Using Supercomputers and Supernetworks to Explore the Ocean of Life

Moore Foundation PI Meeting

Calit2@UCSD

July 17, 2007

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

UCSD

UCIrvine

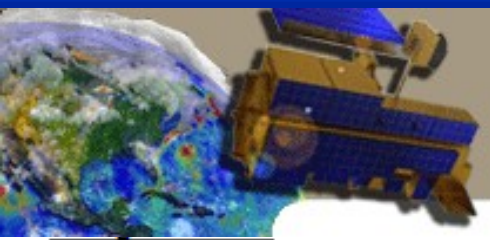
it²

Abstract

Calit2, in partnership with J. Craig Venter Institute in Rockville, MD, and UCSD's SDSC and Scripps Institution of Oceanography, is creating a Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA), funded by the Gordon and Betty Moore Foundation. CAMERA collaborates closely with DoE's Joint Genome Institute. The CAMERA computational and storage cluster containing the metagenomic data can be accessed via the web over novel dedicated 10 Gb/s light pipes (termed "lambdas") through the National LambdaRail, providing direct connection to the scalable Linux clusters in individual user laboratories. These clusters are reconfigured as "OptIPortals," providing the end user with local scalable visualization, computing, and storage. Currently over 1000 web users are registered from over 40 countries and a dozen OptIPortal sites are under construction.

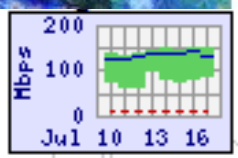
UCSD
UCIrvine

Challenge: Average Throughput of NASA Data Products to End User is 10-100 Mbps

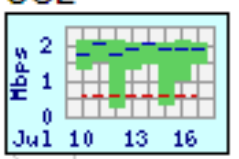


ENSIGHT Active Testing TERRA Destinations

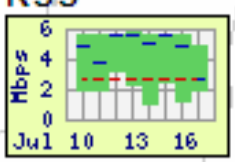
Tested July 2007



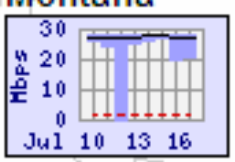
UCL



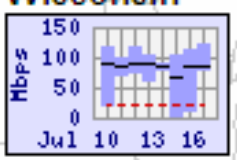
RSS



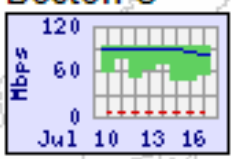
Montana



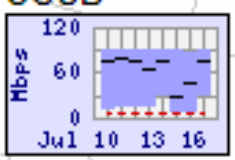
Wisconsin



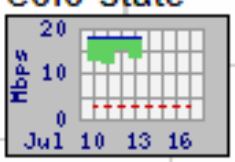
Boston U



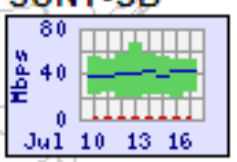
UCSB



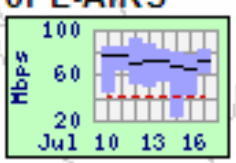
Colo State



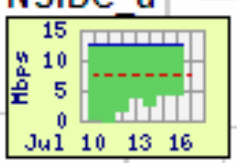
SUNY-SB



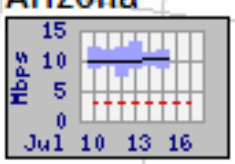
JPL-AIRS



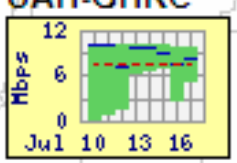
NSIDC_u



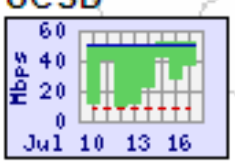
Arizona



UAH-GHRC



UCSD



OptIPuter

http://ensight.eos.nasa.gov/Missions/icesat/index.shtml

it2

Calit2's Direct Access Core Architecture Creates a SuperNetwork Metagenomics Server

Gordon and Betty MOORE FOUNDATION

Powered by Rocks ROCKS

Sargasso Sea Data

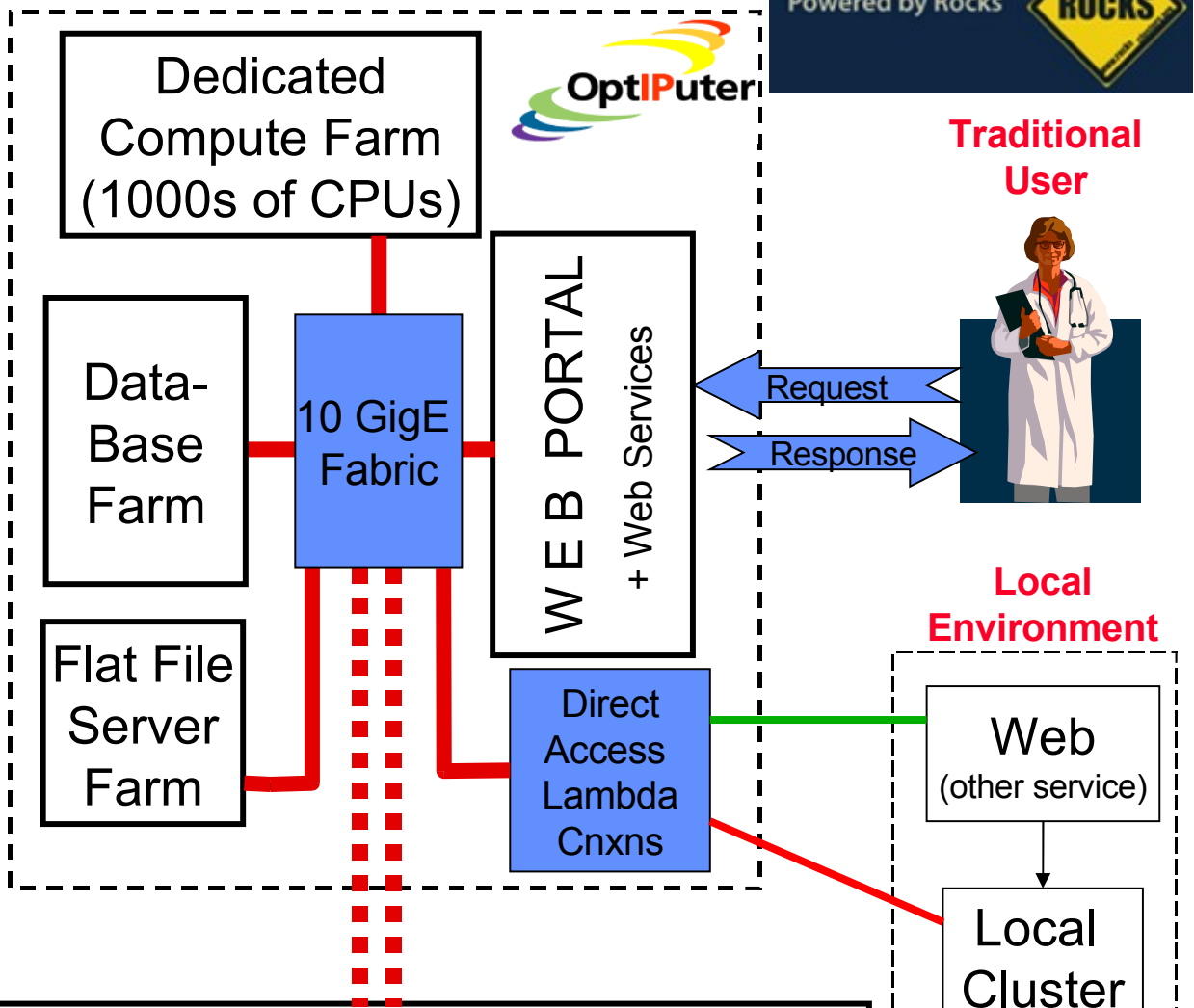
Sorcerer II Expedition (GOS)

JGI Community Sequencing Project

Moore Marine Microbial Project

NASA and NOAA Satellite Data

Community Microbial Metagenomics Data



CENTER FOR EARTH OBSERVATIONS & APPLICATIONS

SCRIPPS INSTITUTION OF OCEANOGRAPHY

SDSC SAN DIEGO SUPERCOMPUTER CENTER

camera Marine Microbial Ecology

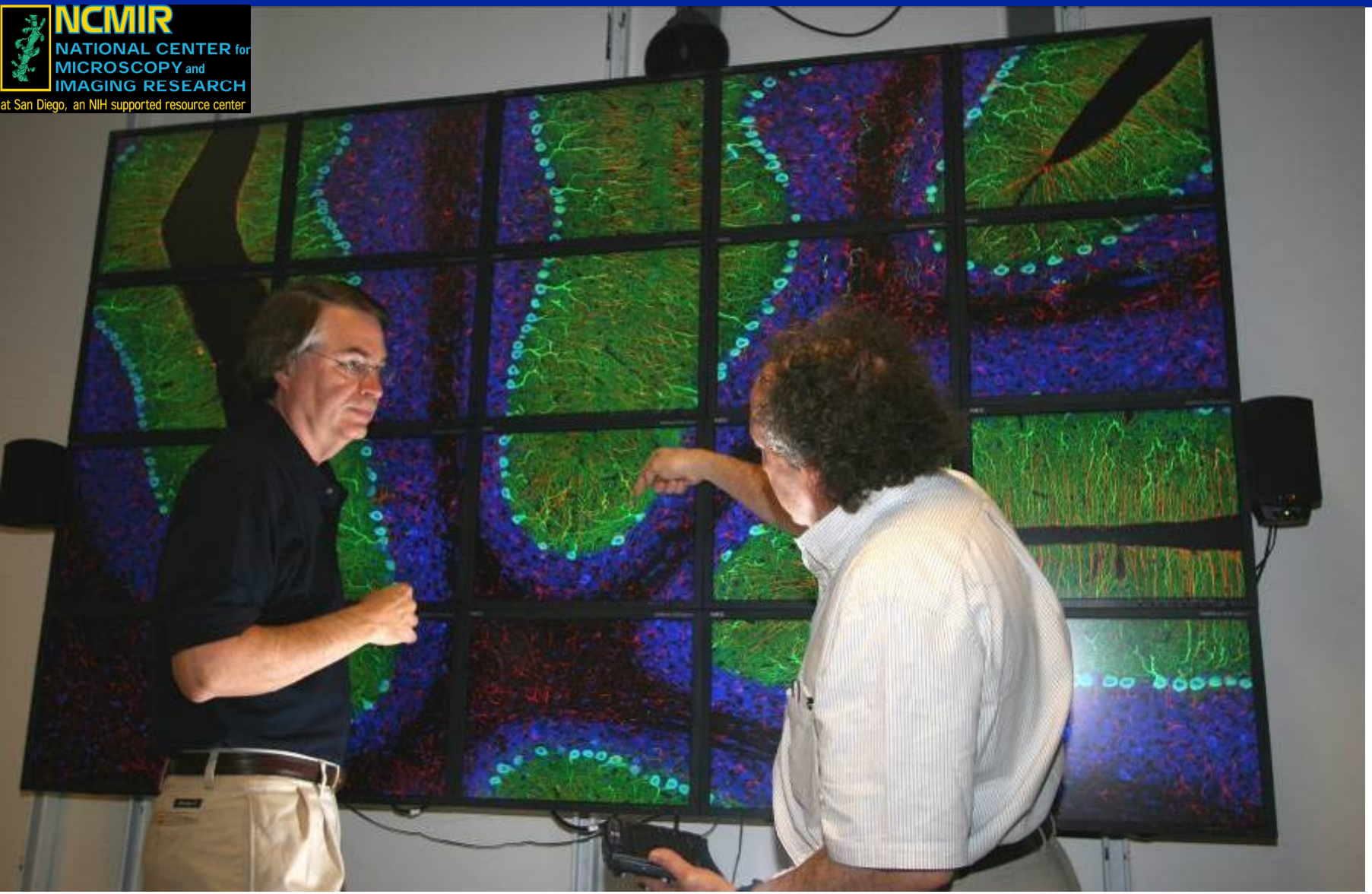
J. Craig Venter INSTITUTE

it2

Source: Phil Papadopoulos, SDSC, Calit2

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



OptIPuter

NSF

$13.5M Over Five Years

Now In the Fifth Year

Picture Source: Mark Ellisman, David Lee, Jason Leigh

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

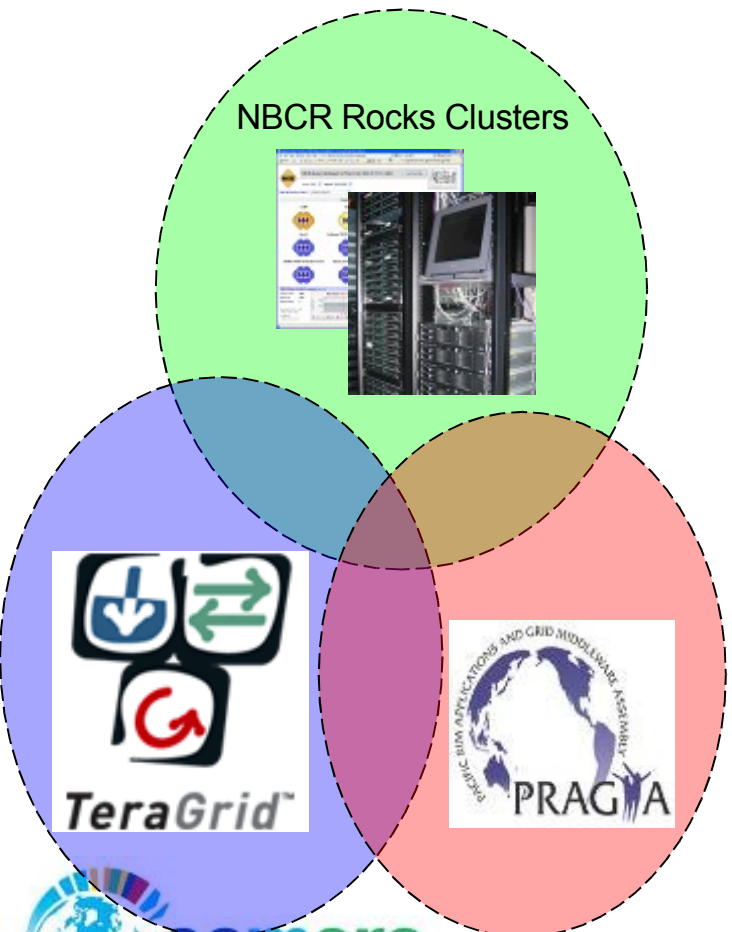
CAMERA Builds on Cyberinfrastructure Grid, Workflow, and Portal Projects in a Service Oriented Architecture

National Biomedical Computation Resource
an NIH supported resource center

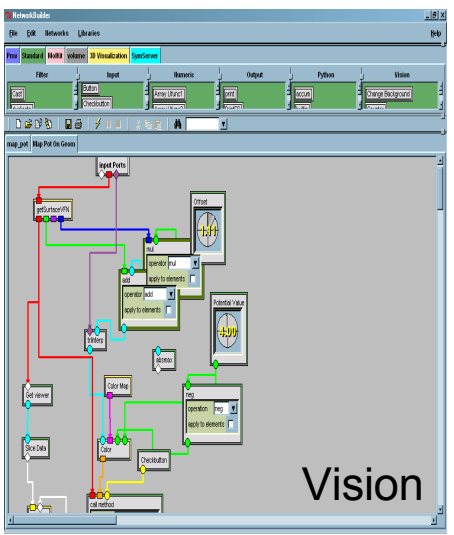
Located in Calit2@UCSD Building

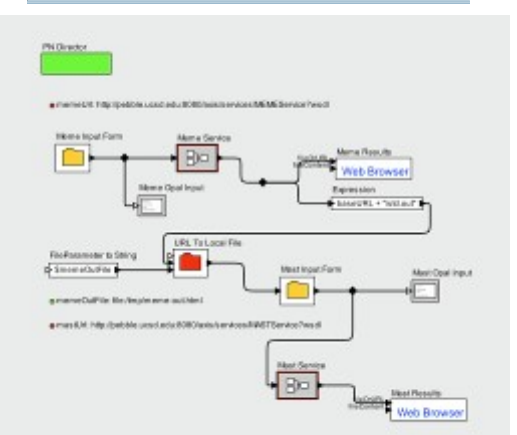
yberinfrastructure: Raw Resources, Middleware & Execution Environment

Virtual Organizations



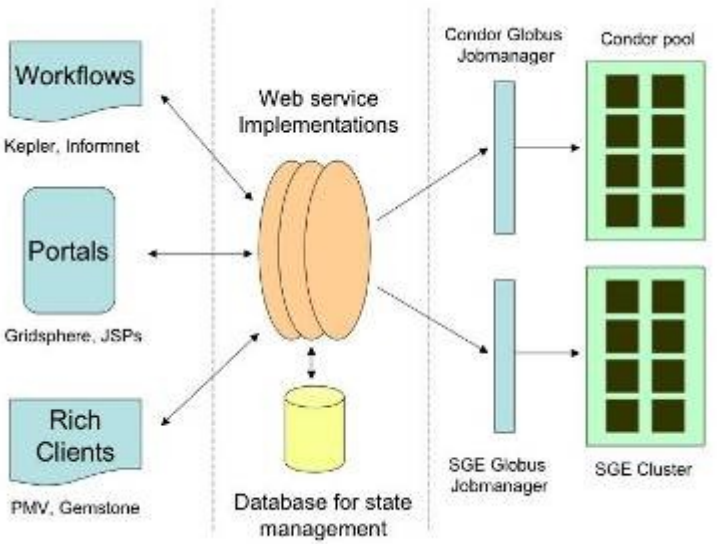
Workflow Management





KEPLER

Web Services



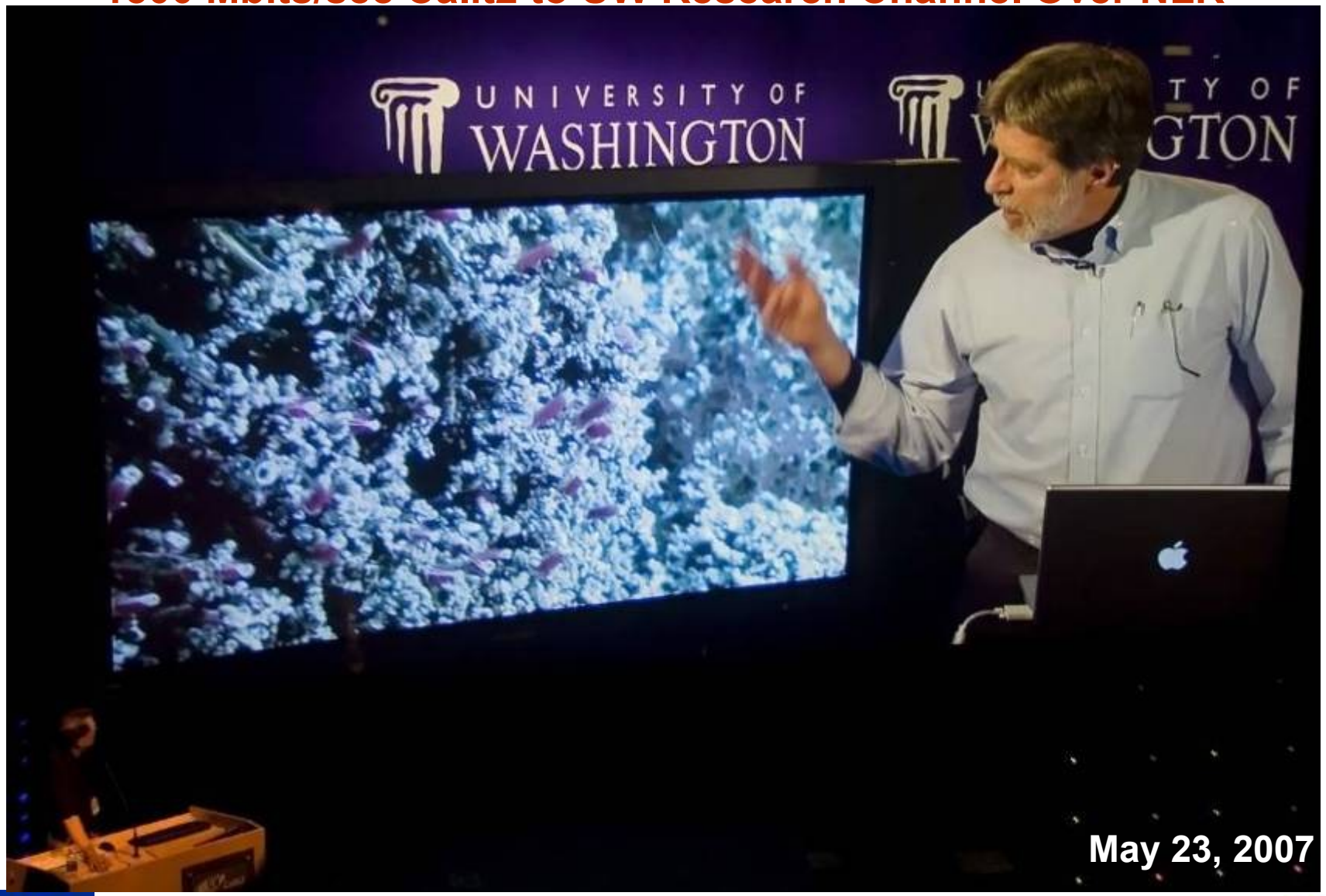
Telescience Portal

telescience project

it²

e-Science Collaboratory Without Walls Enabled by Uncompressed HD Telepresence

1500 Mbits/sec Calit2 to UW Research Channel Over NLR



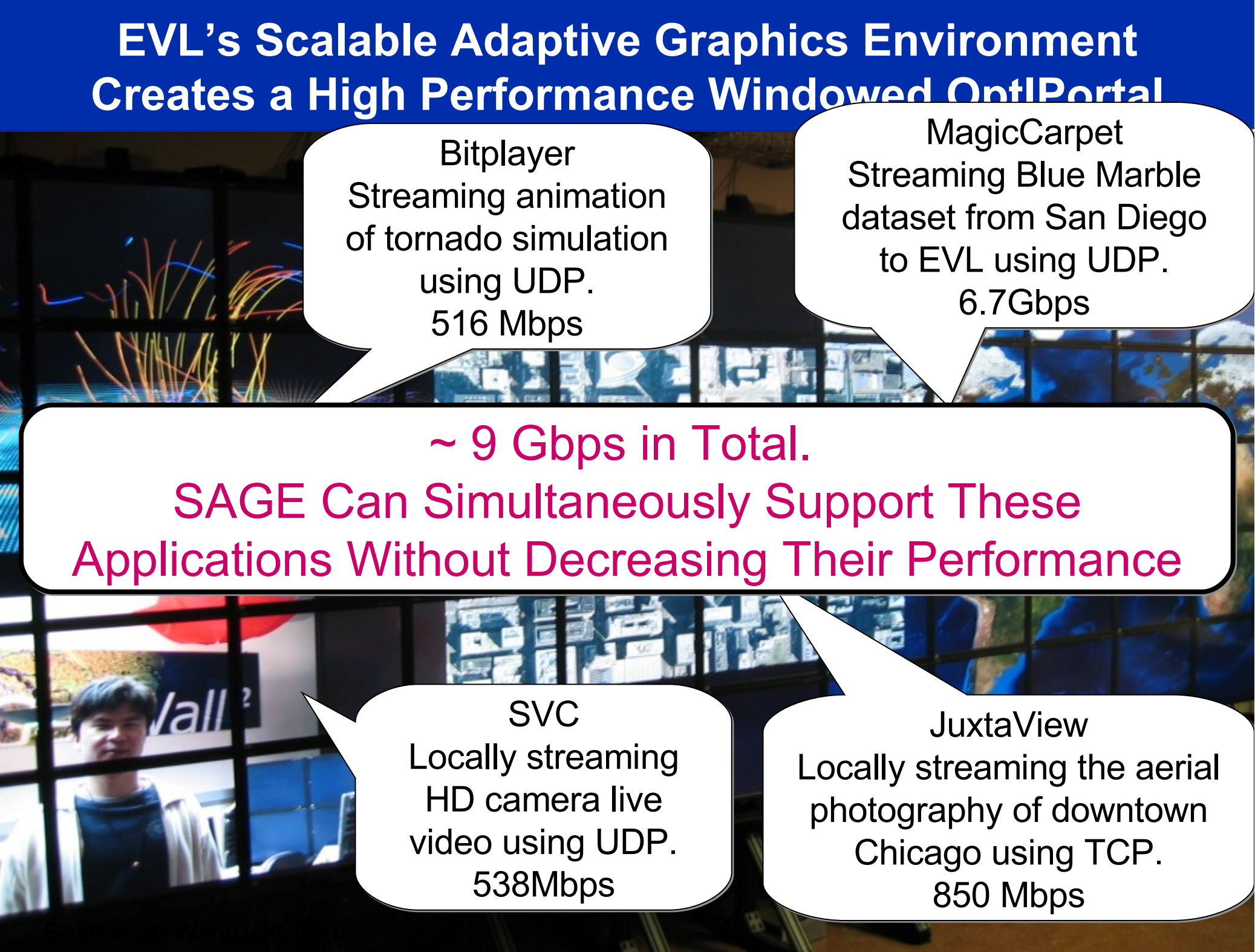
UCSD
UCIrvine

John Delaney, PI LOOKING, Neptune

Photo: Harry Ammons, SDSC

it2

EVL's Scalable Adaptive Graphics Environment Creates a High Performance Windowed OptIPortal



OptIPortal–
Termination Device for the OptIPuter Global Backplane



An Emerging High Performance Collaboratory for Microbial Metagenomics



Interactive Exploration of Marine Genomes Using 100 Million Pixels

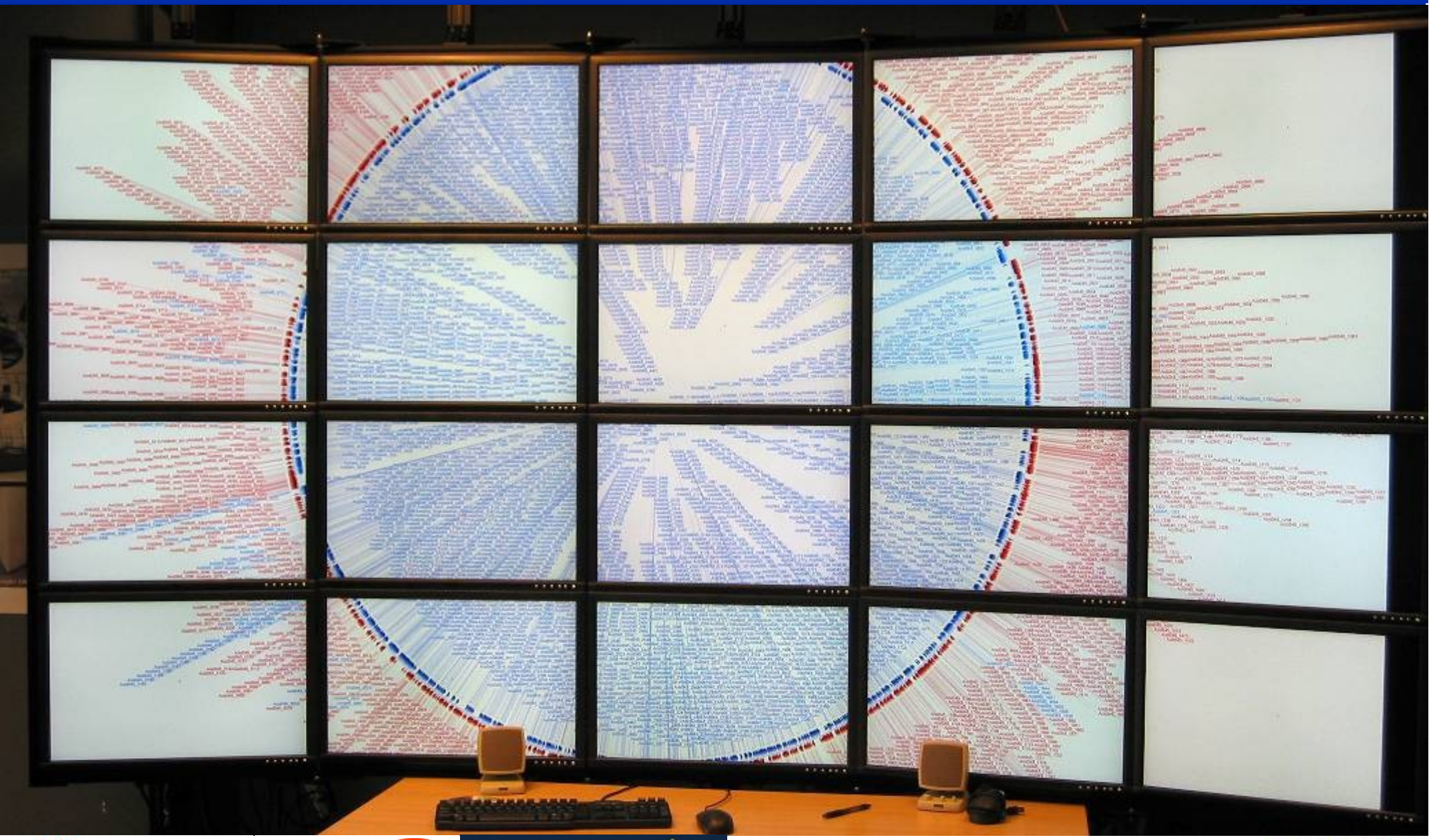


camera
Marine Microbial Ecology

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

it²

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



camera Marine Microbial Ecology

OptIPuter

Powered by Rocks ROCKS

Acidobacteria bacterium Ellin345
Soil Bacterium 5.6 Mb
Source: Raj Singh, UCSD

it2

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



camera Marine Microbial Ecology

OptIPuter

Powered by Rocks ROCKS

Source: Raj Singh, UCSD

it2

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



camera Marine Microbial Ecology

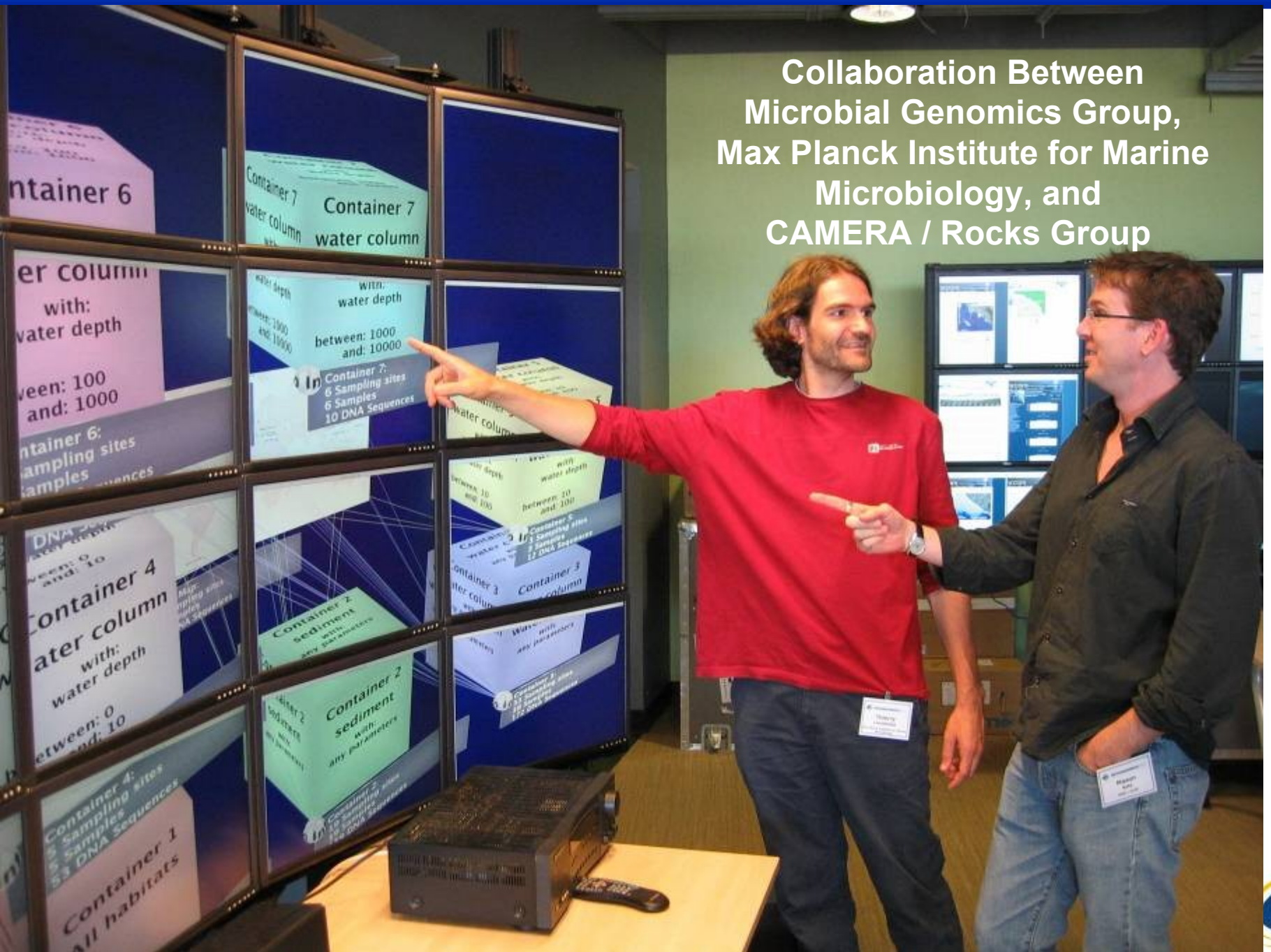
OptIPuter

Powered by Rocks ROCKS

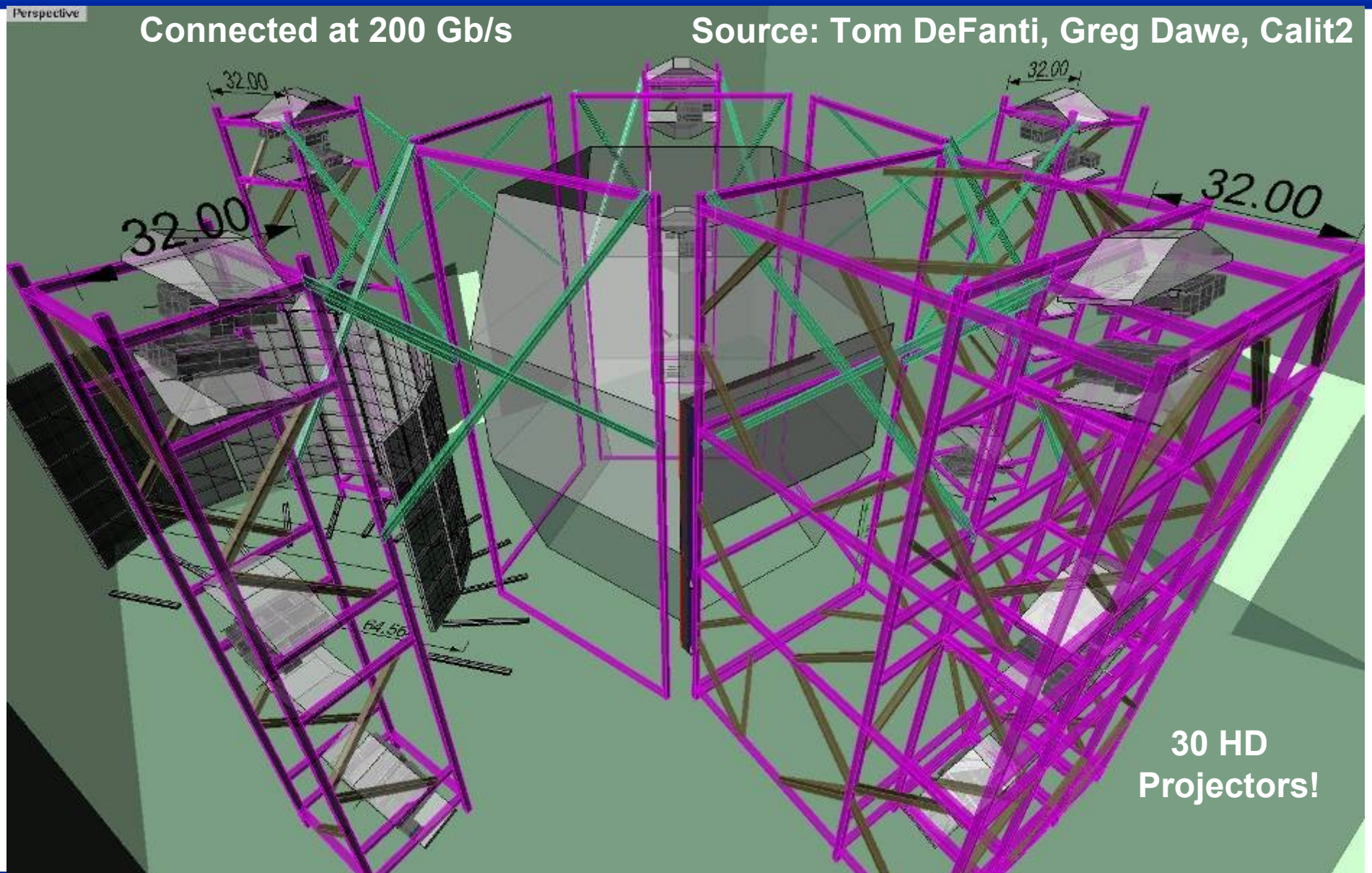
Source: Raj Singh, UCSD

it²

CAMERA is Partnering to Port Metagenomic Community Software to the OptIPortal



3D OptIPortal Calit2 StarCAVE Telepresence "Holodeck"



UCSD
UCIrvine

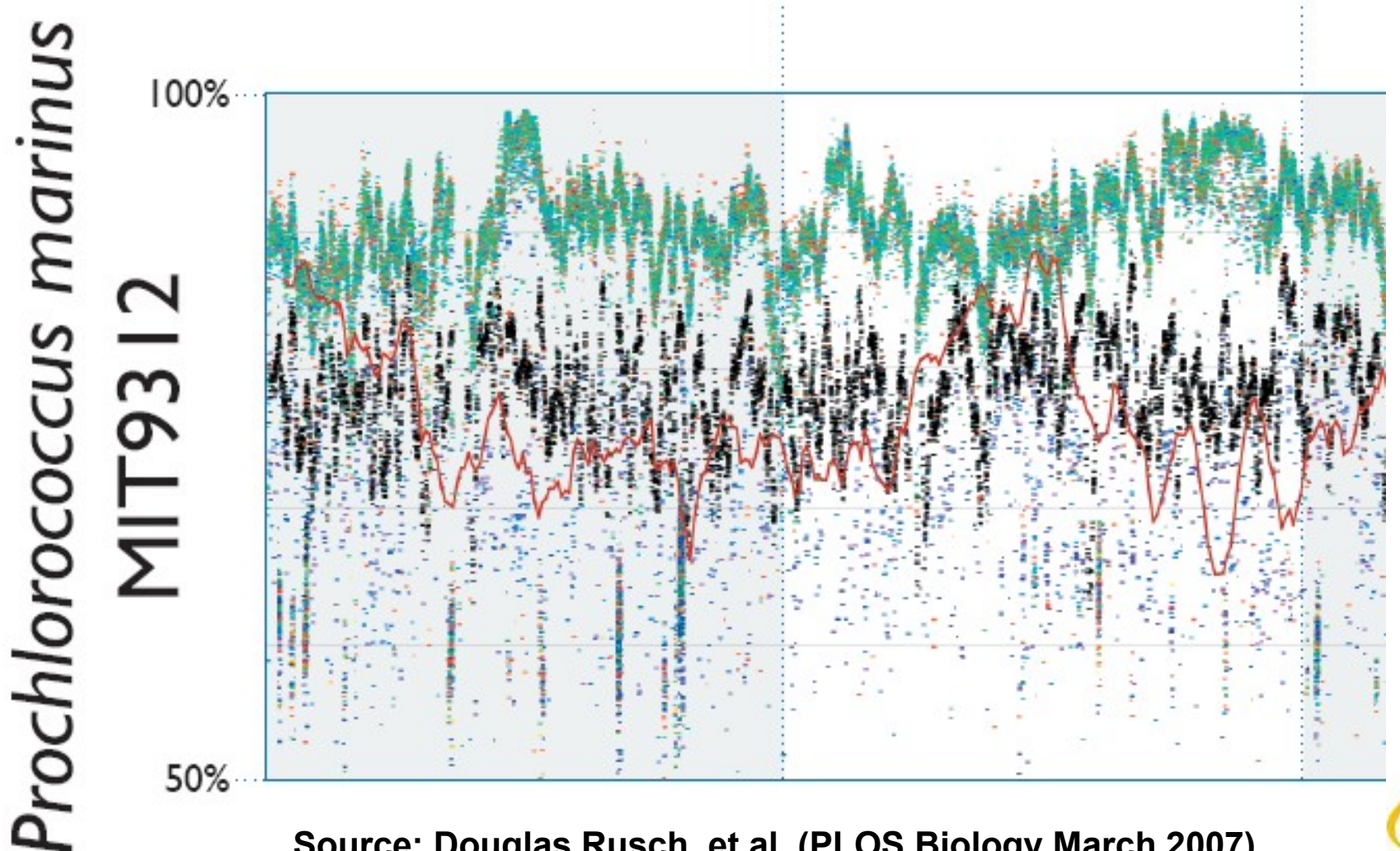
60 GB Texture Memory,
Renders Images 3,200 Times the Speed of Single PC

it2

Metagenomic Challenge--Enormous Biodiversity: Very Little of GOS Metagenomic Data Assembles Well

- **Use Reference Genomes to Recruit Fragments**
  - **Compared 334 Finished and 250 Draft Microbial Genomes**
- **Only 5 Microbial Genera Yielded Substantial and Uniform Recruitment**
  - **Prochlorococcus, Synechococcus, Pelagibacter, Shewanella, and Burkholderia**

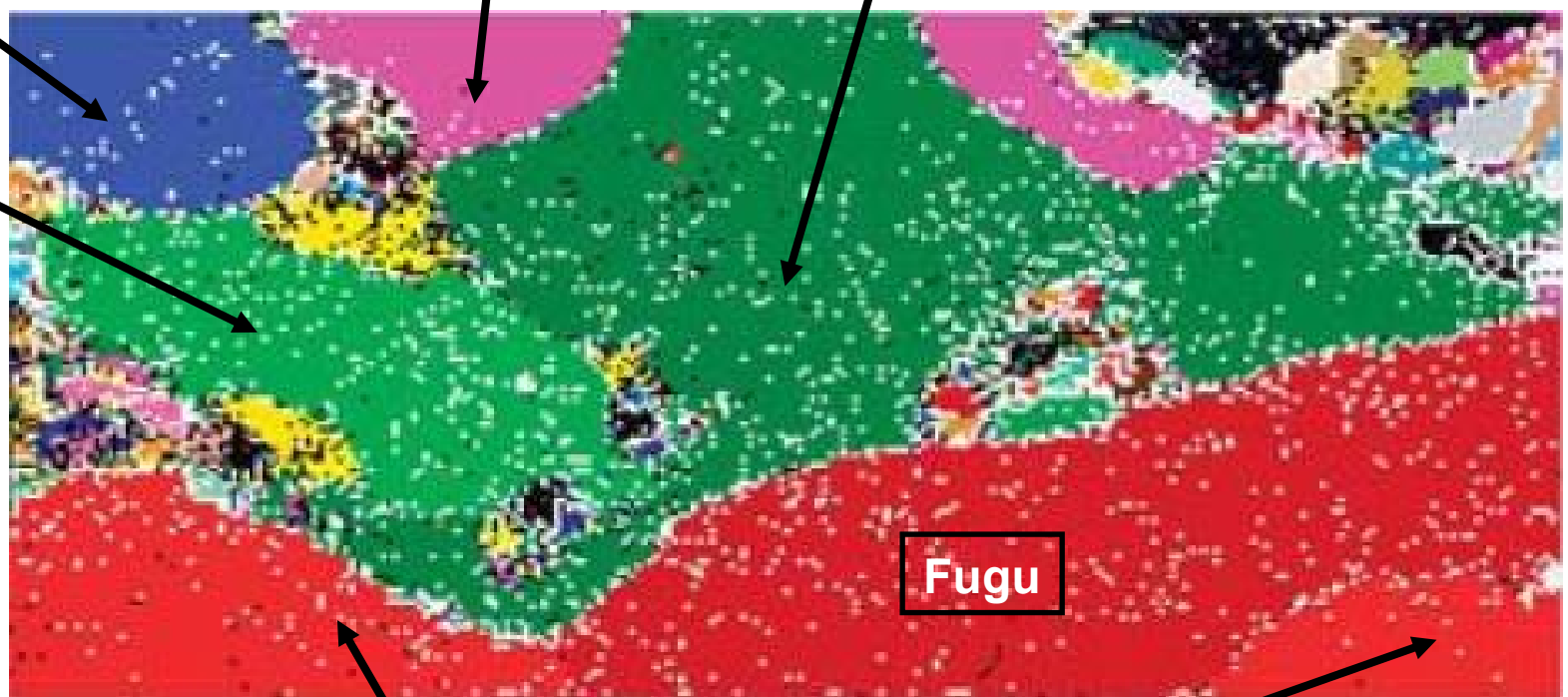
J. Craig Venter INSTITUTE



Source: Douglas Rusch, et al. (PLOS Biology March 2007)

Use of Self Organizing Maps to Identify Species Massive Computation on the Japanese Earth Simulator

ESC The Earth Simulator Center
Japan Agency for Marine-Earth Science and Technology



SOM Created from an Unsupervised Neural Network Algorithm to Analyze Tetranucleotide Frequencies in a Wide Range of Genomes

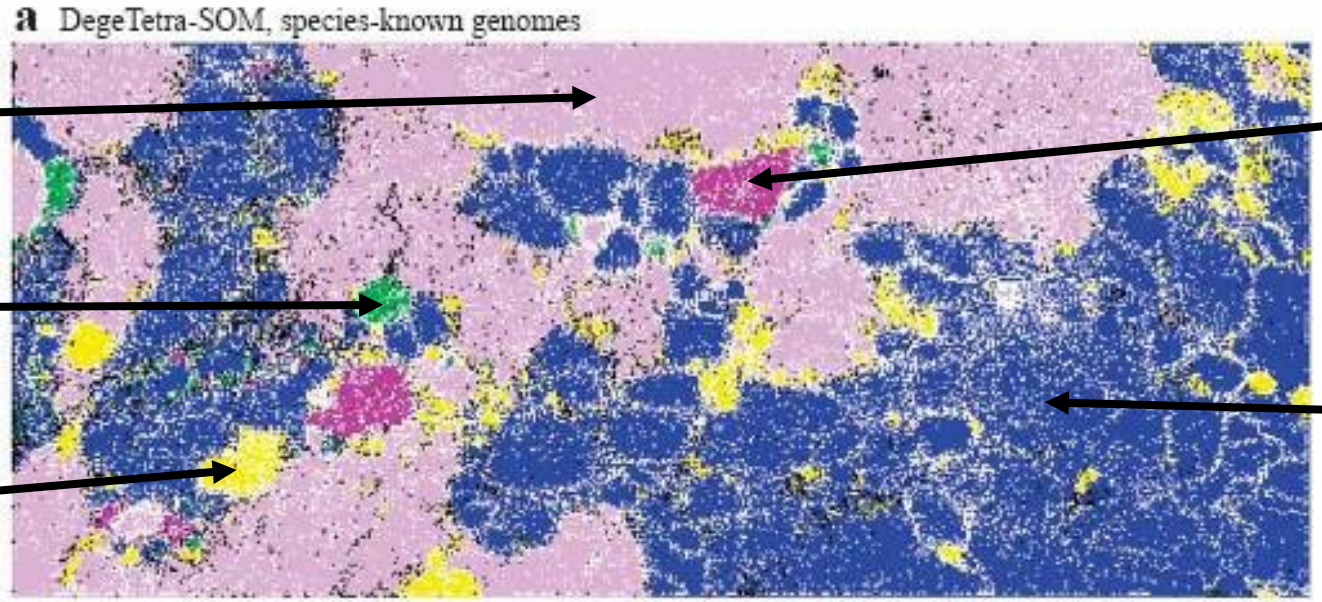
10kb Moving Window

UCSD
UCIrvine

T. Abe, H. Sugawara, S. Kanaya, T. Ikemura
Journal of the Earth Simulator, Volume 6, October 2006, 17–23
www.es.jamstec.go.jp/publication/journal/jes_vol.6/pdf/JES6_22-Abe.pdf

it²

Using SOM, Sargasso Sea Metagenomic Data Yields 92 Microbial Genera !



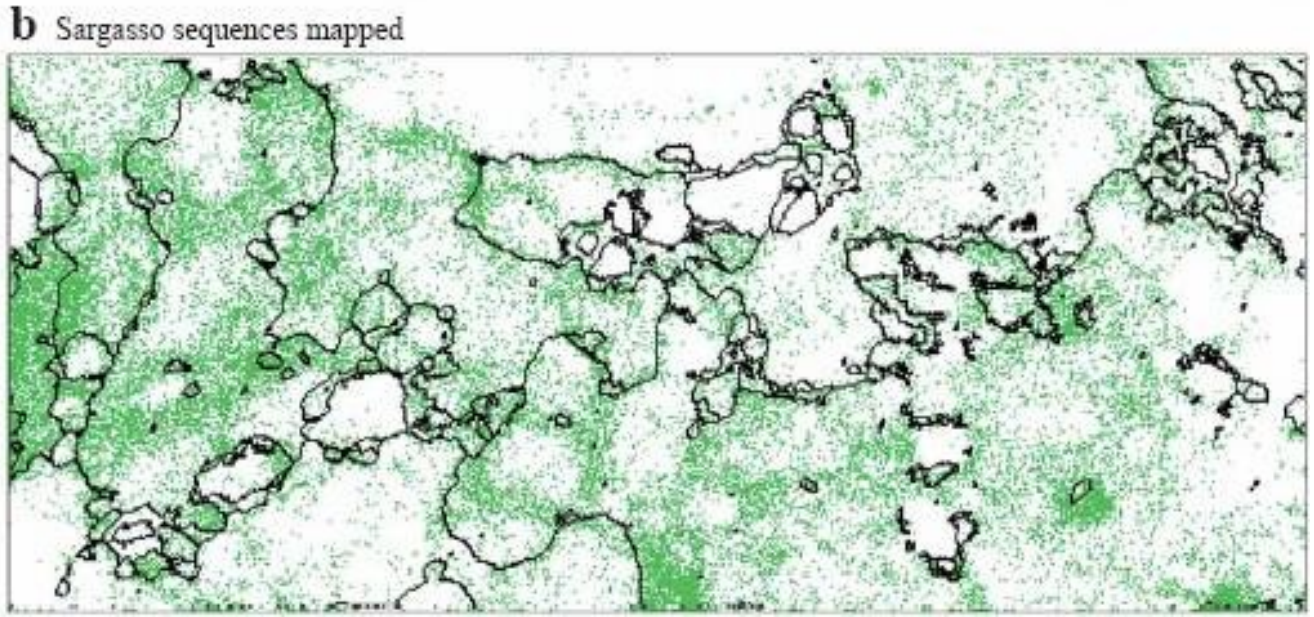
Eukaryotes

Mitochondria

Chloroplasts

Prokaryotes

Viruses



5kb Window

Input Genomes:

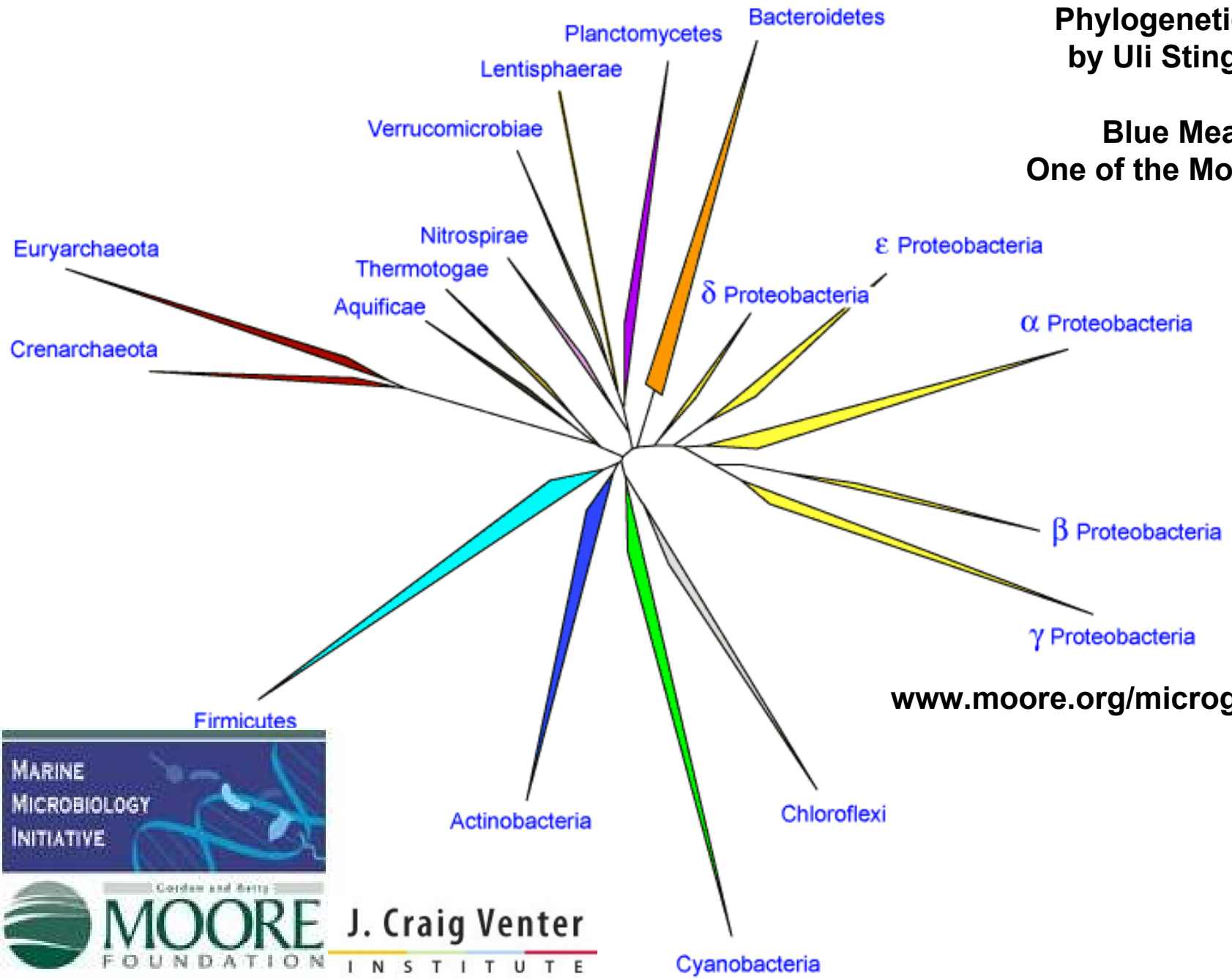
1500 Microbes
40 Eukaryotes
1065 Viruses
642 Mitochondria
42 Chloroplasts

UCSD
UCIrvine

T. Abe, H. Sugawara, S. Kanaya, T. Ikemura
Journal of the Earth Simulator, Volume 6, October 2006, 17–23

it²

Moore Foundation Funded the Venter Institute to Provide the Full Genome Sequence of 155+ Marine Microbes



Phylogenetic Trees Created by Uli Stingl, Oregon State

Blue Means Contains One of the Moore 155 Genomes

www.moore.org/microgenome/trees.aspx

MARINE MICROBIOLOGY INITIATIVE

Gordon and Betty MOORE FOUNDATION

J. Craig Venter INSTITUTE

DOE Genomic Encyclopedia of Bacteria and Archaea (GEBA) / Bergey Solution: Deep Sampling Across Phyla



2007 Goal: Finish ~100 Bacterial and Archaeal Genomes from Culture Collections
Project Lead -- Jonathan Eisen (JGI/UC Davis)

Well sampled phyla

No cultured taxa

Source: Eddie Rubin, DOE JGI

UCSD
UCIrvine

Calit2, SDSC, EVL, and SIO are Creating Environmental Observatory Control Rooms



UCSD UCIrvine evl electronic visualization laboratory SDSC SAN DIEGO SUPERCOMPUTER CENTER SCRIPPS INSTITUTION OF OCEANOGRAPHY it²

Towards a Total Knowledge Integration System for the Coastal Zone—SensorNets Linked to OptIPuter

- Moorings
- Ships
- Autonomous Vehicles
- Satellite Remote Sensing
- Drifters
- Long Range HF Radar
- Near-Shore Waves/Currents
- COAMPS Wind Model
- Nested ROMS Models
- Data Assimilation and Modeling
- Data Systems



Yellow—Proposed Initial OptIPuter Backbone

SOUTHERN CALIFORNIA COASTAL OCEAN OBSERVING SYSTEM

ROADNet

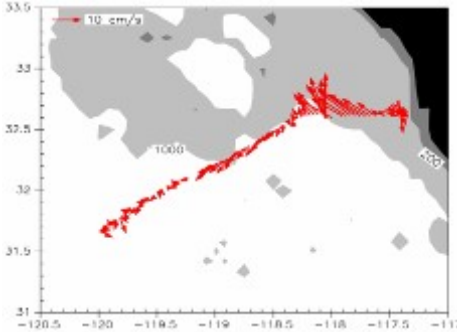
UCSD

UCIrvine

www.sccoos.org/

COCMP

Coastal Ocean Currents Monitoring Program

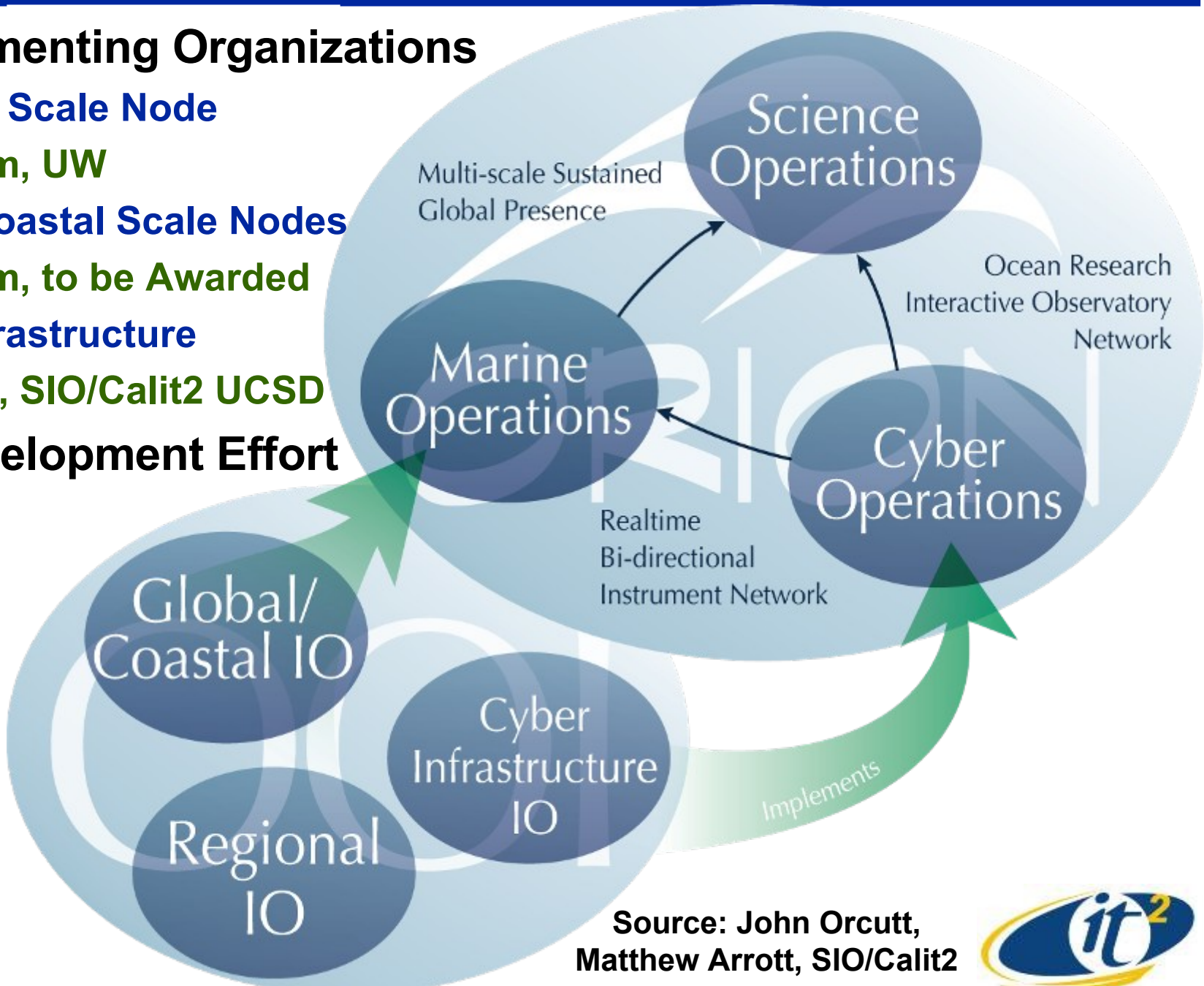




it²

Ocean Observatory Initiative -- Initial Stages

- **OOI Implementing Organizations**
  - **Regional Scale Node**
    - **$150m, UW**
  - **Global/Coastal Scale Nodes**
    - **$120m, to be Awarded**
  - **Cyberinfrastructure**
    - **$30m, SIO/Calit2 UCSD**
- **6 Year Development Effort**



Source: John Orcutt, Matthew Arrott, SIO/Calit2

UCSD UCIrvine