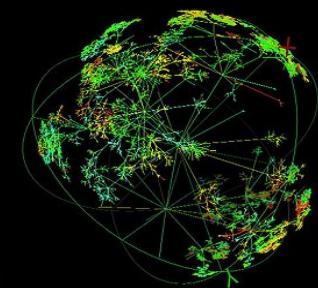


Open Storage Network

**Joe Mambretti, Director, (j-mambretti@northwestern.edu)
International Center for Advanced Internet Research (www.icair.org)
Northwestern University
Director, Metropolitan Research and Education Network (www.mren.org)
Director, StarLight, PI StarLight IRNC SDX, Co-PI Chameleon, PI-iGENI, PI-
OMNINet (www.startap.net/starlight)**

**Alexander Szalay, Bloomberg Distinguished Professor, the Alumni
Centennial Professor of Astronomy, and Professor, Department of
Computer Science, Johns Hopkins University. Director, Institute for Data
Intensive Science, Fellow of the American Academy of Arts and Sciences,
PI Open Storage Network Project**



**Juniper R&E Networking Symposium (J-RENS) 2018
DePaul University
Chicago, Illinois
October 10, 2018**

Introduction to iCAIR:



Accelerating Leading Edge Innovation and Enhanced Global Communications through Advanced Internet Technologies, in Partnership with the Global Community

- **Creation and Early Implementation of Advanced Networking Technologies - The Next Generation Internet All Optical Networks, Terascale Networks, Networks for Petascale and Exascale Science**
- **Advanced Applications, Middleware, Large-Scale Infrastructure, NG Optical Networks and Testbeds, Public Policy Studies and Forums Related to Optical Fiber and Next Generation Networks**
- **Three Major Areas of Activity: a) Basic Research b) Design and Implementation of Prototypes and Research Testbeds, c) Operations of Specialized Communication Facilities (e.g., StarLight, Specialized Science Networks)**

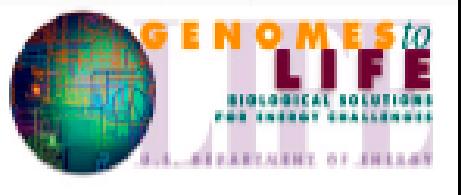


Emerging Topics In Advanced Networking

- Transition From Legacy Networks To Networks That Take Full Advantage of IT Architecture and Technology
- Extremely Large Capacity (Multi-Tbps Streams)
- Specialized Network Services, Architecture and Technologies for Data Intensive Science
- High Degrees of Communication Services Customization
- Highly Programmable Networks
- Network Facilities As Enabling Platforms for Any Type of Service
- Network Virtualization
- Tenet Networks
- Network Virtualization
- Network Programming Languages (e.g., P4) API (e.g., Jupyter)
- Disaggregation
- Orchestrators
- Highly Distributed Signaling Processes
- Network Operations Automation (Including Through AI/Machine Learning)
- SDN/SDX/SDI/OCX/SDC/SDE



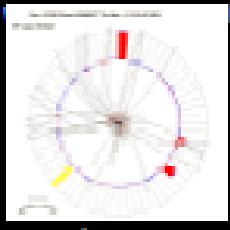
NEPTUNE
A Fiber-Optic Telepresence Under-Sea Observatory



NEON
National Ecological Observatory Network



ALMA
ALMA: Atacama Large Millimeter Array



D0 (DZero)
www-d0.fnal.gov



IVOA:
International Virtual Observatory
wwwivoa.net



ANDRILL:
Antarctic Geological Drilling
www.andrill.org



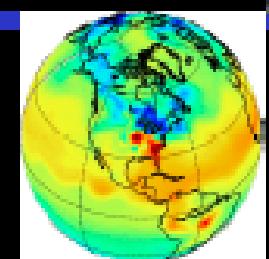
BIRN: Biomedical Informatics Research Network
www.nblm.net



GLEON: Global Lake Ecological Observatory Network



CAMERA
metagenomics
camera.calit2.net



Carbon Tracker
www.esrl.noaa.gov/gmd/ccgg/carbontrack



CineGrid
www.cinegrid.org



LHCONE
www.lhccone.net



CLASS
Comprehensive Large-Array Stewardship System
[www.class.noaa.gov](http://class.noaa.gov)



ISS: International Space Station
www.nasa.gov/station



TeraGrid
www.teragrid.org



XSEDE
www.xsede.org



Open Science Grid
www.opensciencegrid.org



Globus Alliance
www.globus.org



SKA
www.skatelescope.org



Sloan Digital Sky Survey
www.sdss.org



StarLight

www.opensciencegrid.org

Compilation By Maxine Brown

Petascale Computational Science



For Decades, Computational Science
Has Driven Network Innovation
Today –
Petascale Computational Science



National Center for Supercomputing Applications, UIUC

XSEDE

- Extreme Science and Engineering Discovery Environment (XSEDE)
- Goal: Create a Distributed Computational Science Infrastructure to Enable Distributed Data Sharing and High-Speed Computing for Data Analysis and Numerical Simulations
- Builds on Prior Distributed TeraGrid



STARLIGHTSM

Open Science Grid: Selected Investigations



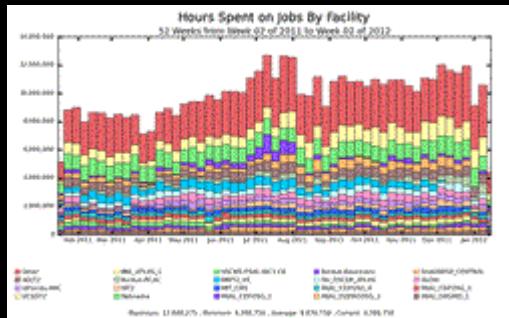
DNA Modeling



Gravity Wave Modeling



Nutrino Studies



Usage



This Distributed Facility
Supports Many Sciences

The Open Science Data Cloud (OSDC) is an **open-source**, **cloud-based** infrastructure that allows scientists to manage, share, and analyze medium to large size scientific datasets.



OPEN SCIENCE DATA CLOUD

Total OSDC Resource Size

TOTAL COMPUTE CORES

7550

COMPUTE RAM

27622 (GB)

RAW STORAGE

10.03 (PB)

USEABLE STORAGE

5.92 (PB)

Public Data Commons

The OSDC hosts a local mirror of **1 PB** of publically available datasets.
The data can also be freely downloaded using rsync or UDR.

EXAMPLE AVAILABLE DATASETS



1000 GENOMES



MODENCODE



EO1



MODIS



NCBI DATASETS



COMPLETE
GENOMICS



US CENSUS

Application for resources available to anyone doing scientific research:

**Open Commons
Consortium**

www.opensciencedatacloud.org



NATIONAL CANCER INSTITUTE
Genomic Data Commons



SCIENTIFIC PROJECT DATA COMMONS



National Institute of
Allergy and
Infectious Diseases



Data Commons & Data Sharing Initiatives



BRAIN Commons



BloodPAC
BLOOD PROFILING & ATLAS IN CANCER



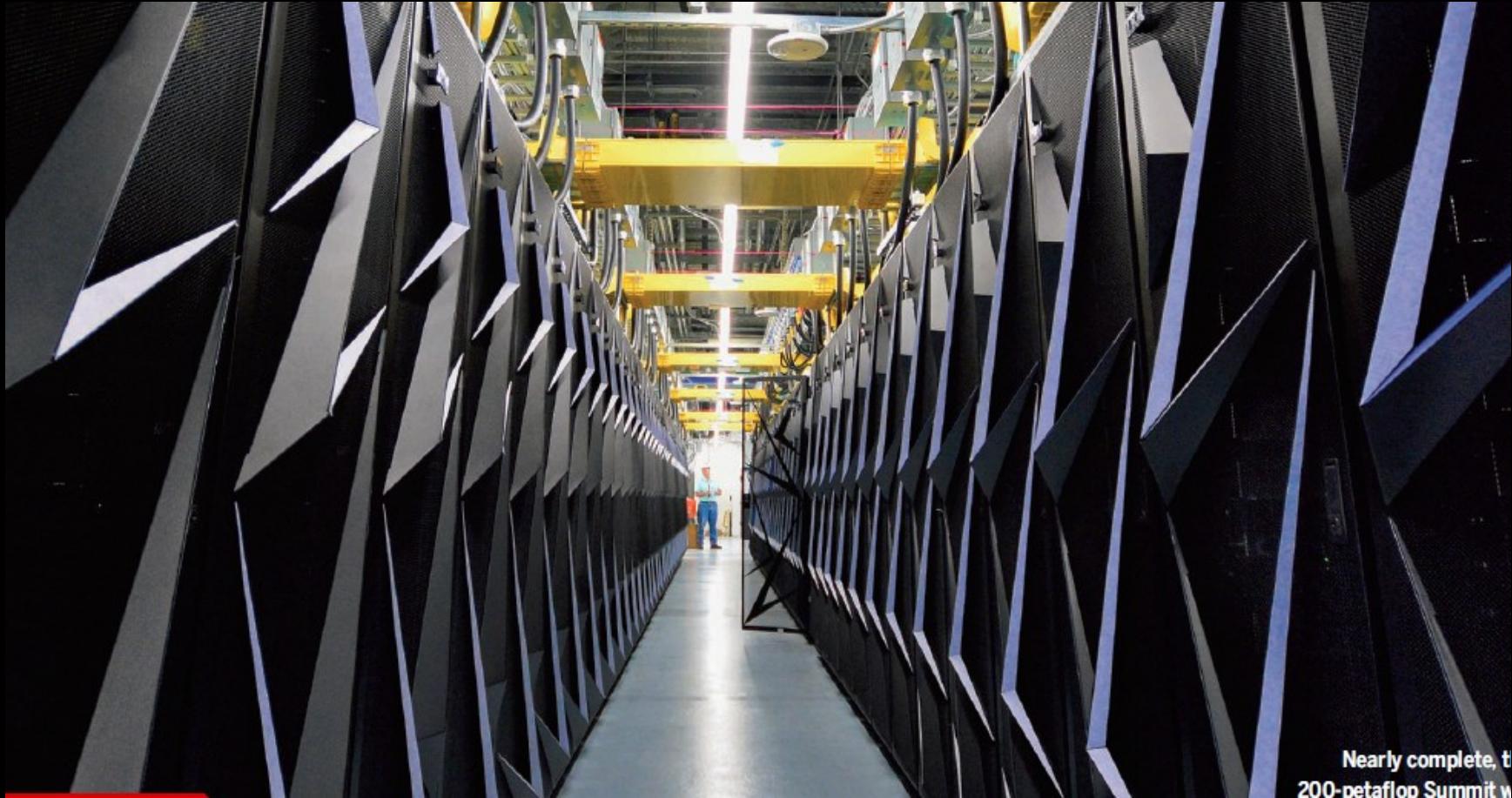
Award# 1445604

jetstream

First NSF Supported Cloud
Infrastructure for Science &
Engineering Research

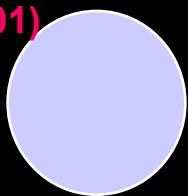
STARLIGHTSM

Summit At Oak Ridge National Laboratory – A Step To A21 Exescale Computer at Argonne National Laboratory (2021)

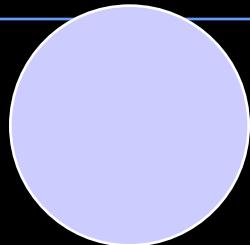


HEP = Staggering Amounts of Data

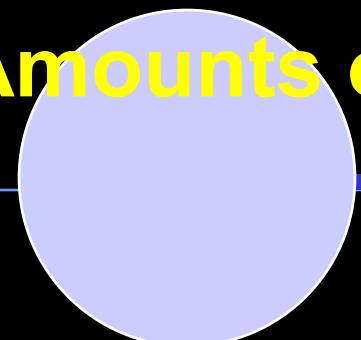
BaBar 0.3
PetaByte/year
(2001)



CDF or D0 Run II
0.5 PetaByte/year
(2003)



LHC Mock Data Challenge
1 PetaByte/year (~2005)



CMS or ATLAS
2 PetaBytes/year
(~2008)



KTeV 50
TeraBytes /year
(1999)



SLD 3 TB /year
(1998)



Run I (CDF or D0)
20 TB /year (1995)



L3 5 TB /year (1993)



E791 50 TB /year
(1991)



• EMC 400GB /year
(1981)

In 1977 the Upsilon (bottom quark) was discovered at Fermilab by experiment E288 led by now Nobel laureate Leon Lederman

The experiment took about 1 million events and recorded the raw data on ~ 300 magnetic tapes for about 6 GB of raw data

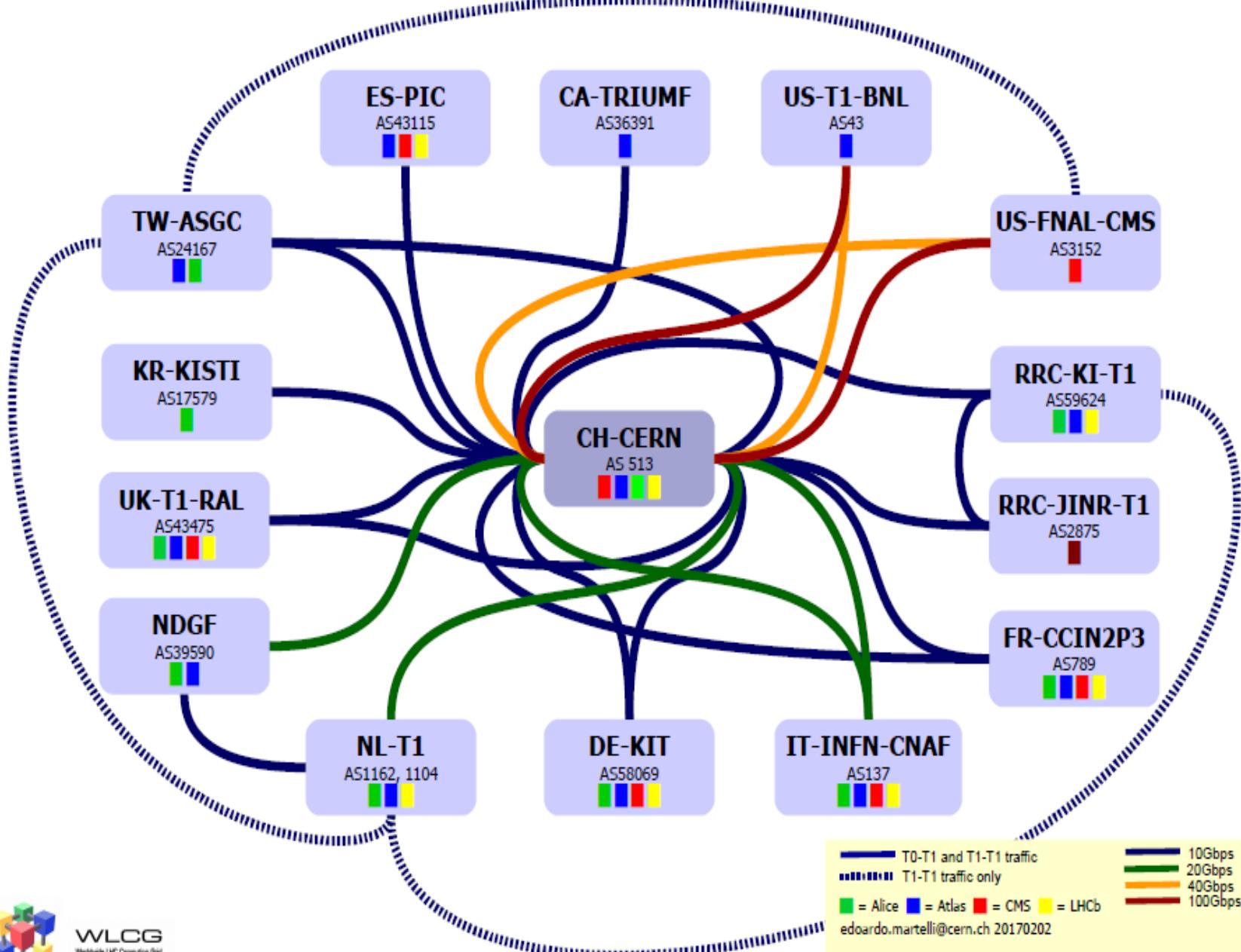


Source: Fermi Lab

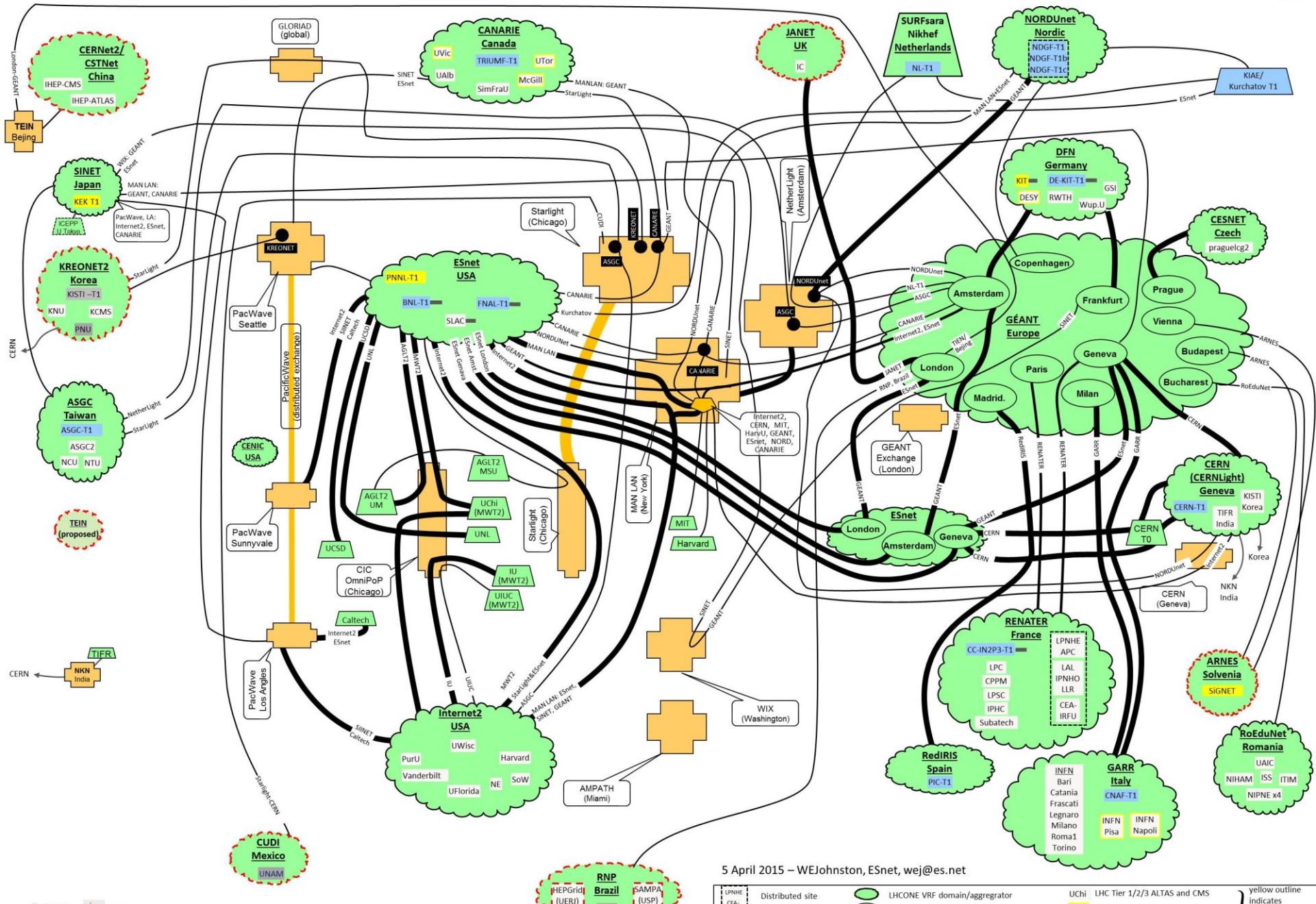
← R L I G H T SM

LHCOPN map

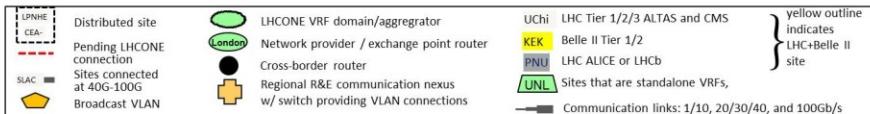
LHCOPN



LHCONE: A global infrastructure for the High Energy Physics (LHC and Belle II) data management



5 April 2015 – WEJohnston, ESnet, wej@es.net



New Science Communities Using LHCONE

- **Belle II Experiment, Particle Physics Experiment Designed To Study Properties of B Mesons (Heavy Particles Containing a Bottom Quark).**
- **Pierre Auger Observatory, Studying Ultra-High Energy Cosmic Rays, the Most Energetic and Rarest of Particles In the Universe.**
- **In August 2017 the PAO, LIGO and Virgo Collaboration Measured a Gravitational Wave Originating From a Binary Neutron Star Merger.**
- **The NOvA Experiment Is Designed To Answer Fundamental Questions In Neutrino Physics.**
- **The XENON Dark Matter Project Is a Global Collaboration Investing Fundamental Properties of Dark Matter, Largest Component Of The Universe.**
- **ProtoNUMA/NUMA Nutrino Research**

Argonne National Laboratory Advanced Photon Source



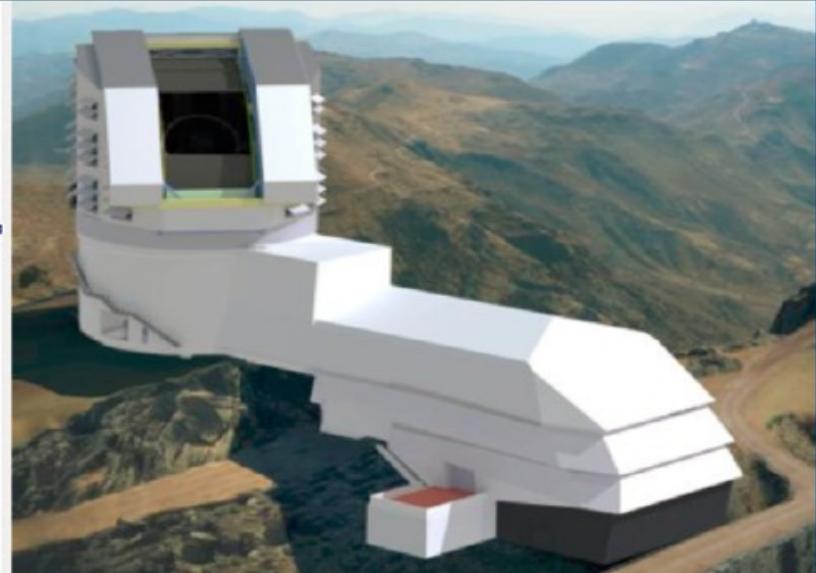
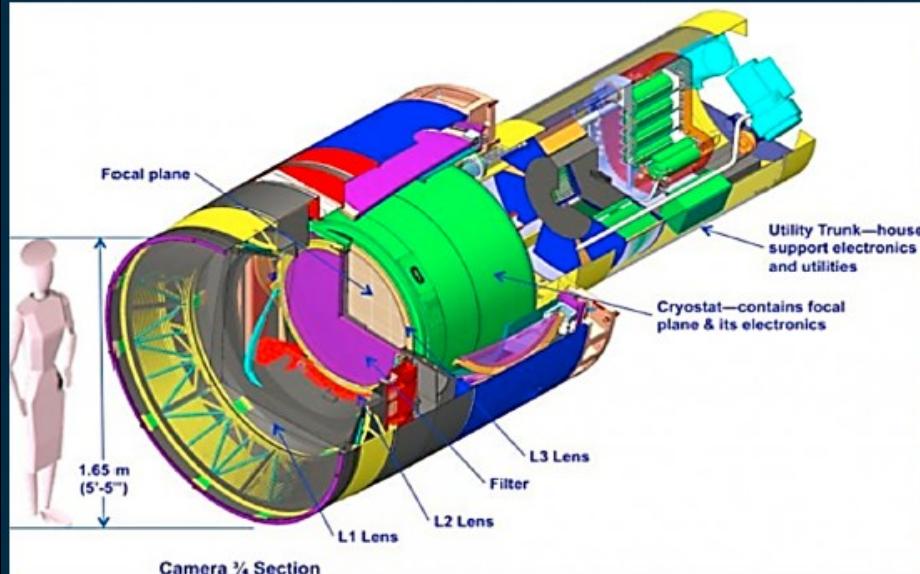
Square Kilometer Array





LSST Data Movement

Upcoming challenges for Astronomy



- **3.2 Gigapixel Camera with calibrated exposures at (10 Bytes / pixel)**
- **Planned Networks: Dedicated 100G for image data, Second 100G for other traffic, and 40G for diverse path**
- **Lossless compressed Image size = 2.7GB
(~5 images transferred in parallel over a 100 Gbps link)**
- **UDP based custom image transfer protocols**

Current Storage Landscape

- **Storage Is Isolated, Difficult To Access**
 - Multiple Isolated Facility, Campus, instrument Systems
 - Incompatibilities, Inefficiencies
 - Especially Problematic for Petabytes – Exabytes Are Also Required
- **Commercial Clouds Not A Solution**
 - Cost Model Makes Collaborative Distribution Prohibitive
 - Operations Prevent Detailed Performance Analytics
 - Limited Data Tools

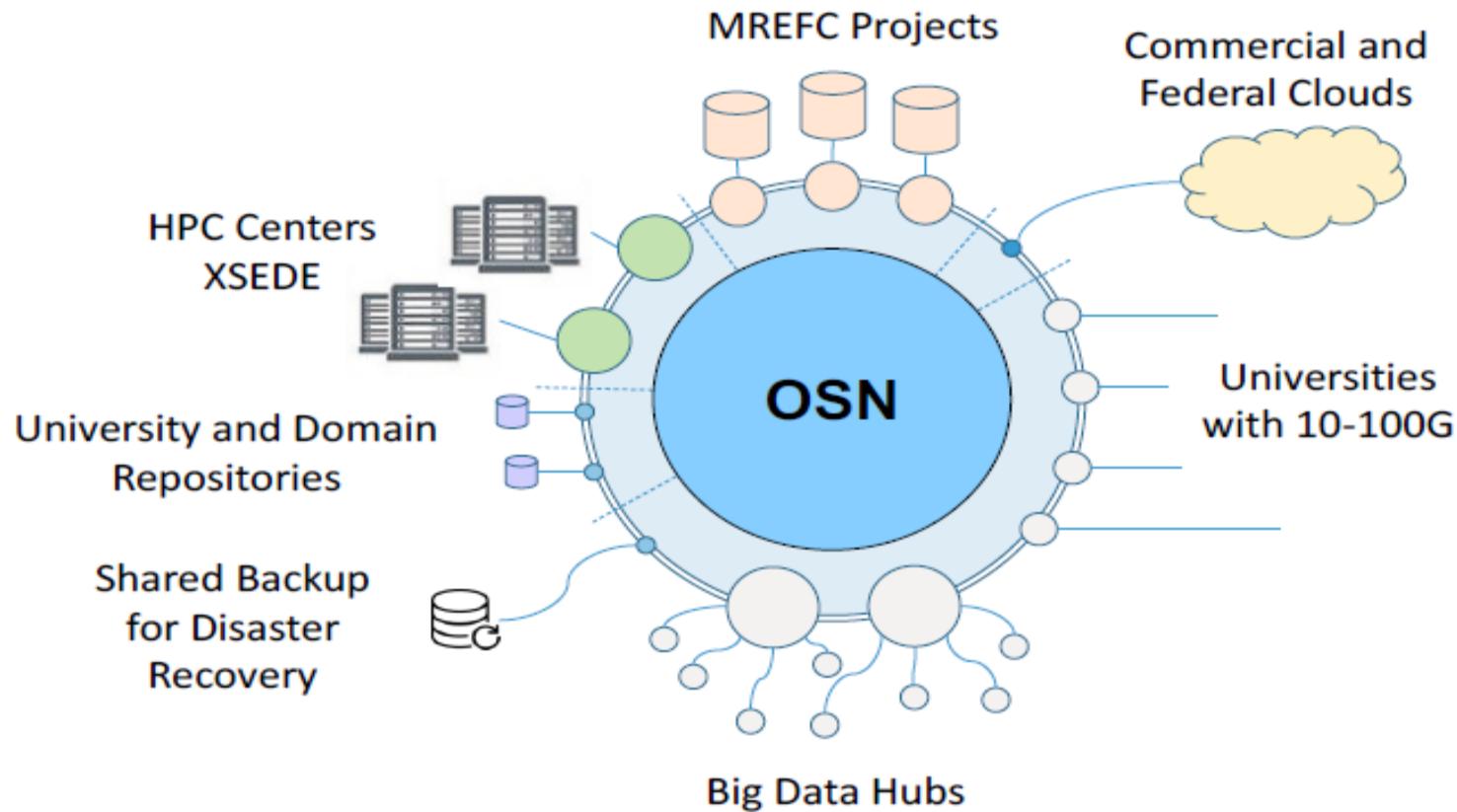
Opportunity: Next Gen Storage For Science

- Large National Distributed Storage System:
 - – *Perhaps 1-2PB Storage Rack On Each CC* Campus (~200PB)*
 - – *Create Redundant Interconnected Storage Substrate*
 - – *Using Industrial Strength Erasure Code Storage*
 - – *Provide High Capacity Aggregate Bandwidth, Easy Data Flow Among Sites (100 Gbps Channels)*
 - – *Potential For Acting As Gateways To Cloud Providers*
 - – *Automatic Compatibility, Simple Standard API (S3)*
 - – *Implement a Set of Simple Policies*
 - – *Enable Sites To Add Additional Storage Locally Funded*
 - – *Provide Variety of Services Built By Communities*

Transformational Impact

- Potential To Totally Change Landscape For Academic Big Data Research (Even At Petabyte Scale)
 - – *Create Homogeneous, Uniform Storage Tier For Science*
 - – *Liberate Communities To Focus On Science, Discovery, Analytics Collaboration, and Preservation*
 - – *Amplify NSF CC* Investment*
 - – *Very Rapidly Spread Best Practices Nationwide*
 - – *Links to Large Science Instruments, Compute Facilities, Data Facilities (Including Big Data Hubs), Analytics Sites, et al*
 - – *Big Data Projects Can Use It For Data Distribution*
 - – *Small Projects Can Build On Existing Infrastructure*
 - – *Enabling Whole Ecosystem of Specialized Services to Flourish*
 - *Major Opportunities for Novel Interdisciplinary Research*

Connections



Many Issues

- **Architecture**
- **Technologies**
- **Services**
- **Mechanisms For Integration With Instruments, Compute Facilities, Data Hubs, Analytic Centers, etc**
- **Policies**
- **Communications/Education**
- **Security**
- **Financial Models**
- **Mechanisms For Innovation and On-Going Extensions and Enhancements**
- **Governance/Management of Data and Facilities**
- **Long vs Short Term Repositories**
- **Etc**

Building Blocks

- Scalable element (SE)
 - *500TB of storage+ single server*
 - *Support 40G interface for sequential read/write*
 - *Should saturate 40G for read, about half for write*
- Stack of multiple SEs
 - *Aggregated to 100G on a fast TOR switch, now becoming quite inexpensive (<\$20K)*
- These can also exist inside the university firewall
 - *But purchased on local funds, storing local data*
- Software stack to be discussed
 - *ZFS, Ceph, Mero,...*
 - *Integrated with Globus “Lite”, with streamlined stack*

Building Blocks 2

- **E2E Services (e.g., BigData Express, SENSE)**
- **APIs**
- **Workflow Managers (e.g., Jupyter)**
- **Data Transfer Nodes**
- **Performance Monitoring/Measurement/Analytics Instrumentation**
- **Ultra High Performance File Systems**
- **Pipelines Based On Direct Connections Between HP File Systems and High Performance Optical Channels (e.g., Lightpaths)**
- **Interdomain Services**

Management

- Who owns it?
 - OSN storage should remain in a common namespace
 - This would enable uniform policies and interfaces
- Software management
 - Central management of software stack (push)
 - Central monitoring of system state
- Hardware management
 - Local management of disk health
 - Universities should provide management personnel
- Policy management
 - This is **hard** and requires a lot more discussion
- Monitoring
 - Two tier, store all events and logs locally, send only alerts up
 - Try to predict disk failures, preventive maintenance
- Establish metrics for success

Initial OSN Facility Sites

- **Johns Hopkins University, Baltimore Maryland**
- **StarLight International/National Communications Exchange Facility, Chicago, Illinois**
- **University of California At San Diego Supercomputing Center, La Jolla, California**
- **Future Sites Under Discussion**
- **International Extensions Possible**



StarLight – “By Researchers For Researchers”

StarLight is an experimental optical infrastructure and proving ground for network services optimized for high-performance applications

Multiple
10GE+100 Gbps
StarWave
Multiple 10GEs
Over Optics –
World’s “Largest”
10G/100G Exchange
First of a Kind
Enabling Interoperability
At L1, L2, L3

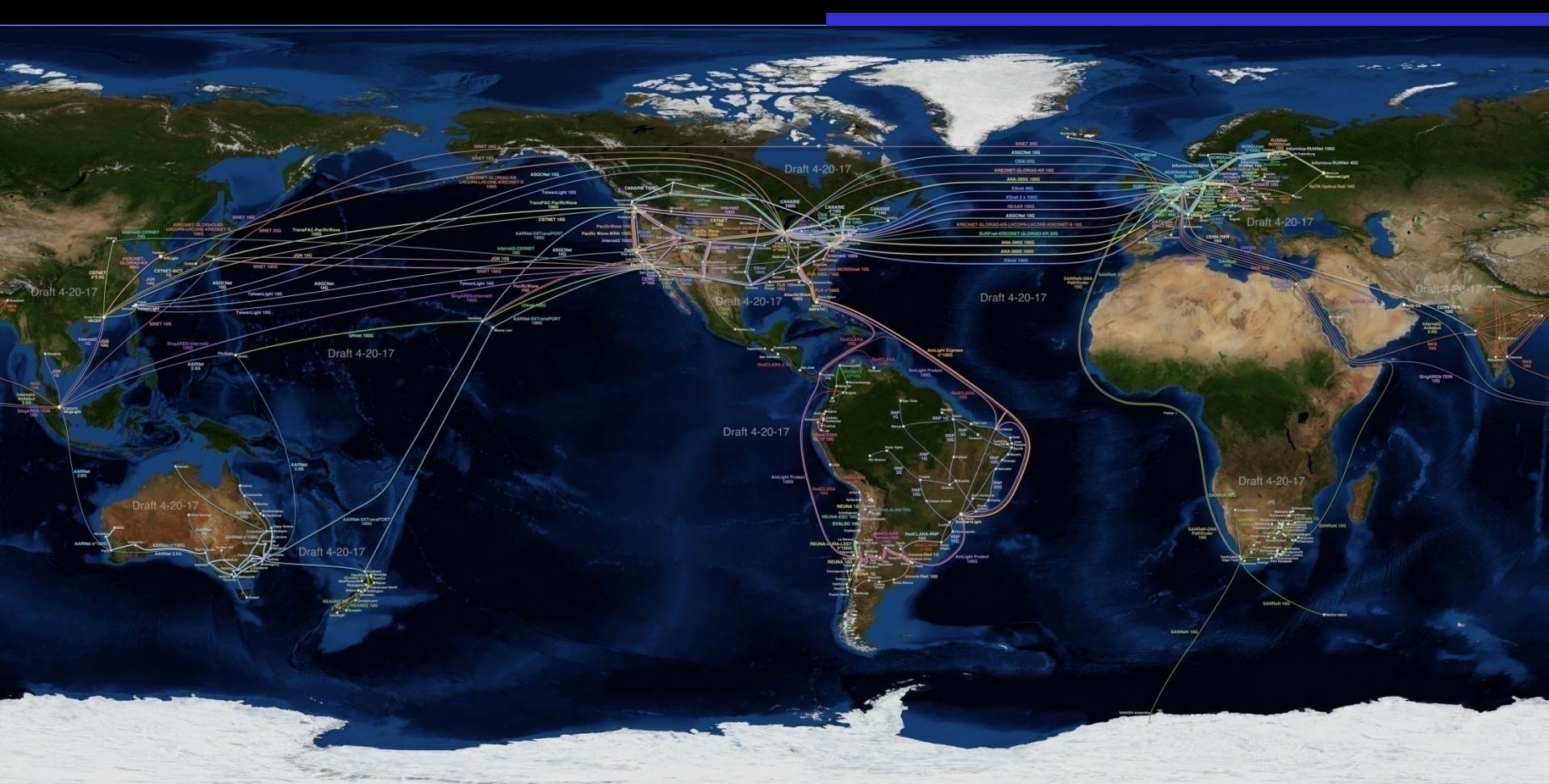


View from StarLight



Abbott Hall, Northwestern University's Chicago Campus

iCAIR: Founding Partner of the Global Lambda Integrated Facility Available Advanced Network Resources



Visualization courtesy of Bob Patterson, NCSA; data compilation by Maxine Brown, UIC.



www.glf.is

STARLIGHTSM

IRNC: RXP: StarLight SDX A Software Defined Networking Exchange for Global Science Research and Education

**Joe Mambretti, Director, (j-mambretti@northwestern.edu)
International Center for Advanced Internet Research (www.icair.org)
Northwestern University**

**Director, Metropolitan Research and Education Network (www.mren.org)
Co-Director, StarLight (www.startap.net/starlight)
PI IRNC: RXP: StarLight SDX**

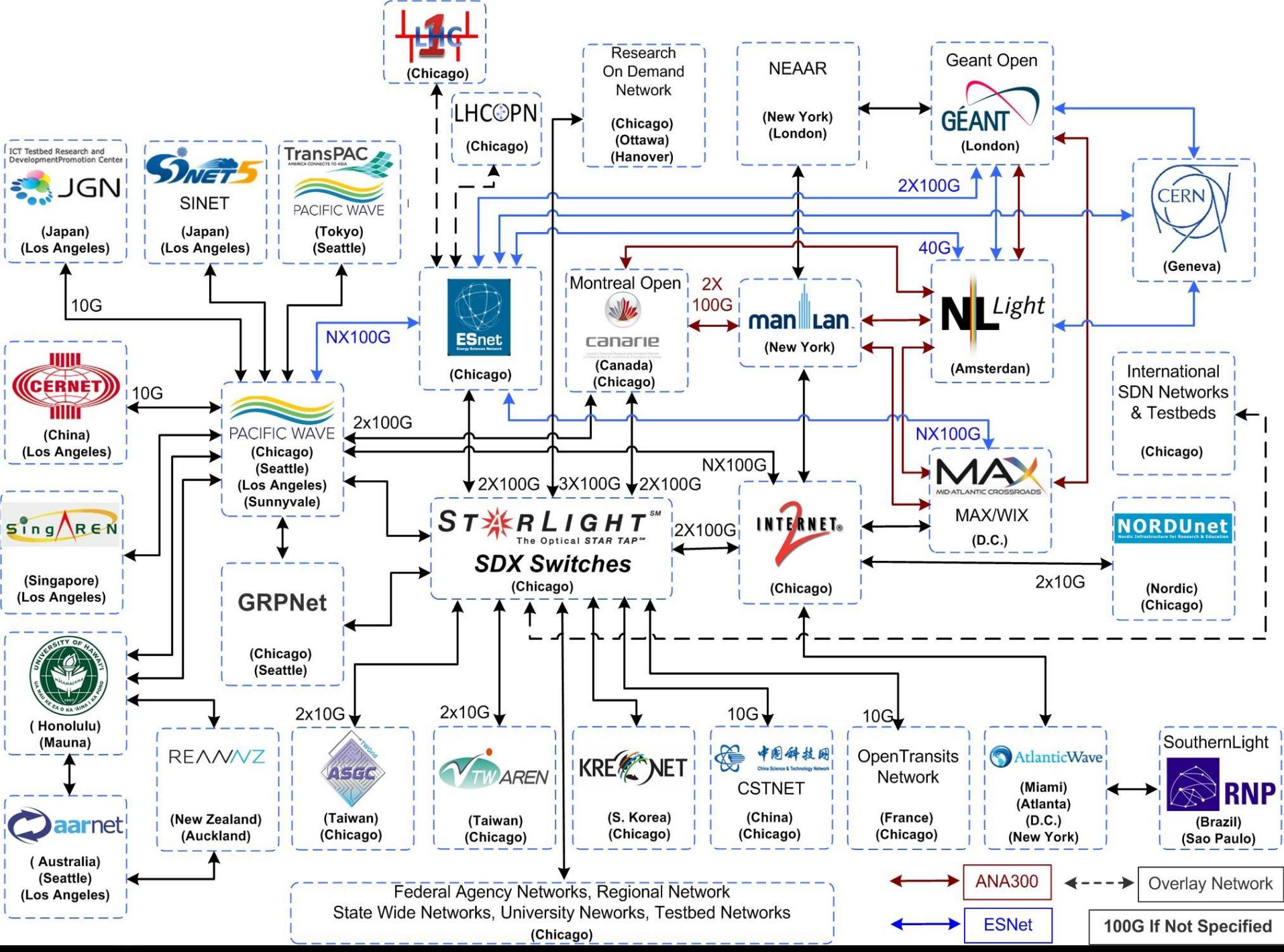
**Co-PI Tom DeFanti, Research Scientist, (tdefanti@soe.ucsd.edu)
California Institute for Telecommunications and Information Technology (Calit2),
University of California, San Diego
Co-Director, StarLight**

**Co-PI Maxine Brown, Director, (maxine@uic.edu)
Electronic Visualization Laboratory, University of Illinois at Chicago
Co-Director, StarLight**

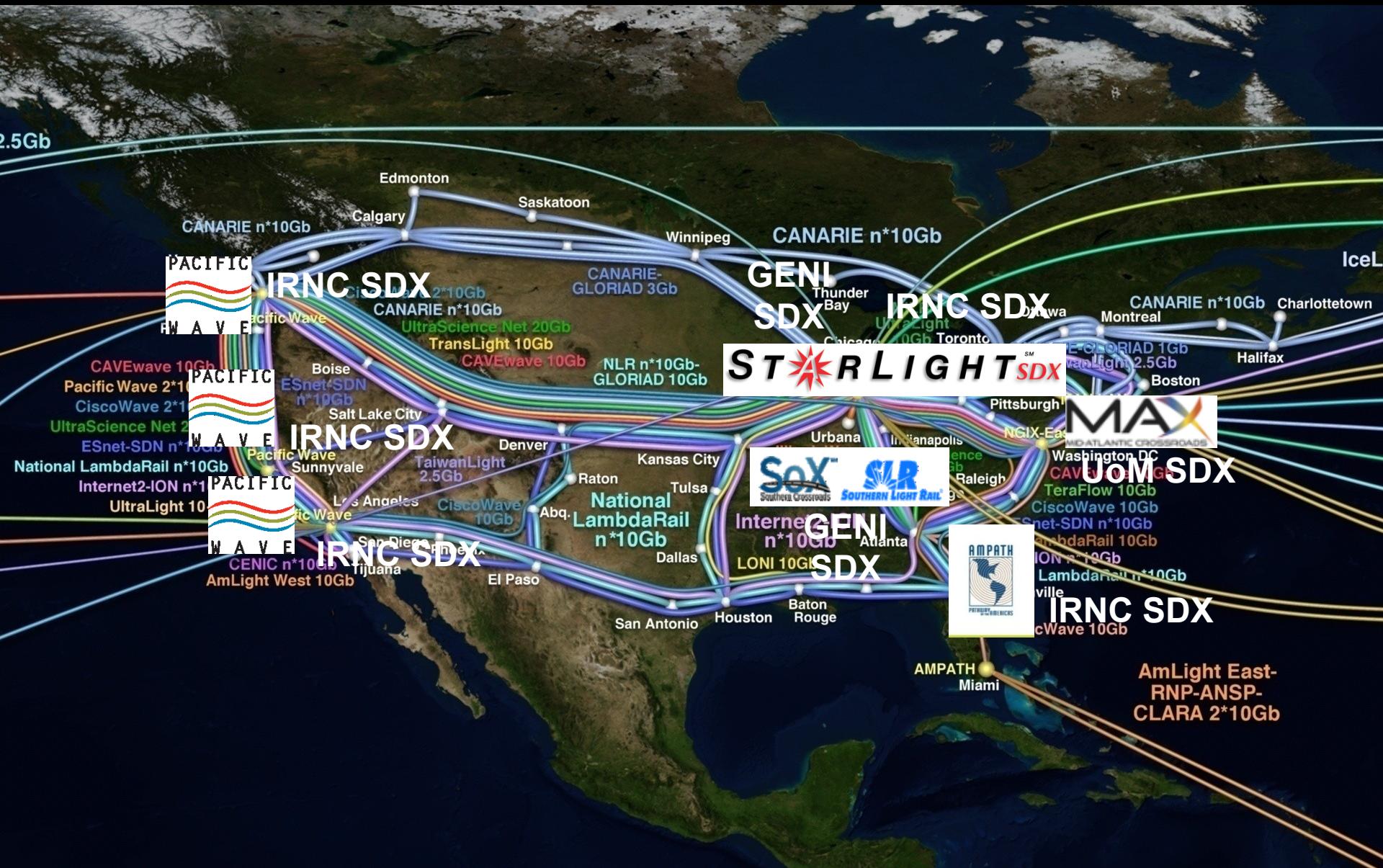
**Jim Chen, Associate Director, International Center for Advanced Internet
Research, Northwestern University**

**National Science Foundation
International Research Network Connections Program**

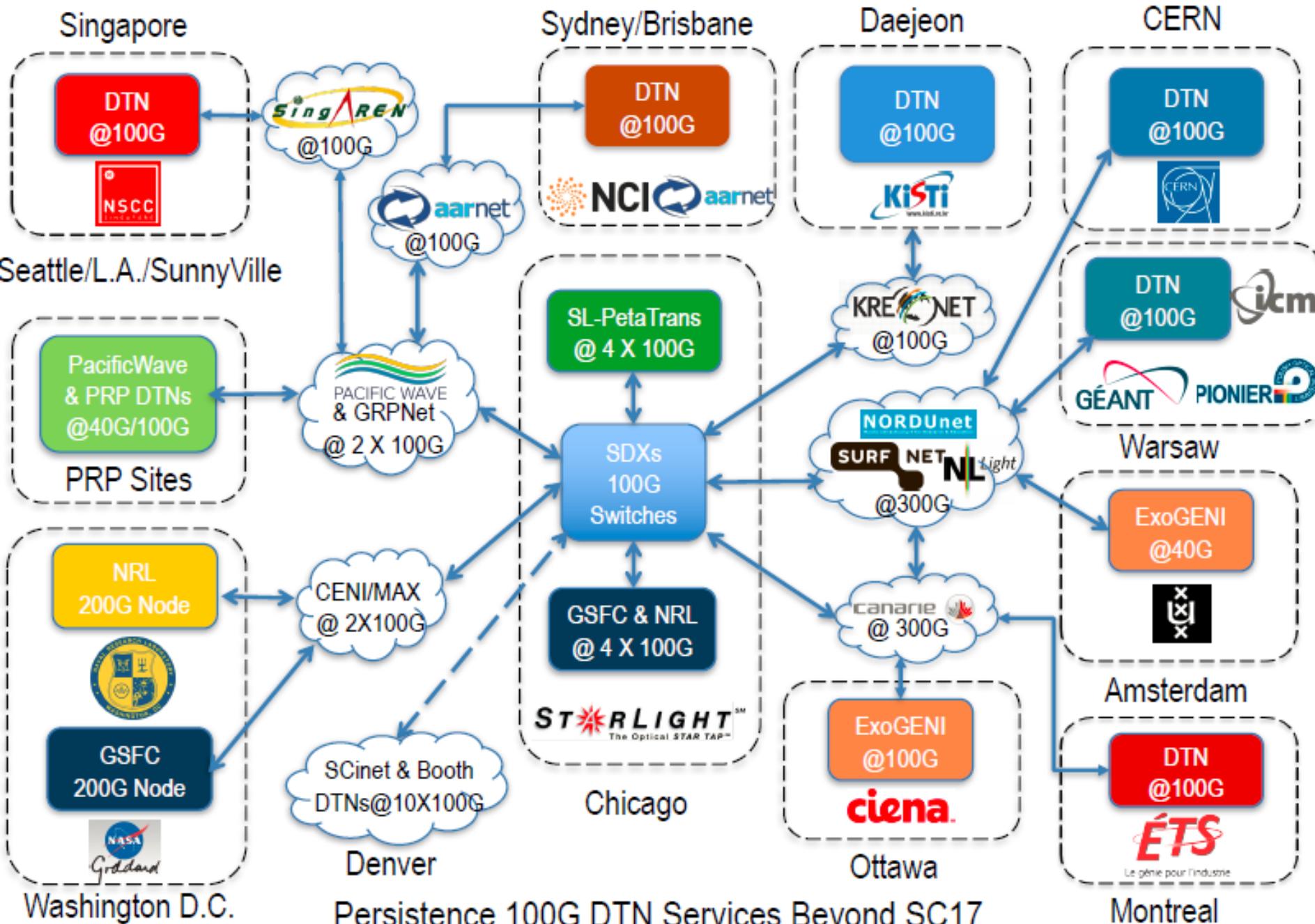




Emerging US SDX Interoperable Fabric



PetaTrans: Petascale Sciences Data Transfer

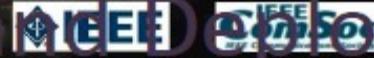


PARIS

March 7 - 9, 2017



Designing and Deploying A



Bioinformatics Software-Defined Network Exchange (SDX): Architecture, Services, Capabilities, and Foundation Technologies

Joe Mambretti, Jim Chen, Fei Yeh

International Center for Advanced Internet Research

Northwestern University

Robert Grossman, Piers Nash, Alison Heath, Renuka Arya, Stuti Agrawal,

Zhenyu Zhang

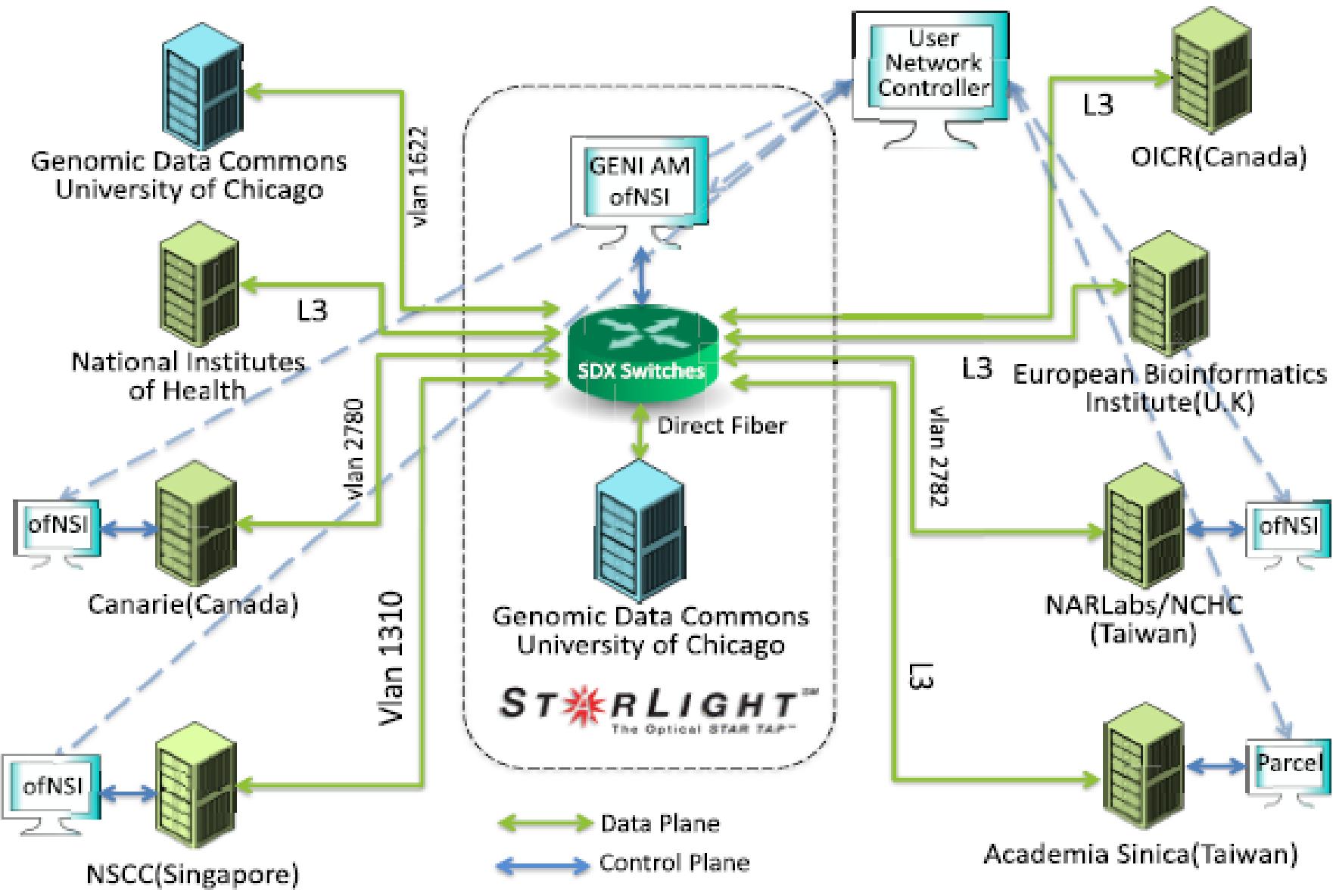
Center for Data Intensive Science

University of Chicago

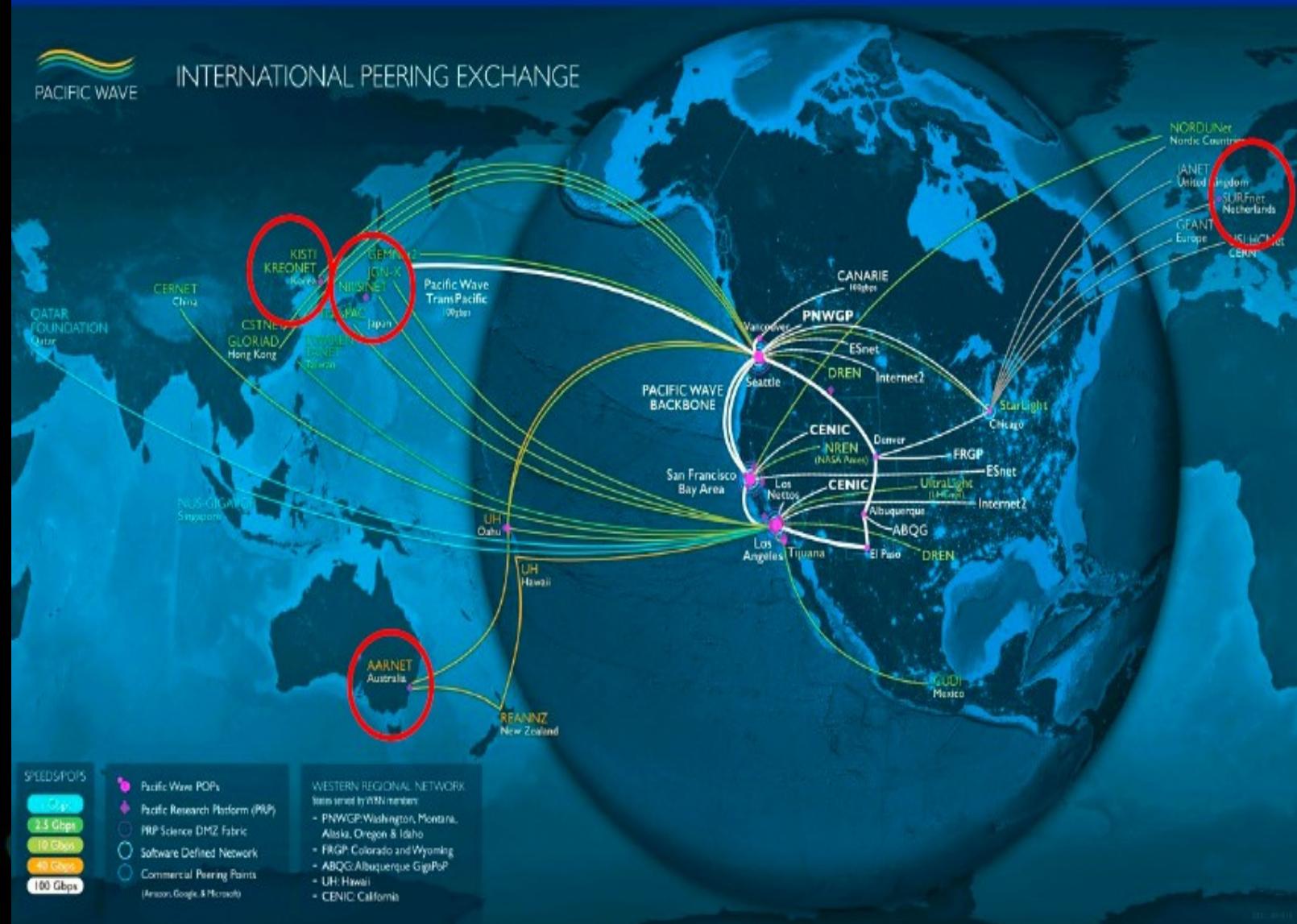
Chicago, Illinois, USA



2016 Bioinformatics SDXs Network



Global Research Platform: Building On CENIC/Pacific Wave, GLIF and GLIF GOLEs (e.g., StarLight et al)



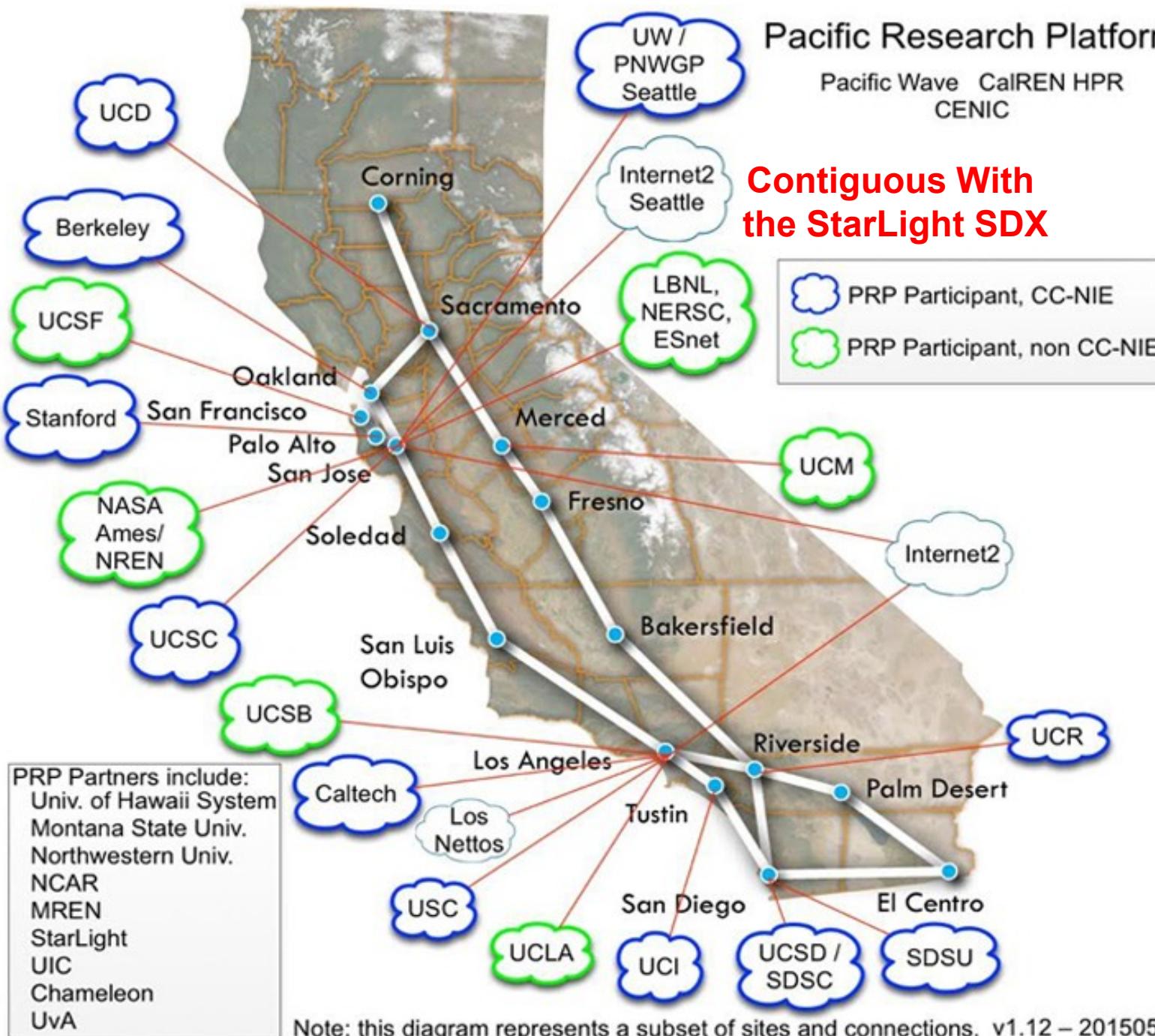
Global Research Platform (GRP)

- A Emerging International Fabric
- A Specialized Globally Distributed Environment/Platform For Science Discovery and Innovation
- Based On State-Of-the-Art-Clouds, Networks, Storage Systems, Data Repositories, etc
- Interconnected With Computational Grids, Supercomputing Centers, Specialized Instruments, et al
- Also, Based On World-Wide 100 Gbps (Soon 100 G+) Networks
- Leveraging Advanced Architectural Concepts, e.g., SDN/SDX/SDI – Science DMZs
- Core Building Blocks Exist Today!
- Ref: 1st Demonstrations @ SC15, Austin Texas November 2015
- Subsequent Demonstrations @ SC16 Salt Lake City Utah, November 2016, Global LambdaGrid Workshop 2016 and 2017,
- Major Demonstrations at SC17 in Denver, Colorado, Planned Demonstrations for SC18 in Dallas Texas in November

Pacific Research Platform

Pacific Wave CalREN HPR
CENIC

Contiguous With
the StarLight SDX





JOINT BIG DATA TESTBED

DRAFT

Still Largely
Reflects SC17
Config in SL
booth..!!

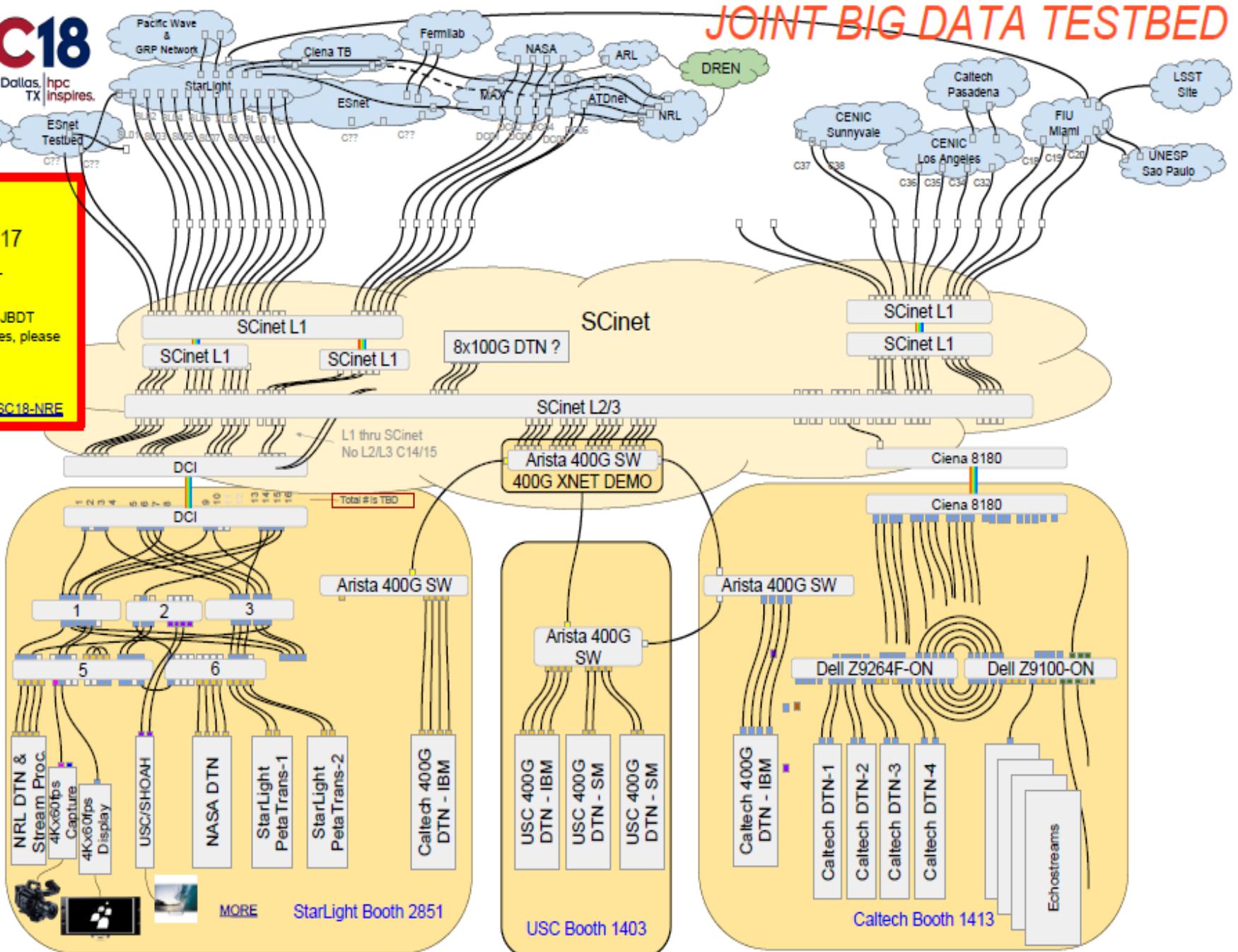
tinyurl.com/SC18-JBDT
To request changes, please
leave a comment

See also

<http://tinyurl.com/SC18-NRF>

- 4000 - LR8
- 1000 - CLR4
- 1000 - CLR4
- 1000 - LR4
- 1000 - SR4
- 1000 - DAC
- 40G - SR4
- 40G - QSFP+
- 10G
- 10G

09/02/2018



MORE

StarLight Booth 2851

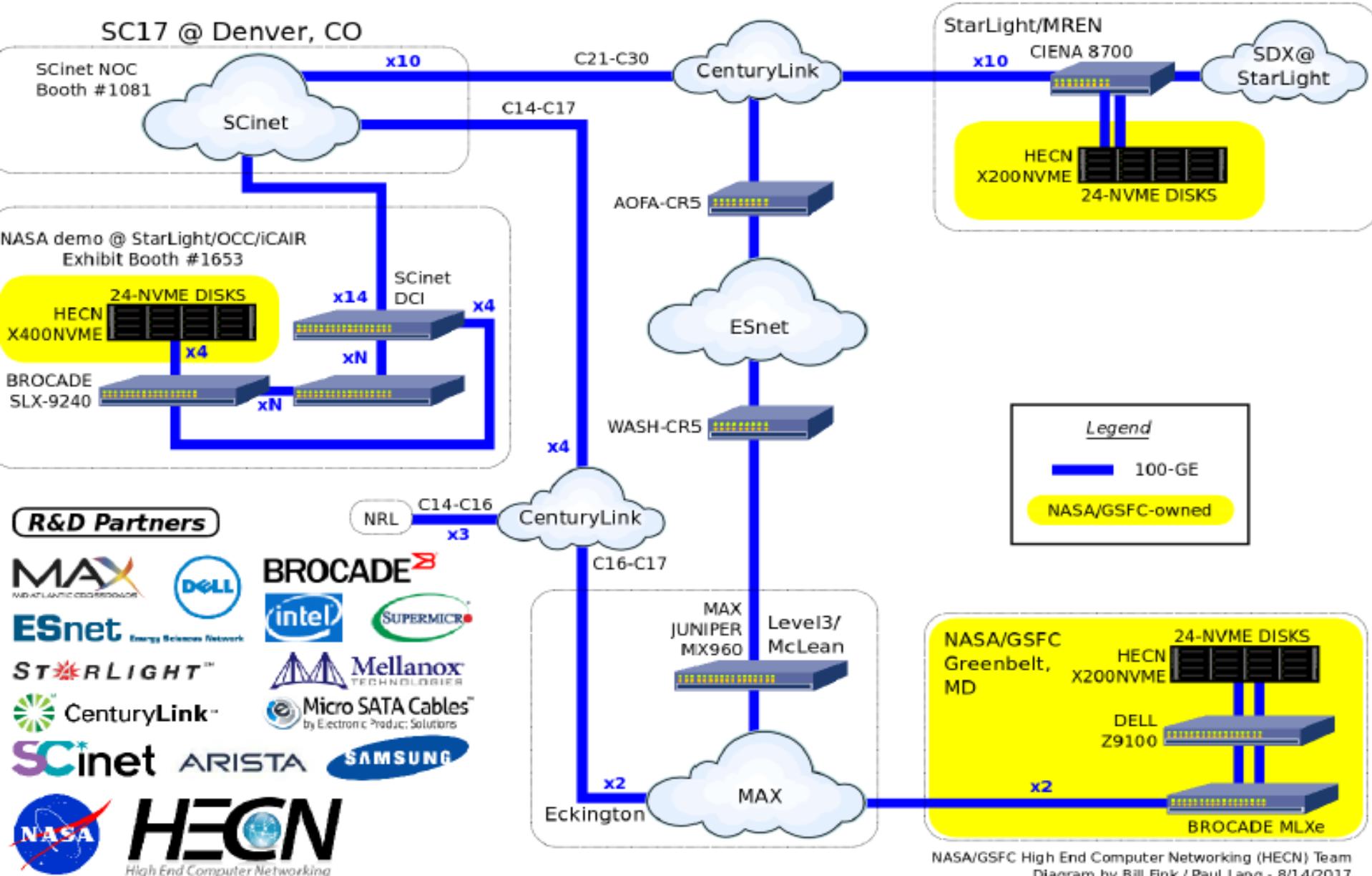
USC Booth 1403

Caltech Booth 1413

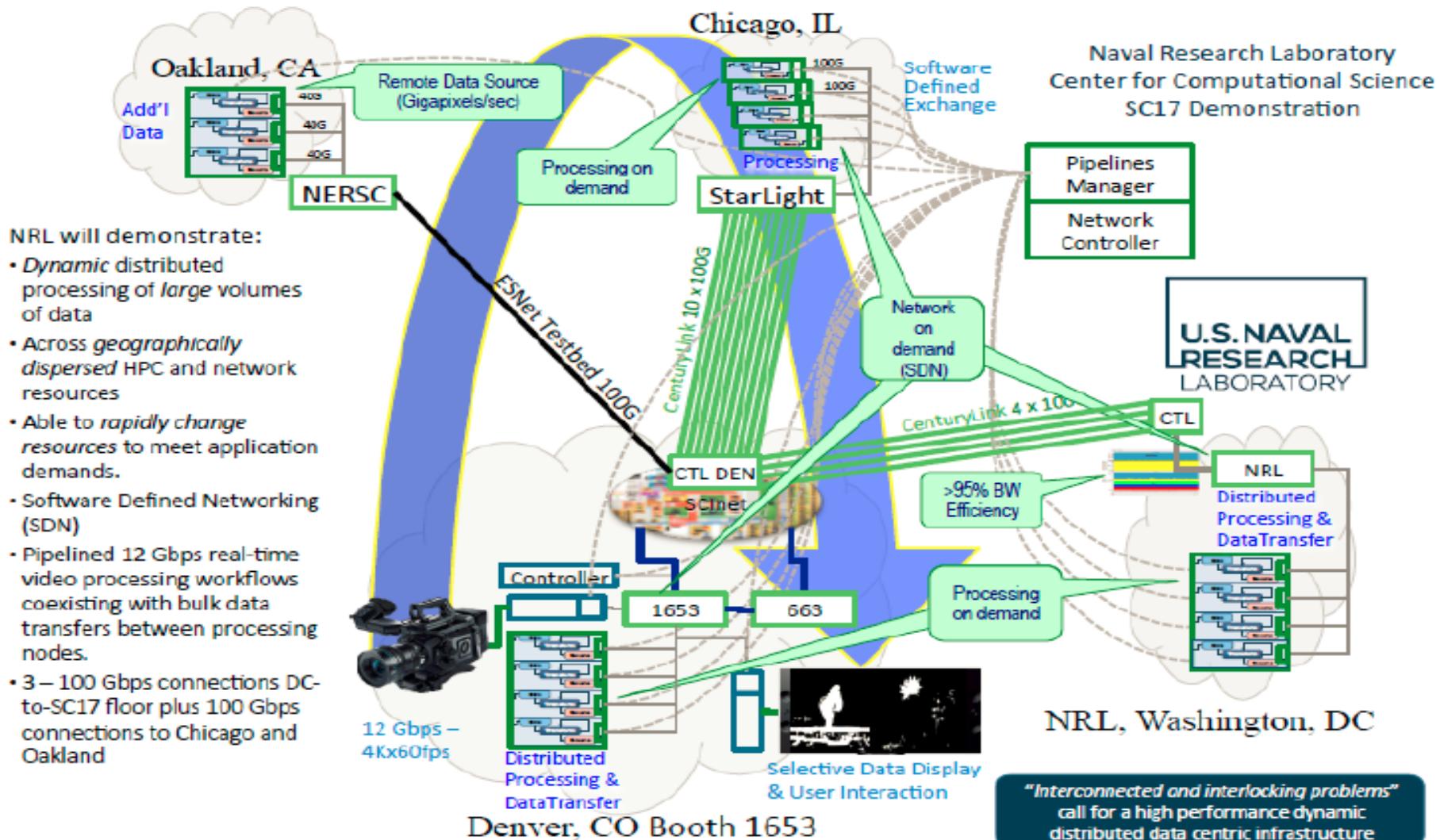
STARLIGHT

Demonstrations of 400 Gbps Disk-to-Disk WAN File Transfers using iWARP and NVMe Drives

An SC17 Collaborative Initiative Among NASA and Several Partners



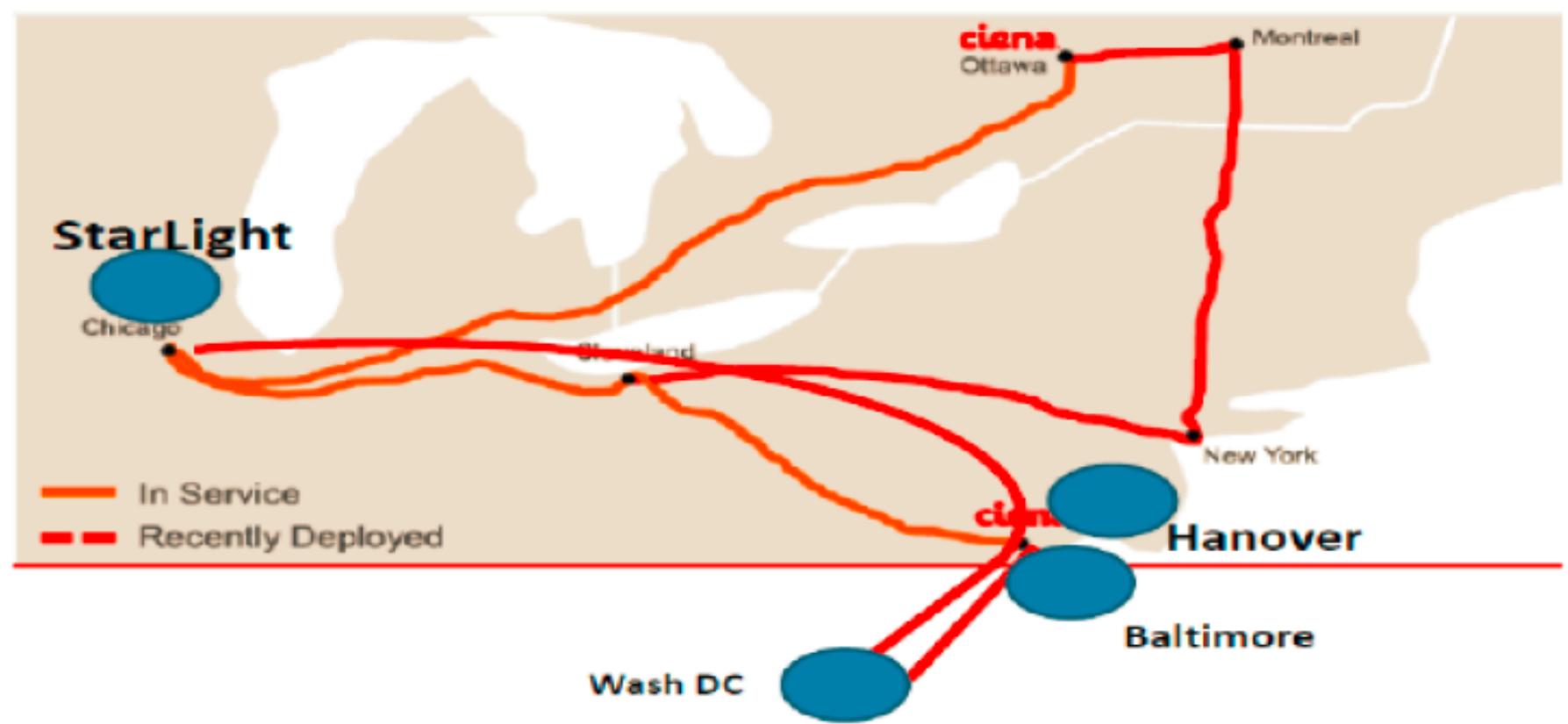
Dynamic Distributed Data Processing



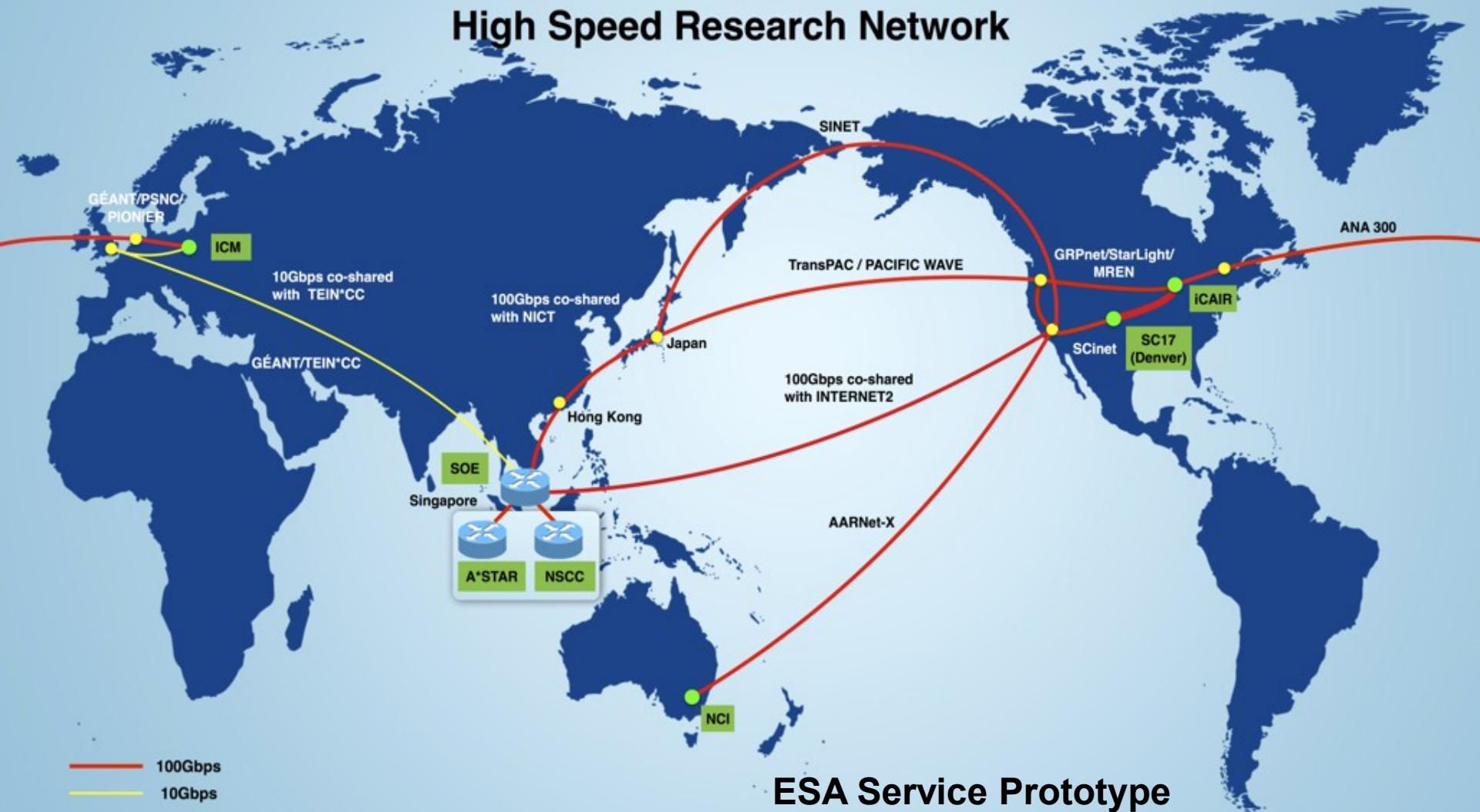
STARLIGHTSM

100 Gbps DTN Optical Testbed

Ciena's OPⁿ research network testbed

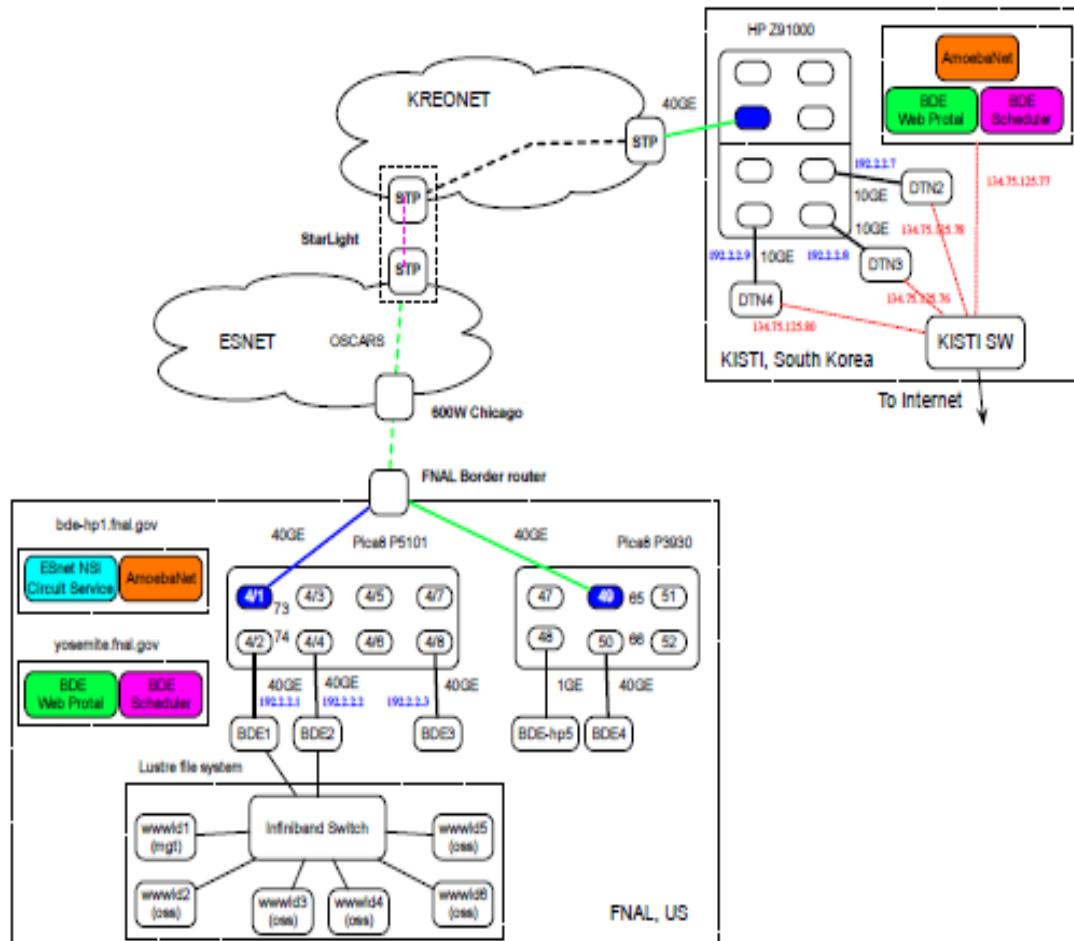


High Speed Research Network





A Cross-Pacific SDN Testbed





www.chameleondcloud.org

CHAMELEON: A LARGE SCALE, RECONFIGURABLE EXPERIMENTAL INSTRUMENT FOR COMPUTER SCIENCE

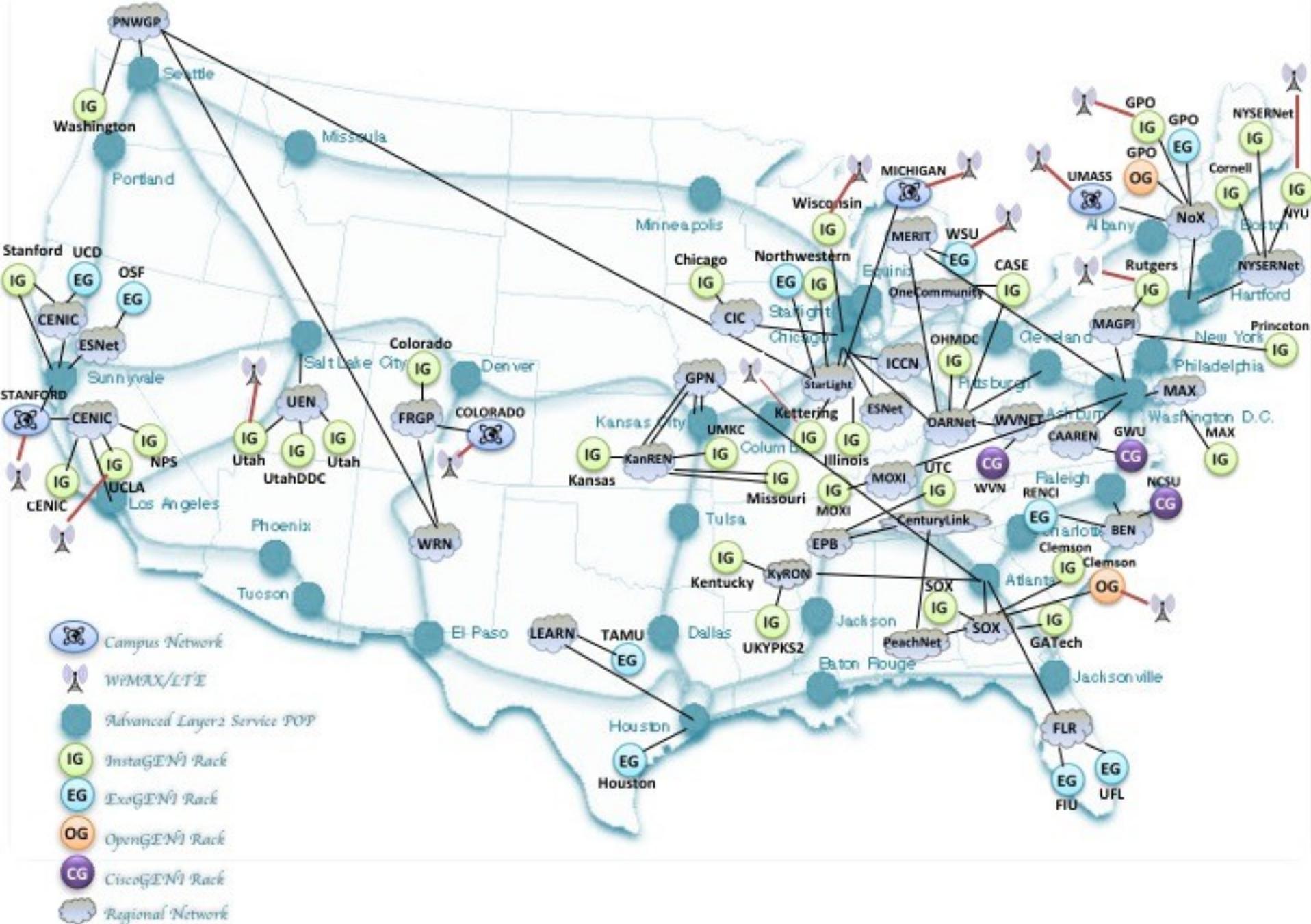
Kate Keahey

Joe Mambretti, Pierre Riteau, Paul Ruth, Dan Stanzione

SEPTEMBER 28, 2017



STARLIGHTSM



GENI – iCAIR P4 Testbed – Integrated With GENI StarLight SDX

- In Partnership With The GENI Initiative, iCAIR Is Developing a P4 Testbed for Computer Science Research.
- The Testbed Will Be Integrated With the GENI SDX At StarLight
- P4 (Programming Protocol-Independent Packet Processors).
- An Emerging Networking Programming Language,
- A Domain Specific Language for Network Protocols.
- Highly Flexible In Contrast To OpenFlow
- Testbed Based on Tofino (Barefoot Networks) Switches
- Compiler (V16) Enables Rules To Be Dynamically Implemented In Chip

www.startap.net/starlight

**Thanks to the NSF, DOE, DARPA,
NIH, USGS, NASA,
Universities, National Labs,
International Partners,
and Other Supporters**

