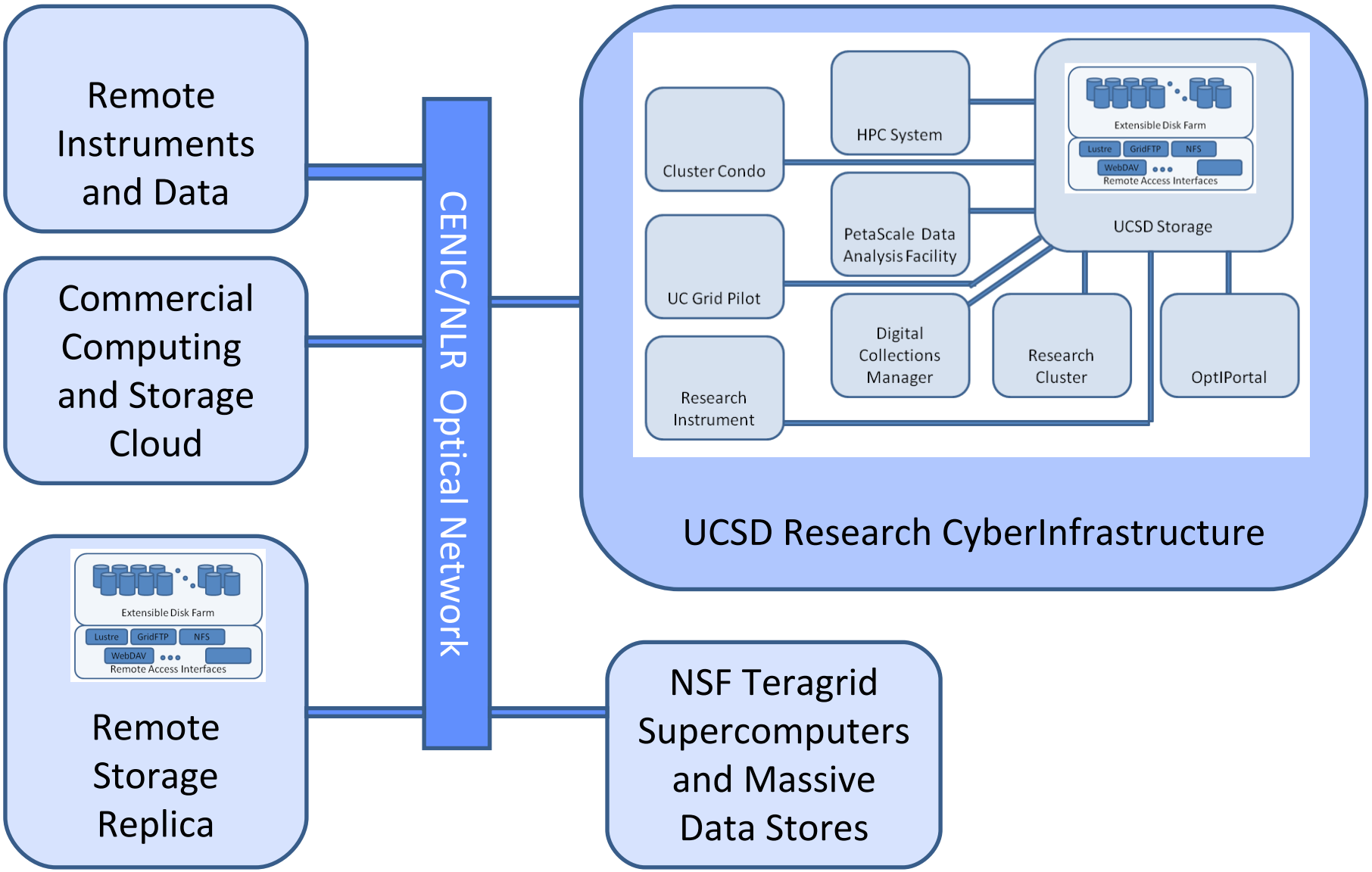
Links Two Dozen State and Regional Optical Networks

UCSD
UCIrvine

NLR 4 x 10Gb Lambdas Initially
Capable of 40 x 10Gb wavelengths at Buildout

it2

CENIC/NLR/GLIF Extend Optical Networks Outside Campus Boundaries to Remote Resources



UCSD
UCIrvine

Source:Phil Papadopoulos, SDSC/Calit2

it²

Instrument Control Services: UCSD/Osaka Univ. Link Enables Real-Time Instrument Steering and HDTV

Most Powerful Electron Microscope in the World -- Osaka, Japan





TS telescience project



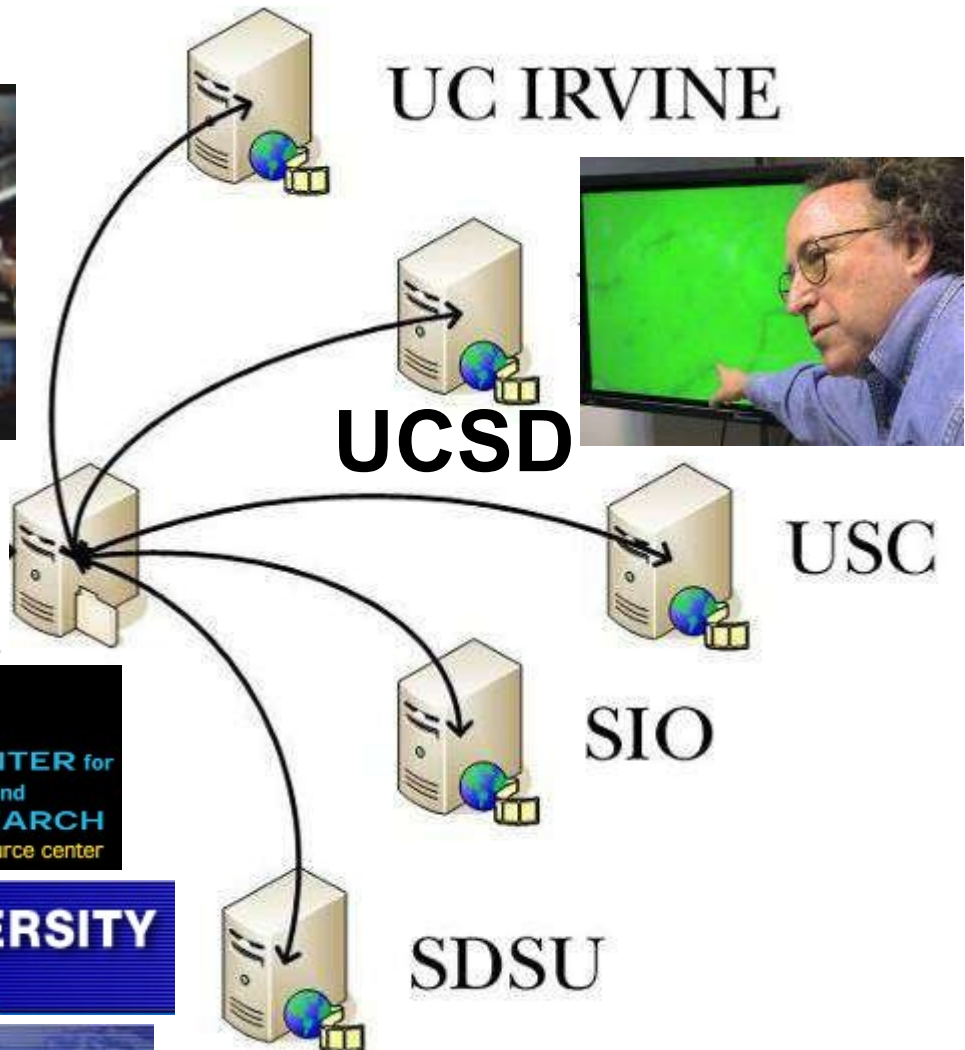
HDTV

NBCR

NCMIR NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH at San Diego, an NIH supported resource center

OSAKA UNIVERSITY

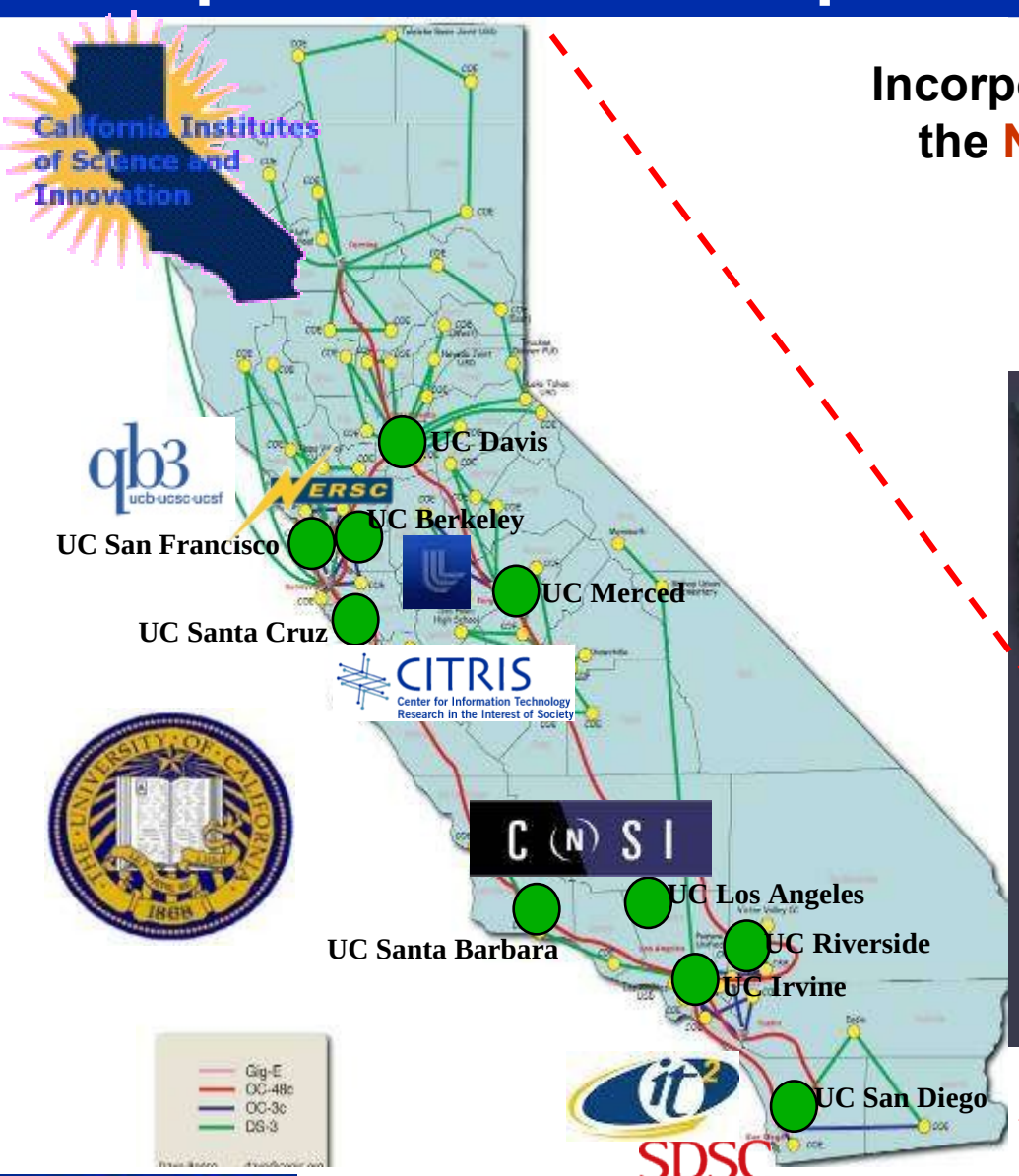
KDDI GROUP KDDI Labs USA



Source: Mark Ellisman, UCSD

it²

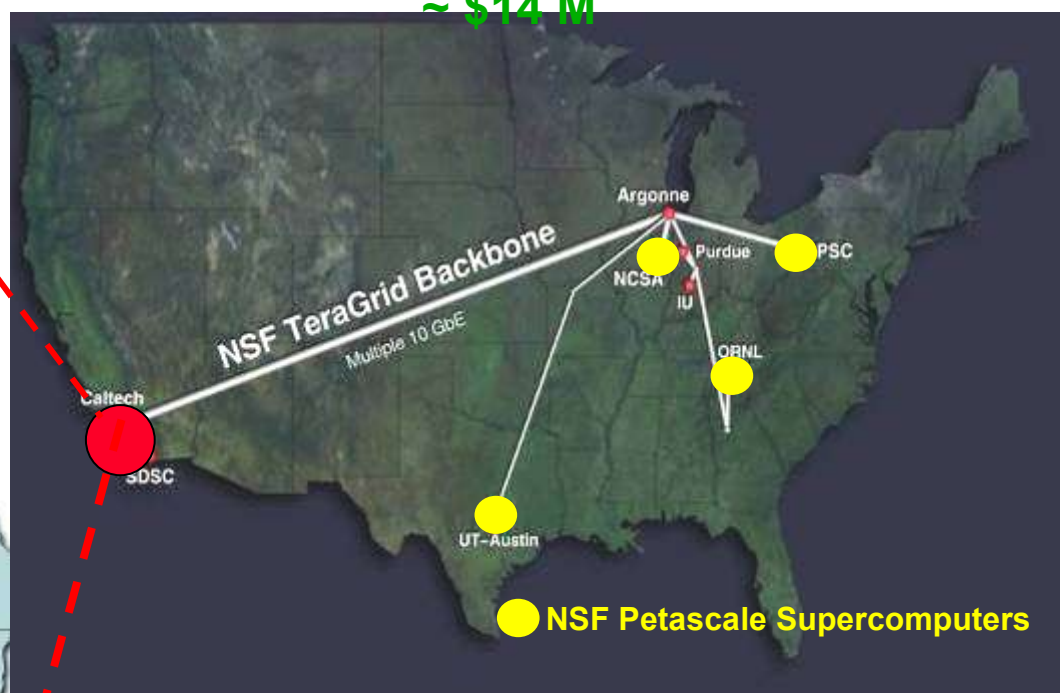
Calit2/SDSC Proposal to Create a UC Cyberinfrastructure of OptIPuter "On-Ramps" to NLR & TeraGrid Resources



CENIC "Hybrid Network"
Incorporating Traditional Routed IP Service and the New Frame and Optical Circuit Services:

Layer 3: Routed IP Network
Layer 2: Switched Ethernet Network
Layer 1: Switched Optical Network

~ $14 M



Creating a Critical Mass of End Users on a Secure LambdaGrid

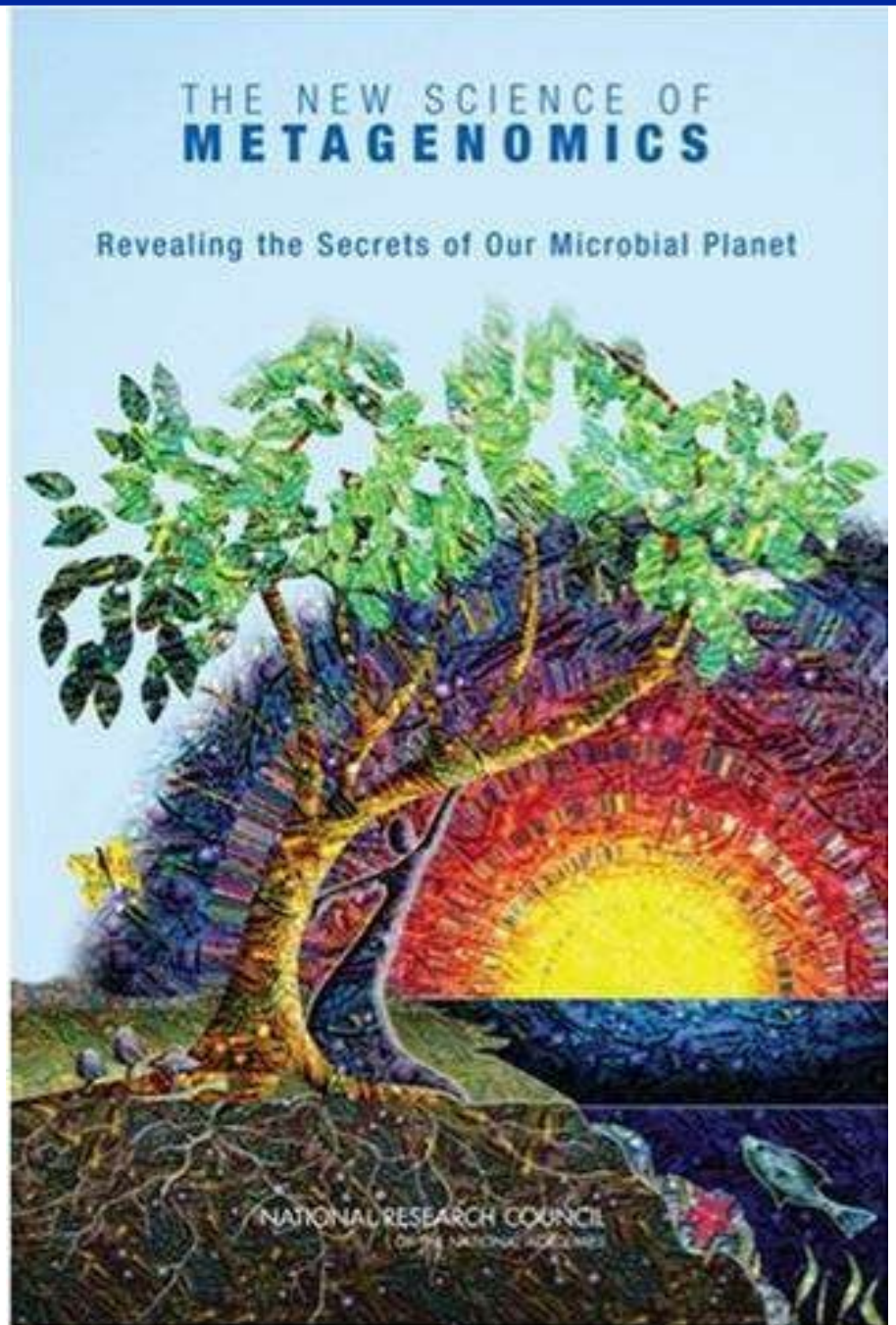
UCSD
UCIrvine

Source: Fran Berman, SDSC , Larry Smarr, Calit2

An OptIPuter Worked Example From The New Science of Metagenomics

NRC Report:

Metagenomic data should be made publicly available in international archives as rapidly as possible.

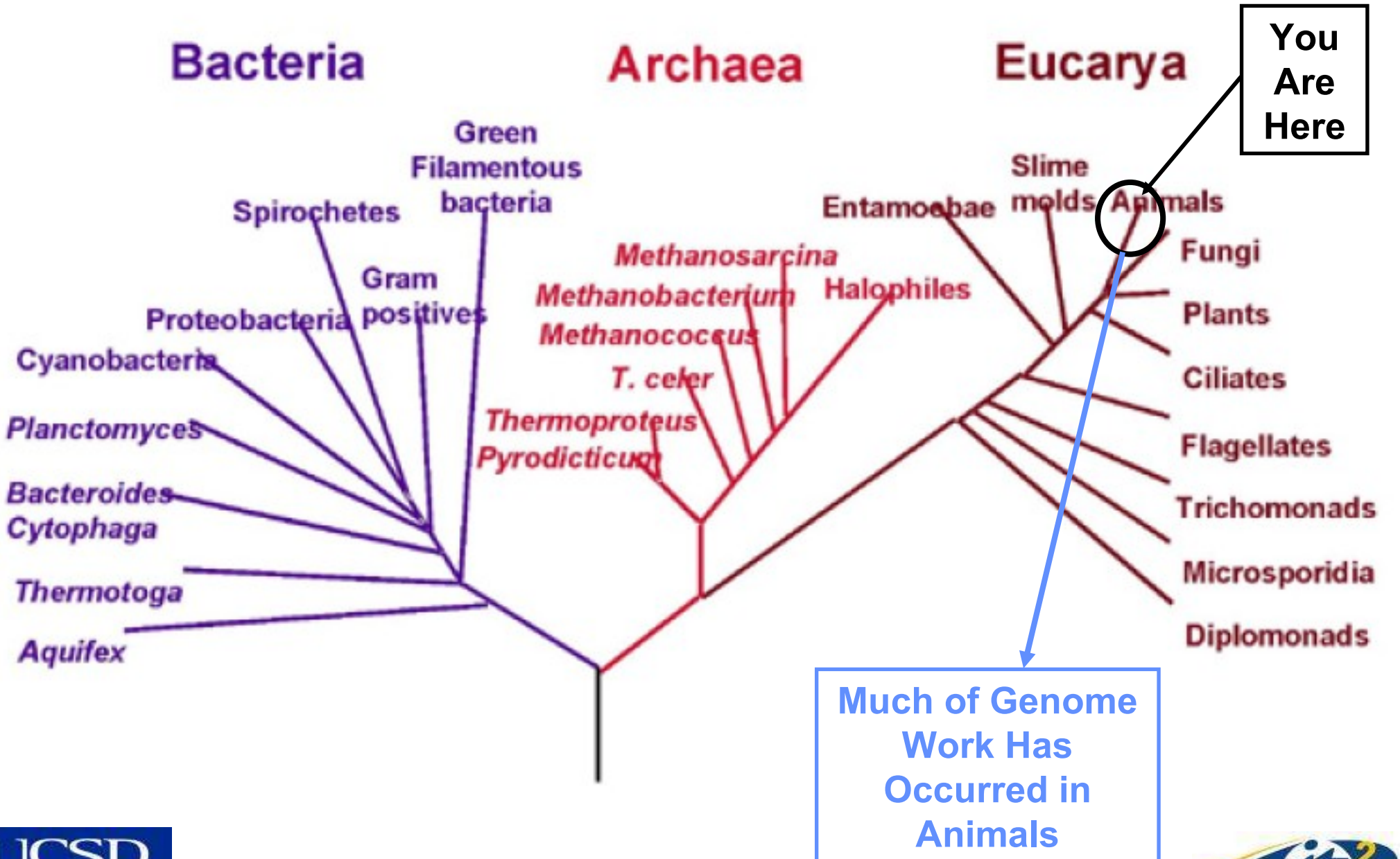


"The emerging field of metagenomics, where the DNA of entire communities of microbes is studied simultaneously, presents the greatest opportunity -- perhaps since the invention of the microscope – to revolutionize understanding of the microbial world." –

National Research Council
March 27, 2007

UCSD
UCIrvine

Evolution is the Principle of Biological Systems: Most of Evolutionary Time Was in the Microbial World



UCSD
UCIrvine

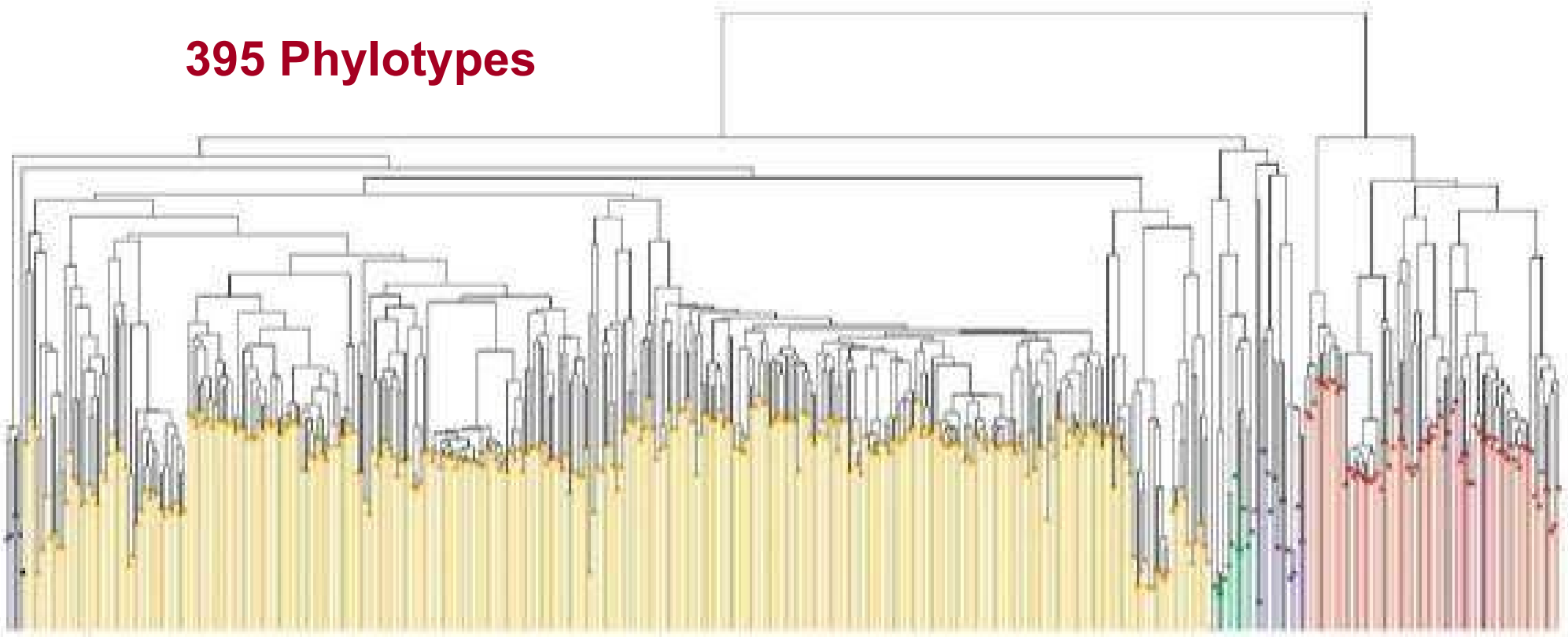
Source: Carl Woese, et al

it²

The Human Microbiome is the Next Large NIH Drive to Understand Human Health and Disease

- "A majority of the bacterial sequences corresponded to uncultivated species and novel microorganisms."
- "We discovered significant inter-subject variability."
- "Characterization of this immensely diverse ecosystem is the first step in elucidating its role in health and disease."

395 Phylotypes



UCSD
UCIrvine

"Diversity of the Human Intestinal Microbial Flora"
Paul B. Eckburg, et al *Science* (10 June 2005)

it²

Marine Genome Sequencing Project – Measuring the Genetic Diversity of Ocean Microbes

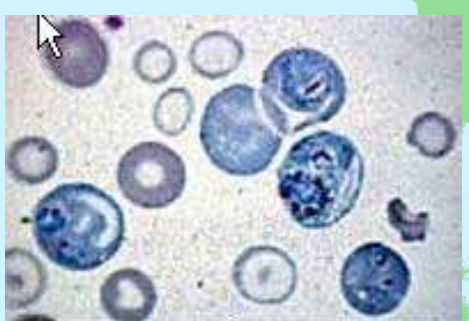


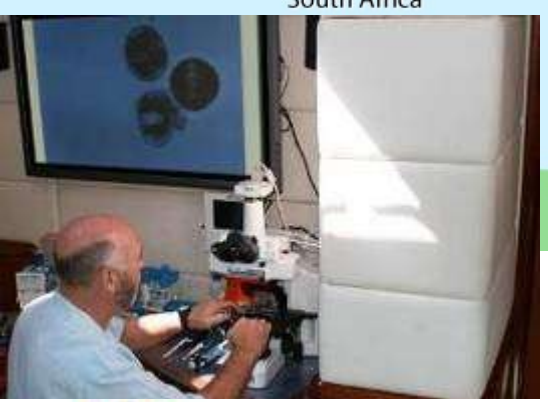
Each Sample ~2000 Microbial Species

Specify Ocean Data

Canada
United States
Mexico
Puerto Rico
Honduras
Costa Rica
Panama
Galapagos Islands
x 10 samples
French Polynesia
Cook Islands
Vanuatu
Fiji
Tonga
New Caledonia
x 17 samples
Australia
x 8 samples
Cocos-Keeling Is.
Seychelles
Tanzania/ Zanzibar
x 7 samples
Madagascar
South Africa

Analyzed
Sampled, but not Analyzed





MARINE MICROBIOLOGY INITIATIVE

camera
Marine Microbial Ecology

Gordon and Betty MOORE FOUNDATION

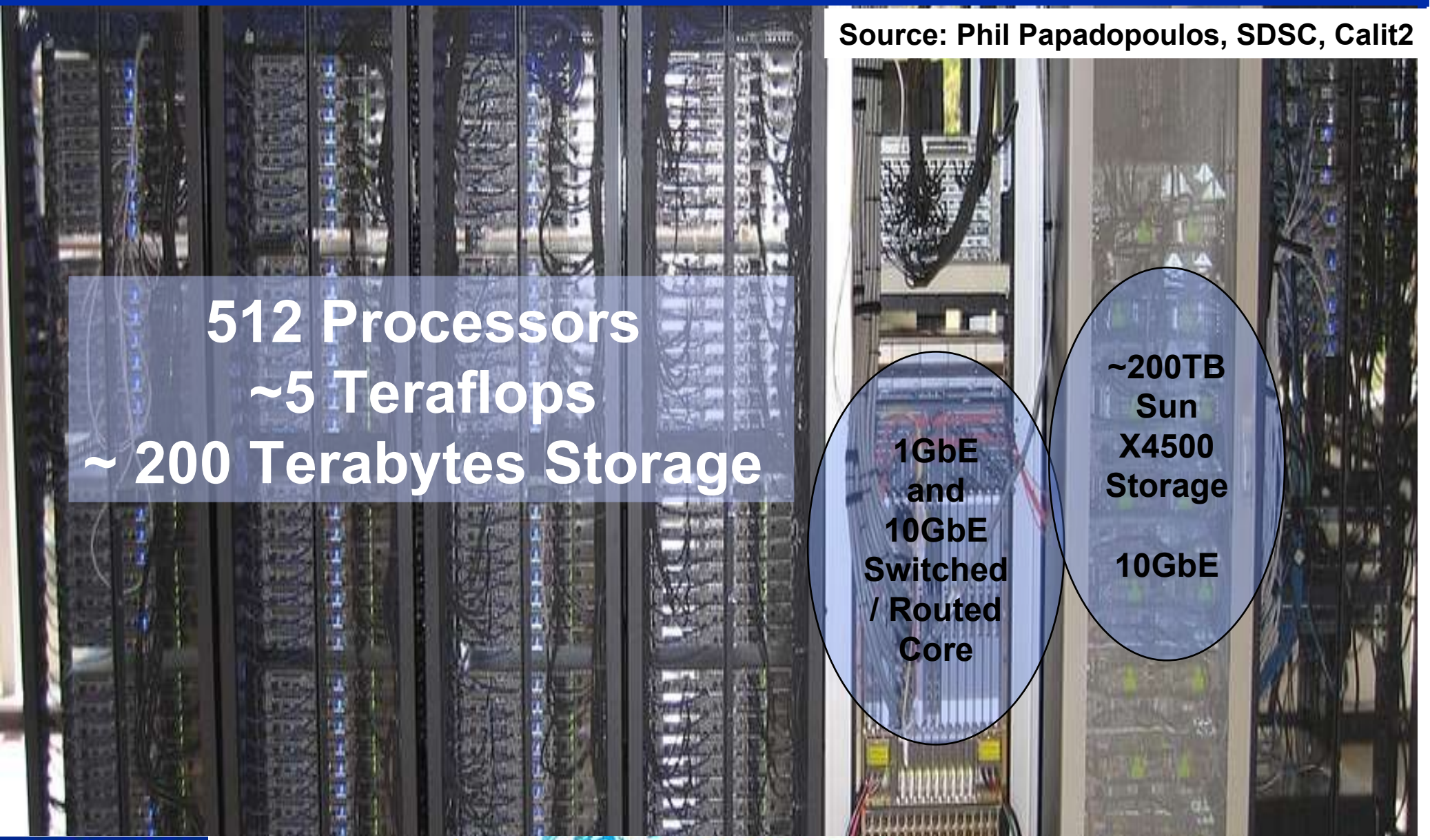
Sorcerer II Data Will Double Number of Proteins in GenBank!

J. Craig Venter INSTITUTE

it²

Calit2 Microbial Metagenomics Cluster- Next Generation Optically Linked Science Data Server

Source: Phil Papadopoulos, SDSC, Calit2



UCSD UCIrvine SDSC ROCKS camera Marine Microbial Ecology J. Craig Venter INSTITUTE it2

CAMERA's
Global Microbial Metagenomics CyberCommunity

Over 2010 Registered Users From Over 50 Countries

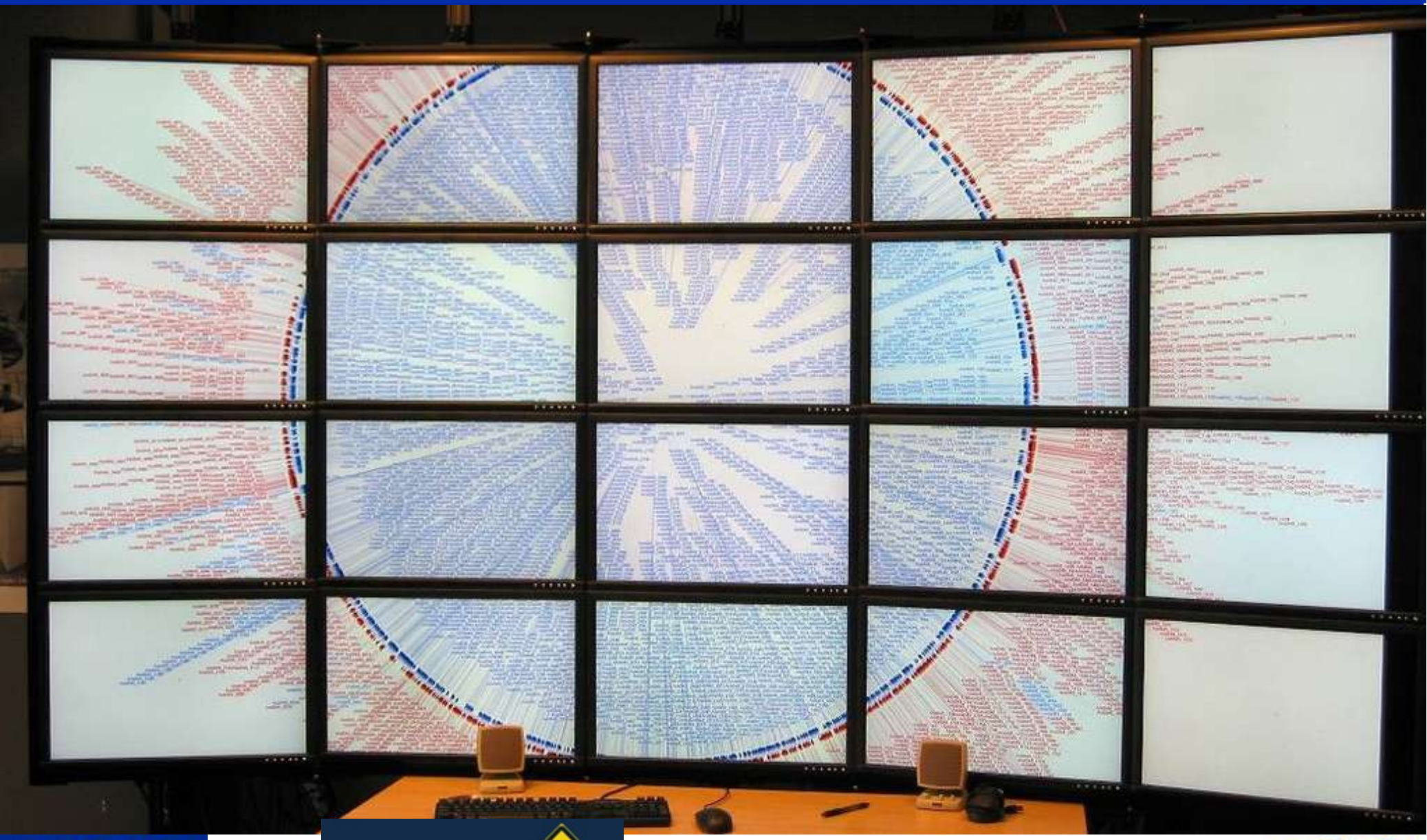


camera
Marine Microbial Ecology

J. Craig Venter INSTITUTE

it²

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



UCSD
UCIrvine

Powered by Rocks

ROCKS

Acidobacteria bacterium Ellin345
Soil Bacterium 5.6 Mb

it²

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



UCSD
UCIrvine

Source: Raj Singh, UCSD

it²

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



UCSD
UCIrvine

Source: Raj Singh, UCSD

it²

Interactive Exploration of Marine Genomes Using 100 Million Pixels

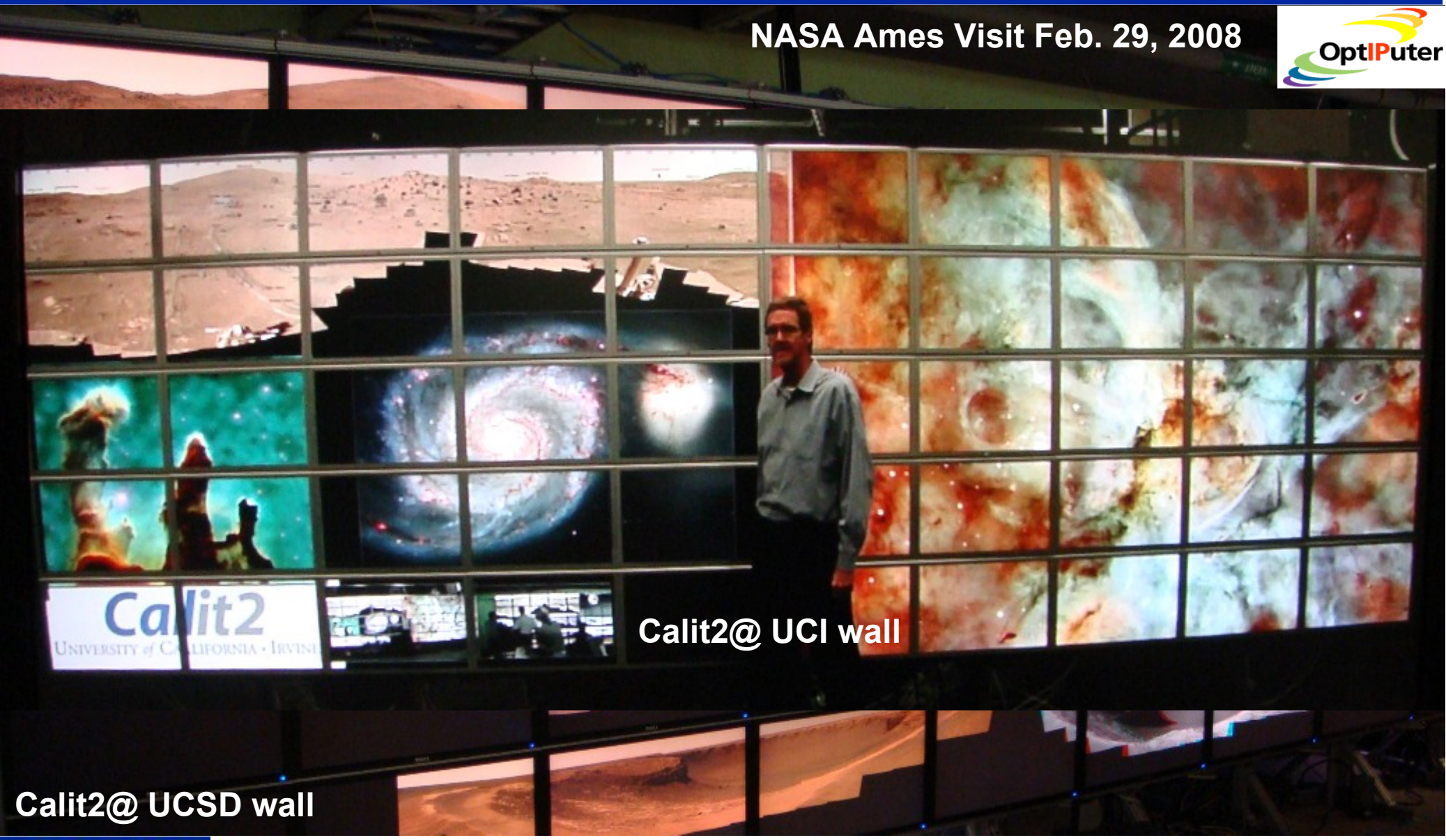


camera
Marine Microbial Ecology

Ginger Armburst (UW), Terry Gaasterland (UCSD SIO)

it²

The Calit2 200 Megapixel OptIPortals at UCSD and UCI Are Now a Gbit/s HD Collaboratory



UCSD
UCIrvine

NASA Ames is Completing a 245 Mpixel Hyperwall as Project Columbia Interface

it2

OptIPlanet Collaboratory Persistent Infrastructure Supporting Microbial Research



OptIPuter

Ginger Armbrust's Diatoms: Micrographs, Chromosomes, Genetic Assembly

ROCKS



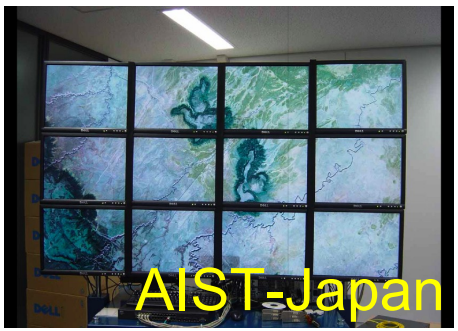




UW's Research Channel Michael Wellings

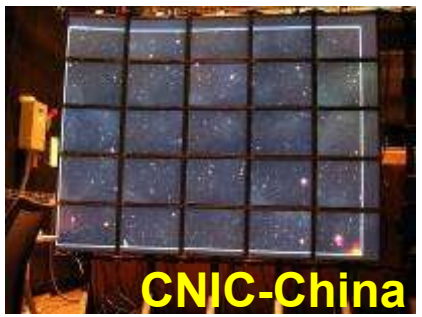
it2

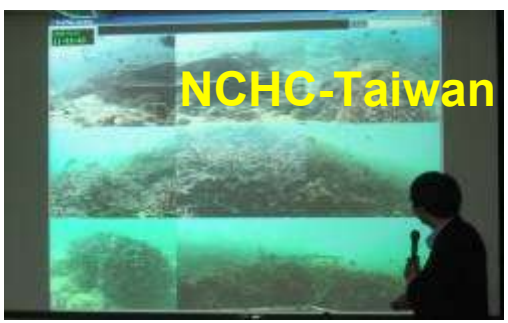
OptIPortals
Are Being Adopted Globally

























UCSD
UCIrvine

Powered by Rocks ROCKS

evl electronic visualization laboratory

it2