



# University of California, San Diego



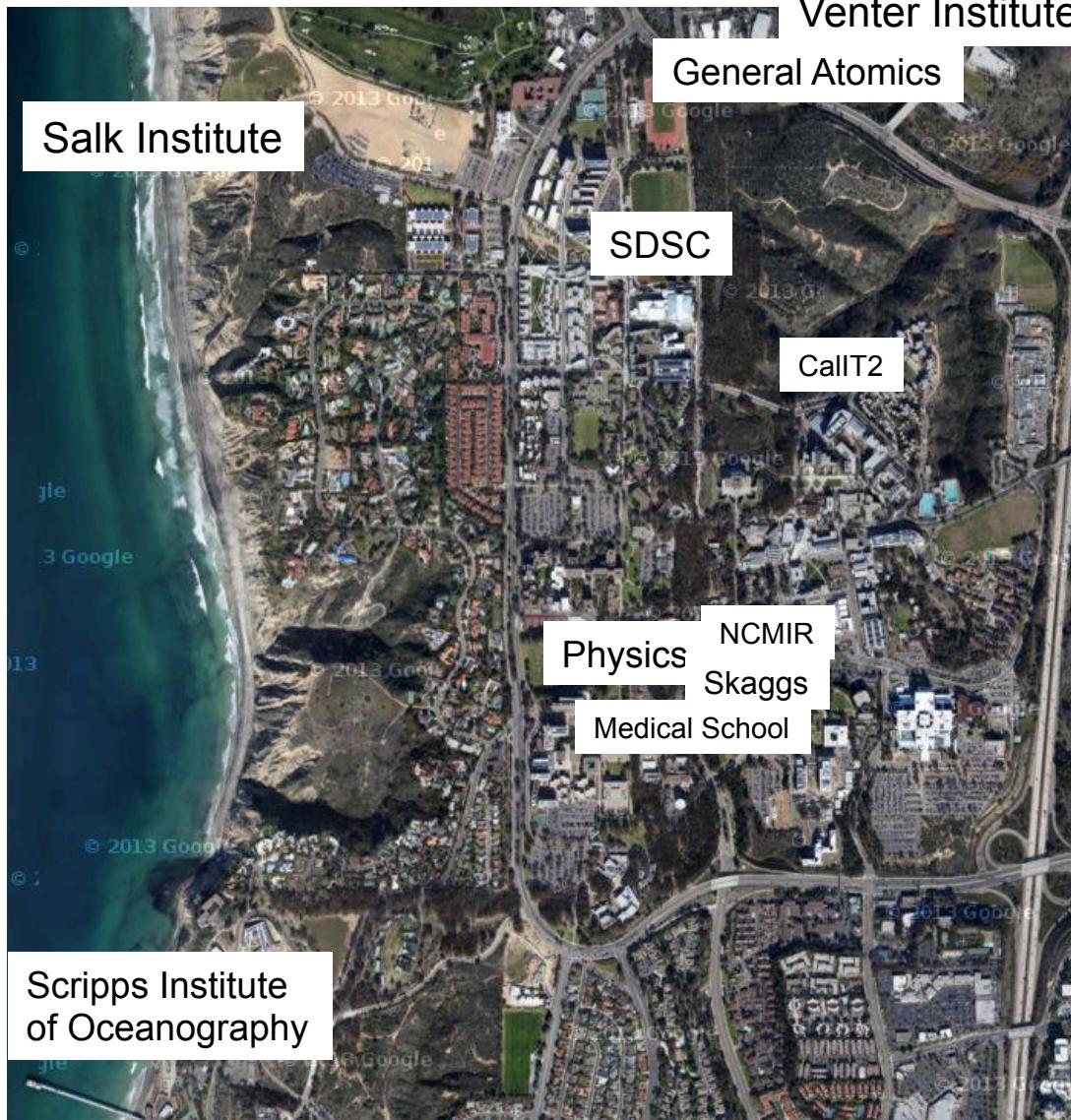


# Overview

- UCSD and its environment
- My roles within this environment
- Vision
- Computing facilities at UCSD
- Past joint ventures with OSG
- Future opportunities
- Conclusion



# UCSD and its environment



In addition to the 3 main  
UCSD units:

General Campus  
Medical School  
Scripps I.O.

there are many other  
research organizations  
on and around campus.



# UCSD by Numbers

- 22,700 undergraduates
  - 38% Social Sciences, 25% Biological Sciences, 18% Engineering, 8% Science & Math, 4% Humanities, 3% Arts
- 6,300 graduate students
- 2,588 Researchers and post-docs
- 2,505 Professors
- \$1 Billion extramural funding annually across three major organizations:
  - General Campus
  - Medical School
  - Scripps Institute of Oceanography

# Mike Norman's roles

- Director of SDSC
- PI of Gordon Data-Intensive Computer project for XSEDE
- Co-Chair of Campus Research Cyberinfrastructure Oversight Committee
- Member of Chancellor's Council
- Professor of Physics
  - Formation of the first galaxies and reionization
  - HPC code development



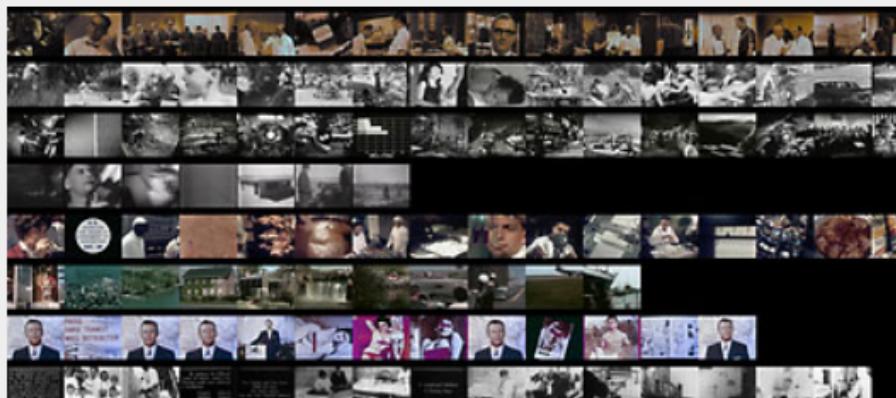
# Vision

- Two main areas I'd like to learn from and contribute to OSG

- Big Data
  - what are the application needs in scientific research?
  - What are the appropriate HW and SW technologies?
- Long Tail of Science
  - CI to support the broad spectrum of researchers who are not HPC experts
  - Persistent data platforms for research and discovery

# Big Data on Gordon

## Large-scale video analytics



Comparative visualization reveals similarities and differences between multiple films within an archive. *Image: Virginia Kuhn, USC*

## Energy $\mu$ -grid analytics

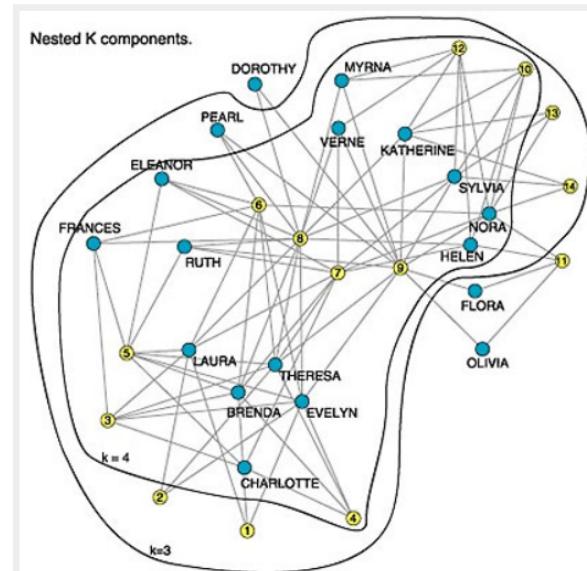


6/11/13

Characterized by

- Data streams
- persistent data sets
- Data mining tools

## Social network cohesion



7

02/28/2013

## SDSC Coordinates Effort to Establish the BigData Top100 List

*Benchmarking Initiative Outlined in Inaugural Issue of 'Big Data Journal'*



The San Diego Supercomputer Center (SDSC) at the University of California, San Diego, today announced plans for a community-based effort to create the BigData

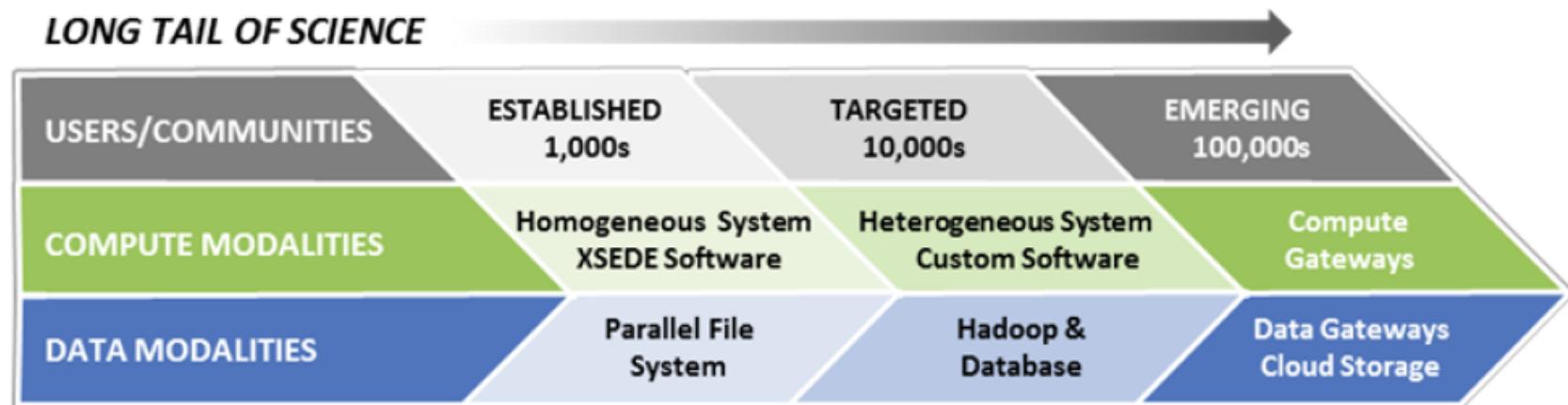
Top100 List, the first global ranking of its kind for systems designed for big data applications.

The BigData Top100 List will rank systems according to their performance on an application-level workload specification, while also reporting on system efficiencies in terms of price/performance. As an application-level benchmark, the list will complement other rankings of high-performance computing (HPC) systems, such as the [Top500](#) and [Graph500](#).

Preliminary information about the new list is at [www.bigdatatop100.org](http://www.bigdatatop100.org), which also includes information for those interested in joining this consortium and supporting its development.

# CI for the Long Tail of Science

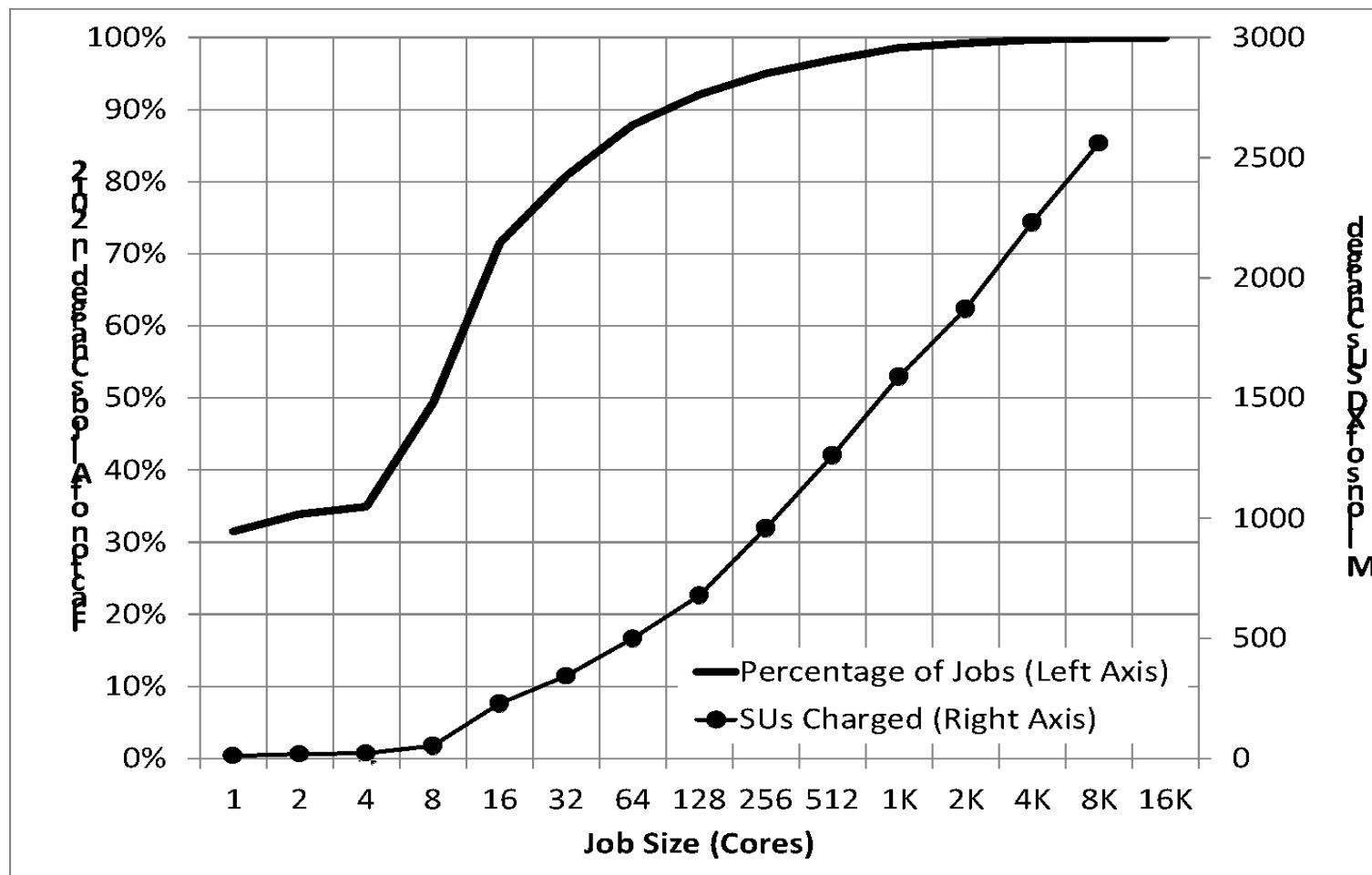
- New compute and data modalities needed to support the long tail



*Figure 1.1 The Way to a More Inclusive Computing Environment*

From SDSC proposal to NSF 13-528

# Job Size Distribution on XSEDE





# Computing Infrastructure at UCSD and SDSC



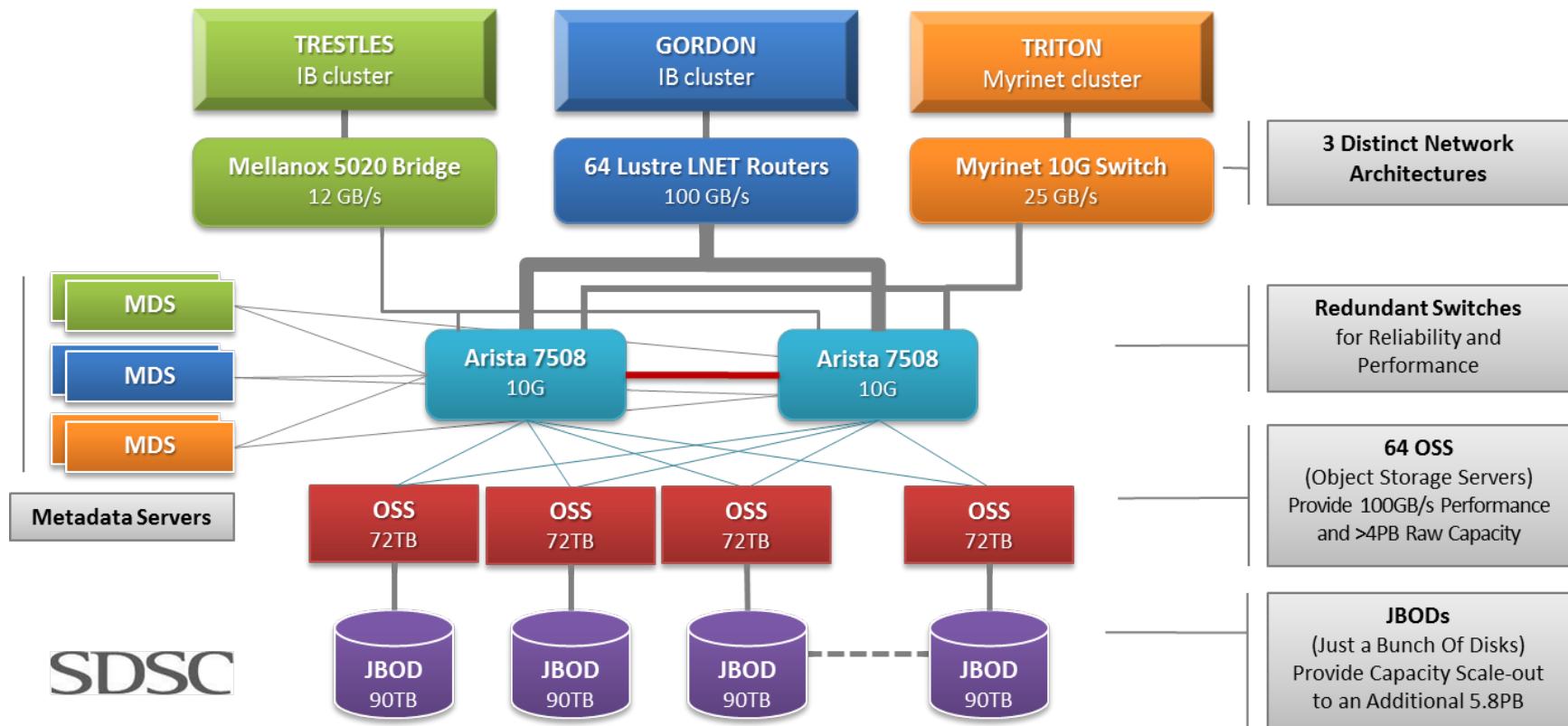
XSEDE resources  
Gordon (16k Cores)  
Trestles (10k Cores)  
4GB RAM per core

Campus Condo  
Triton Shared  
Compute Cluster  
(CPU+GPU)



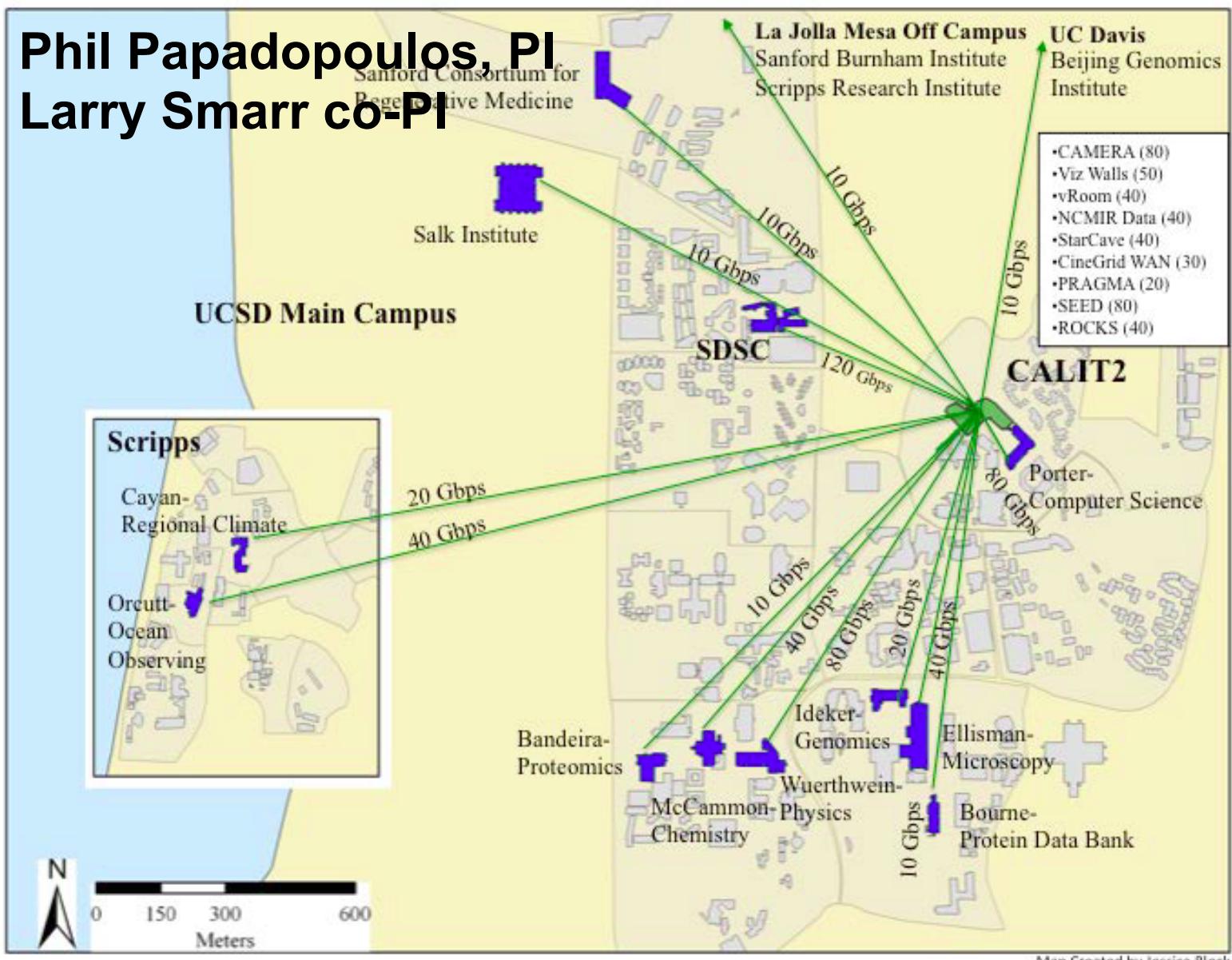
# SDSC Storage: 10G PFS

## Data Oasis Heterogeneous Architecture





# Creating a “Big Data Freeway” System Connecting Instruments, Computers, & Storage



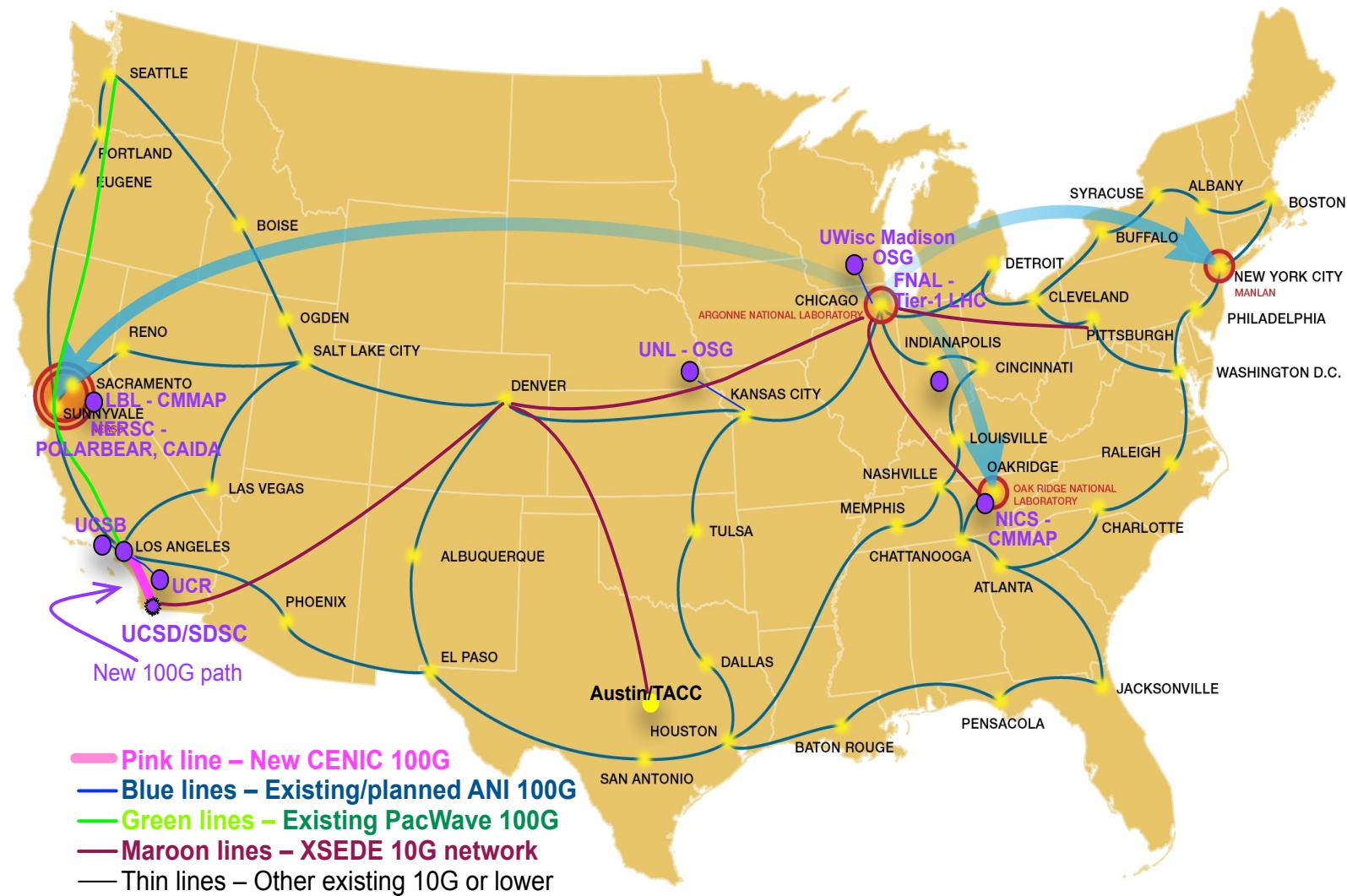
**PRISM  
@UCSD**

**Start Date  
1/1/13**



**SDSC**

# UCSD Connected to I2/EStnet at 100G





# SDSC Storage: Open Stack SWIFT





# UCSD Science on OSG

Past & Present

- Protein Data Base calculating pairwise similarities of proteins.
- NEES Earthquake Engineering
- Analysis of WMAP data for Cosmology
- Study of cloud formation in Climate Science
- Study of protein structure:
  - Skaggs School of Pharmacy
  - Center for Theoretical Biological Physics
- Particle Physics with CMS



# OSG Activities at UCSD

- CMS Tier-2 as an OSG site
- FKW's management roles in OSG:
  - Chair of Interim Executive board (12/2004 – 2006)
  - Application Coordinator (since 2006)
  - Resources Manager (since 2010)
  - CoPI of NSF grant (since 2012)
- UCSD group commitments to OSG
  - gfactory operations
  - Frontend for XSEDE on OSG, and Engage VO
  - Login platform in support of UCSDGrid, Baker Lab, PDB, Skaggs, ...



# OSG Satellites at UCSD

- Anydata, Anytime, Anywhere
  - with UNL, Wisconsin
- dV/dt
  - with Notre Dame, Wisconsin, USC
- CorralWMS
  - with USC, FNAL
- Advanced Network and Distributed Storage Laboratory
  - With Wisconsin, SDSC, Cal(IT)<sup>2</sup>

# Future Opportunities

- Bringing new customers to SDSC that benefit from SDSC having OSG interfaces to Trestles, Gordon, and future systems.
  - Surge capacity for LHC data processing
  - Support data-intensive and/or large RAM computing in OSG
- Allow for existing SDSC customers and science at UCSD in general to transparently compute across SDSC and OSG
  - PDB protein structure alignment (Bourne)
  - Marine Metagenomics (Ellisman)
  - Institute for Genomic Medicine
  - Brain Mapping Institute

# Conclusion

- Propose to add Mike Norman as member to OSG Council to represent UCSD and its affiliated organizations.
- UCSD's life sciences, biomedical sciences, and geosciences foci will broaden OSG's role at the frontiers of scientific research
- SDSC resources complementary to OSG
- Interesting case study in “campus bridging”