Software Defined Network Exchanges (SDXs): Services, Architecture, Technology, and Future Directions

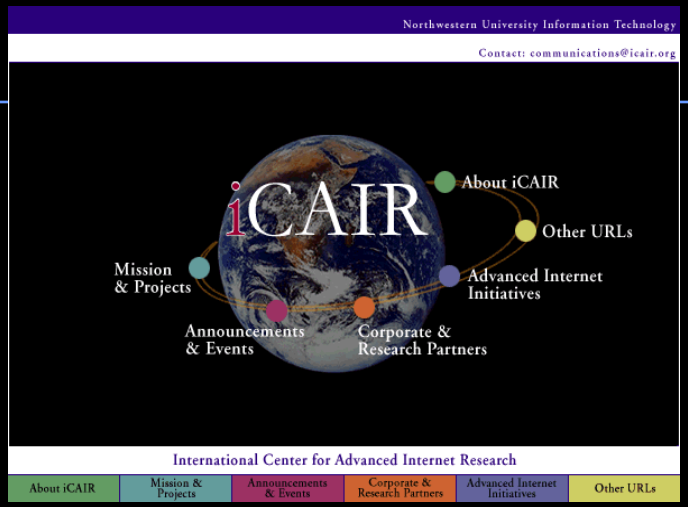
Joe Mambretti, Director, (j-mambretti@northwestern.edu)
International Center for Advanced Internet Research (www.icair.org)
Northwestern University
Director, Metropolitan Research and Education Network (www.mren.org)
Director, StarLight, PI- StarLight SDX, PI-iGENI, PI-OMNINet, Co-PI Chameleon, (www.startap.net/starlight)

University of Tokyo
Tokyo, Japan
February 16, 2016

iCAIR

STARLIGHT℠

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 Science**
- **Advanced Applications, Middleware, Large-Scale Infrastructure, NG Optical Networks and Testbeds, Public Policy Studies and Forums Related to NG Networks**
- **Three Major Areas of Activity: a) Basic Research b) Design and Implementation of Prototypes c) Operations of Specialized Communication Facilities (e.g., StarLight)**

GriPhyN
Data Intensive Science

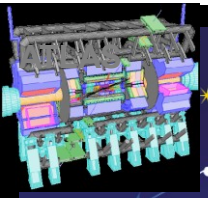
NEPTUNE
A Fiber-Optic Telescope to Inner Space

NEES

GENOMES to LIFE
BIOLOGICAL SOLUTIONS FOR ENERGY CHALLENGES
U.S. DEPARTMENT OF ENERGY

NG Digital Sky Survey

NEON
National Ecological Observatory Network

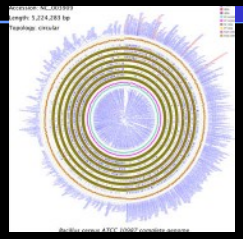


ALMA: Atacama Large Millimeter Array

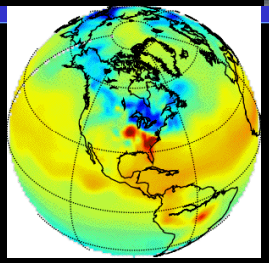
ANDRILL: Antarctic Geological Drilling
www.andrill.org

BIRN

BIRN: Biomedical Informatics Research Network
www.nbirn.net



CAMERA metagenomics
camera.calit2.net



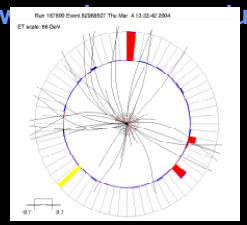
Carbon Tracker
www.esrl.noaa.gov/gmd/ccgg/carbontracker

CineGrid

CineGrid
www.cinegrid.org

LHC

LHCONE
www.lhcone.net



DØ (DZero)
www-d0.fnal.gov

GEON

GEON: Geosciences Network
www.geongrid.org

GLEON: Global Lake Ecological Observatory Network

OOI
OCEAN OBSERVATORIES INITIATIVE
CYBERINFRASTRUCTURE
Providing a link between ocean research and discovery

OOI-CI
ci.oceanobservatories.org

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

CLASS

Comprehensive Large-Array Stewardship System
www.class.noaa.gov

LIGO
www.ligo.org

WLCG
Worldwide LHC Computing Grid

WLCG
lcg.web.cern.ch/LCG/public/

PRAGMA

Pacific Rim Applications and Grid Middleware Assembly
www.pragma-grid.net

TeraGrid

TeraGrid
www.teragrid.org

IVOA

IVOA: International Virtual Observatory
www.ivoa.net

Open Science Grid

OSG
www.opensciencegrid.org

the globus alliance

Globus Alliance
www.globus.org

SKA
SQUARE KILOMETRE ARRAY

SKA
www.skatelescope.org

Sloan Digital Sky Survey
www.sdss.org

XSEDE

XSEDE
www.xsede.org

Compilation By Maxine Brown

STARLIGHT SM

StarLight International/National Communications Exchange Facility

Abbott Hall, Northwestern University's Chicago Campus

STARLIGHT℠

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

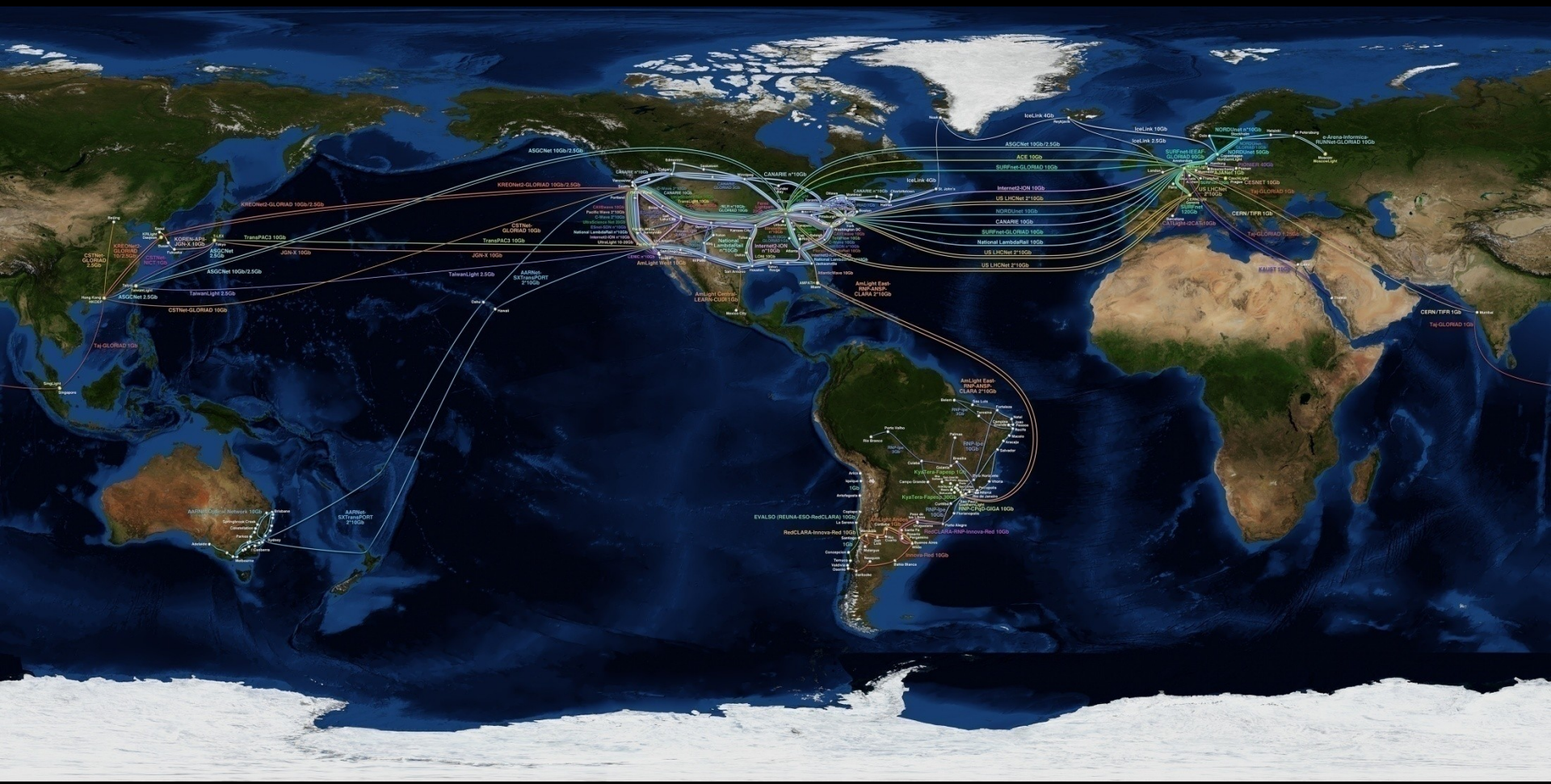
View from StarLight

Abbott Hall, Northwestern University's Chicago Campus

iCAIR

STARLIGHT℠

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

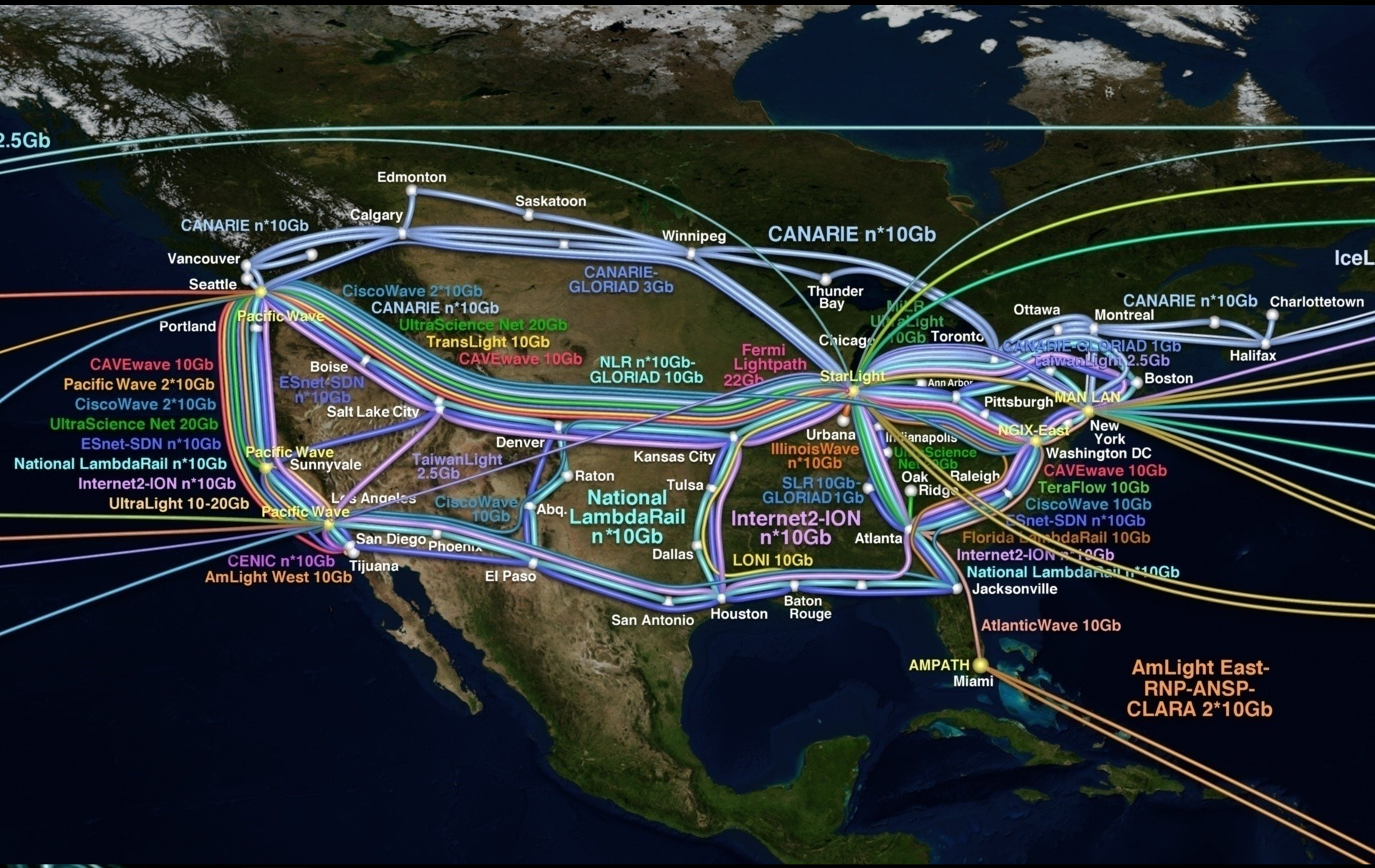


Visualization courtesy of Bob Patterson, NCSA; data compilation by Maxine Brown, UIC.

iCAIR

www.glif.is

STARLIGHT℠



GLIF 2011

iCAIR

STARLIGHT SM

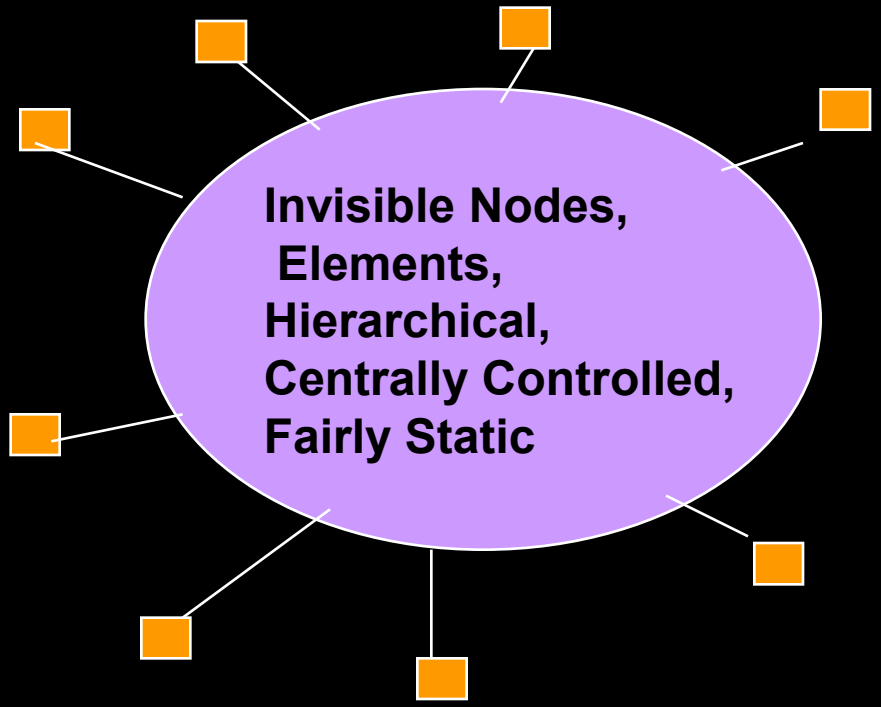
PacificWave

Macro Network Science Themes

- Transition From Legacy Networks To Networks That Take Full Advantage of IT Architecture and Technology
- Extremely Large Capacity (Multi-Tbps Streams)
- High Degrees of Communication Services Customization
- Highly Programmable Networks
- Network Facilities As Enabling Platforms for Any Type of Service
- Network Virtualization
- Highly Distributed Processes
- SDN/SDX/SDI/OCX/SDC/SD*

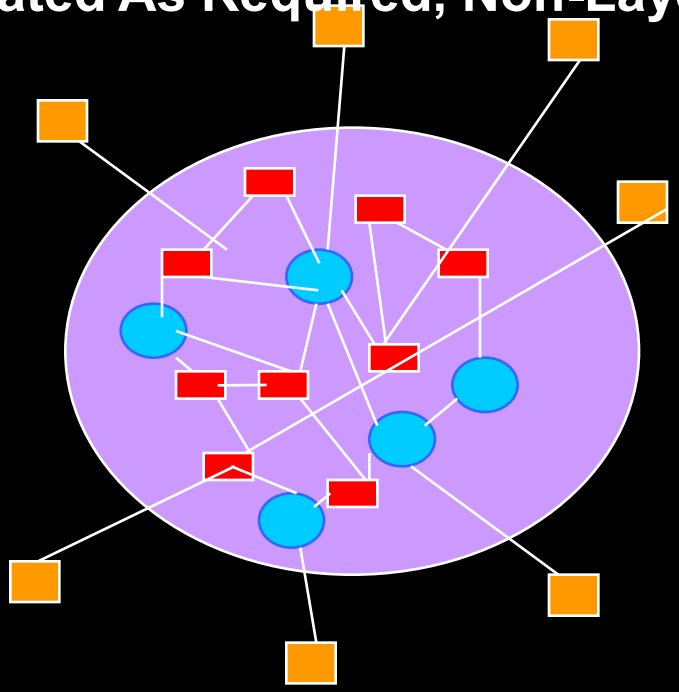
Paradigm Shift – Ubiquitous Services Based on Large Scale Distributed Facility vs Isolated Services Based on Separate Component Resources

Traditional Provider Services:
Invisible, Static Resources,
Centralized Management,
Highly Layered



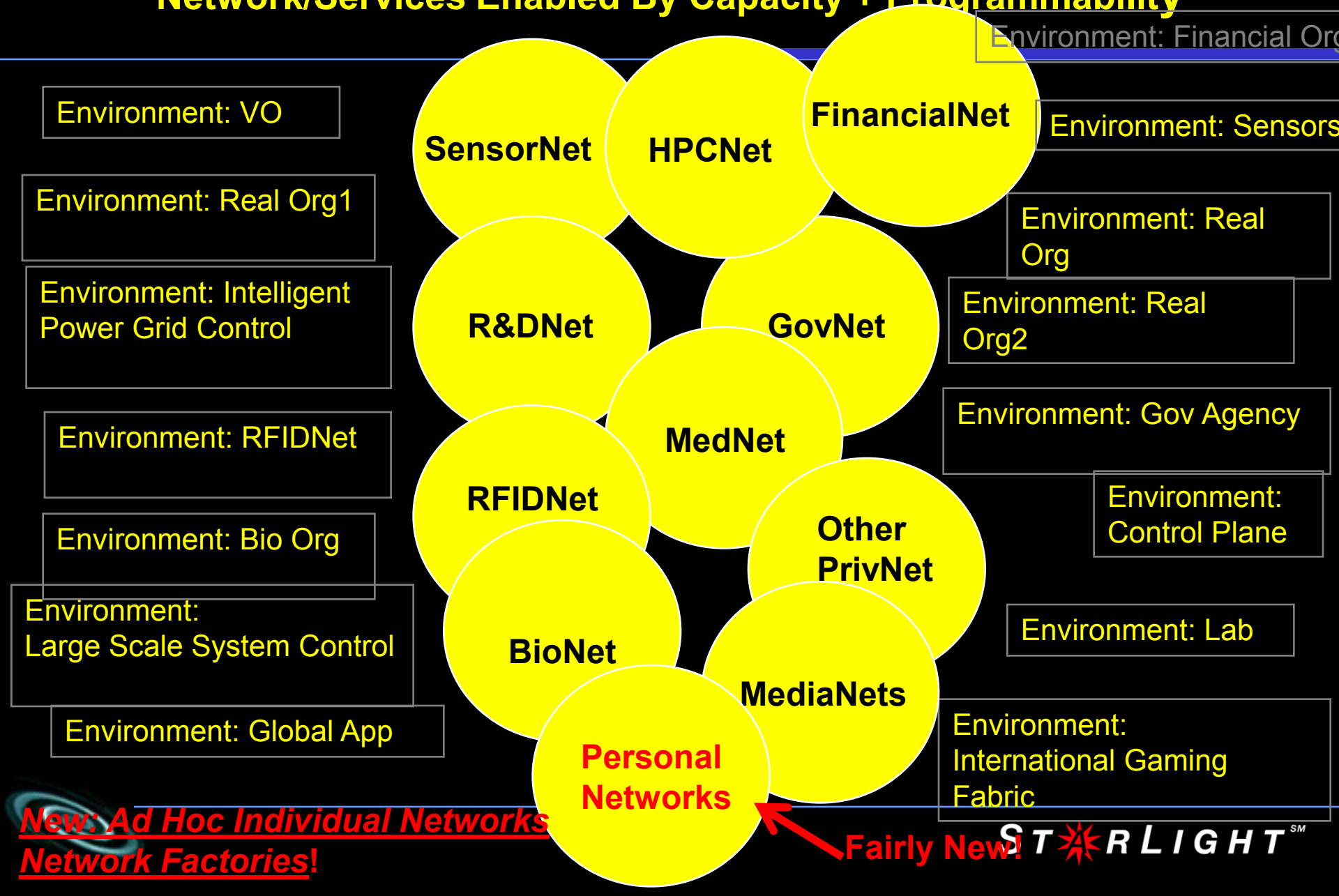
Limited Services, Functionality,
Flexibility, Expandability

Distributed Programmable Resources,
Dynamic Services,
Visible & Accessible Resources,
Integrated As Required, Non-Layered



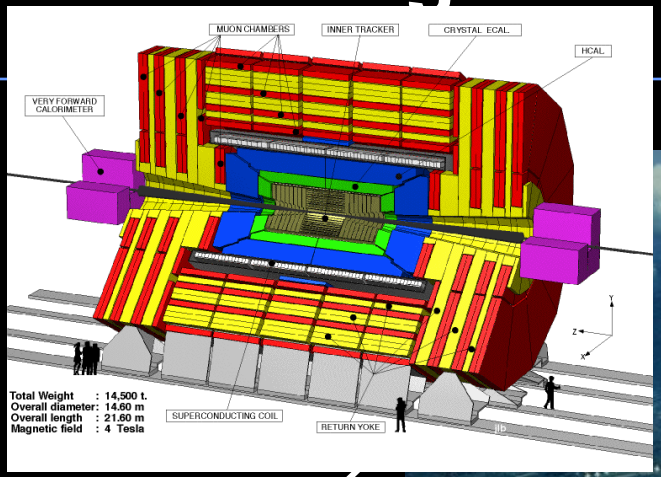
Unlimited Services, Functionality,
Flexibility, Expandability

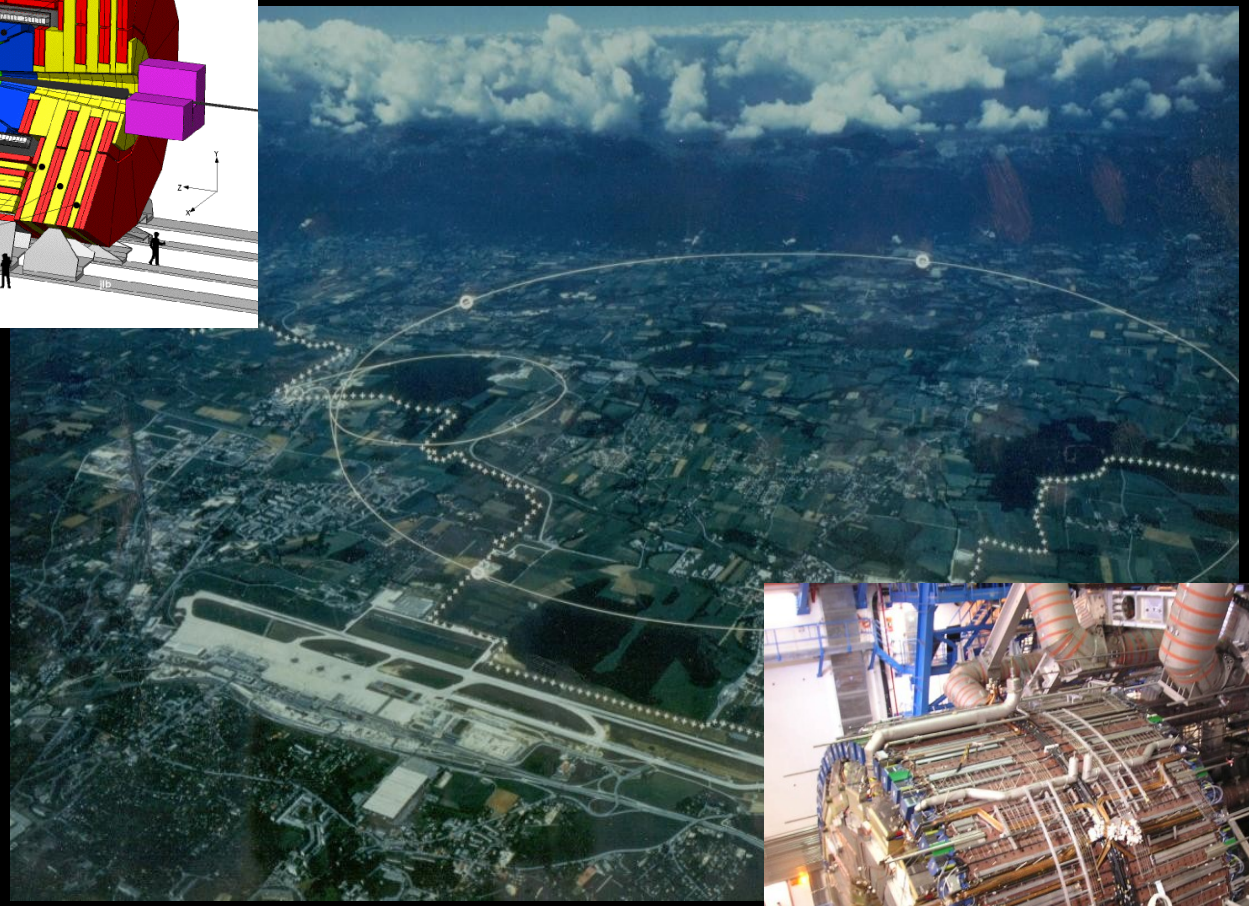
A Next Generation Architecture: *Distributed Facility* Enabling Many Types Network/Services Enabled By Capacity + Programmability

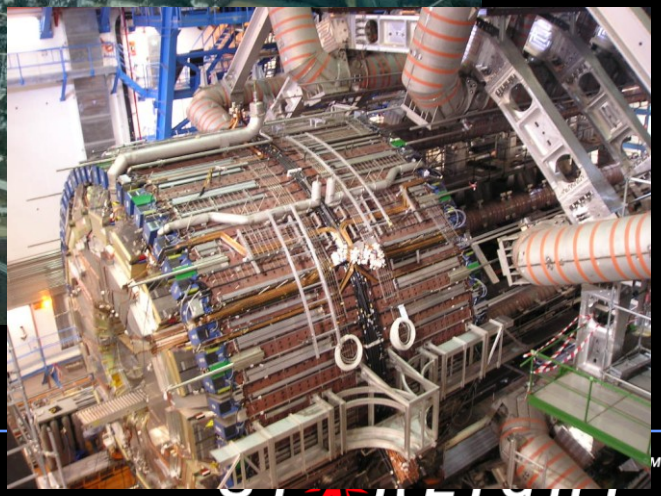


STARLIGHT℠

Large Hadron Collider at CERN





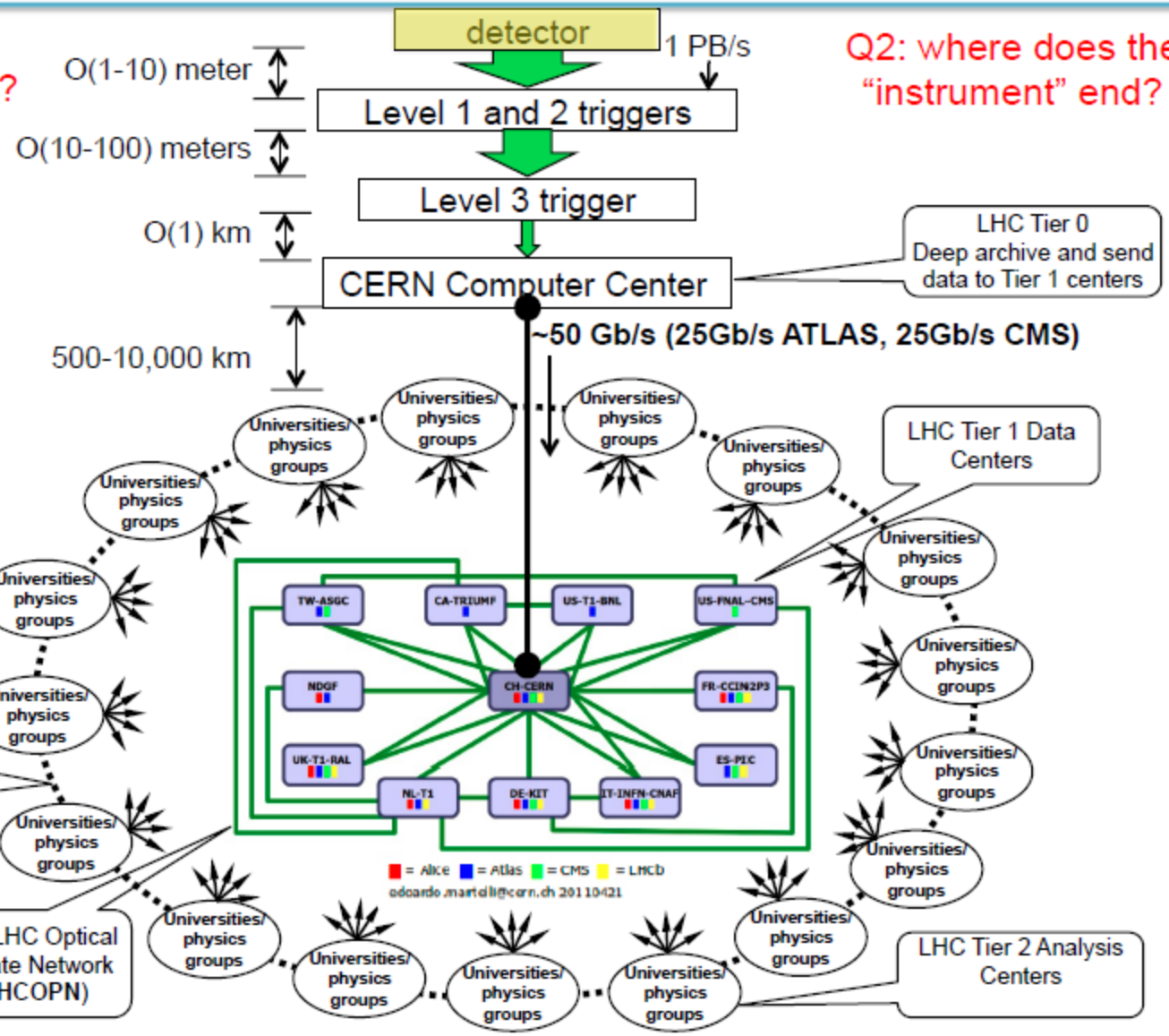


iCAIR

Network-Centric View of Large Hadron Collider (@CERN)

Q1: where does "discovery" occur?

CERN →T1	miles	kms
France	350	565
Italy	570	920
UK	625	1000
Netherlands	625	1000
Germany	700	1185
Spain	850	1400
Nordic	1300	2100
USA – New York	3900	6300
USA - Chicago	4400	7100
Canada – BC	5200	8400
Taiwan	6100	9850

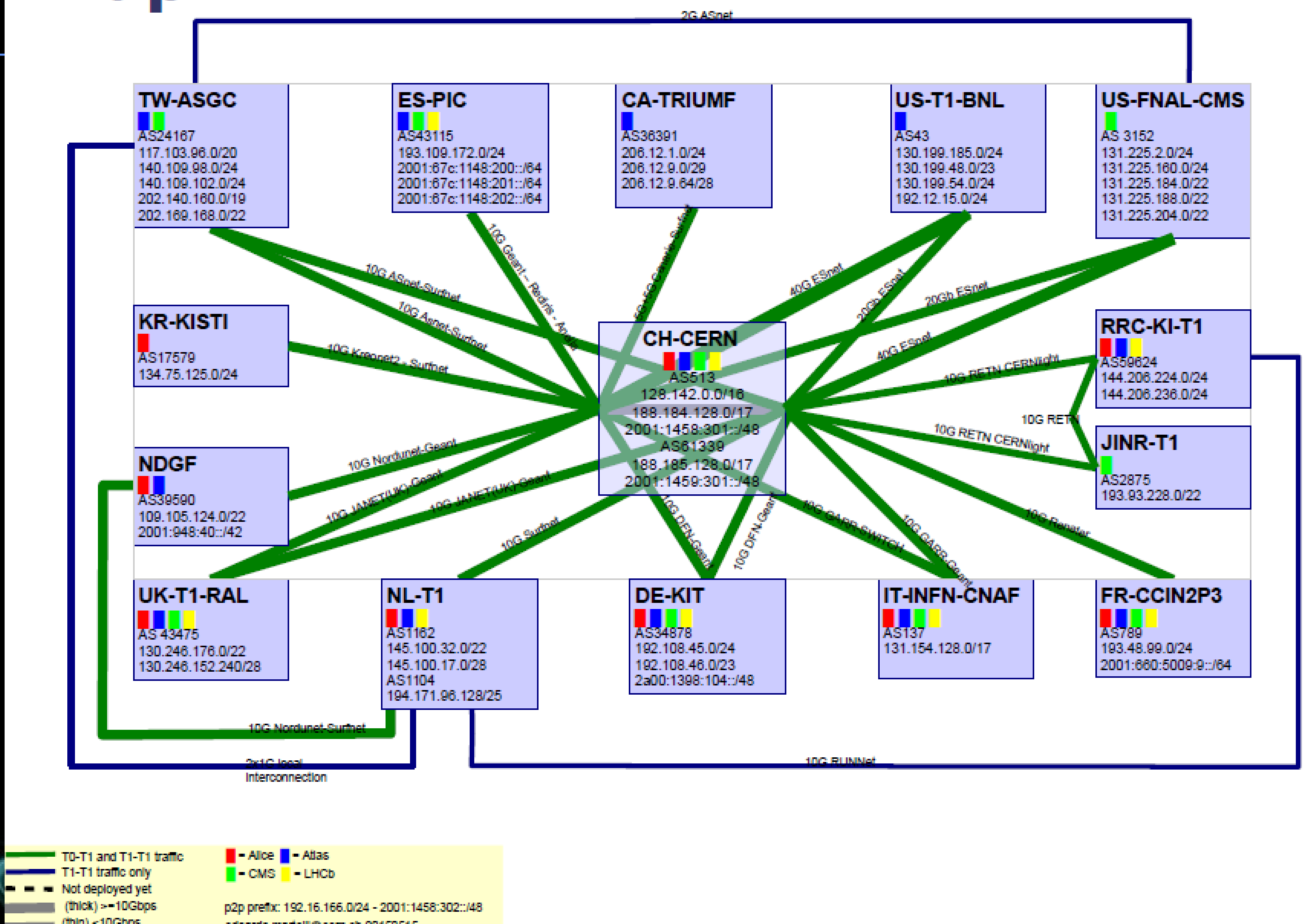


Q2: where does the "instrument" end?

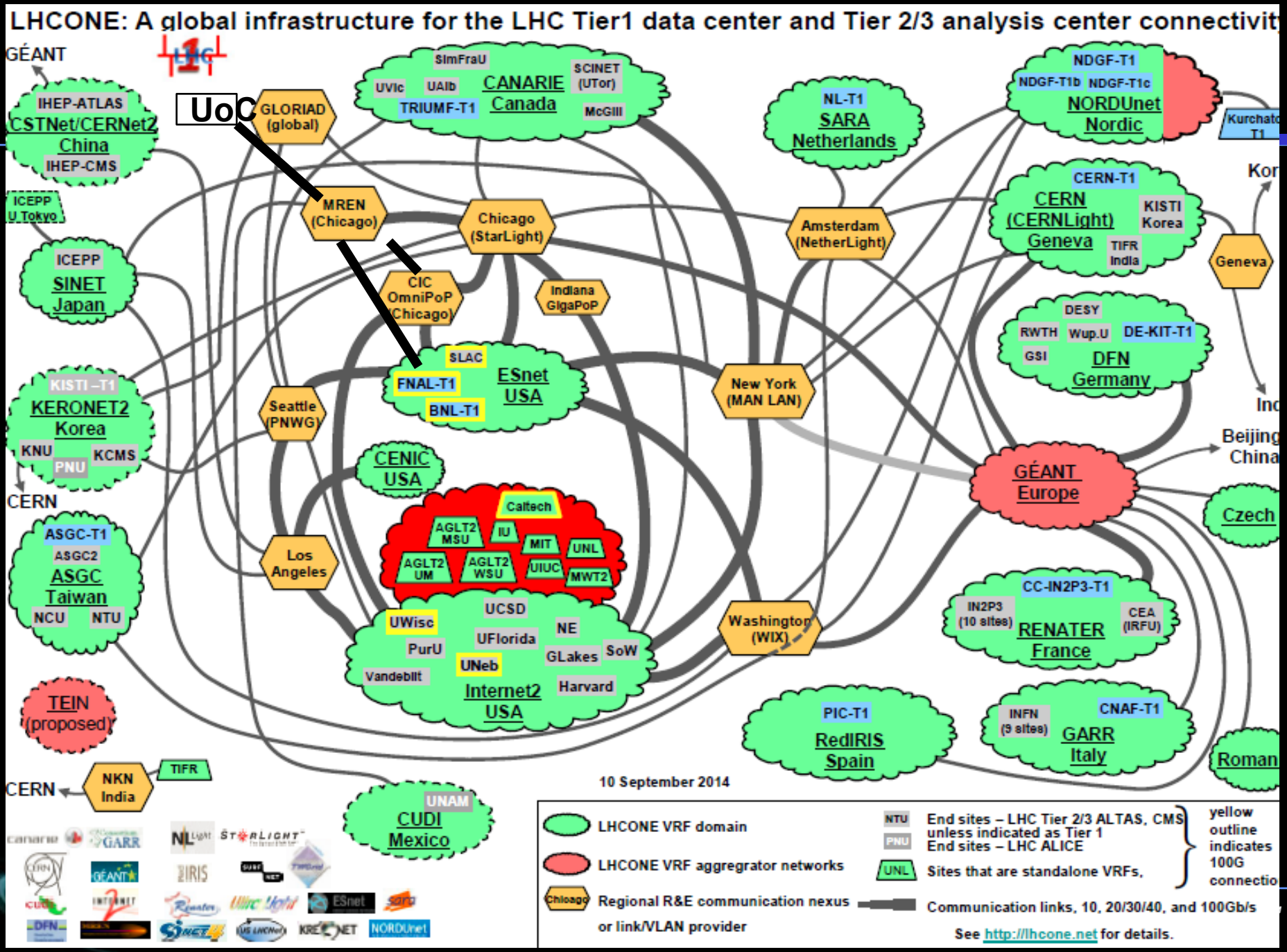
Source: Bill Johnston

Map

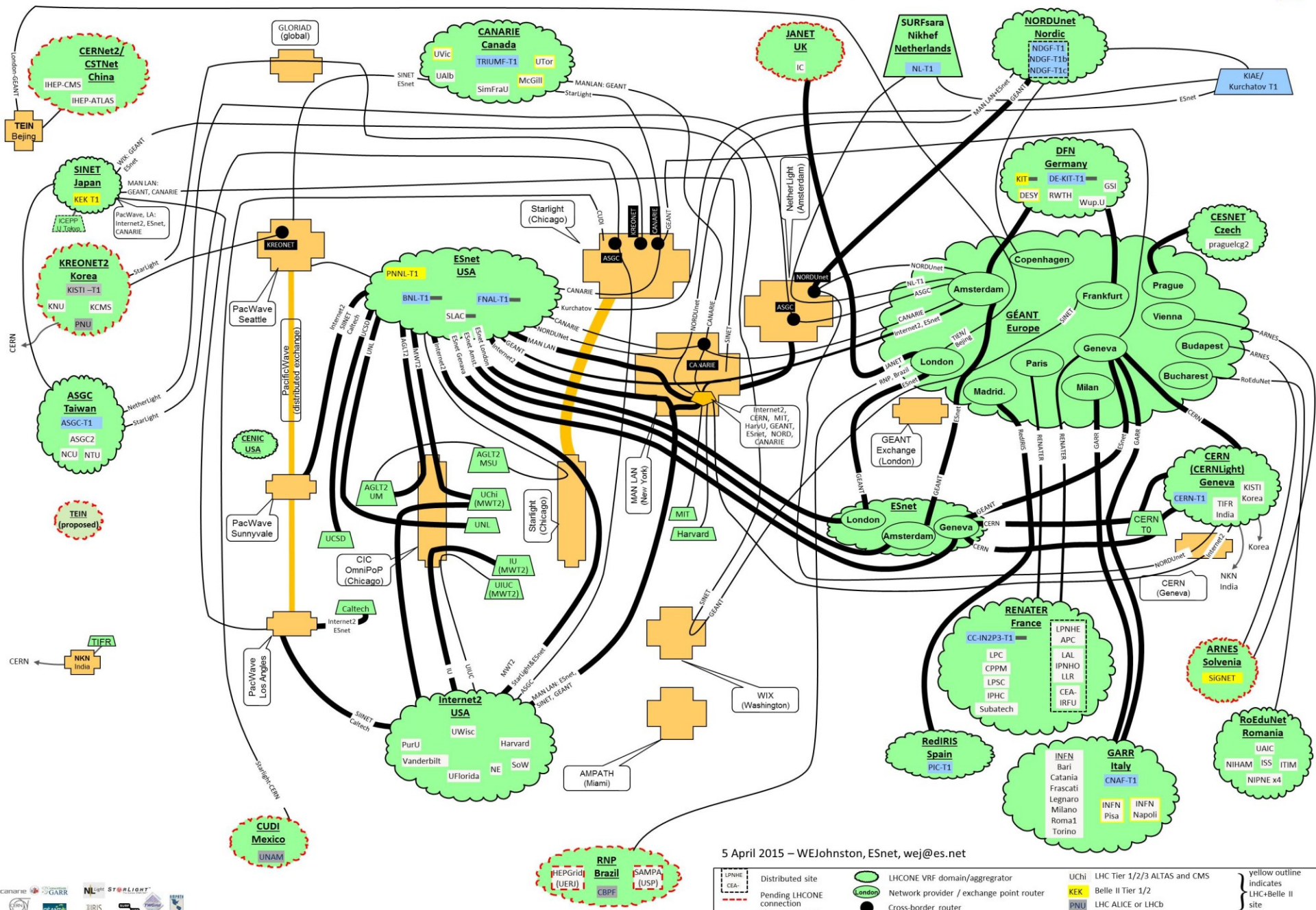
LHCOPN



2G ASnet
TW-ASGC
AS24167
117.103.96.0/20
140.109.98.0/24
140.109.102.0/24
202.140.160.0/19
202.169.168.0/22
ES-PIC
AS43115
193.109.172.0/24
2001:67c:1148:200::/64
2001:67c:1148:201::/64
2001:67c:1148:202::/64
CA-TRIUMF
AS36391
206.12.1.0/24
206.12.9.0/29
206.12.9.64/28
US-T1-BNL
AS43
130.199.185.0/24
130.199.48.0/23
130.199.54.0/24
192.12.15.0/24
US-FNAL-CMS
AS 3152
131.225.2.0/24
131.225.160.0/24
131.225.184.0/22
131.225.188.0/22
131.225.204.0/22
KR-KISTI
AS17579
134.75.125.0/24
CH-CERN
AS513
128.142.0.0/16
188.184.128.0/17
2001:1458:301::/48
AS61339
188.185.128.0/17
2001:1459:301::/48
RRC-KI-T1
AS59624
144.206.224.0/24
144.206.236.0/24
JINR-T1
AS2875
193.93.228.0/22
NDGF
AS39590
109.105.124.0/22
2001:948:40::/42
UK-T1-RAL
AS 43475
130.246.176.0/22
130.246.152.240/28
NL-T1
AS1162
145.100.32.0/22
145.100.17.0/28
AS1104
194.171.96.128/25
DE-KIT
AS34878
192.108.45.0/24
192.108.46.0/23
2a00:139b:104::/48
IT-INFN-CNAF
AS137
131.154.128.0/17
FR-CCIN2P3
AS789
193.48.99.0/24
2001:660:5009:9::/64
10G ASnet-Surfnet
10G Asnet-Surfnet
10G Geant - Rediris - Anella
5G - 10G Canarie-Surfnet
40G ESnet
20Gb ESnet
20Gb ESnet
40G ESnet
10G Kreonet2 - Surfnet
10G RETN CERNlight
10G RETN
10G RETN CERNlight
10G Nordunet-Geant
10G JANET(UK)-Geant
10G JANET(UK)-Geant
10G Surfnet
10G DFN-Geant
10G DFN-Geant
10G GARR-SWITCH
10G GARR-Geant
10G Renater
10G Nordunet-Surfnet
2x1G local interconnection
10G RUNNet
T0-T1 and T1-T1 traffic
T1-T1 traffic only
Not deployed yet
(thick) >=10Gbps
(thin) <10Gbps
- Alice
- Atlas
- CMS
- LHCb
p2p prefix: 192.16.166.0/24 - 2001:1458:302::/48
edoardo.martelli@cern.ch 20150515


LHCONE: A global infrastructure for the LHC Tier1 data center and Tier 2/3 analysis center connectivity
GÉANT
IHEP-ATLAS
CSTNet/CERNet2
China
IHEP-CMS
ICEPP
U Tokyo
ICEPP
SINET
Japan
KISTI –T1
KERONET2
Korea
KNU
PNU
KCMS
CERN
ASGC-T1
ASGC2
ASGC
Taiwan
NCU
NTU
TEIN
(proposed)
CERN
NKN
India
TIFR
UoC
GLORIAD
(global)
MREN
(Chicago)
CIC
OmniPoP
Chicago)
Chicago
(StarLight)
Indiana
GigaPoP
SLAC
FNAL-T1
BNL-T1
ESnet
USA
Seattle
(PNWG)
CENIC
USA
Los
Angeles
Caltech
AGLT2
MSU
IU
MIT
UNL
AGLT2
UM
AGLT2
WSU
UIUC
MWT2
UCSD
UWisc
NE
UFlorida
PurU
GLakes
SoW
UNeb
Vandebilt
Harvard
Internet2
USA
UNAM
CUDI
Mexico
SInFraU
UVic
UAlb
CANARIE
Canada
SCINET
(UTor)
TRIUMF-T1
McGill
NL-T1
SARA
Netherlands
Amsterdam
(NetherLight)
New York
(MAN LAN)
Washington
(WIX)
NDGF-T1
NDGF-T1b
NDGF-T1c
NORDUnet
Nordic
CERN-T1
CERN
(CERNLight)
Geneva
KISTI
Korea
TIFR
India
DESY
RWTH
Wup.U
DE-KIT-T1
GSI
DFN
Germany
GÉANT
Europe
PIC-T1
RedIRIS
Spain
CC-IN2P3-T1
IN2P3
(10 sites)
RENATER
France
CEA
(IRFU)
INFN
(9 sites)
CNAF-T1
GARR
Italy
Kurchato
T1
Geneva
Beijing
China
Czech
Roman
10 September 2014
LHCONE VRF domain
LHCONE VRF aggregator networks
Chicago
Regional R&E communication nexus
or link/VLAN provider
NTU
PNU
End sites – LHC Tier 2/3 ALTAS, CMS
unless indicated as Tier 1
End sites – LHC ALICE
UNL
Sites that are standalone VRFs,
yellow
outline
indicates
100G
connectio
Communication links, 10, 20/30/40, and 100Gb/s
See http://lhcone.net for details.

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



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

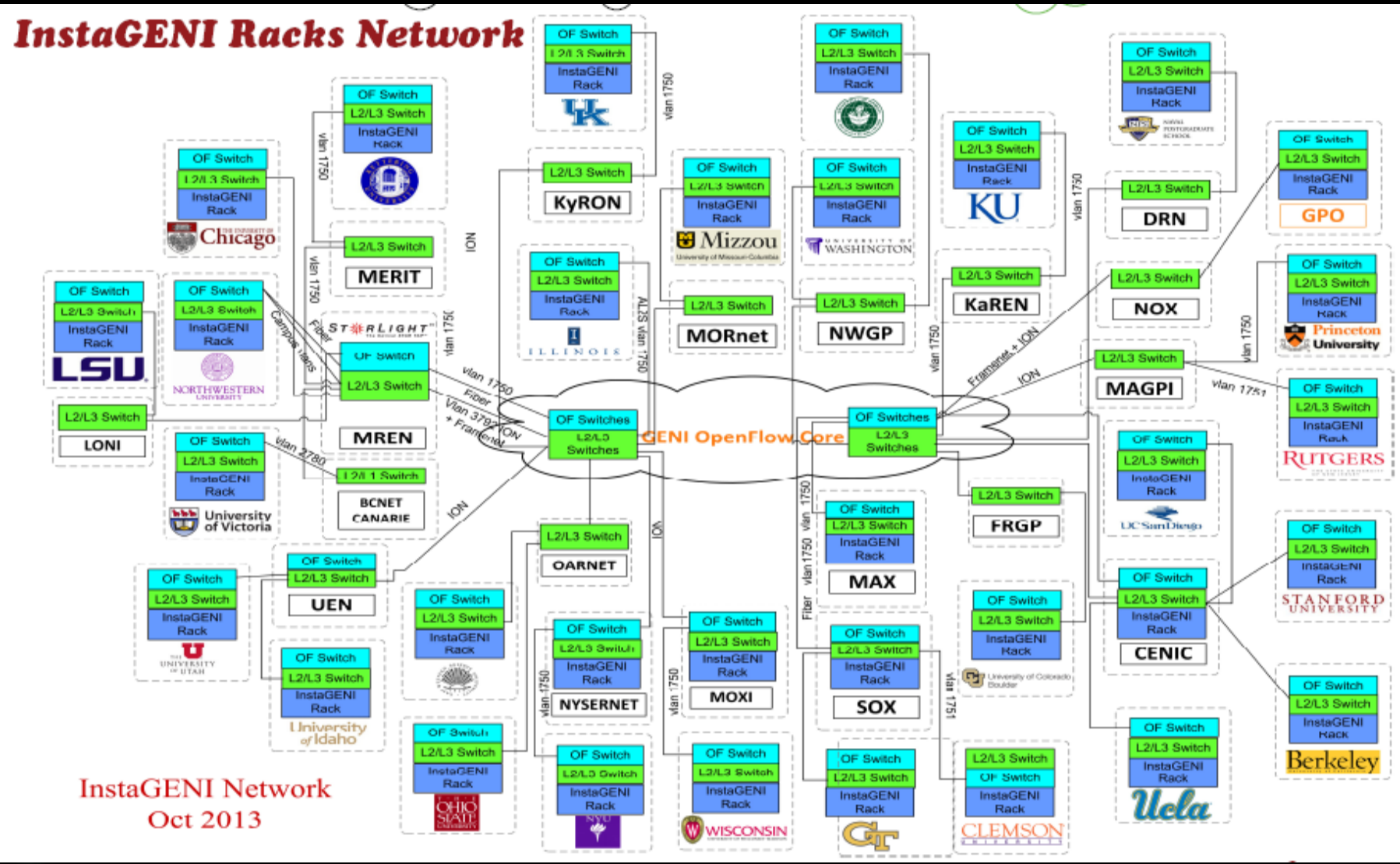
Also see http://lhcone.net for details.

- Distributed site
- Pending LHCONE connection
- Sites connected at 40G-100G
- Broadcast VLAN
- LHCONE VRF domain/aggregator
- Network provider / exchange point router
- Cross-border router
- Regional R&E communication nexus w/ switch providing VLAN connections
- UChi LHC Tier 1/2/3 ALTAS and CMS
- KEK Belle II Tier 1/2
- PNU LHC ALICE or LHCb
- UNL Sites that are standalone VRFs,
- Communication links: 1/10, 20/30/40, and 100Gb/s

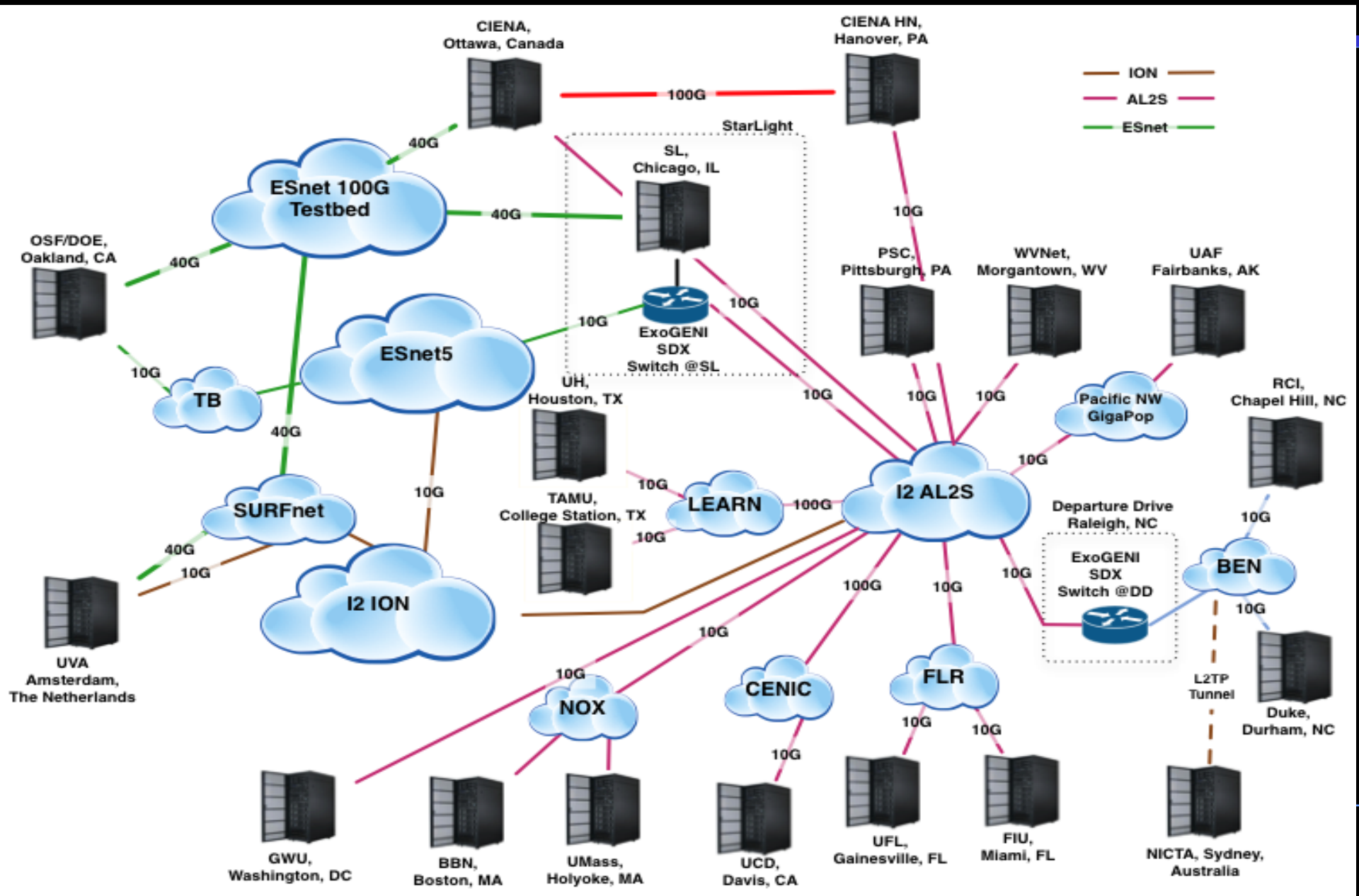
yellow outline indicates LHC+Belle II site


KUAS
NetFPGA
NOX
OpenFlow
NCKU
NetFPGA
NOX
OpenFlow
NOX
OpenFlow
NetFPGA
ETRI
NOX
OpenFlow
NetFPGA
KiSTi
University of Victoria
TOMATO
OpenFlow
InstaGENI
CRC
NOX
OpenFlow
NetFPGA
cenic
InstaGENI
OpenFlow
JGN-X
NSI
東京大学
THE UNIVERSITY OF TOKYO
VNode
PACIFIC GIGAPOP
aarnet
NICTA
ExoGENI
OpenFlow
NCHC
NetFPGA
NOX
OpenFlow
ofNSI
KRLight
NSI
canarie
APPLIED RESEARCH CENTER FOR COMPUTER NETWORKS
OpenFlow
GLORIAD
NSI
SingLight
SingAREN
OpenFlow
NetFPGA
Northern Wave
SoX
Southern Crossroads
GENI AM
OpenFlow
InstaGENI
renci
ExoGENI
OpenFlow
MREN
STARLIGHT
The Optical STAR TAP
GENI AM
ExoGENI
OpenFlow
NSI
ofNSI
INTERNET2
GENI AM
OpenFlow
BEN
Breakable Experimental Network
ExoGENI
Duke
UNIVERSITY
ExoGENI
OpenFlow
CNC
OpenFlow
NORTHWESTERN
UNIVERSITY
NetFPGA
NOX
OpenFlow
TOMATO
InstaGENI
OpenFlow
AMPATH
NSI
OpenFlow
MAGPI
uen
UTAH EDUCATION NETWORK
WWW.UEN.ORG
GENI AM
OpenFlow
THE UNIVERSITY OF UTAH
VNode
OpenFlow
InstaGENI
CPqD
OpenFlow
RNP
NSI
NL Light
NSI
NOX
OpenFlow
NetFPGA
UNIVERSITY OF AMSTERDAM
NSI
ExoGENI
Princeton University
OpenFlow
InstaGENI
USP
InstaGENI
TOMATO
TECHNISCHE UNIVERSITÄT KAISERSLAUTERN
DFN
Deutsches Forschungsnetz
GEANT2
NSI
ja.net
University of BRISTOL
OpenFlow
University of Essex
OpenFlow
grnet
NSI
Belnet
dedicated connectivity
iMinds
CONNECT.INNOVATE.CREATE
InstaGENI

National Science Foundation Global Environment for Network innovations

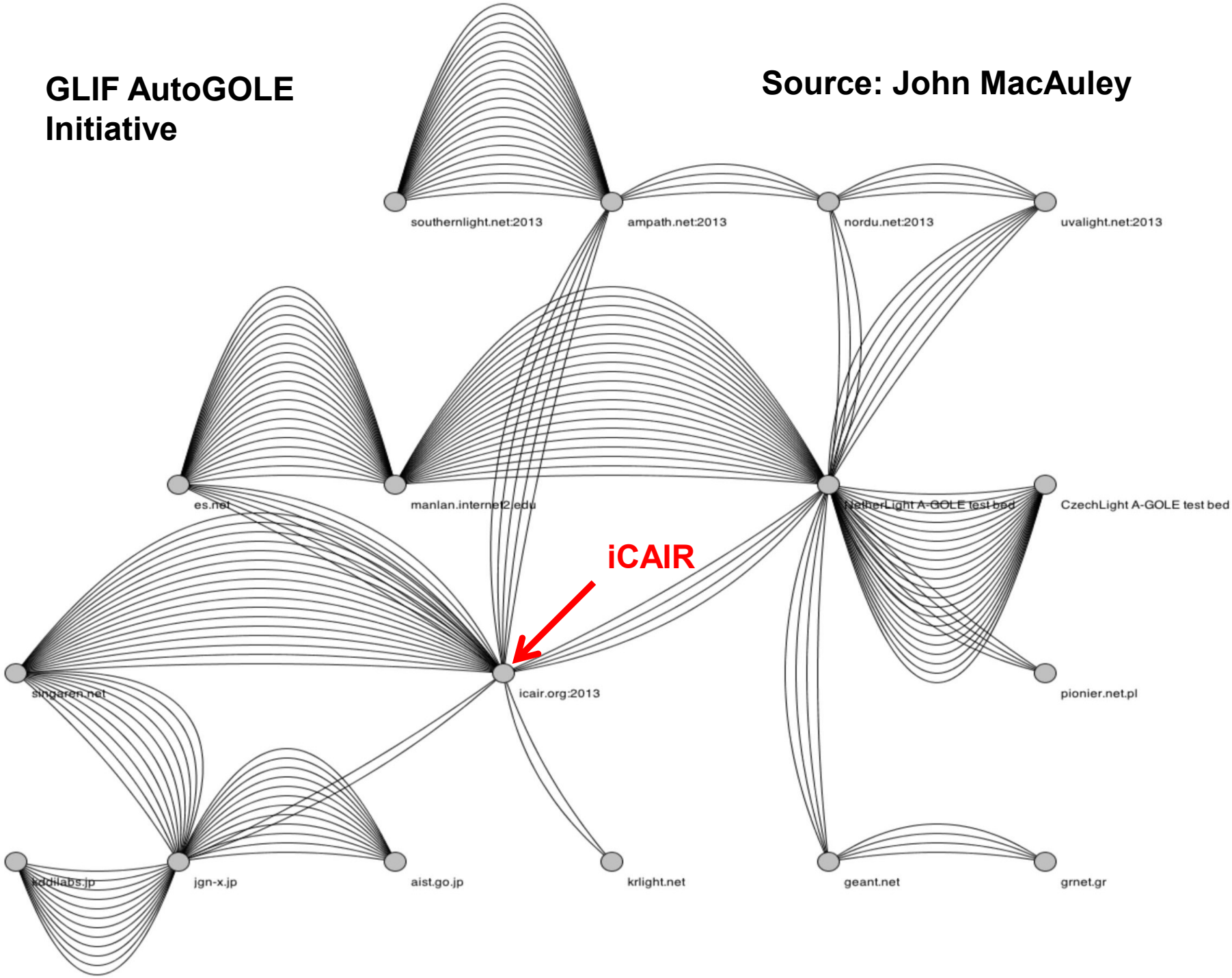


International 40G and 100 G ExoGENI Testbed



GLIF AutoGOLE Initiative

Source: John MacAuley



Software Defined Networking Exchanges (SDXs)

- With the Increasing Deployment of SDN In Production Networks, the Need for an SDN Exchange (SDX) Has Been Recognized.
- Current SDN Architecture Is Single Domain Centralized Controller Oriented
- Required Capabilities for Multi-Domain Distributed SDN Resource Discovery, Signaling Provisioning, Operations, and Fault Detection and Recovery Are Fairly Challenging.
- Nonetheless – Many Motivations Exist for SDXs

Motivations for SDXs

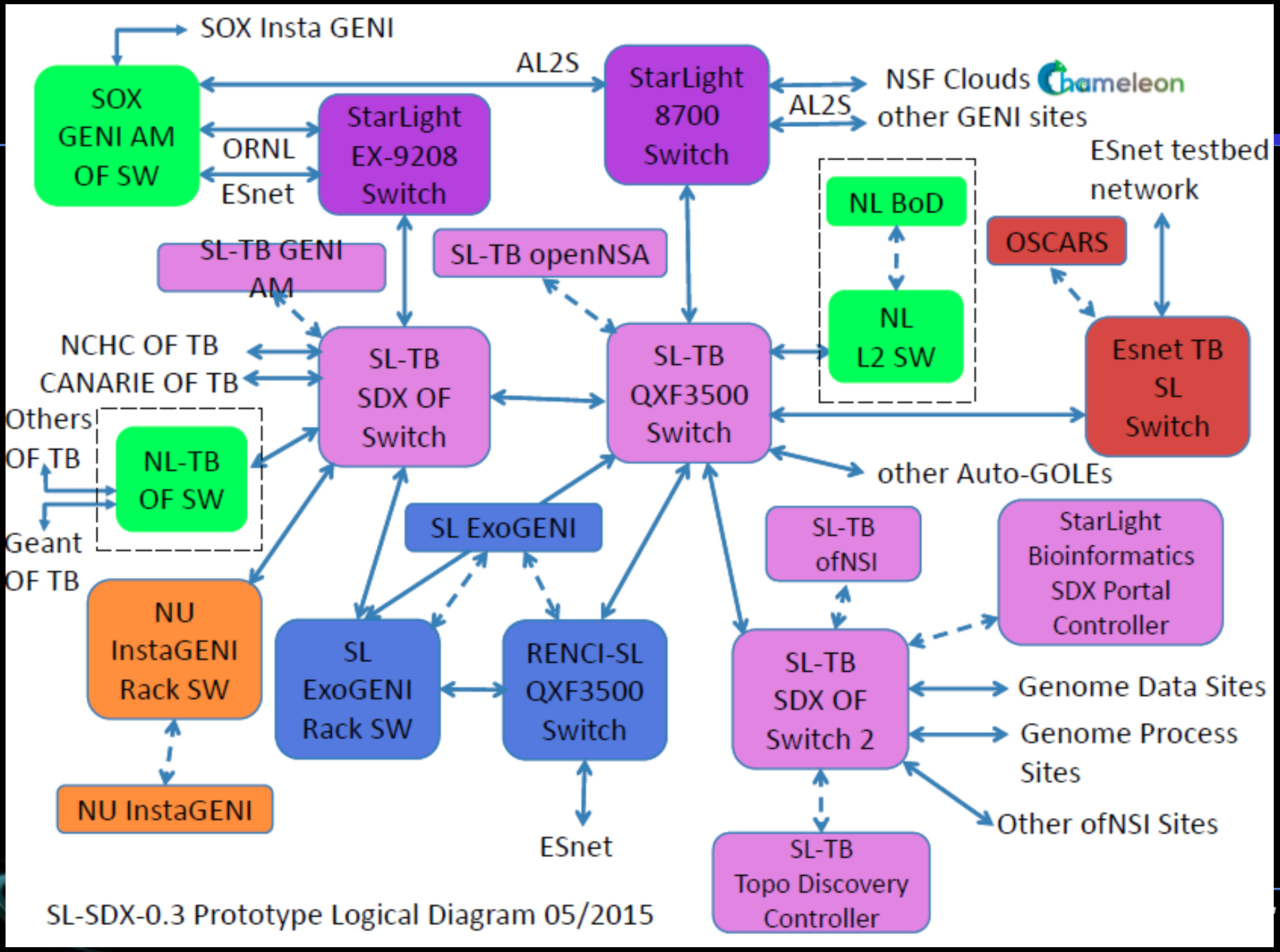
- WH Office of Science and Technology Policy – Large Scale Science Instrumentation
- Large Scale Ultra High Resolution Digital Media Services
- Multi-Domain Networks Interconnecting Data Centers (SDN Is Already in Production Within Large Scale Data Centers)
- Multi-Domain SDN Services
- Providing Capabilities for Edge Control
- Democratization Of Exchange Facilities
- Many New Types of Services and Capabilities
- Etc.

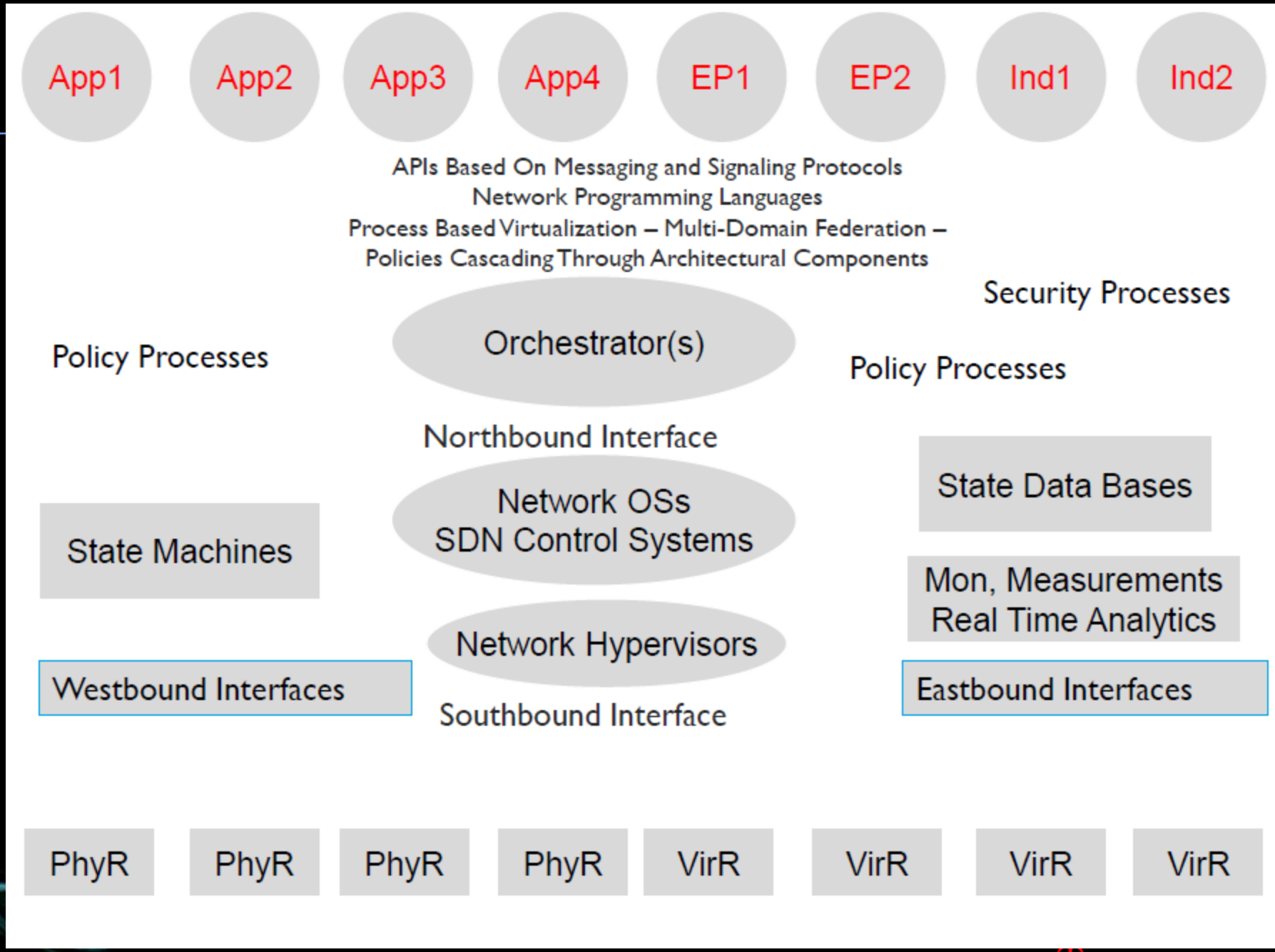
Software Defined Networking Exchanges (SDXs)

- Today, No Production SDX Exists.
- However, With Support From the GENI Project, the International Center for Advanced Internet Research (iCAIR) and Its Research Partners Have Designed and Implemented a Prototype SDX at the StarLight International/National Communications Exchange Facility
- This SDX Is a Multi-Domain Service Enabling Federated Controllers To Exchange Signaling and Provisioning Information.

Selected SDX Architectural Attributes

- Control and Network Resource APIs
- Multi Domain Integrated Path Controller
- Controller Signaling, Including Edge Signaling
- SDN/OF Multi Layer Traffic Exchange
- Multi Domain Resource Advertisement/Discovery
- Topology Exchange
- Multiple Service Levels At All Layers
- Granulated Resource Access (Policy Based), Including Through Edge Processes
- Foundation Resource Programmability
- Various Types of Gateways To Other Network Environments
- Integration of OF and Non-OF Paths, Including 3rd Party Integration
- Programmability for Large Scale Large Capacity Streams


SOX Insta GENI
SOX GENI AM OF SW
AL2S
StarLight 8700 Switch
NSF Clouds
Chameleon
AL2S
other GENI sites
ORNL
ESnet
StarLight EX-9208 Switch
ESnet testbed network
NL BoD
OSCARS
SL-TB GENI AM
SL-TB openNSA
NL L2 SW
NCHC OF TB
CANARIE OF TB
SL-TB SDX OF Switch
SL-TB QXF3500 Switch
Esnet TB SL Switch
Others OF TB
NL-TB OF SW
other Auto-GOLEs
Geant OF TB
SL ExoGENI
SL-TB ofNSI
StarLight Bioinformatics SDX Portal Controller
NU InstaGENI Rack SW
SL ExoGENI Rack SW
RENCI-SL QXF3500 Switch
SL-TB SDX OF Switch 2
Genome Data Sites
Genome Process Sites
Other ofNSI Sites
NU InstaGENI
ESnet
SL-TB Topo Discovery Controller
SL-SDX-0.3 Prototype Logical Diagram 05/2015


App1
App2
App3
App4
EP1
EP2
Ind1
Ind2
APIs Based On Messaging and Signaling Protocols
Network Programming Languages
Process Based Virtualization – Multi-Domain Federation –
Policies Cascading Through Architectural Components
Security Processes
Policy Processes
Orchestrator(s)
Policy Processes
Northbound Interface
State Data Bases
Network OSs
SDN Control Systems
State Machines
Mon, Measurements
Real Time Analytics
Network Hypervisors
Westbound Interfaces
Eastbound Interfaces
Southbound Interface
PhyR
PhyR
PhyR
PhyR
VirR
VirR
VirR
VirR

Architectural Components

- Hybrid Networking Services (Multi-Service, Multi-Layer, Multi-Domain)
- Network Programming Languages (e.g., P4, Frenetic)
- Abstraction Definitions
- APIs
- AP/Service Signaling and Policy Bundling
- Policy Bundle Distribution
- Primitives
- BGP Extensions and Substitutes
- NDL Schema
- Orchestration Processes

STARLIGHT SDX

Other Architectural Components 2

- Northbound Interfaces
- Network OSs
- Network Hypervisors
- State Information Data Bases
- Data Modeling Languages (e.g., YANG)
- Controller Federation Processes
- Hybrid Services/Services Federation/Services Chaining
- Southbound Interfaces
- Eastbound Interfaces
- Westbound Interfaces

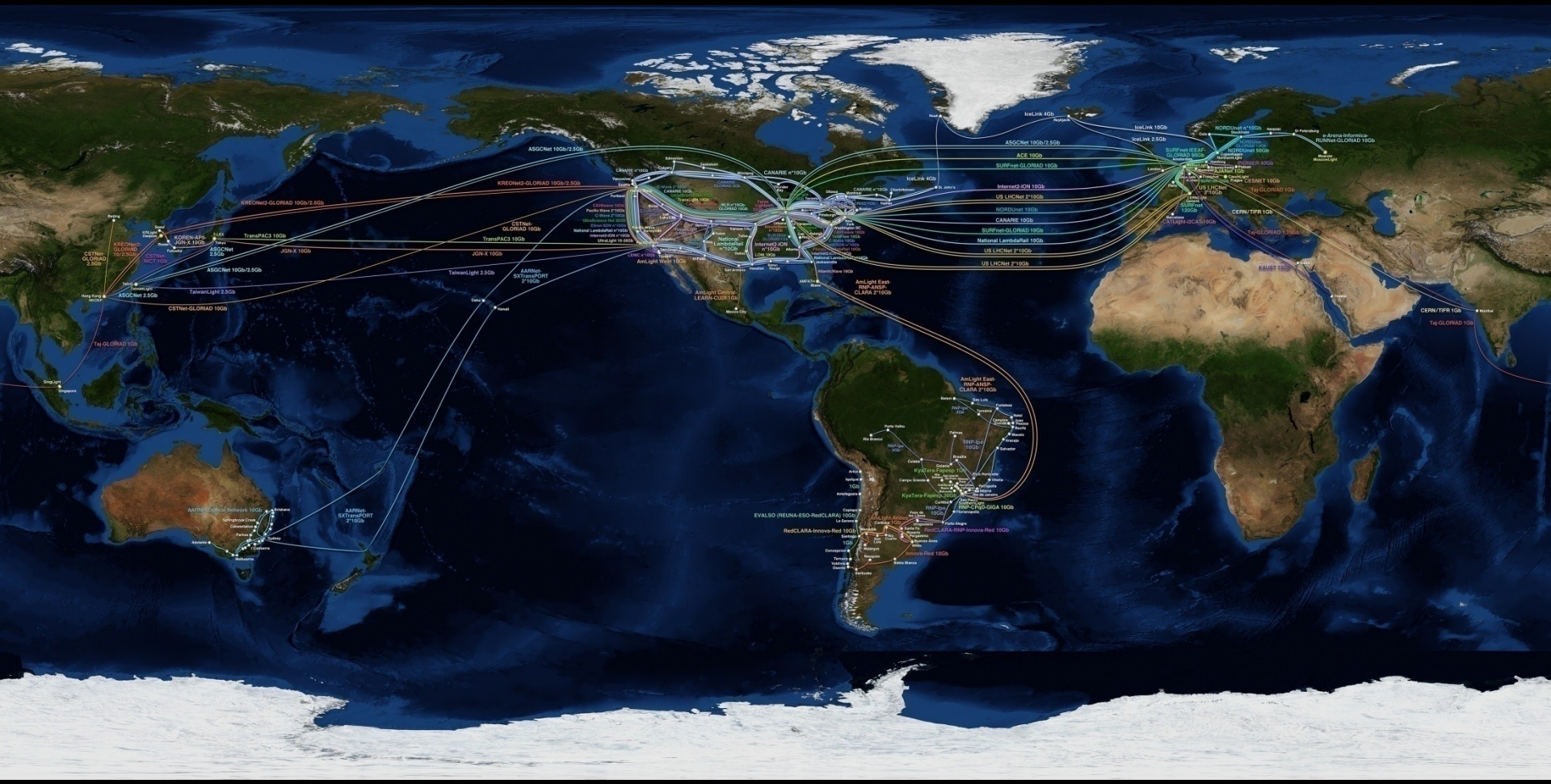
Other Architectural Components 3

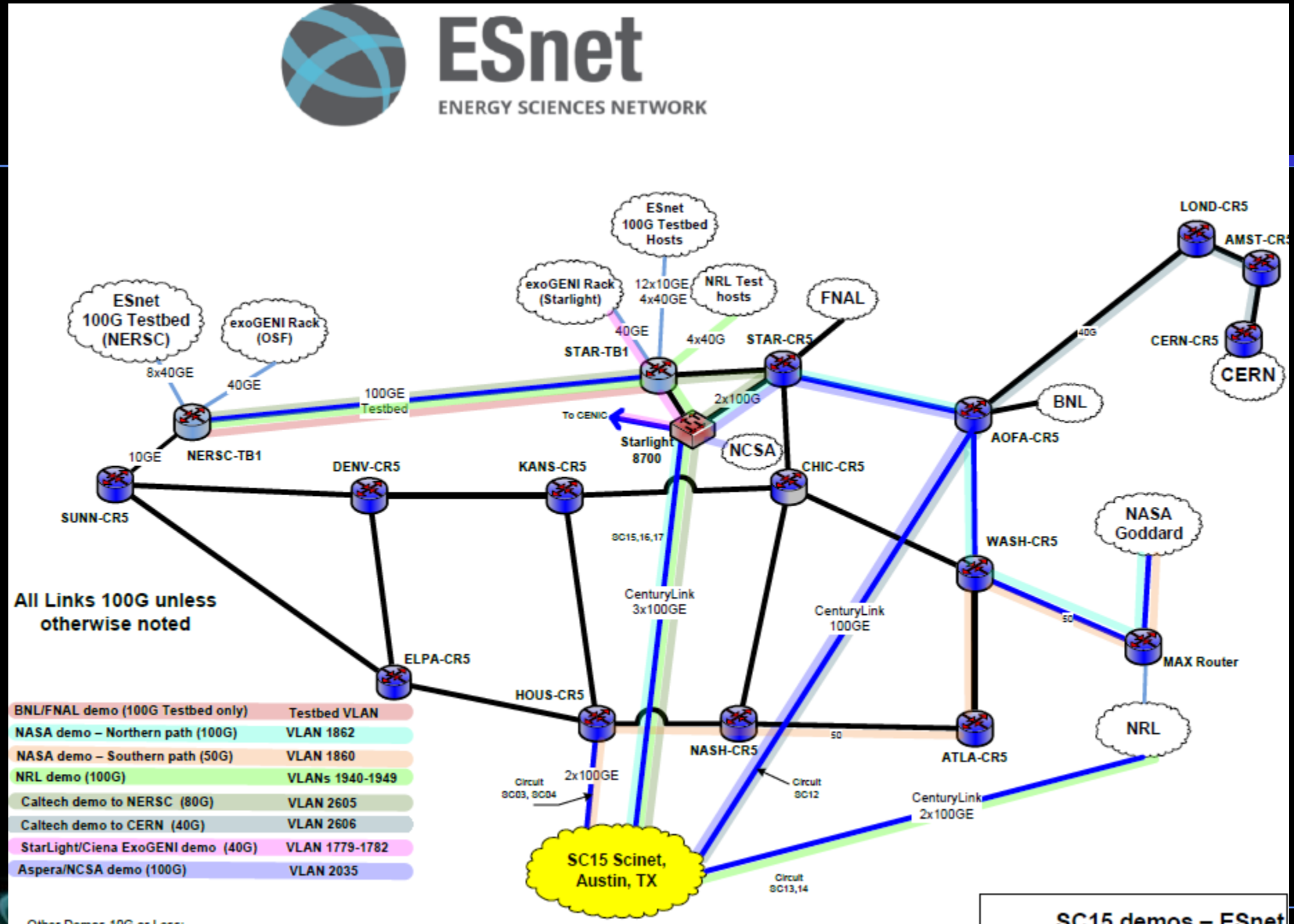
- Data Plane Processes
- Network Function Virtualization (NFV)
- Measurements
- Real Time Analytics
- Distributed Virtual NOC Operations

StarLight International Software Defined Networking Exchange (SDX)

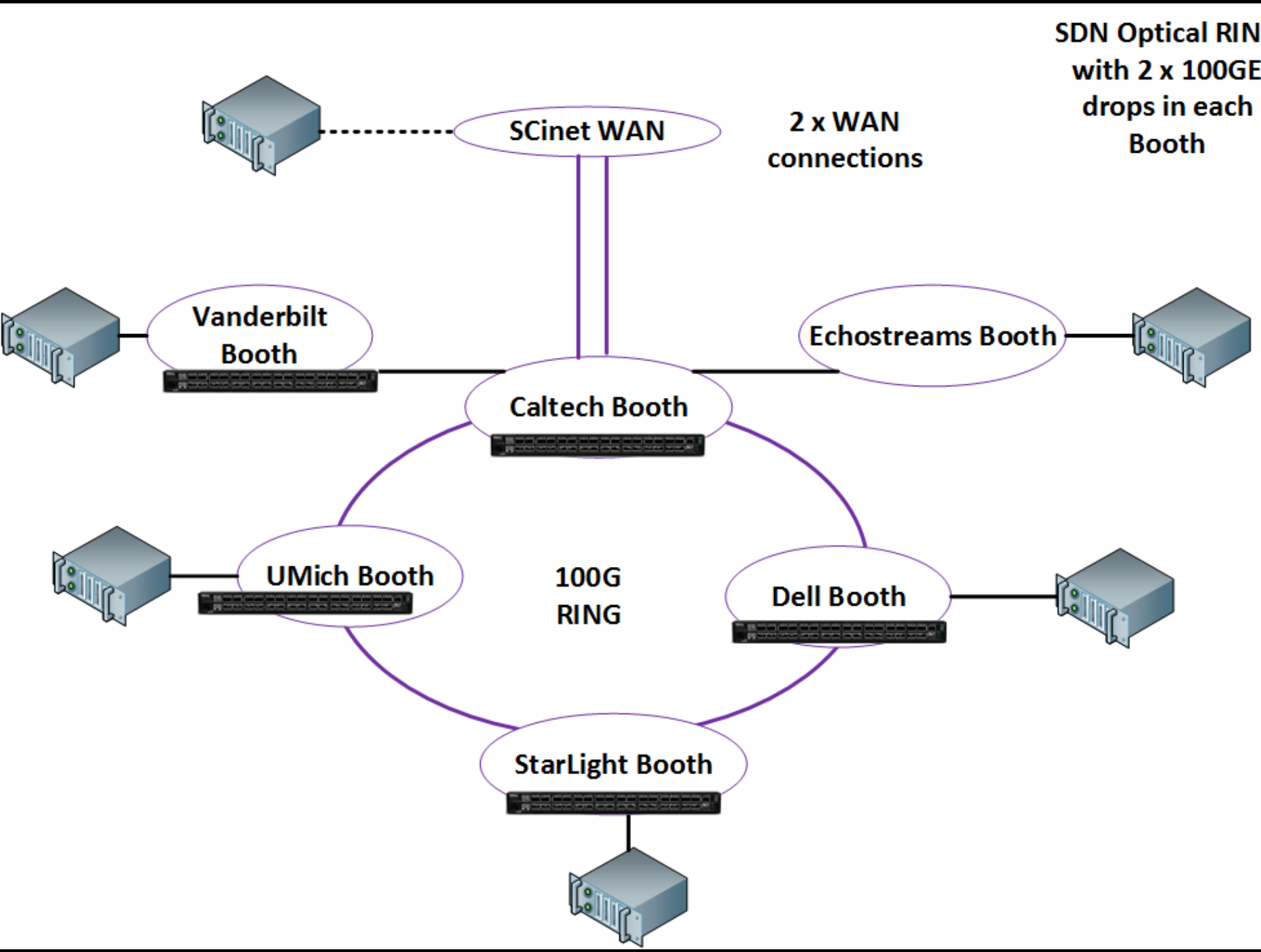
- The National Science Foundation-Funded StarLight International SDX Will Provide The Services, Architecture, and Technologies Designed To Provide Scientists, Engineers, and Educators With Highly Advanced, Diverse, Reliable, Persistent, and Secure Networking Services, Enabling Them to Optimally Access Resources in North America, South America, Asia, South Asia (including India), Australia, New Zealand, Europe, the Middle East, North Africa, And Other Sites Around the World.
- The StarLight SDX Initiative Undertakes Continued Innovation and Development of Advanced Networking Services and Technologies.
- Funded By the NSF International Research Network Connections (IRNC) Program

GLIF is a consortium of institutions, organizations, consortia and country National Research & Education Networks who voluntarily share optical networking resources and expertise to develop the *Global LambdaGrid* for the advancement of scientific collaboration and discovery – a *Federation!.*




ESnet
ENERGY SCIENCES NETWORK
ESnet 100G Testbed Hosts
exoGENI Rack (Starlight)
12x10GE 4x40GE
NRL Test hosts
FNAL
LOND-CR5
AMST-CR5
ESnet 100G Testbed (NERSC)
exoGENI Rack (OSF)
40GE
4x40G
STAR-CR5
STAR-TB1
40G
CERN-CR5
CERN
8x40GE
40GE
100GE Testbed
2x100G
To CENIC
BNL
Starlight 8700
NCSA
AOFA-CR5
10GE
NERSC-TB1
DENV-CR5
KANS-CR5
CHIC-CR5
SUNN-CR5
SC15,16,17
NASA Goddard
WASH-CR5
All Links 100G unless otherwise noted
CenturyLink 3x100GE
CenturyLink 100GE
50
MAX Router
ELPA-CR5
HOUS-CR5
NASH-CR5
ATLA-CR5
NRL
50
Circuit SC03, SC04
2x100GE
Circuit SC12
CenturyLink 2x100GE
SC15 Scinet, Austin, TX
Circuit SC13,14
BNL/FNAL demo (100G Testbed only) Testbed VLAN
NASA demo – Northern path (100G) VLAN 1862
NASA demo – Southern path (50G) VLAN 1860
NRL demo (100G) VLANs 1940-1949
Caltech demo to NERSC (80G) VLAN 2605
Caltech demo to CERN (40G) VLAN 2606
StarLight/Ciena ExoGENI demo (40G) VLAN 1779-1782
Aspera/NCSA demo (100G) VLAN 2035
Other Demos 10G or Less:
ESnet/RENCI demo: NERSC to ANL
ESnet ENOS Demo: WASH, AMST, CERN
ANL QoS Demo: DENV, ATLA
SC15 demos – ESnet
Brian Tierney, ESnet 11/6/2015
FILENAME SC15-DEMOS-V9.VSD

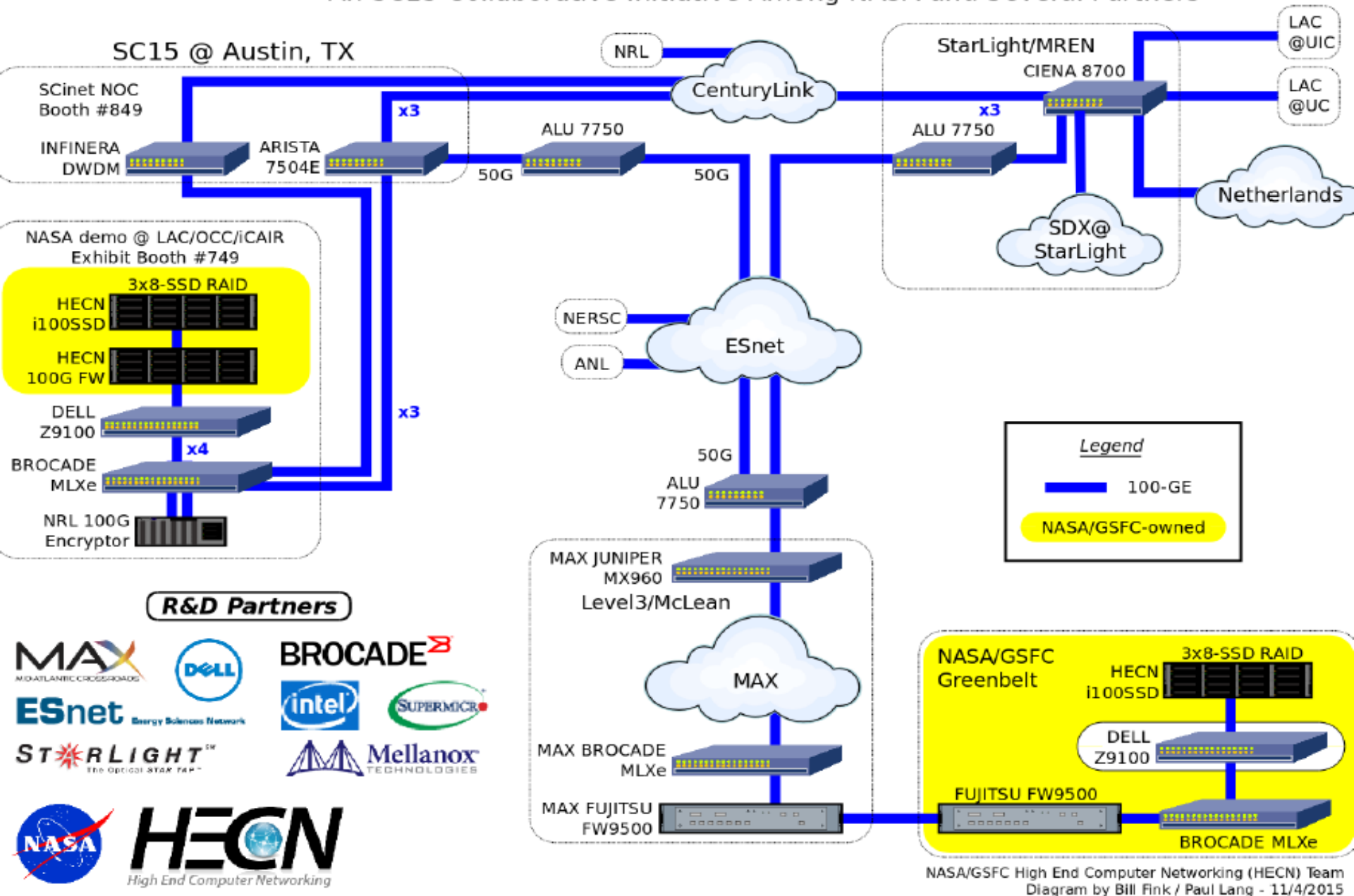

NetherLight
Caltech
Fermilab
DE-KIT
GEANT
DFN
StarLight
Vanderbilt
Austin
AmLight
RNP
UNESP
WAN providers
ESNet
CENIC
CenturyLink
NetherLight
GEANT
DFN
RNP
ANESP


SDN Optical RING with 2 x 100GE drops in each Booth
SCinet WAN
2 x WAN connections
Vanderbilt Booth
Echostreams Booth
Caltech Booth
UMich Booth
100G RING
Dell Booth
StarLight Booth

SC15

Demonstrations of 100 Gbps Disk-to-Disk WAN File Transfer Performance via SDX and 100G FW

An SC15 Collaborative Initiative Among NASA and Several Partners



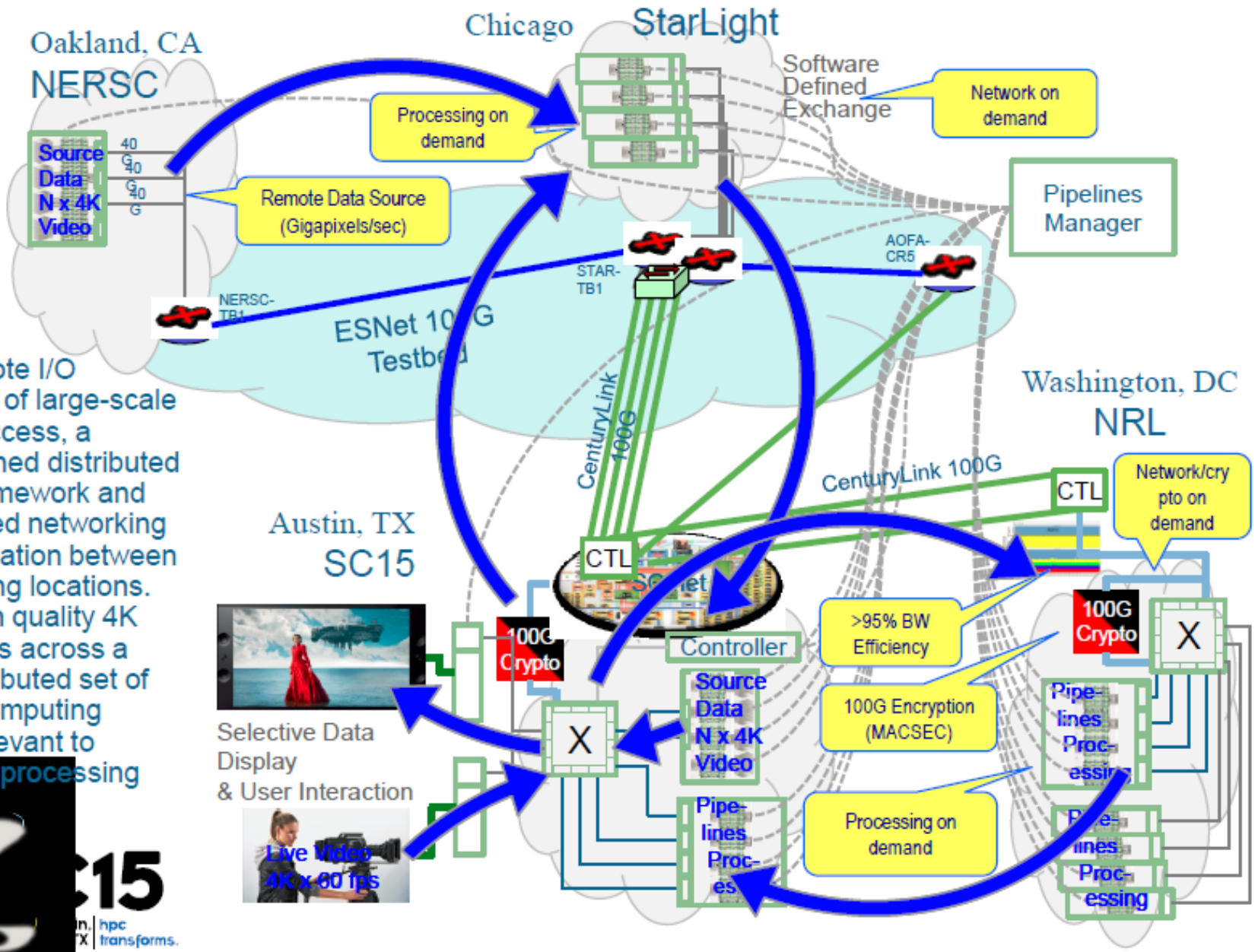
R&D Partners

MAX MID-ATLANTIC CROSSROADS · DELL · BROCADE · ESnet Energy Sciences Network · intel · SUPERMICRO · STARLIGHT The Optical STAR TAP · Mellanox TECHNOLOGIES

NASA · HECN High End Computer Networking

NASA/GSFC High End Computer Networking (HECN) Team
Diagram by Bill Fink / Paul Lang - 11/4/2015

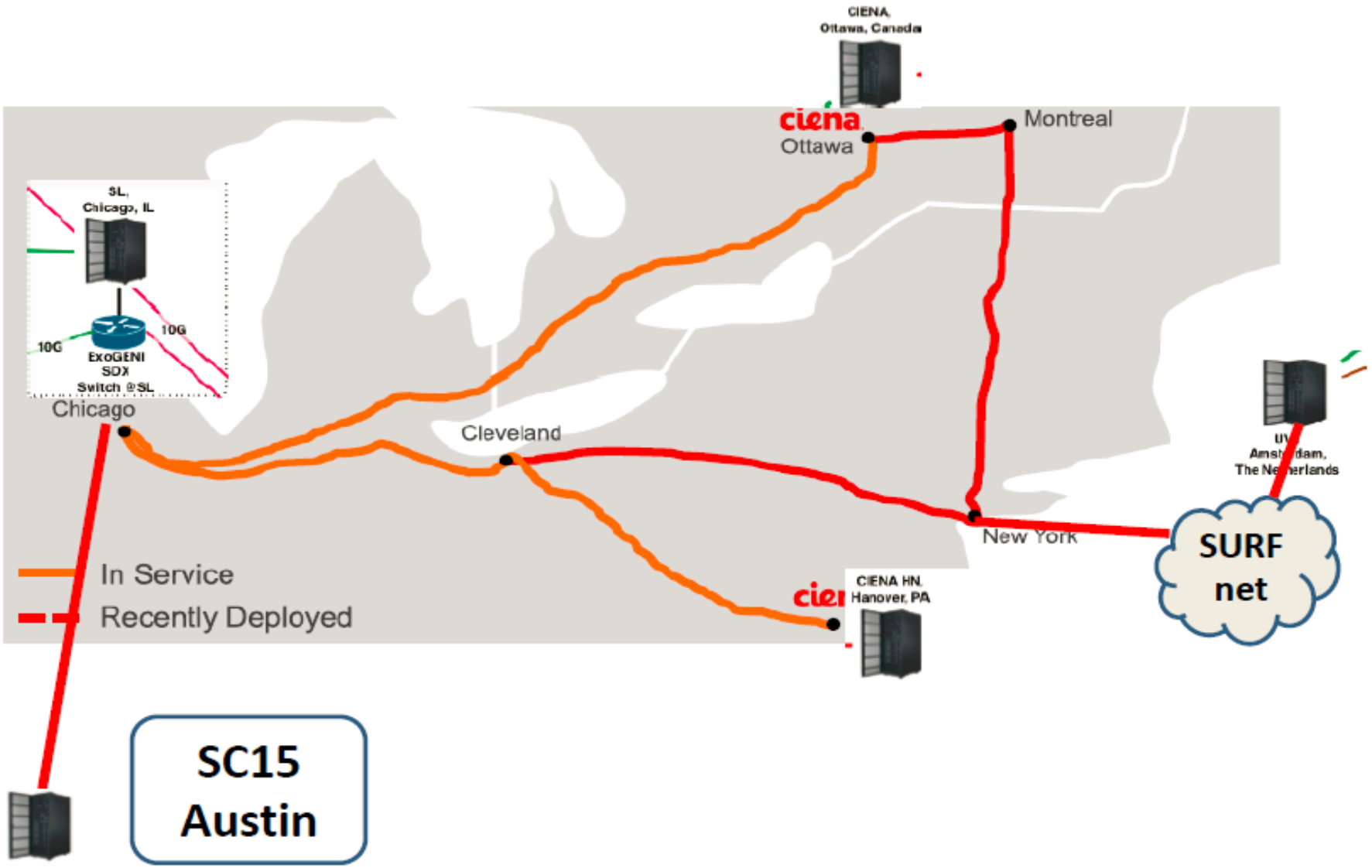
Dynamic Remote I/O Network



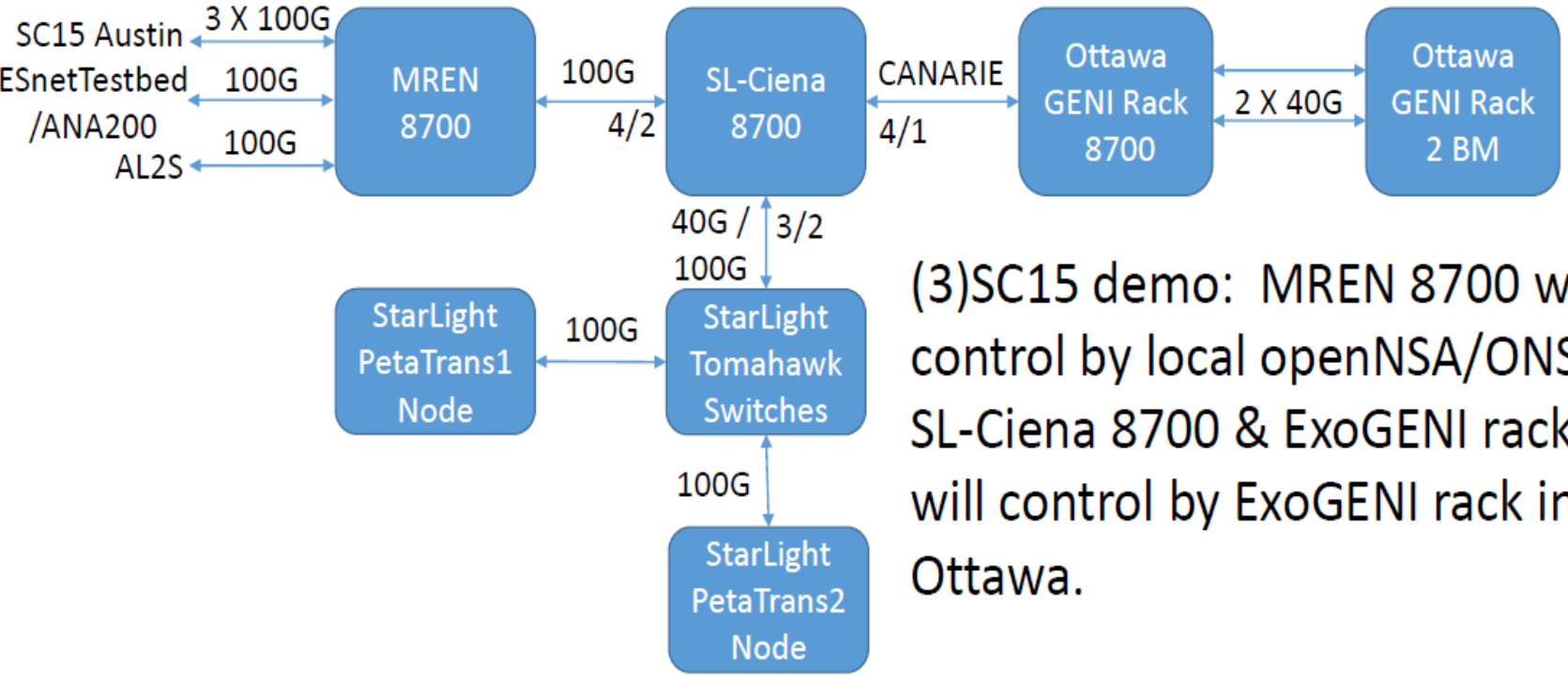
Dynamic Remote I/O Demonstration of large-scale remote data access, a dynamic pipelined distributed processing framework and software defined networking enabled automation between distant operating locations. Live production quality 4K video workflows across a nationally distributed set of storage and computing resources - relevant to emerging data processing challenges.

SC15

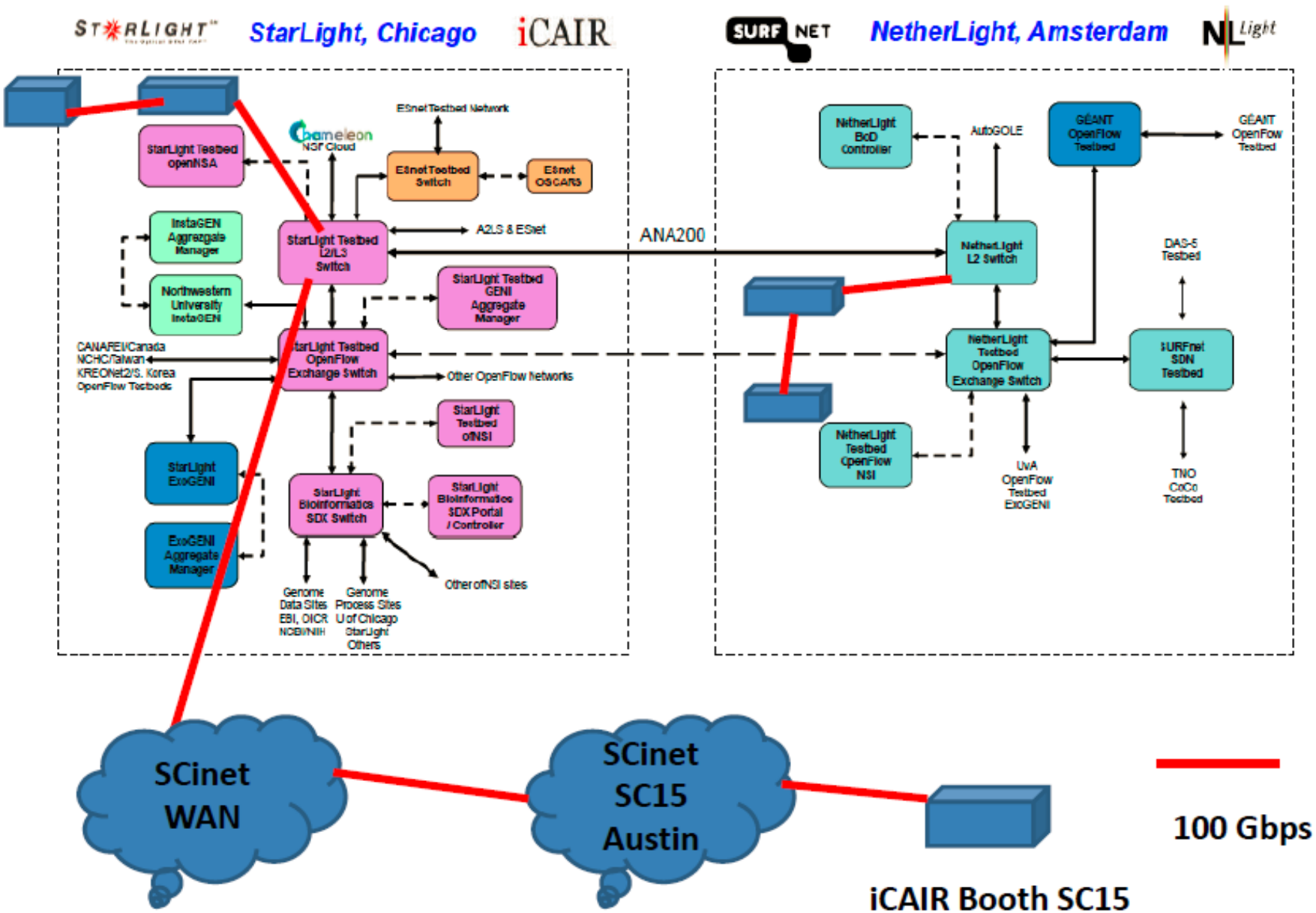
hpc transforms.


CIENA,
Ottawa, Canada
ciena
Ottawa
Montreal
SL,
Chicago, IL
10G
10G
ExoGENI
SOX
Switch @SL
Chicago
Cleveland
New York
SURF
net
In Service
Recently Deployed
CIENA HN,
Hanover, PA
SC15
Austin

Multi-Tenant 100GE Science Network Exchange SC15 NRE Testing Phase(Sep/20 – Oct/20 2015)



(3)SC15 demo: MREN 8700 will control by local openNSA/ONSA, SL-Ciena 8700 & ExoGENI rack of will control by ExoGENI rack in Ottawa.


StarLight, Chicago
iCAIR
NetherLight, Amsterdam
StarLight Testbed openNSA
Chameleon NSF Cloud
ESnet Testbed Network
ESnet Testbed Switch
ESnet OSCARS
InstaGEN Aggregate Manager
StarLight Testbed L2/L3 Switch
A2LS & ESnet
ANA200
Northwestern University InstaGENI
StarLight Testbed GENI Aggregate Manager
CANARIE/Canada
NCHC/Taiwan
KREONet2/S. Korea
OpenFlow Testbeds
StarLight Testbed OpenFlow Exchange Switch
Other OpenFlow Networks
StarLight Testbed ofNSI
StarLight ExoGENI
StarLight Bioinformatics SDX Switch
StarLight Bioinformatics SDX Portal / Controller
ExoGENI Aggregate Manager
Other ofNSI sites
Genome Data Sites EBI, OICR NCBI/NIH
Genome Process Sites U of Chicago StarLight Others
NetherLight BoD Controller
AutoGOLE
GÉANT OpenFlow Testbed
GÉANT OpenFlow Testbed
NetherLight L2 Switch
DAS-5 Testbed
NetherLight Testbed OpenFlow Exchange Switch
SURFnet SDN Testbed
NetherLight Testbed OpenFlow NSI
UvA OpenFlow Testbed ExoGENI
TNO CoCo Testbed
SCinet WAN
SCinet SC15 Austin
100 Gbps
iCAIR Booth SC15

Multiple HPC Cloud Computing Testbeds Specifically Designed for Science Research

=> Open Commons Consortium

At Scale Experimentation
Integerated With High Performance Networks

iCAIR

STARLIGHT℠

The OCC And Next Generation Networking

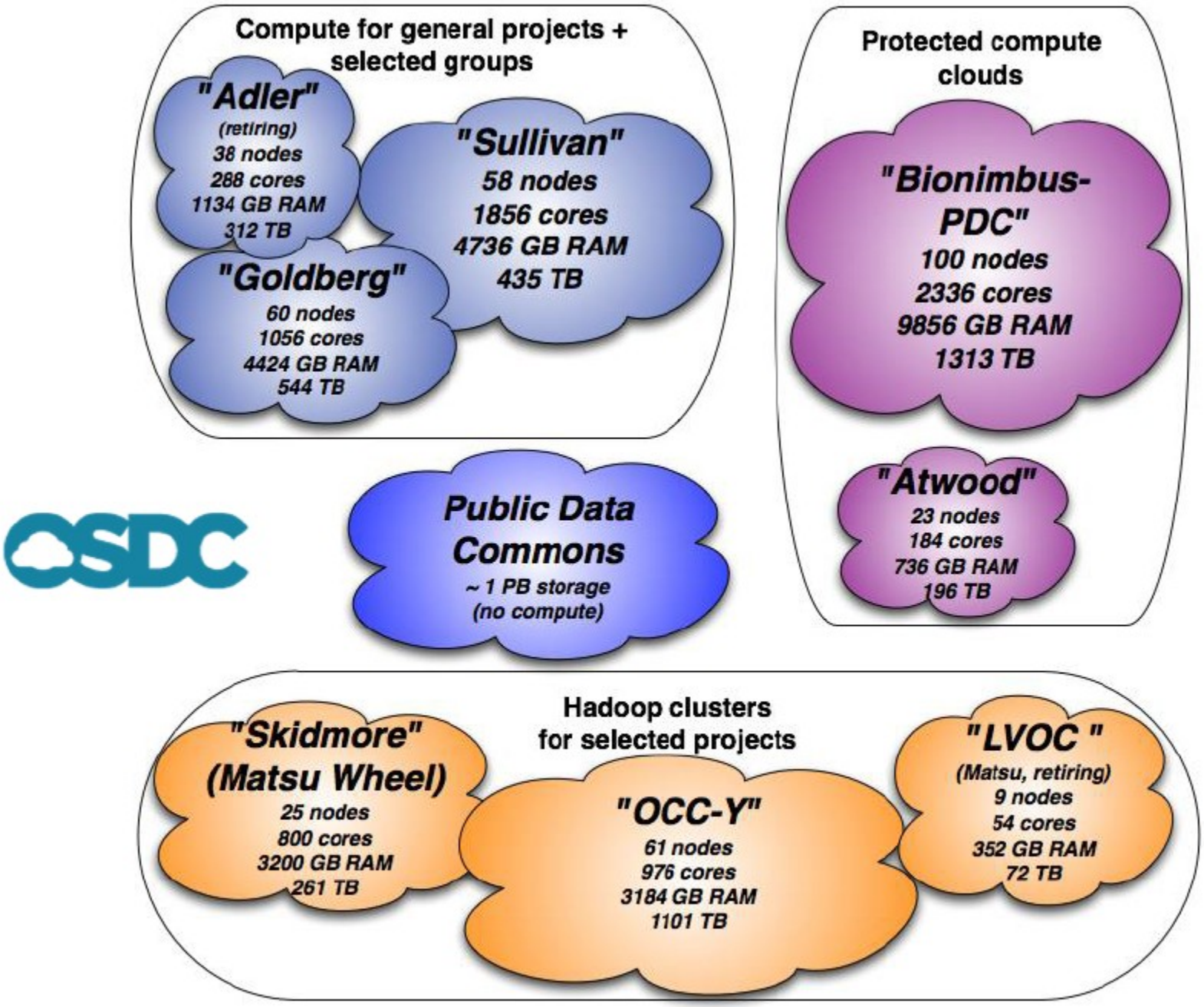
- The OCC Has Formed a Partnership With The Advanced Networking Community To Created Large Scale (Global) Distributed Environments For Data Intensive Scientific Research – Based On a Foundation of Programmable, Deterministic, Edge Customizable High Performance Large Capacity Networks (e.g., 100 G and 100 G +)

Creating A Data Commons

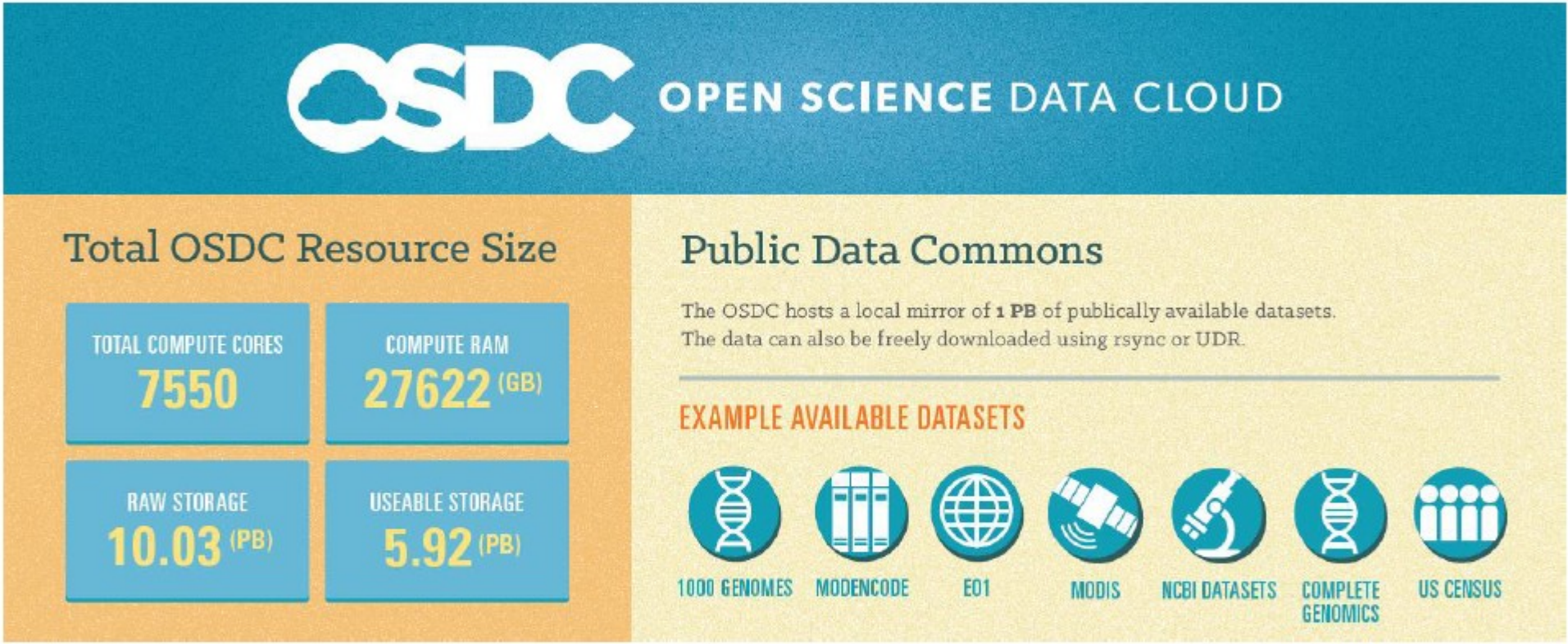
- Clouds and Data Commons For Global Data Intensive Science
- Vast Amounts Of Data Are Being Created By Many New Powerful Sophisticated Instruments – Zettabytes!
- Few Individual Scientists Have the Resources To Manage, Store, Analyze, and Transport This Data
- A Data Commons Is A Facility Designed To Co-locate Data, Storage, Computing, and Networking Infrastructure, Along With Commonly Used Tools For Analyzing and Sharing Data As A Research Resource.

Core Data Commons Services

- Policy Based Access Procedures, Digital IDs
- Metadata Services
- Storage Services
- High Performance Transport, Including Specialized Transport Services For Large Scale Streams
- Policy Based Data Export Tools/Services
- Compute Services with Images/Containers Integrating Commonly Used Tools, Applications and Services, Specialized for Individual Science Domain Research Communities


Compute for general projects + selected groups
"Adler"
(retiring)
38 nodes
288 cores
1134 GB RAM
312 TB
"Sullivan"
58 nodes
1856 cores
4736 GB RAM
435 TB
"Goldberg"
60 nodes
1056 cores
4424 GB RAM
544 TB
Protected compute clouds
"Bionimbus-PDC"
100 nodes
2336 cores
9856 GB RAM
1313 TB
"Atwood"
23 nodes
184 cores
736 GB RAM
196 TB
OSDC
Public Data Commons
~ 1 PB storage
(no compute)
Hadoop clusters for selected projects
"Skidmore" (Matsu Wheel)
25 nodes
800 cores
3200 GB RAM
261 TB
"OCC-Y"
61 nodes
976 cores
3184 GB RAM
1101 TB
"LVOC "
(Matsu, retiring)
9 nodes
54 cores
352 GB RAM
72 TB

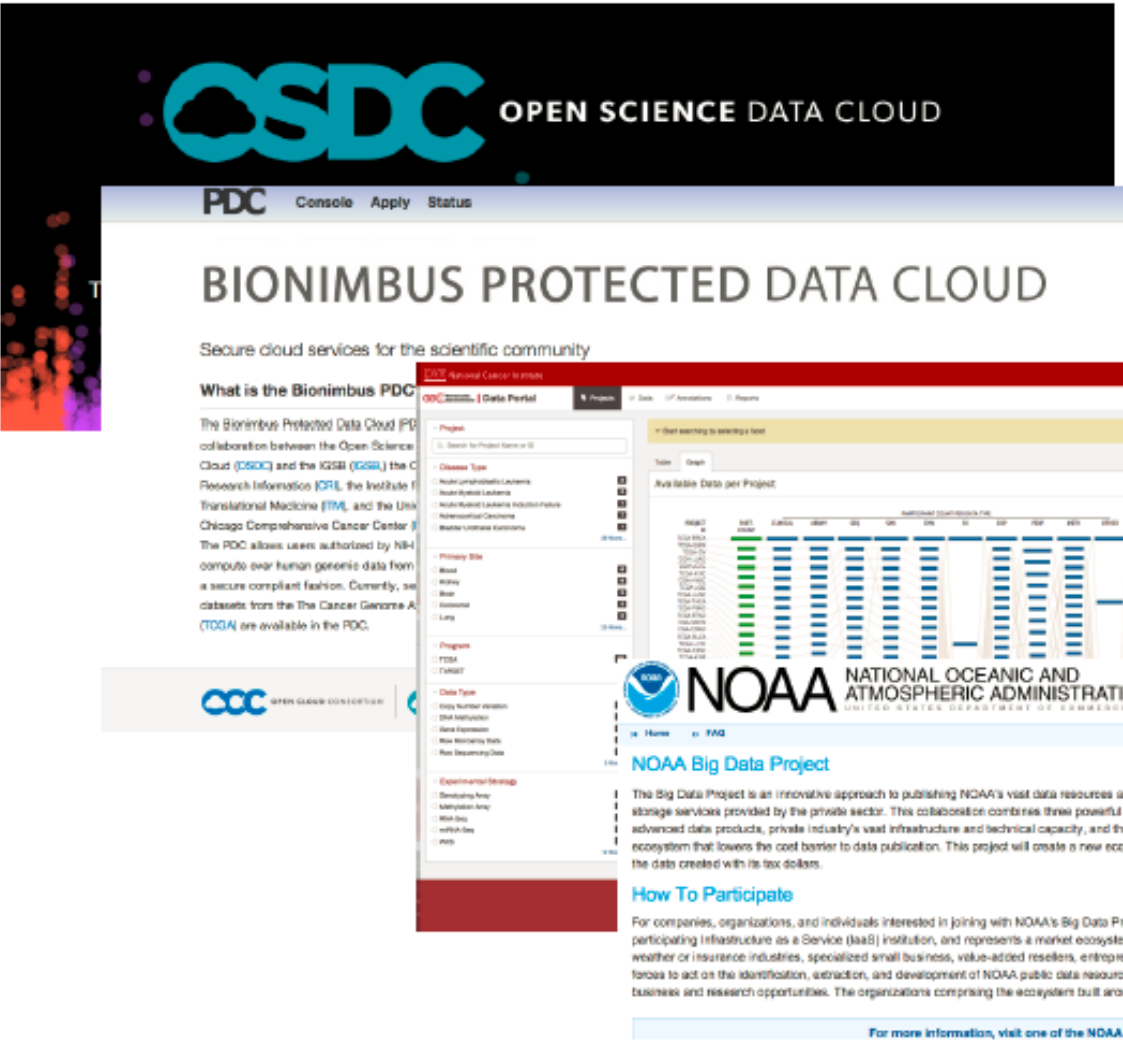
The Open Science Data Cloud (OSDC) is an **open-source**, **cloud-based** infrastructure that allows scientists to manage, share, and analyze medium to large size scientific datasets.



Application for resources available to anyone doing scientific research:

www.opensciencedatacloud.org

Maria Patterson (mtpatter@uchicago.edu) *Center for Data Intensive Science, University of Chicago*



Open Science Data Cloud (2010)

Bionimbus Protected Data Cloud (2013)

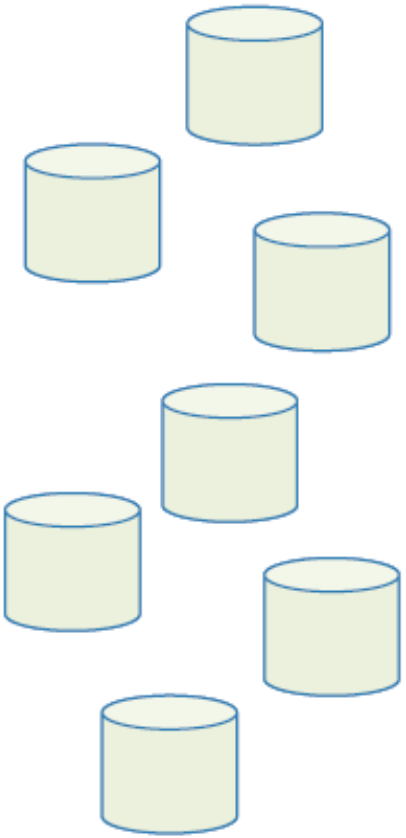
NCI Data Commons (Nov 2015)

OCC-NOAA Data Commons (Oct 2015)

- Over 650+ total unique user accounts
- Users from 100+ different institutions
- Each month, on average: about 180 unique users and 1.8 million core hours

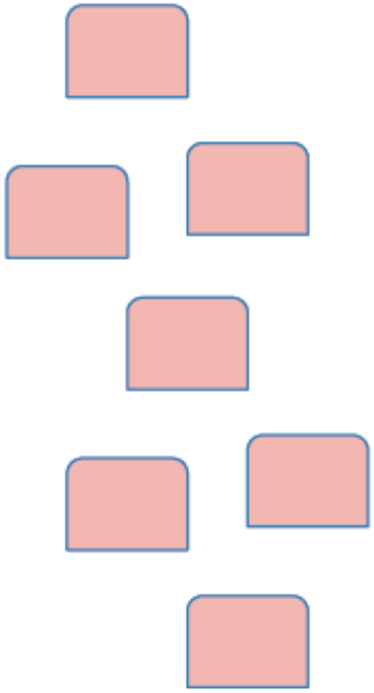
OSDC Commons Architecture

Digital ID Service
& Metadata Service



Object storage
(permanent
data archives)

Scalable light
weight workflow



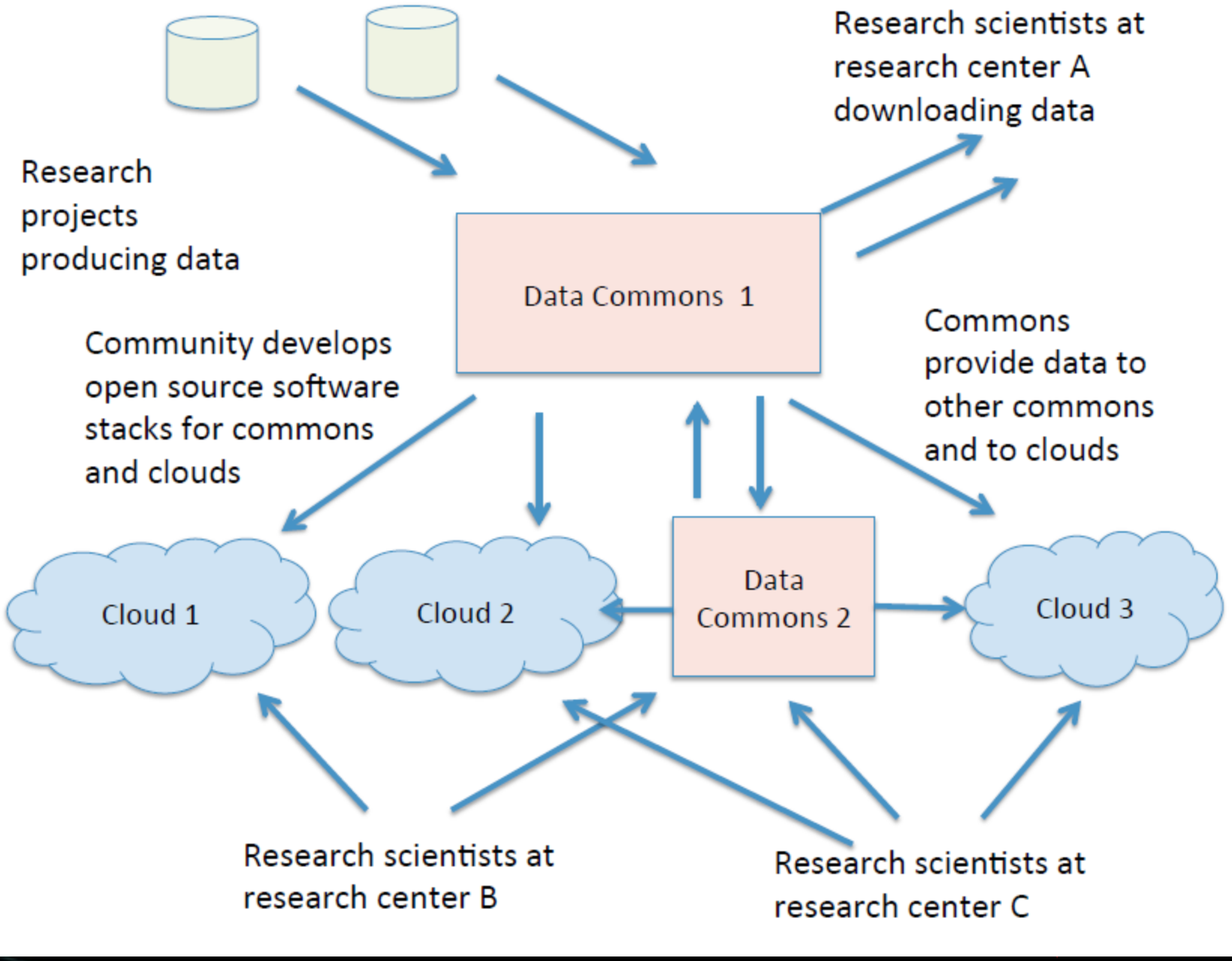
Community data products
(data harmonization)

Co-located
“pay for
compute”

Open APIs
for data
access and
data access
portal

Data
submission
portal

Devops supporting virtual machines and containers


Research scientists at research center A downloading data
Research projects producing data
Data Commons 1
Community develops open source software stacks for commons and clouds
Commons provide data to other commons and to clouds
Cloud 1
Cloud 2
Data Commons 2
Cloud 3
Research scientists at research center B
Research scientists at research center C

- The OSDC provides the scientific community with resources for storing, sharing and analyzing TB- & PB-scale datasets.
- In 2014, the OSDC supported over 338 distinct users from 101 distinct organizations. These organizations were from 14 different countries.
- The OSDC specializes in supporting data intensive projects. The table on the right depicts the number of times users exceeded a certain number of core hours during a single month in 2014.

# of Core Hours	# of Users
10,000	444
20,000	142
50,000	46
100,000	23
200,000	7

Available Resources

- OSDC, PDC, and GDC resources currently offer over 21,000 cores, 22.1PB of raw storage, and 10.01 PB of useable storage
- This is spread over ~25 racks of equipment on the OSDC/PDC and ~17 racks for the GDC.

Type	Resource	Cores	Storage (PB)	
			Raw	Useable
OSDC	Sullivan, Goldberg, Skidmore, OCC-Y, Root	5008	6.3	3.36
PDC	Bionimbus PDC, Atwood PDC	3104	5.2	1.4
GDC	NCI Genomic Data Commons	13760	10.6	5.25
TOTAL		**21872**	**22.1**	**10.01**

Overview

GEVADIS (http://www.geuvadis.org/)
Genetic European Variation in Health and Disease
A European Medical Sequencing Consortium

RNAseq data produced from a subset of the same individuals used
To produce the 1000 genomes (http://www.1000genomes.org)

Samples in the GEUVADIS dataset represent 5 European Populations that were sequenced by 7 “performers”.

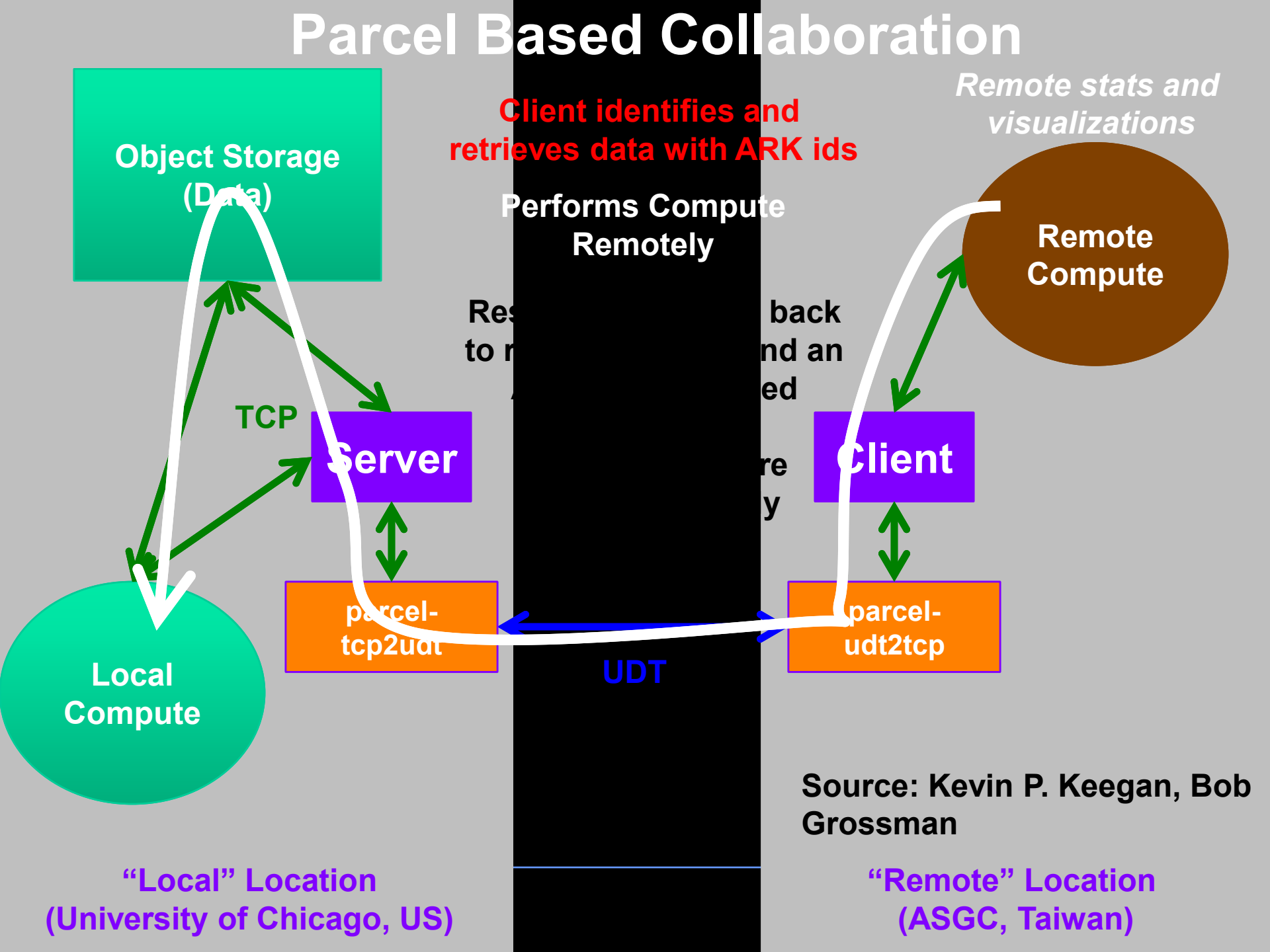
We examined a subset of ~50 samples to explore differences among the populations as well as potential bias in sequencing.

- Stuti’s docker does this
- Study FPKM abundances with visual and statistical tools

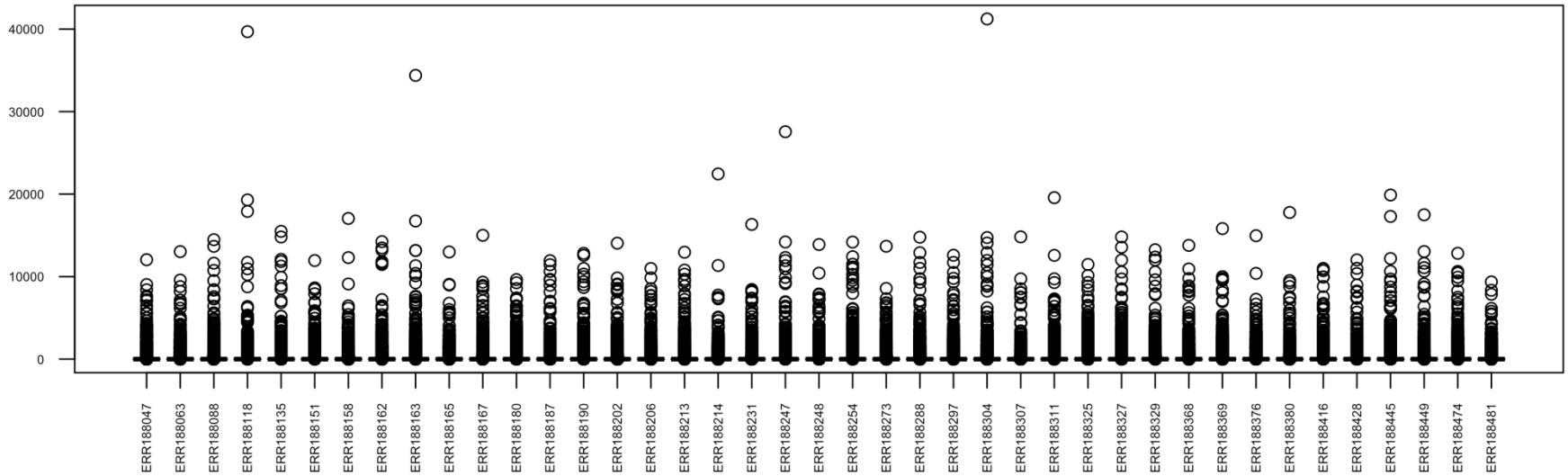
Original data were housed at UoC, analysis occurred at both locations. Data and results were shared via Parcel access to the UoC object store.

Source: Kevin P. Keegan, Bob Grossman

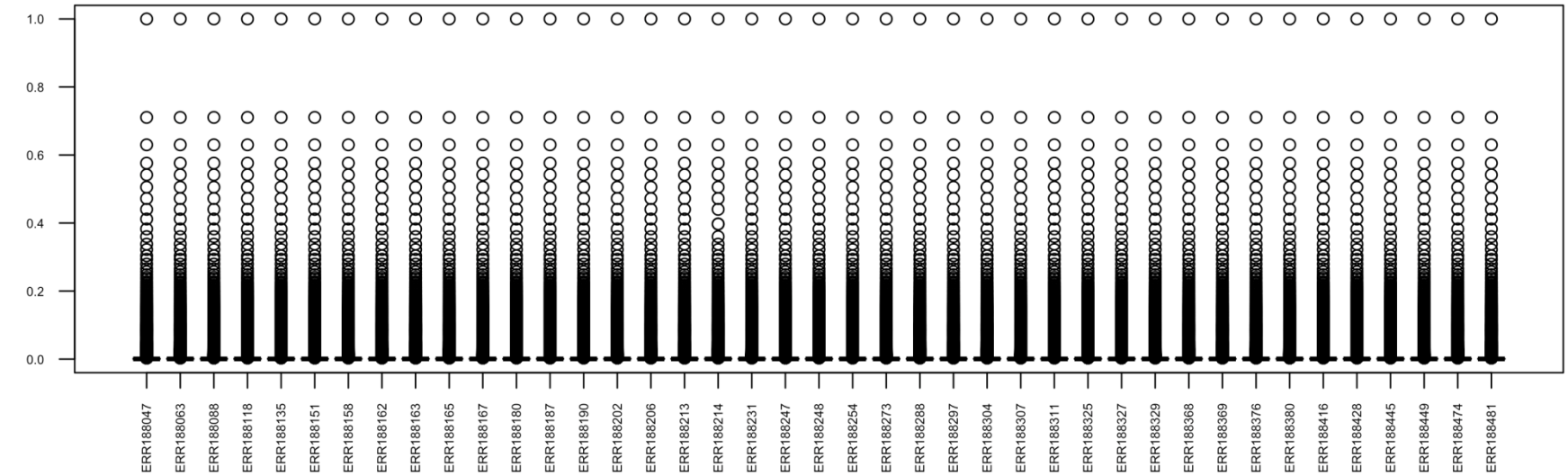
Parcel Based Collaboration



combined_FPKM.txt RAW

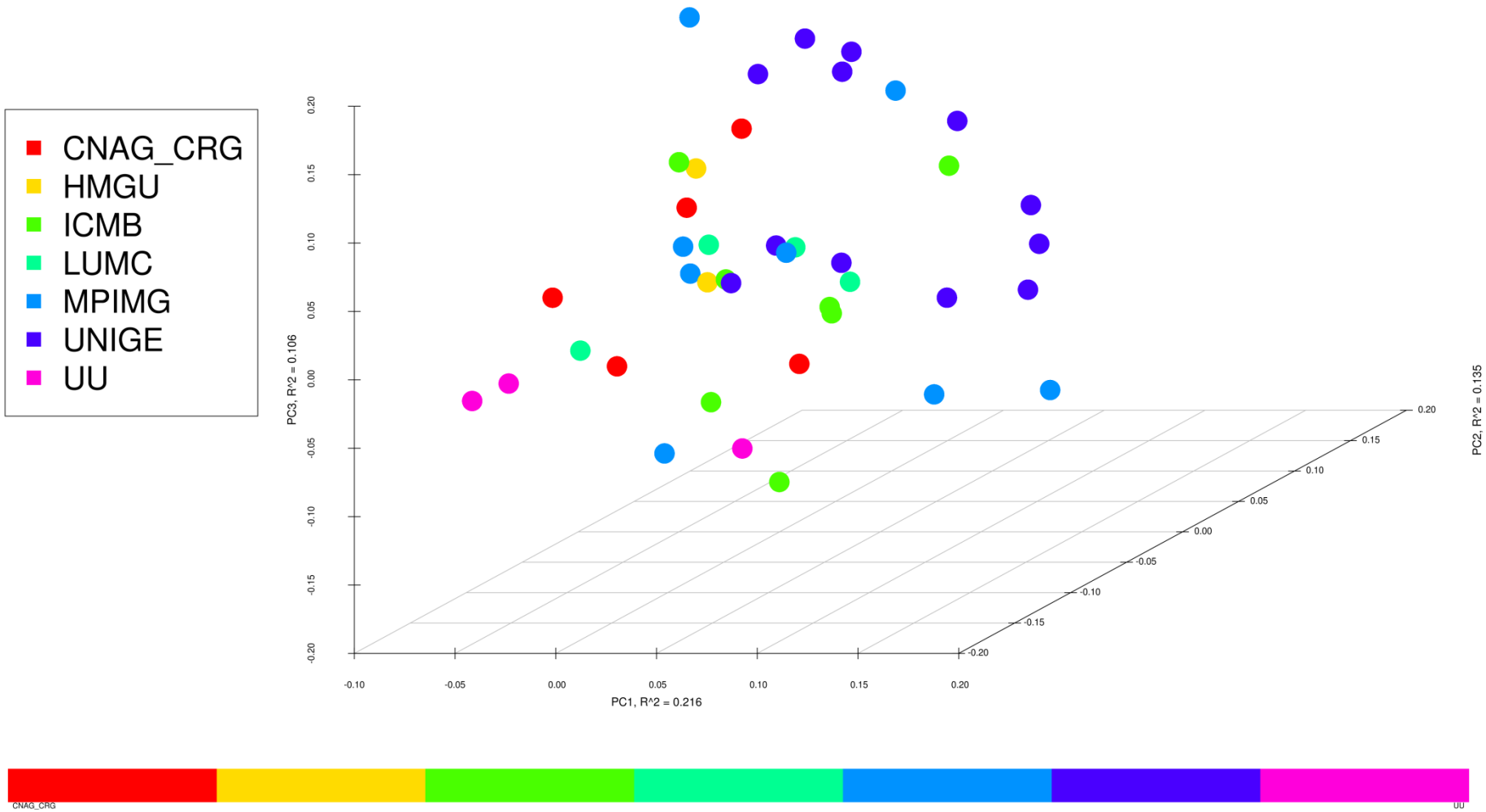


combined_FPKM.txt PREPROCESSED (quantile norm)



combined_FPKM.txt.quantile.PREPROCESSED.txt.euclidean.PCoA.Performer.PCoA

Source: Kevin P. Keegan, Bob Grossman



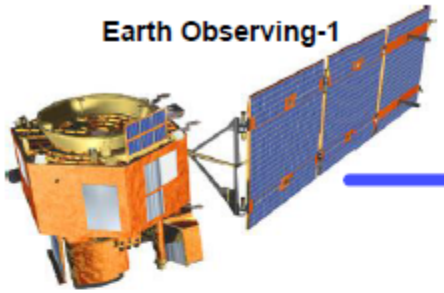
Project Matsu

- A Data Commons Open Source Project With A Goal Of Supporting Earth Science Research Communities By Processing Satellite Imaging Generated By NASA's Earth Orbiting Satellites
- This Is a Joint Project Established By the OCC And NASA's EO-1 Mission (Dan Mandl Is the Lead)

EO-1 L1G Image

STARLIGHT SM

Matsu Analytic Wheel



New data observed by EO-1 and downloaded to NASA

NASA Goddard Space Flight Center

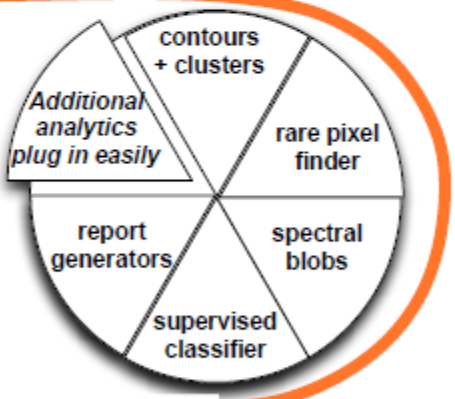
NASA images sent to OSDC Public Data Commons cloud for permanent storage

OSDC Public Data Commons (GlusterFS)

Data read into HDFS only once

HDFS

Wheel analytics run over data using MapReduce

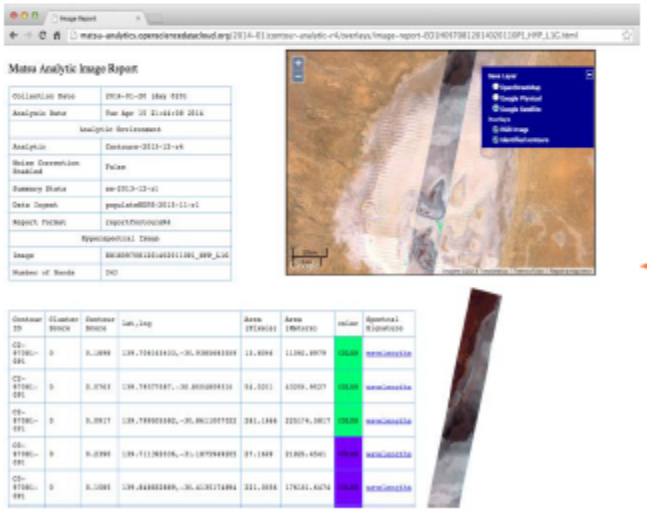


Metadata stored

NoSql Database (Accumulo)

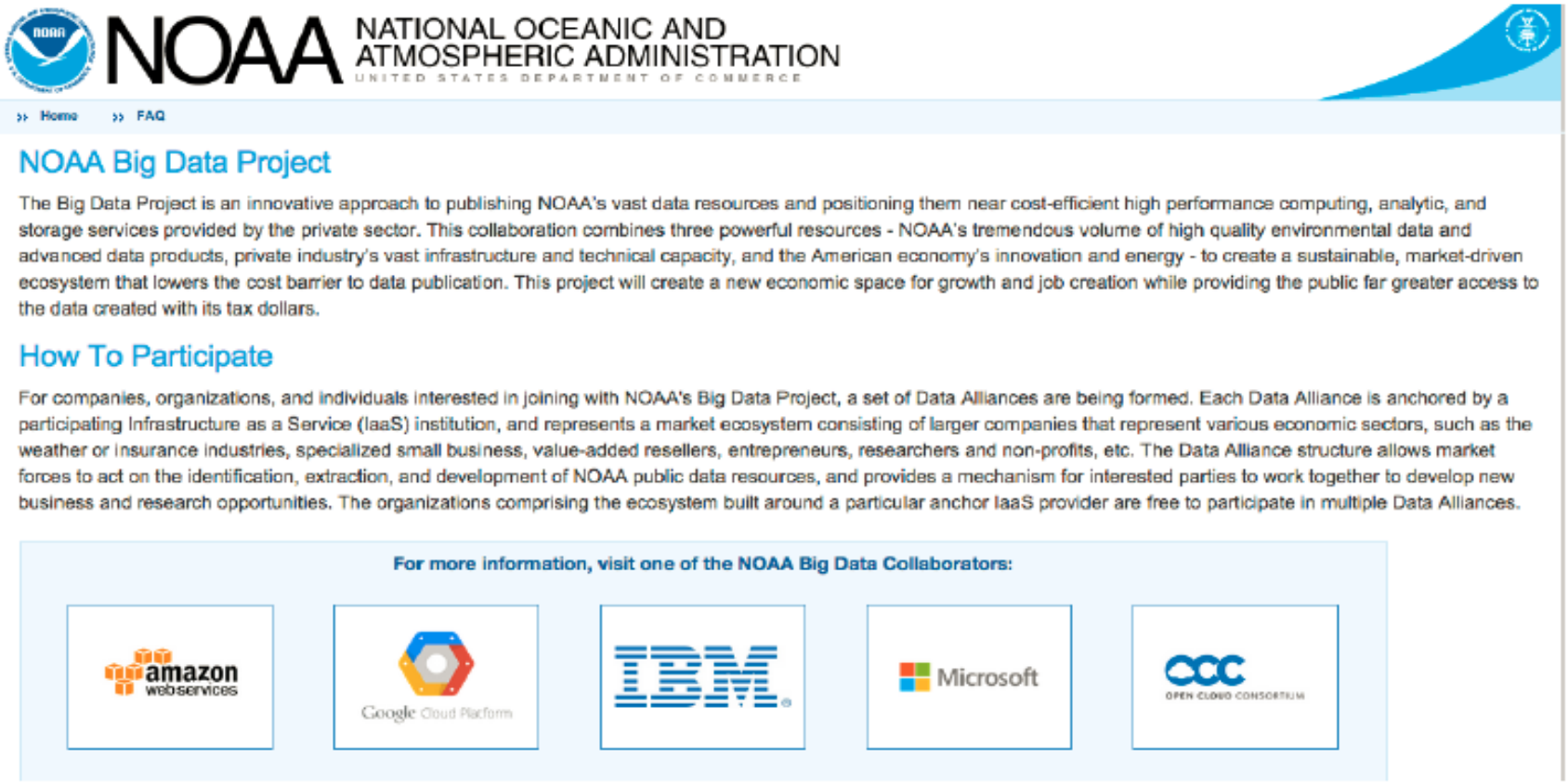
Analytic results stored

Secondary analysis can be done from analytic database



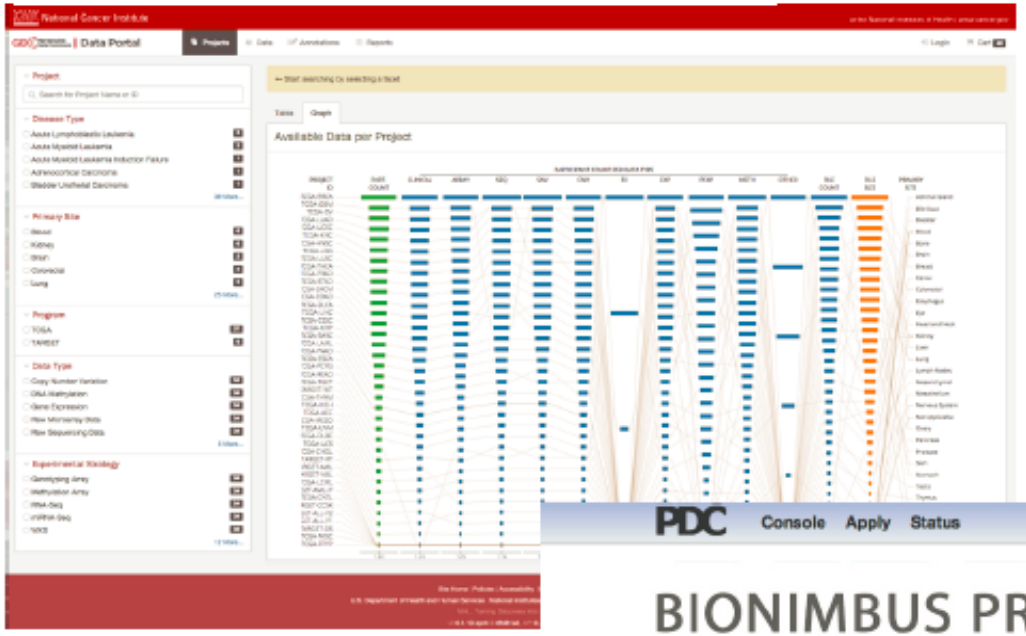
Analytic reports generated by Wheel are accessible via web browser

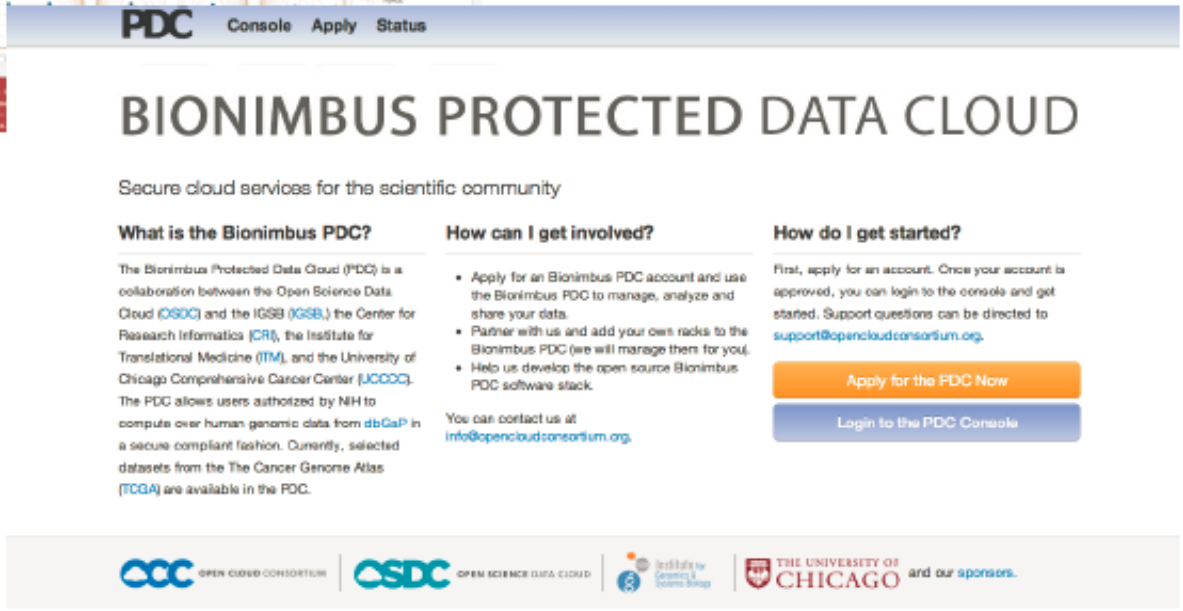
Maria Patterson (mtpatter@uchicago.edu)



- Public-private data collaborative announced April 21, 2015 by US Secretary of Commerce Pritzker.
- AWS, Google, IBM, Microsoft and Open Cloud Consortium will form five collaborations.
- We will develop an OCC/NOAA Data Commons.

NCI Genomic Data Commons





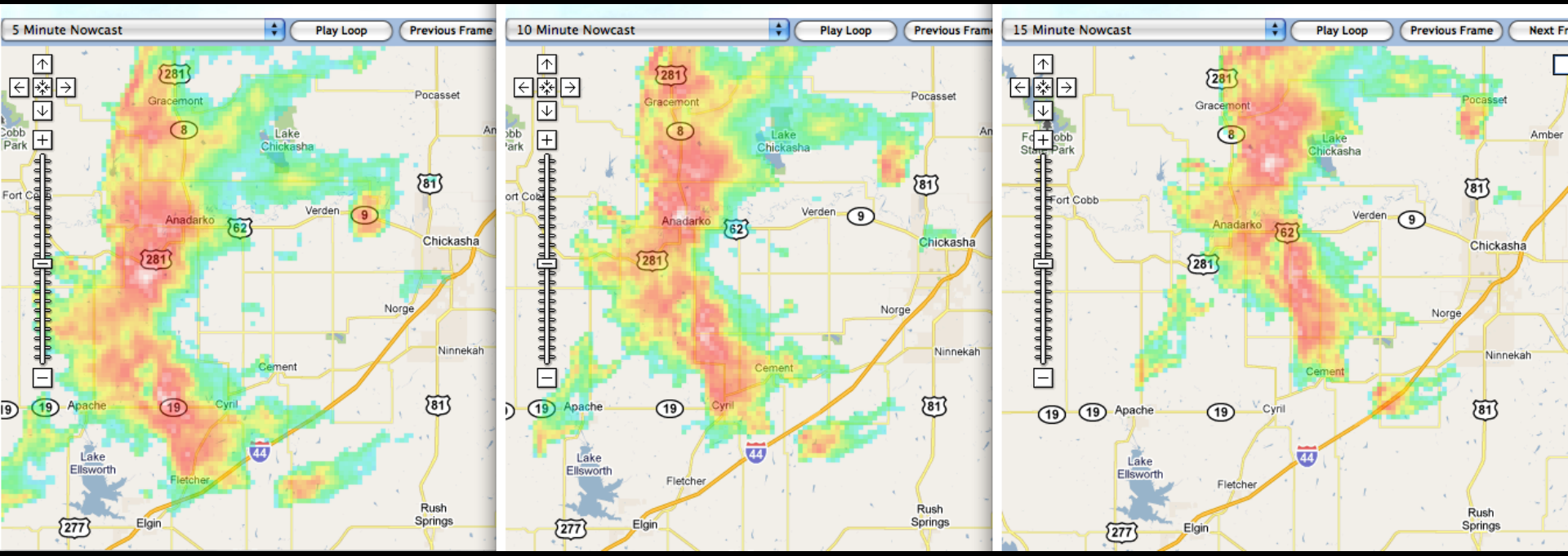
University of Chicago biomedical data commons developed in collaboration with the OCC.

STARLIGHT℠

GENI Engineering Conference (GEC 19) SDX Demonstration Atlanta March 18-20, 2014

- Initial SDX Capability Between GENI Sites (StarLight and SOX) Was Demonstrated
- Motivation: To Share a Vision of Interconnected US Nationwide SDN Infrastructure, With Multiple SDN Capable Networks and Domains
- SDX Benefits Was Showcased Through a Compelling Application – Nowcast – Developed By Mike Zink and His Colleagues at University of Massachusetts, Amherst
- (GEC 19 Was Co-Located With the GLIF Tech Workshop, March 19-20)

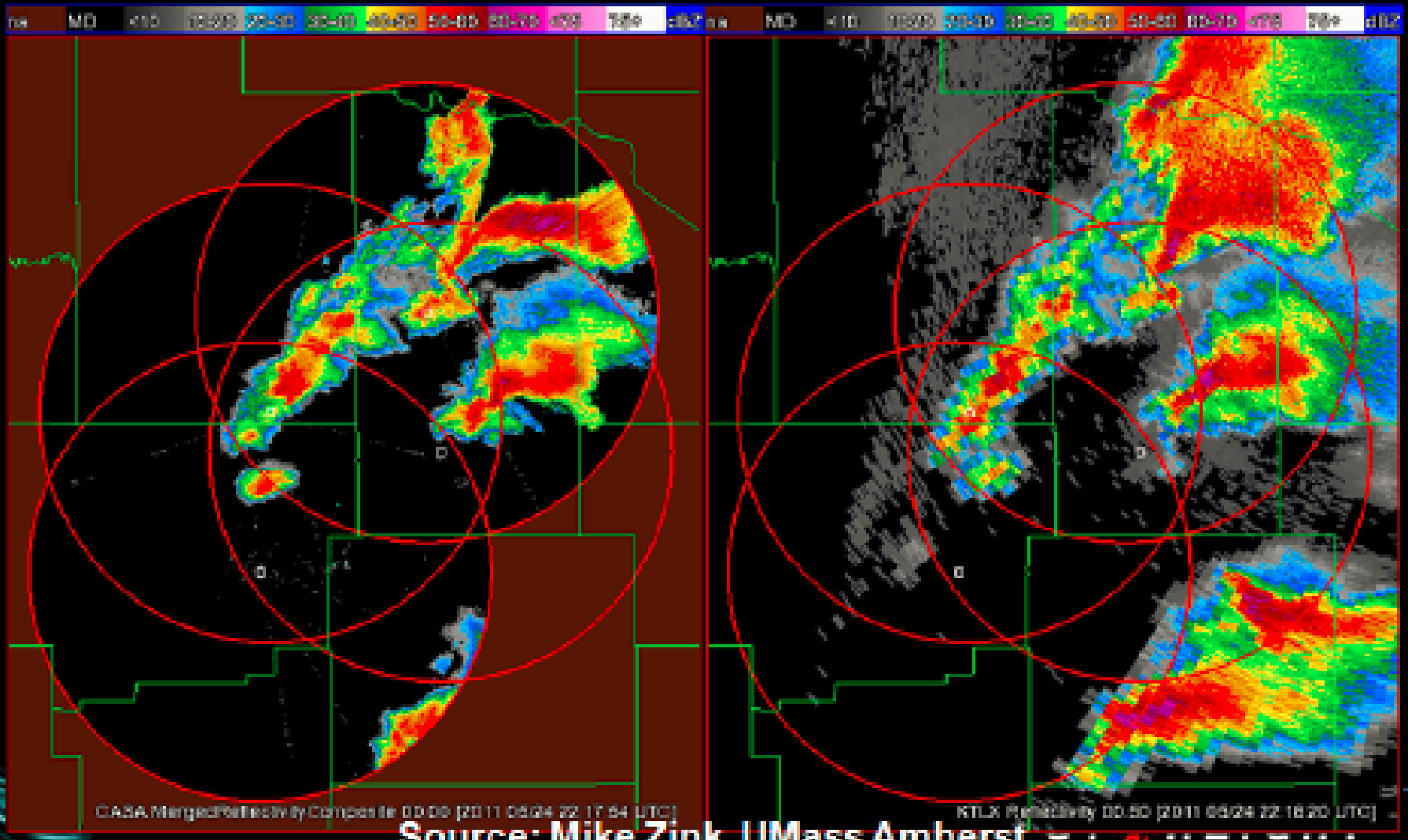
Nowcast Example



Source: Mike Zink, UMass Amherst

STORMLIGHT℠

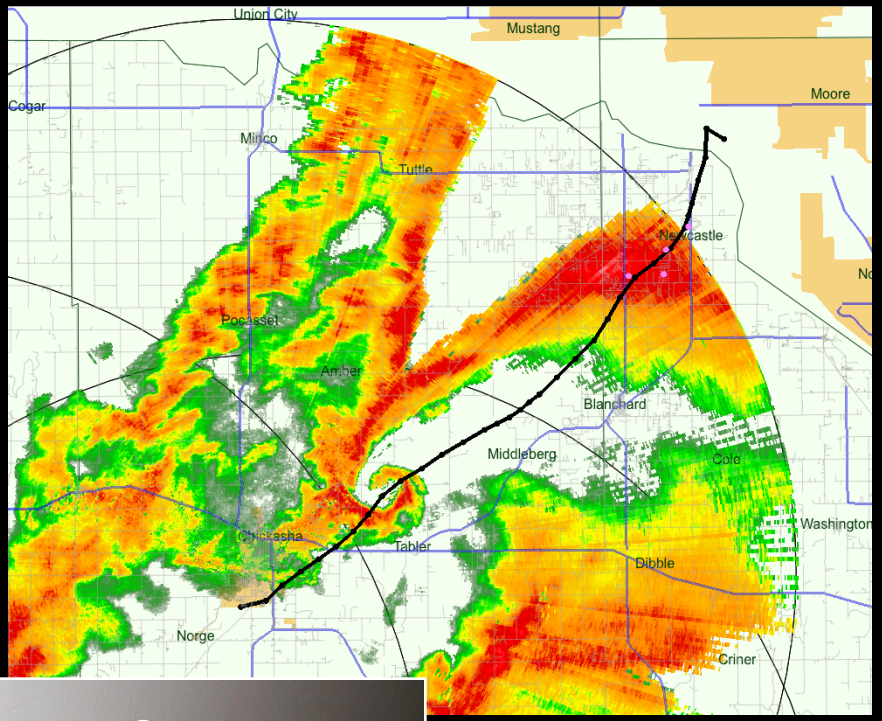
Comparison With Existing System

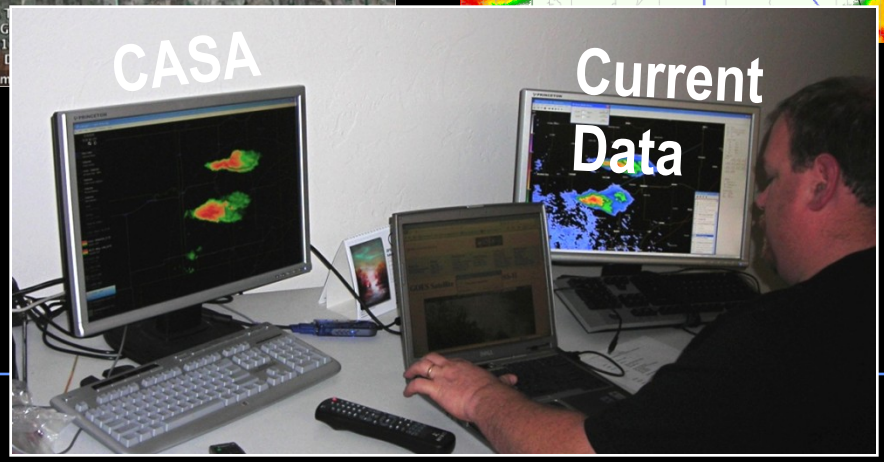


Source: Mike Zink, UMass Amherst

Potential

Source: Mike Zink, UMass Amherst

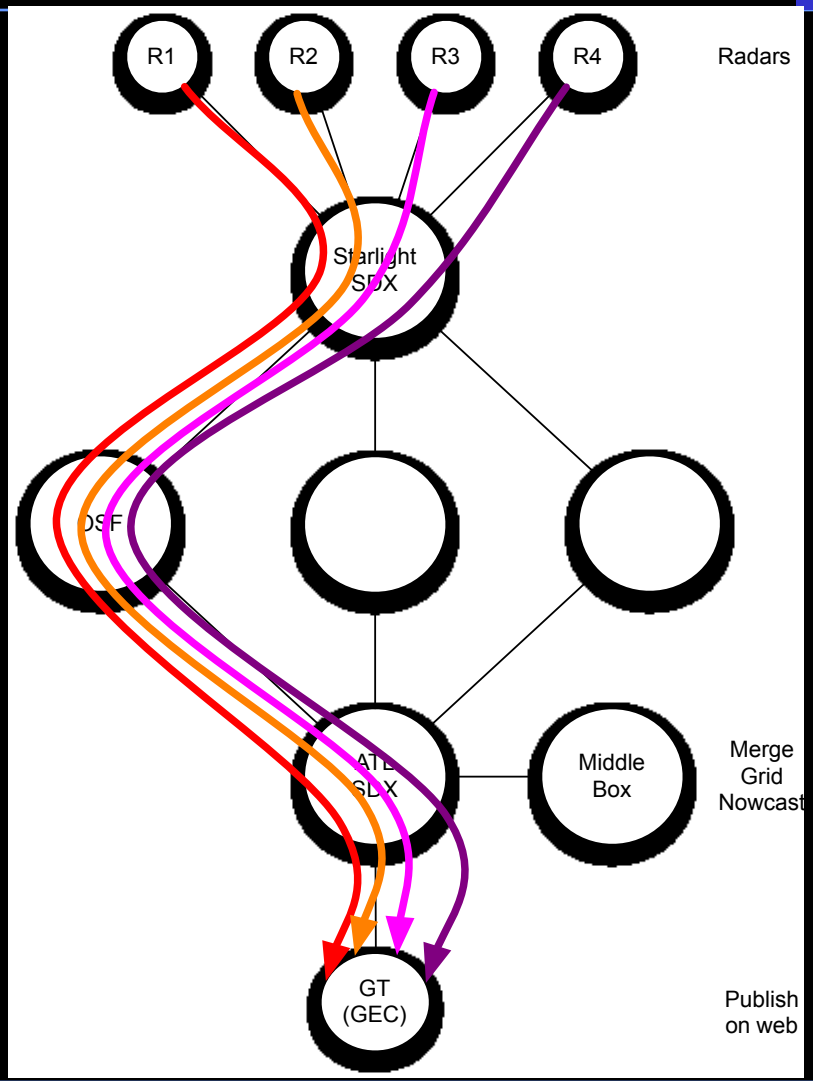


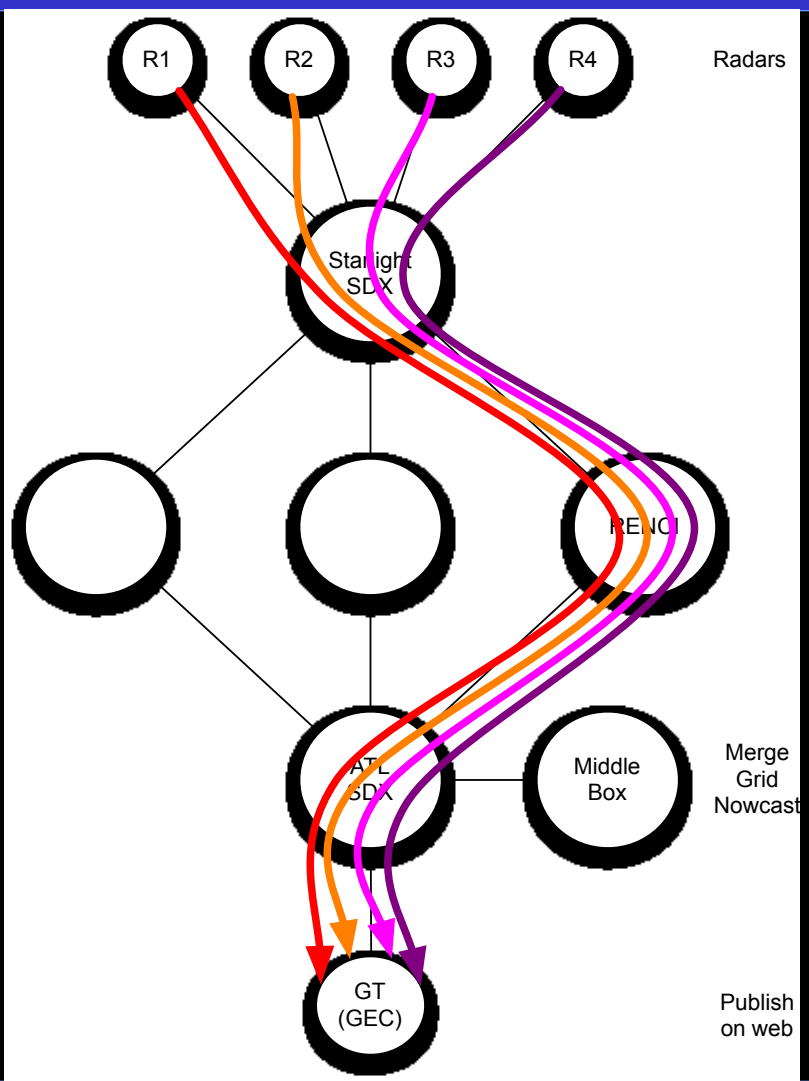


CASA Data, EM Decision-Making Protects First Responders and Public

STORMLIGHT℠

GENI SDX Demo Scenario 1



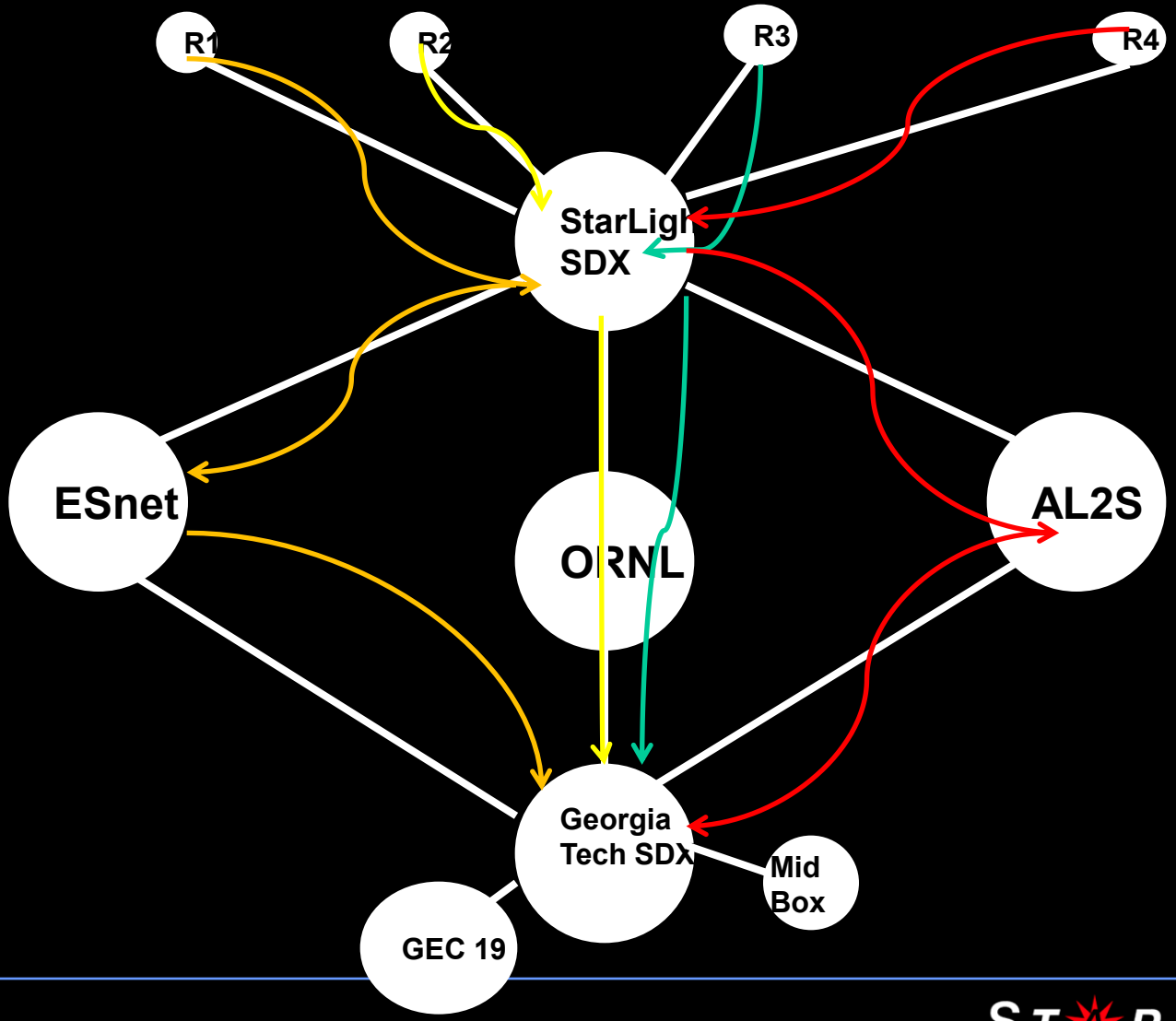


Slide by Mike Zink, UMass Amherst

STARLIGHT℠

GENI SDX Demo Scenario 2

Simulated Radar (4)



STARLIGHT℠

Global LambdaGrid Workshop (GLIF)

- International SDX (iSDX) Demonstrations Showcase a World-wide Prototype Environment That Could Be Used for Modeling Major Weather Systems, Including The Depiction of Severe Weather Patterns.
- Application Based on Nowcast System Being Developed by the NSF Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere (Led By University of Massachusetts at Amherst), Which Is Being Designed for Next Generation Weather Prediction and Visualization Systems.

Global LambdaGrid Workshop Sept 30-Oct 1 2014 In New Zealand

- Demonstrations of A Prototype Implemented Across the Globe Using the GLIF To Interconnect Sites World-Wide, Supported By Interoperable International SDXs, Including Prototype SDXs at
  - a) the StarLight International/National Communications Exchange Facility, Designed by iCAIR
  - b) Auckland, New Zealand, Designed by REANNZ and Google,
  - c) Taiwan, Designed by High Performance Computing Center/TWAREN
  - d) SOX in Atlanta, Designed by Georgia Tech
  - d) NetherLight in Amsterdam, Designed by SURFnet,
  - e) Ottawa, Designed by CANARIE and Cybera,
  - f) Tokyo, Designed by the University of Tokyo
- Each Site Had Its Own SDN/OpenFlow Controllers, Which Were Federated To Enable Cross-Domain Interoperability.
- Controllers Used Distributed Control Plane To Directly Address and Dynamically Manage Multiple Paths Among Sites Via Distributed Data Plane - Transporting Nowcast Instrumentation Traffic Among Sites.

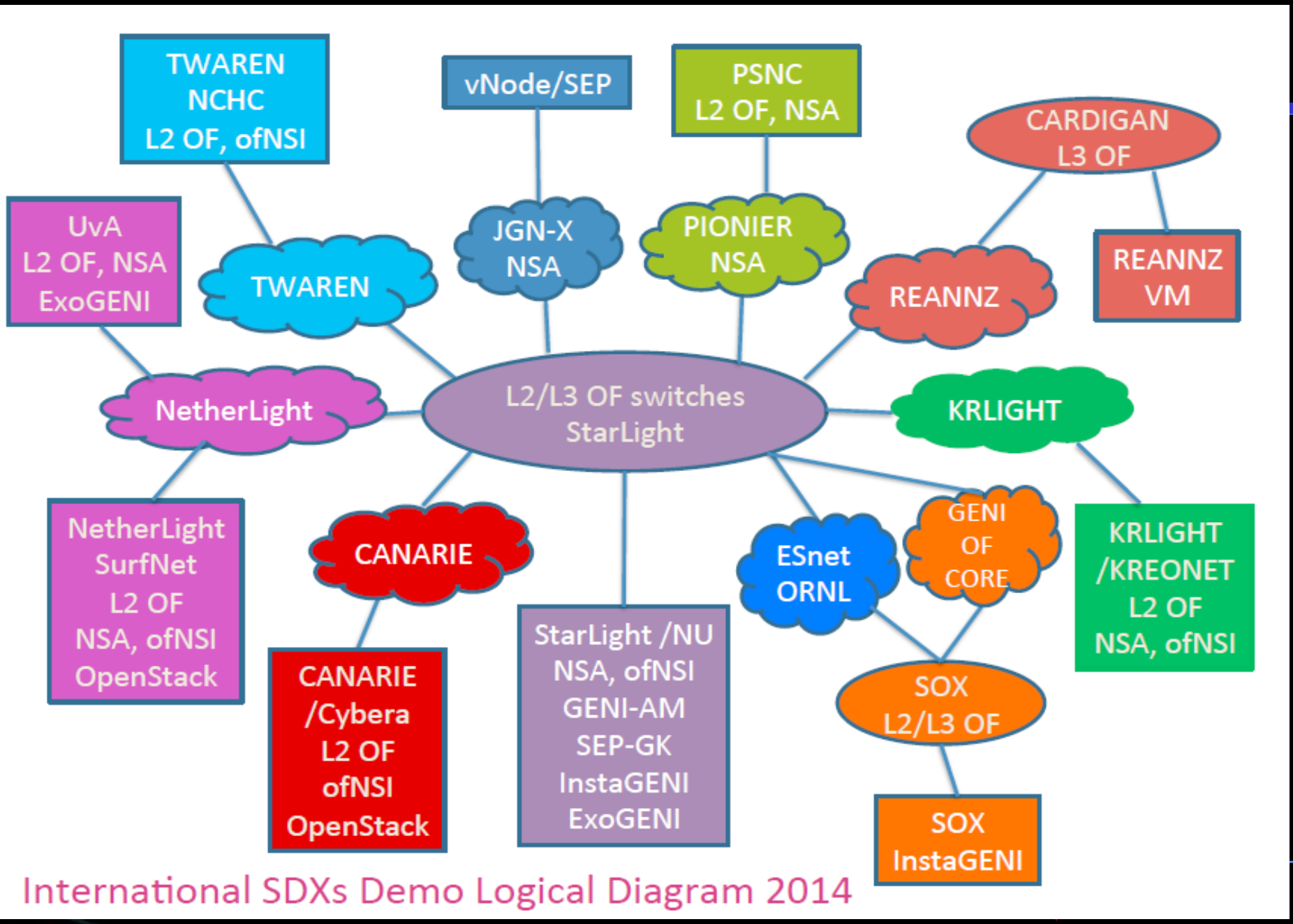
International Software-Defined Network Exchanges (iSDXs): A Demonstration of Global Capabilities

Joe Mambretti, Jim Chen, Fei Yeh
International Center for Advanced Internet Research
Northwestern University, USA
Mike Zink, Divyashri Bhat
University of Massachusetts, Amherst, USA
Ronald Van der Pol
Surfnet, Netherlands
Grace Lee, WunYuan Huang, Te-Lung Liu
NARLabs, National Center for High Performance Computing,Taiwan
Thomas Tam, Herve Guy,
CANARIE, Canada
Alex Valiushko, John Shillington,
Cybera, Canada
Buseung Cho, KISTI
Republic of Korea
Michiaki Hayashi, KDDI Labs, Japan
Toshiaki Tarui, Hitachi, Japan
Aki Nakao, University of Tokyo, Japan
Steve Cotter, T. Charles Yun, Jamie Curtis, Andrej Ricnik
REANNZ, New Zealand
Josh Bailey, Google, New Zealand
Artur Binczewski Belter Bartosz Miłosz Przywecki Piotr Rydlichowski
Poznan Supercomputing and Networking Center, Poland
Russ Clark, Georgia Tech, USA

Global LambdaGrid Workshop
Queenstown, New Zealand
September 30-October 1, 2014

iCAIR

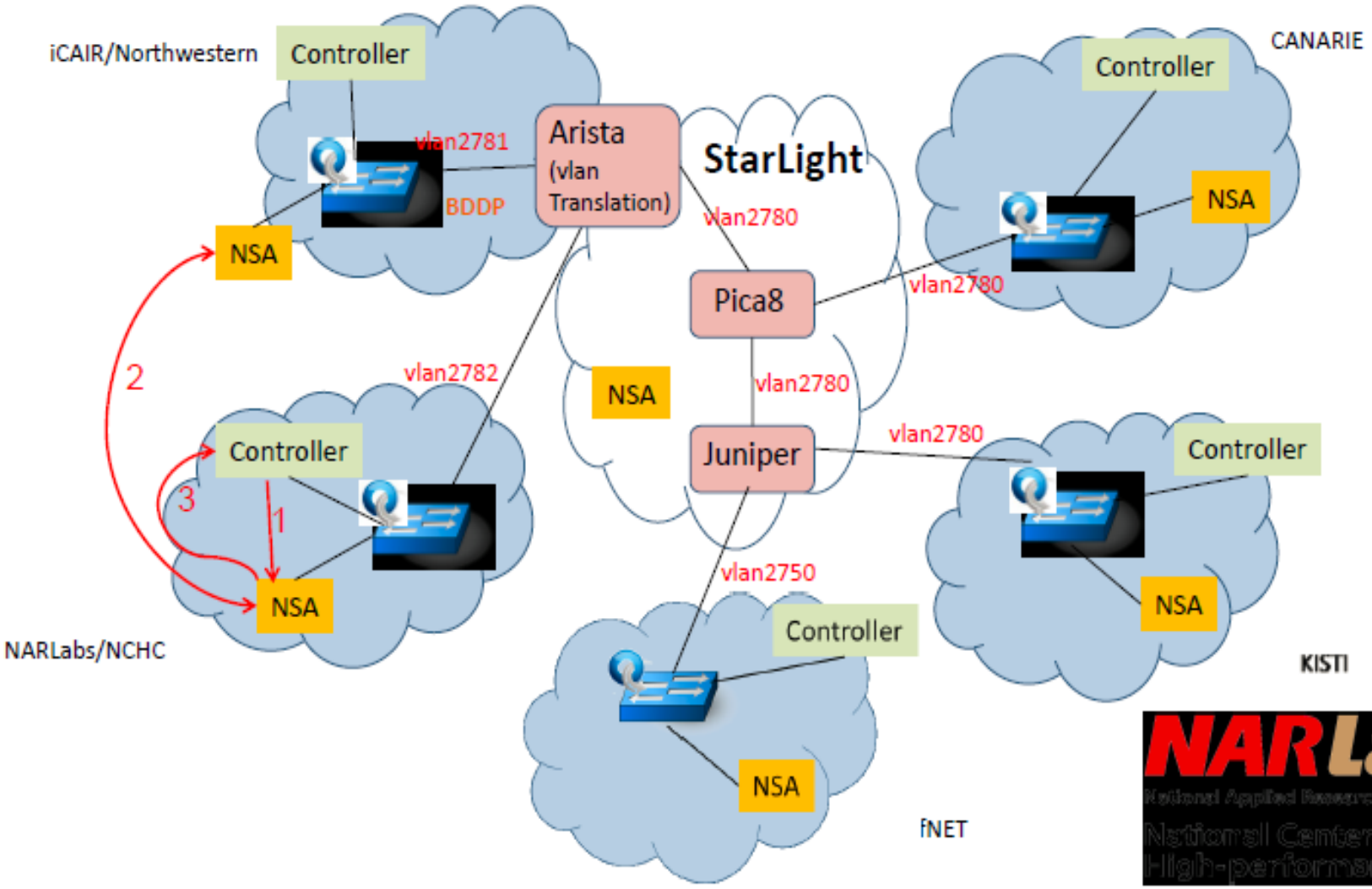
STARLIGHT℠



International SDXs Demo Logical Diagram 2014


Nowcast - SDX Demo
iCAIR
g1t

NSI-OpenFlow Hybird Topology Exchange



Slice Exchange Showcase at GEC 21: Special Thanks To Aki Nakao

Japan-US Slice Exchange over SDX

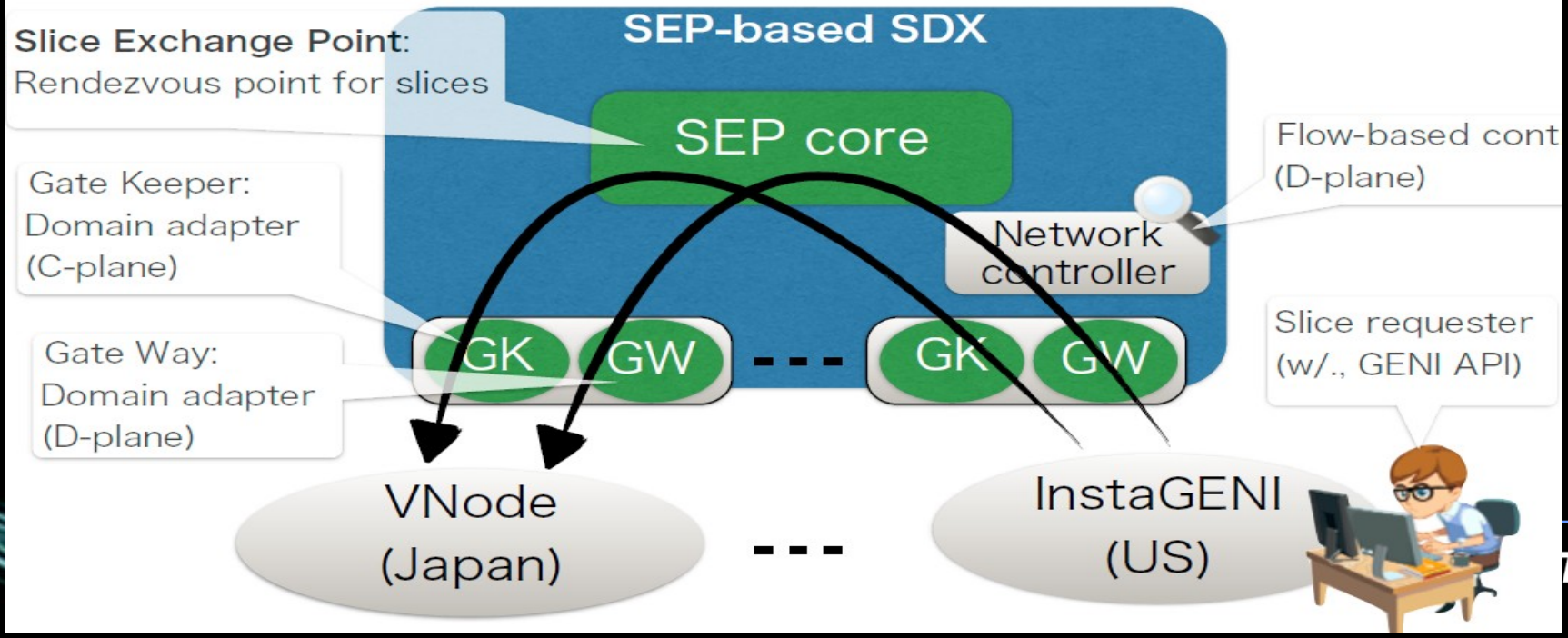
KDDI R&D LABS

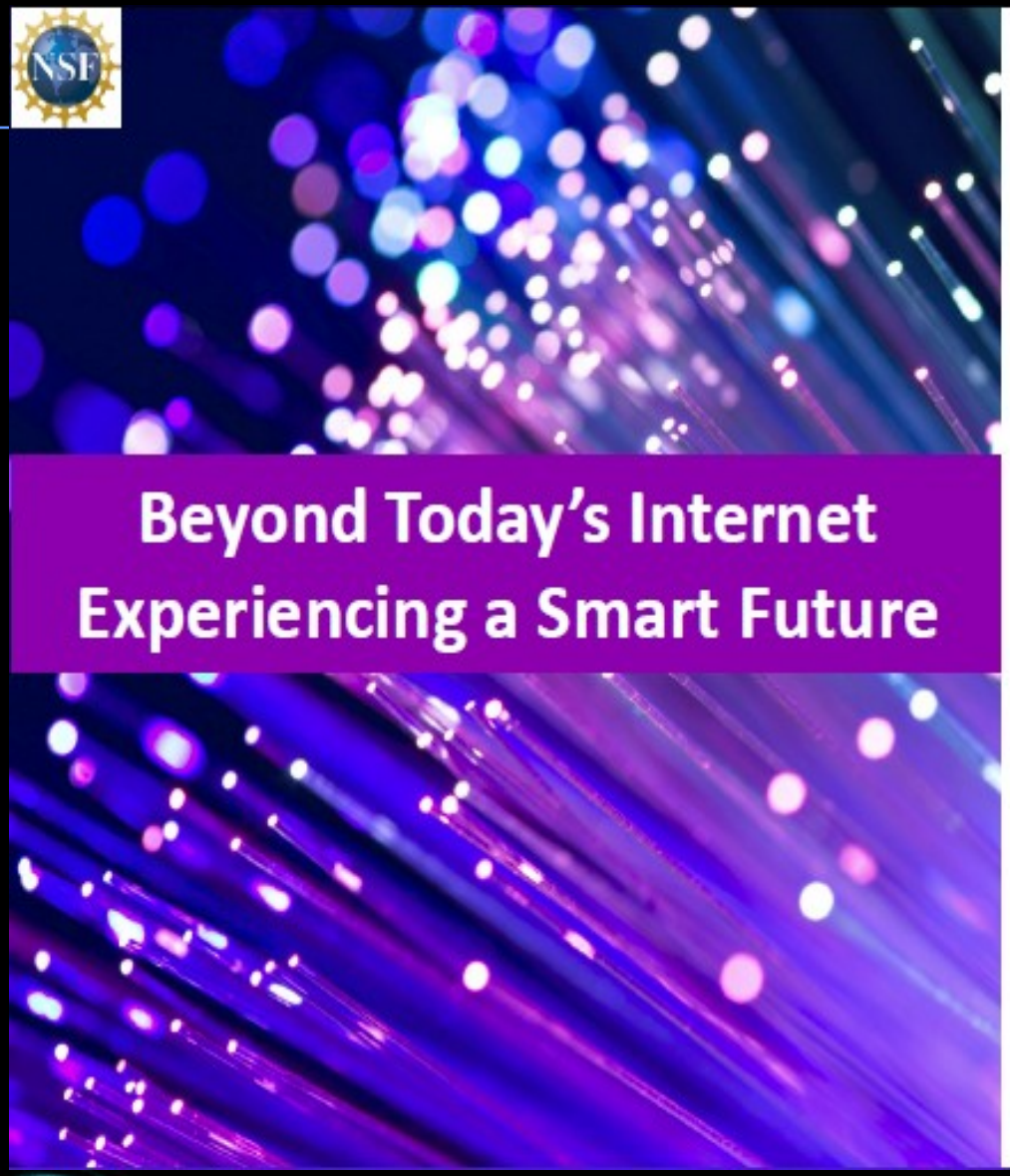
HITACHI Inspire the Next

東京大学 THE UNIVERSITY OF TOKYO

NORTHWESTERN UNIVERSITY

Slice Exchange Architecture



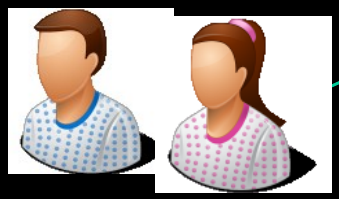


Prototype SDX Bioinformatics Exchange: Demonstrating an Essential Use-Case for Personalized Medicine

Robert Grossman, Piers Nash, Allison Heath, Renuka Arya
University of Chicago

Joe Mambretti, Jim Chen
Northwestern University

Future Vision:
A Nationwide Virtual Comprehensive Cancer Center



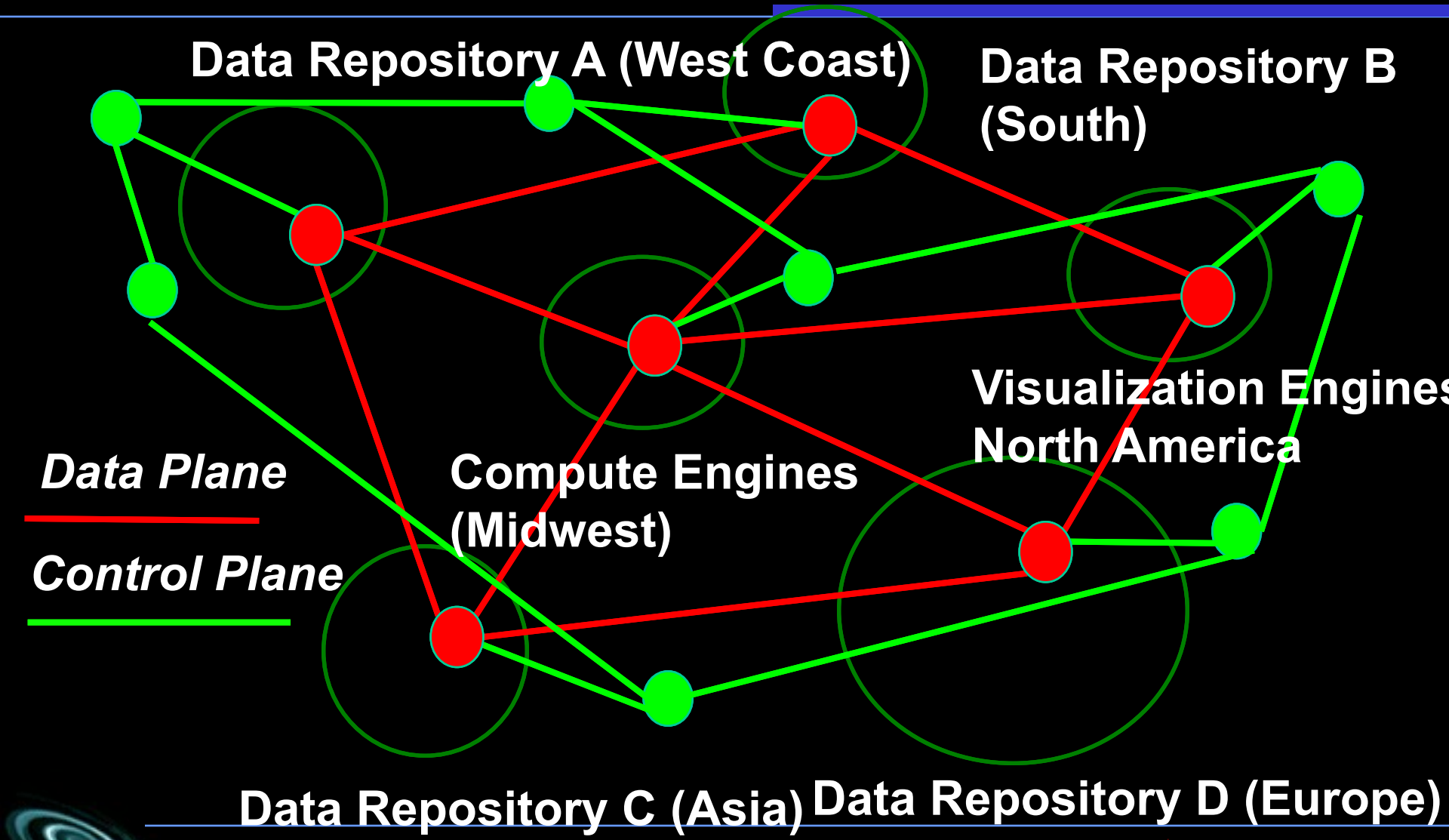
Patients

Hospitals,
Doctors

Cloud Computation
Genomic Data Commons

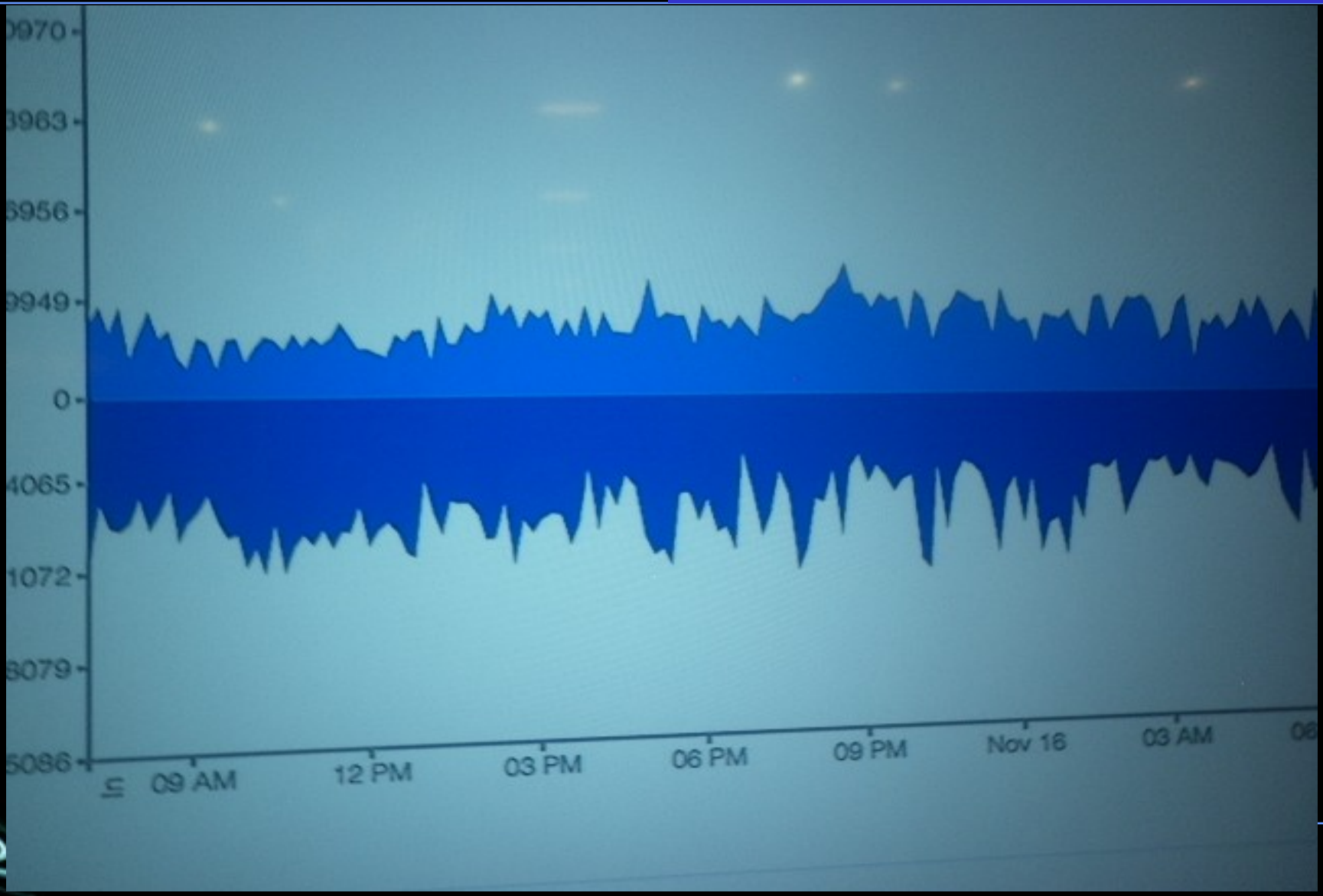
Output: Data-Aware, Analytics-Informed Diagnosis, Prognosis, Optimal Treatment

Biomedical Data Commons:
Flow Orchestration: Control Plane + Data Plane

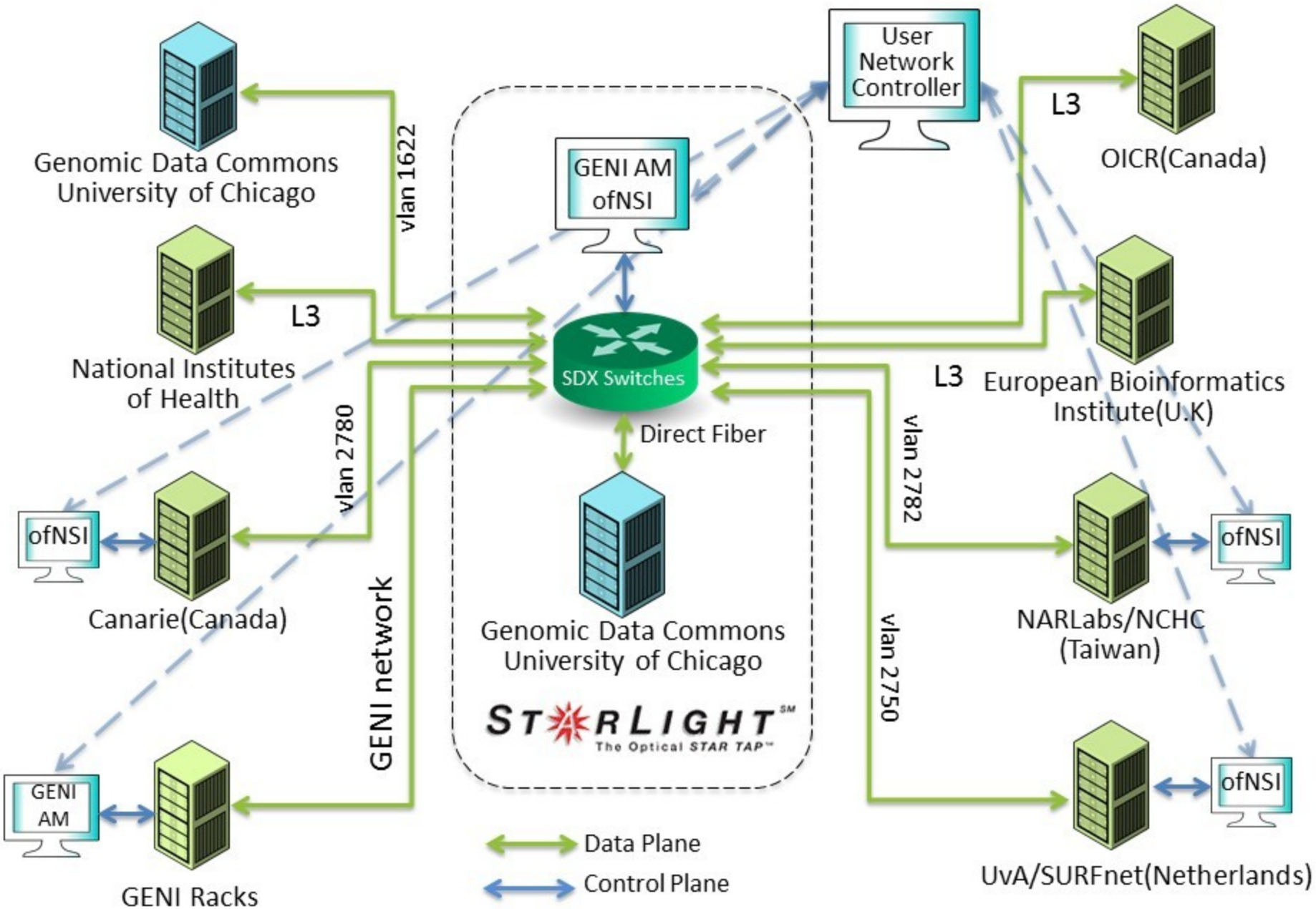


STARLIGHT℠

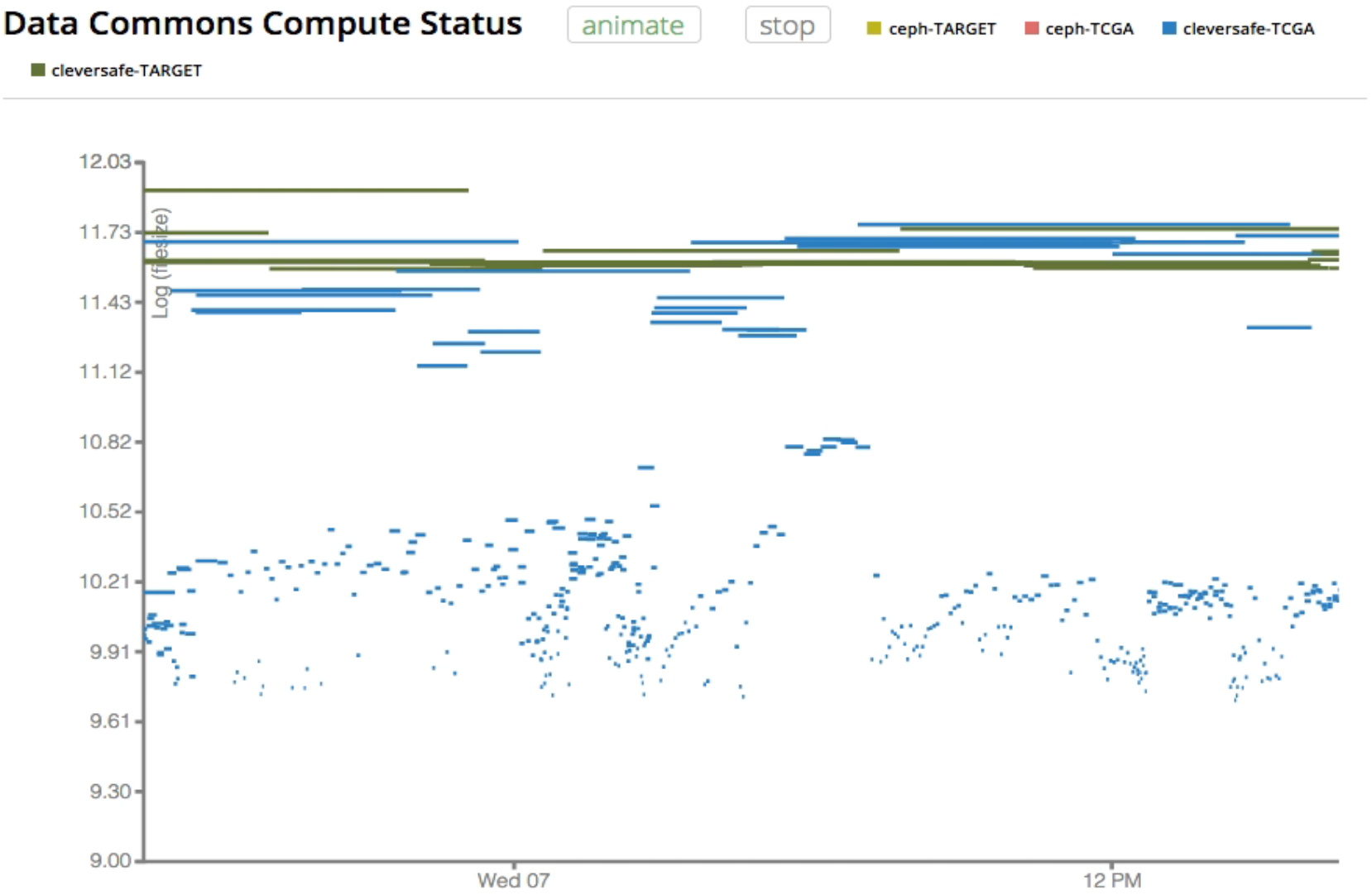
BI Data Flow Visualization (Inbound-Outbound) From SDSC To UoC



GEC22 Bioinformatics SDXs Demo Network



Genomic Data Commons Data Transfer



An Experimental Testbed For Computer Science Research

www.chameleoncloud.org

CHAMELEON:
A LARGE-SCALE, RECONFIGURABLE EXPERIMENTAL ENVIRONMENT FOR CLOUD RESEARCH

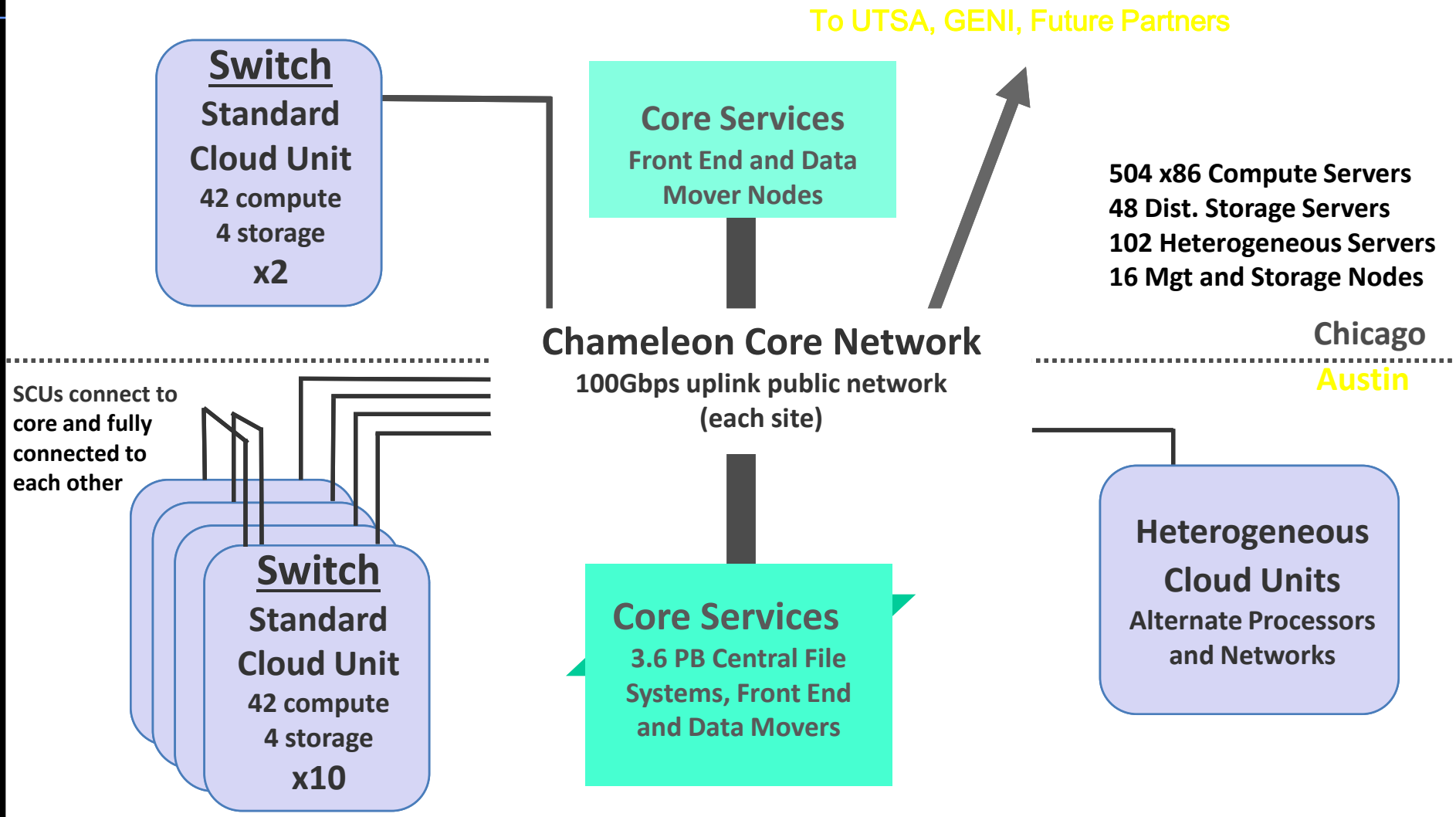
Principal Investigator: Kate Keahey

Co-PIs: J. Mambretti, D.K. Panda, P. Rad, W. Smith, D. Stanzione

Chameleon Design Strategy

- **Large-scale:** "Big Data, Big Compute, Big Instrument Research"
  - ~650 nodes (~14,500 cores), 5 PB disk over two sites, 2 sites connected with 100G network
- **Reconfigurable:** "As close as possible to having it in your lab"
  - From bare metal reconfiguration to clouds
  - Support for repeatable and reproducible experiments
- **Connected:** "One stop shopping for experimental needs"
  - Workload and Trace Archive
  - Partnerships with production clouds: CERN, OSDC, Rackspace, Google, and others
  - Partnerships with users
- **Complementary**: Partnerships With Many Other Projects
  - Complementing GENI, Grid'5000, and other experimental testbeds

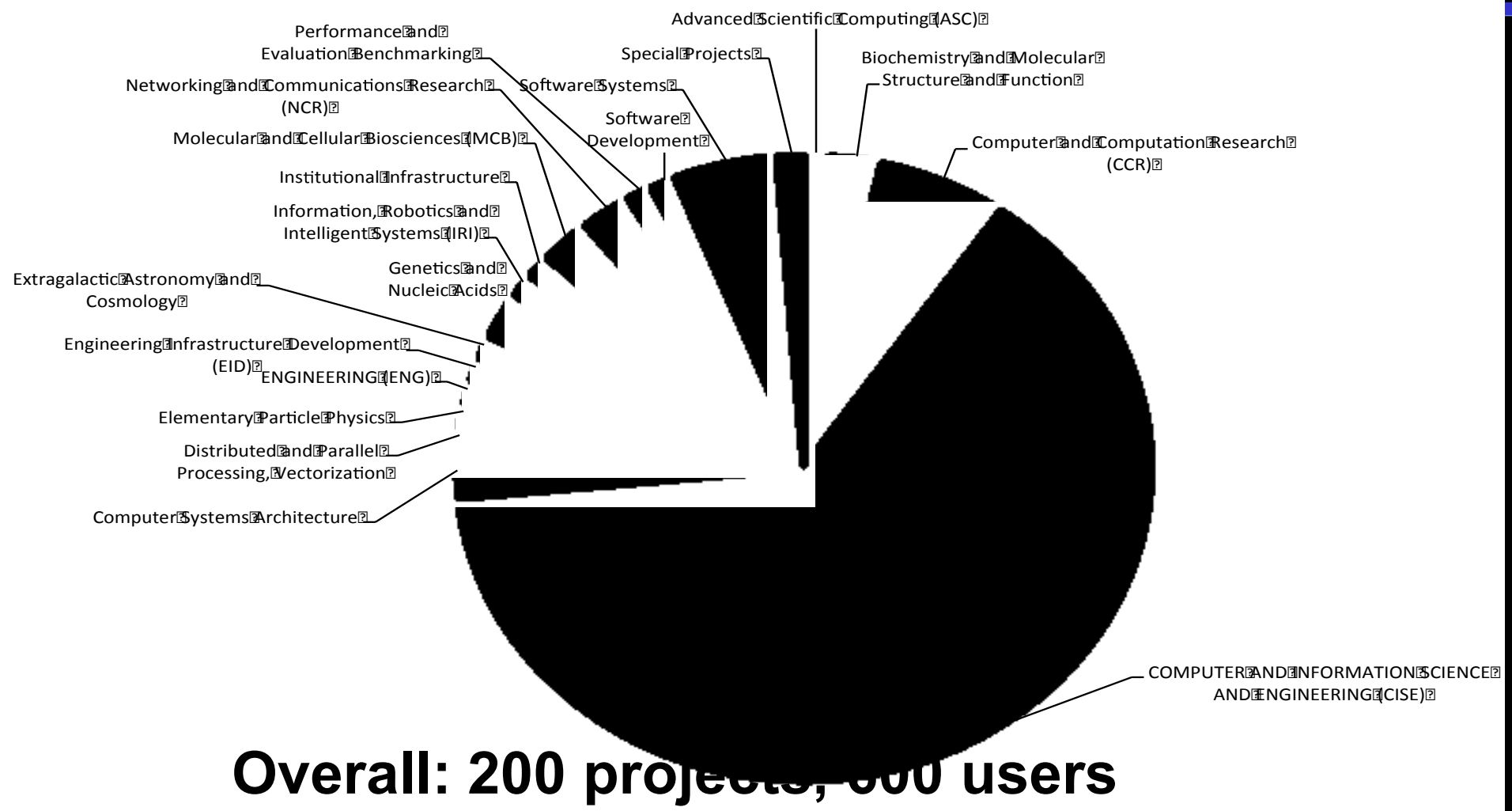
Chameleon Hardware



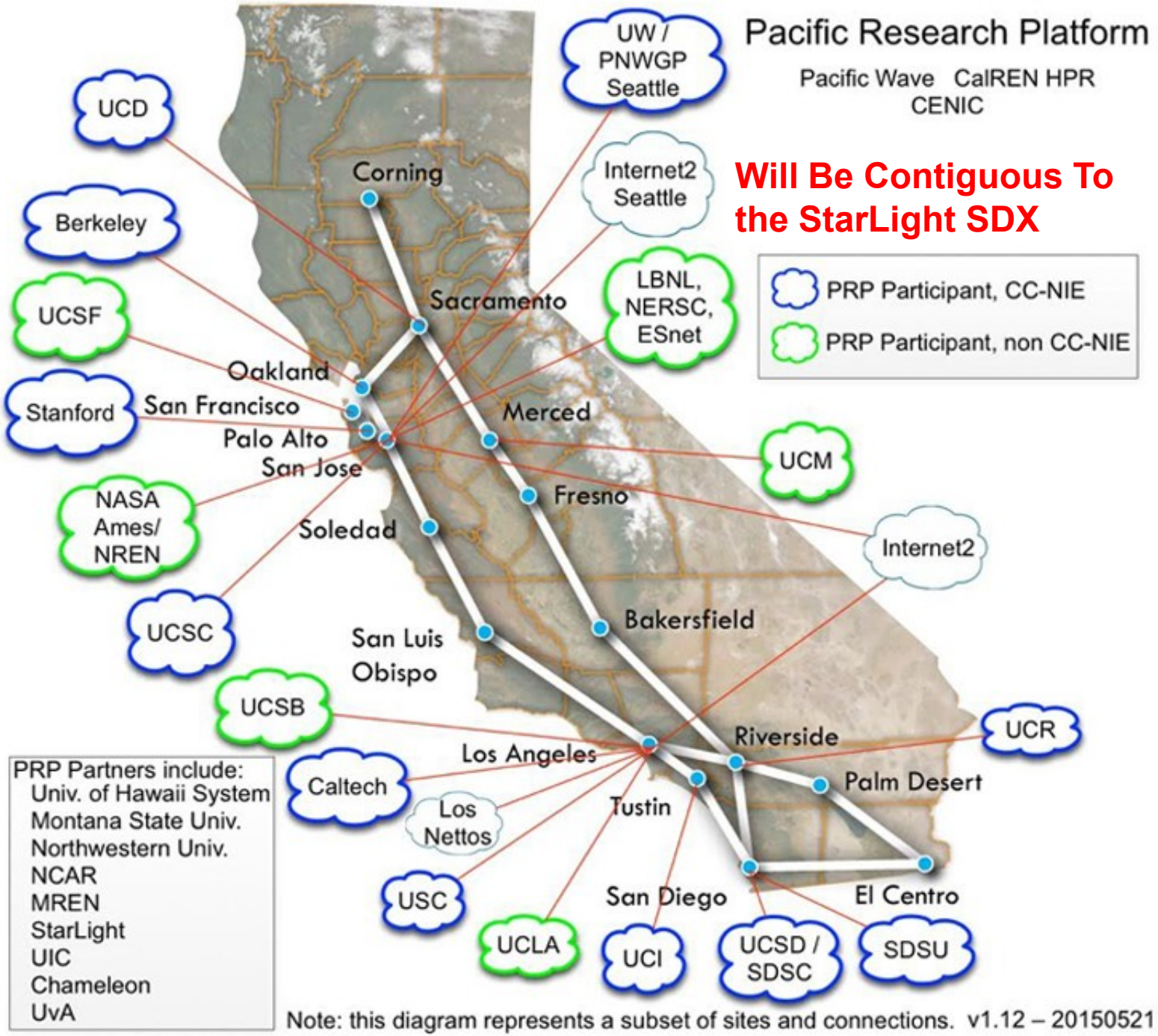
Standard Cloud Unit

- Each of the 12 SCUs is comprised of a single 48U rack
  - Allocations can be an entire SCU, multiple SCUs, or within a single one.
- A single 48 port Force10 s6000 OpenFlow-enabled switch connects all nodes in the rack (with an additional network for management/control plane).
  - 10Gb to hosts, 40Gb uplinks to Chameleon core network
- An SCU has 42 Dell R630 compute servers, each with dual-socket Intel Xeon (Haswell) processors and 128GB of RAM
- In addition, each SCU has 4 DellFX2 storage servers, each with a connected JBOD of 16 2TB drives.
  - Can be used as local storage within the SCU, or allocated separately (48 total available for Hadoop configurations)

Chameleon Projects



Overall: 200 projects, 600 users


Pacific Research Platform
Pacific Wave CalREN HPR
CENIC
Will Be Contiguous To
the StarLight SDX
PRP Participant, CC-NIE
PRP Participant, non CC-NIE
UW / PNWGP Seattle
UCD
Corning
Internet2 Seattle
Berkeley
LBNL, NERSC, ESnet
Sacramento
UCSF
Oakland
Stanford
San Francisco
Merced
Palo Alto
San Jose
UCM
NASA Ames/ NREN
Fresno
Soledad
Internet2
UCSC
Bakersfield
San Luis Obispo
UCSB
Riverside
UCR
Los Angeles
Caltech
Palm Desert
Tustin
Los Nettos
USC
San Diego
El Centro
UCLA
UCI
UCSD / SDSC
SDSU
PRP Partners include:
Univ. of Hawaii System
Montana State Univ.
Northwestern Univ.
NCAR
MREN
StarLight
UIC
Chameleon
UvA
Note: this diagram represents a subset of sites and connections. v1.12 – 20150521

Global Research Platform

- Currently – A Concept
- A Specialized Globally Distributed Platform For Science Discovery and Innovation
- Based On State-Of-the-Art-Clouds
- Interconnected With Computational Grids, Supercomputing Centers, Specialized Instruments, et al
- Also, Based On World-Wide 100 Gbps Networks
- Leveraging Advanced Architectural Concepts, e.g., SDN/SDX/SDI – Science DMZs
- Ref: Demonstrations @ SC15, Austin Texas November 2015

www.startap.net/starlight

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

iCAIR