

Global Research Platform Workshop: An Overview

Joe Mambretti, Director, (j-mambretti@northwestern.edu)

International Center for Advanced Internet Research (www.icaair.org)

Northwestern University

Director, Metropolitan Research and Education Network (www.mren.org)

**Director, StarLight International/National Communications Exchange Facility
(www.startap.net/starlight),**

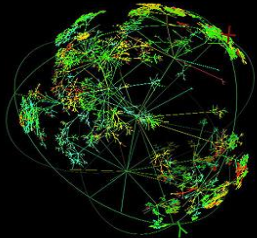
PI IRNC: RXP: StarLight SDX, Co-PI Chameleon, PI-iGENI, PI-OMNINet

Global Research Platform Workshop

Co-Located With IEEE International Conference On eScience

Limassol, Cyprus

October 9-10, 2023



Annual Global Research Platform Workshop – Co-Located With IEEE International Conference On eScience Oct 9-10

'23 eScience

October 9-13, 2023

Limassol, Cyprus

IEEE eScience 2023 brings together leading interdisciplinary research communities, developers and users of eScience applications and enabling IT technologies. The objective of the eScience Conference is to promote and encourage all aspects of eScience and its associated technologies, applications, algorithms and tools with a strong focus on practical solutions and challenges. eScience 2023 interprets eScience in its broadest meaning that enables and improves innovation in data- and compute-intensive research across all domain sciences ranging from traditional areas in physics and earth sciences to more recent fields such as social sciences, arts and humanities, and artificial intelligence for a wide variety of target architectures including

Important Dates

~~February 10, 2023~~ **Friday, February 24, 2023**
Workshop Submissions

~~February 24, 2023~~ **Friday, March 10, 2023**
Workshop Acceptance Notification

Friday, May 26, 2023
Paper Submissions

Friday, June 30, 2023
Notification of Paper Acceptance



Next Generation Distributed Environment For Global Science



NSF's Cyberinfrastructure Framework for the 21st Century (CIF21)

- ***“Across the full range of NSF---supported fields increasingly sophisticated instrumentation and expanded computational resources are opening new windows onto phenomena from the universe to the human brain, from the largest scales to the smallest. Across all domains, data play the key role in a profound transformation of the culture and conduct of science and society.***
- ***This Revolution Will Transform Research, Practice, And Education In Science and Engineering As Well As Advance Innovation In Society***
- ***This vision of the near future shows clearly the urgent need for a comprehensive, scalable, cyberinfrastructure that bridges diverse scientific communities and integrates high---performance computing, data, software, and facilities in a manner that brings theoretical, computational, experimental, and observational approaches together to advance the frontier.”***



Global Collaborative Research Communities

- **Science Is Global**
- **Open Information Sharing, A Cornerstone of The Science Process Is A Key Motivation For This Forum**
- **Concepts, Experiments, Instruments, Methods, Techniques, Data, Technologies And Results Are Openly Communicated and Shared Among Collaborative Science Communities World-Wide**
- **The Global Research Platform Is An International Collaborative Partnership Creating A Distributed Environment for International Data Intensive Science**
- **The GRP Facilitates High Performance Data Gathering, Analytics, Transport (100 Gbps-Tbps E2E), Computing, And Storage**
- **www.theglobalresearchplatform.net**

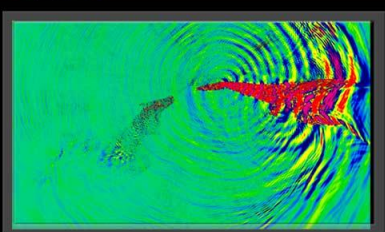
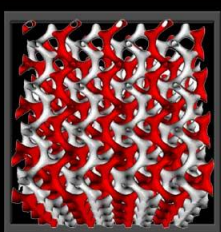
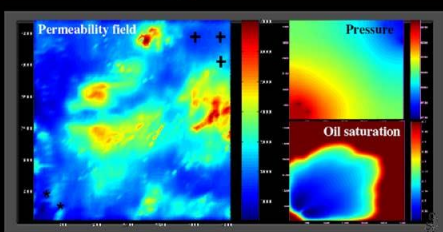
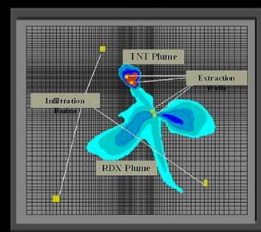
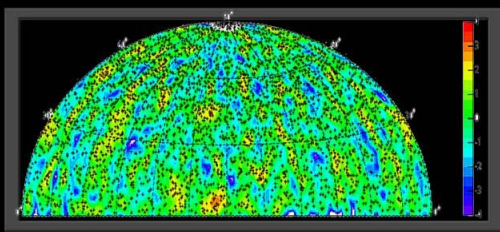


Large Scale Science Ecosystems

- **Science Domains Create Cyberinfrastructure Ecosystems, Some Distributed World Wide, Some Devoted To Domains, Some Shared Among Domains**
- **GRP Provides Opportunities For Information Sharing: Cyberinfrastructure Architecture, Implementation, Technologies and Operations Among Projects (Especially Useful For Cross Disciplinary Research)**
- **Projection/Definition of Future, Specialized Requirements, Architecture, Services, Techniques, Technologies, Processes Described In Cyberinfrastructure “Blueprints”**
- **Cambrian Explosion Of Requirements and Innovations**
- **Techniques and Technologies Emerge from Multiple Sources (Academic, Commercial, Government Labs, Utilitarian Imperatives, e.g., Commercial Clouds)**
- **Macro-Trend: “Software Eating The World” - Software Defined Everything**
- ***Multiple Software Building Blocks For Data-Intensive Science (Modules/Components) Are Emerging***

TERAGRID

Extensible Terascale Facility



An Early Example: TeraGrid

STARLIGHTSM

Selected Applications



GENI
www.geni.net



Open Storage Network
www.openstorage.network.org



OSIRIS
www.osris.org



XSEDE
www.xsede.org



Blue Waters
bluewaters.ncsa.illinois.edu



PRAGMA
www.pragma-grid.net



CENTRA
www.globalcentra.org



OSG
www.openscience.grid.org



GRP
theglobalresearchplatform.net/



PRP
pacificresearchplatform.org



CHASE-CI
www.calit2.net/newsroom/article.php?id=2910



SAGE2
sage2.sagecommons.org



Polar Geospatial Center
www.pgc.umn.edu



IceCube
icecube.wisc.edu



Chameleon
www.chameleoncloud.org



Jetstream
www.jetstream-cloud.org



Genomic Science Program
genomicscience.energy.gov



LSST
www.lsst.org



Pierre Auger Observatory
www.auger.org



Belle II
www.belle2.org



LBNF/DUNE/ProtoDUNE
lbnf.fnal.gov



ISS
www.nasa.gov/station



SKA
www.skatelescope.org



XENON
xenon.astro.columbia.edu



NOVA
novaexperiment.fnal.gov



Virgo
www.virgo-gw.eu



LIGO
www.ligo.caltech.edu



SDSS
www.sdss.org



ALMA
www.almaobservatory.org



LHC
home.cern/science/accelerators/large-hadron-collider



LHCONE
twiki.cern.ch/twiki/bin/view/LHCONE/WebHome



LHCOPN
twiki.cern.ch/twiki/bin/view/LHCOPN/WebHome



IVOA
www.ivoa.net

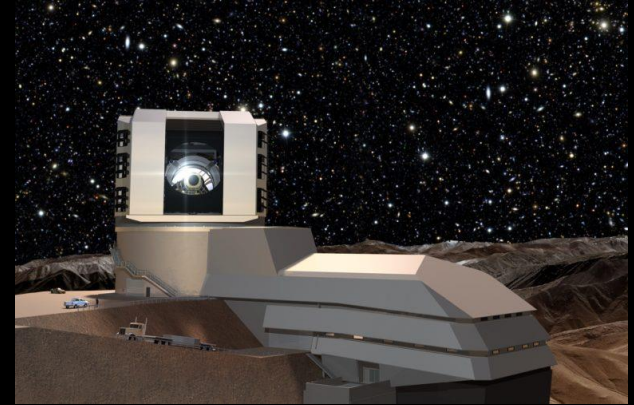
Instruments: Exebytes Of Data



High Luminosity LHC



SKA Australia Telescope Facility



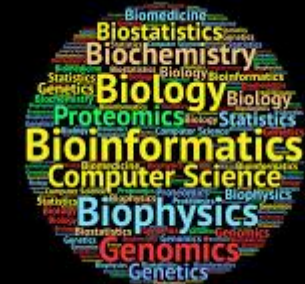
Vera Rubin Observatory



KSTAR Korea Superconducting Tokamak



Next Gen Advanced Photon Source



Bioinformatics/Genomics

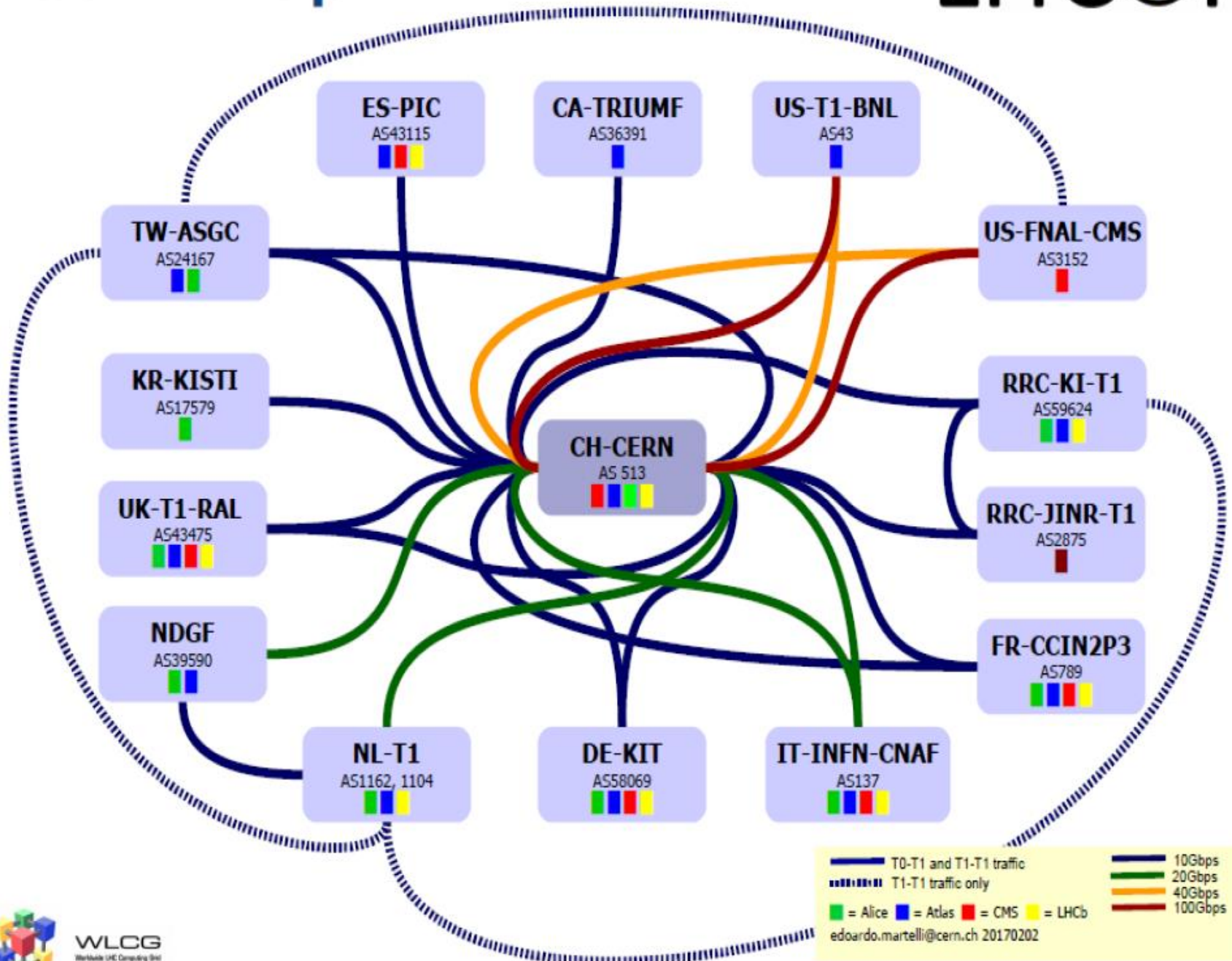
Global Scale Science Highlighted At Prior GRP Workshops

- **The Square Kilometer Array: Data Transport, Processing, Archiving and Access, Shaun Amy, Australia Telescope National Facility**
- **Large Synoptic Survey Telescope Distributed Computing and Networks, Jeff Kantor, LSST**
- **Korean Fusion Program: KSTAR, ITER and K-DEMO and International Collaborators, Si-Woo Yoon, National Fusion Research Institute**
- **Square Kilometer Array (SKA), Richard Hughes-Jones, GÉANT**
- **Vera C. Rubin Observatory, Large Synoptic Survey Telescope (LSST), Nate Lust, LSST/Rubin Observatory**
- **Belle II, Super B-Factory Experiment, Silvio Pardi, National Institute for Nuclear Physics, (INFN)**
- **Deep Underground Neutrino Experiment (DUNE) – Kenneth Herner, Fermi National, Accelerator Laboratory**
- **Distributed Computing Operations For HL-LHC With Operational Intelligence, Federica Legger, National Institute of Nuclear Physics (INFN)**
- **Next-Generation Cyberinfrastructures for LHC, High-Luminosity LHC and Data Intensive Sciences, Harvey Newman, Caltech**
- **KAUST Genomics Cloud, Alex Moura, KAUST**

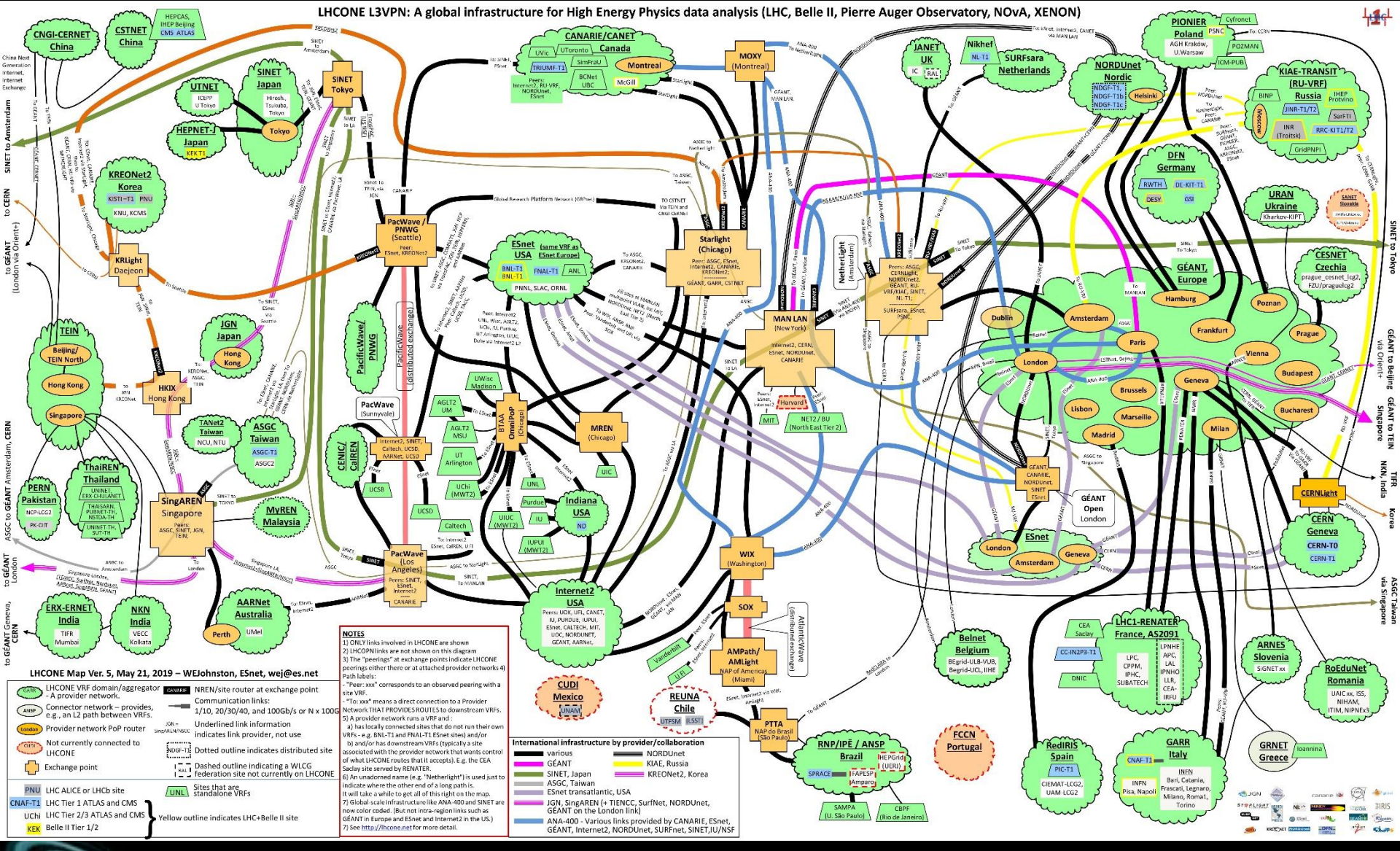
Questions

- **Can Services, Architecture, Technologies, and Techniques Be Replicated Across Large Scale Science Ecosystems?**
- **To What Degree Can Large Scale Science Ecosystems Be Generalized To Support Multiple Science Domains?**
- **An Example: The Worldwide LHC Computing Grid (WLCG): Global Collaboration Of ~ 170 Computing Centres In More Than 40 Countries, Integrating National and International Grid Infrastructures.**
- **WLCG Provides Global Resources To Gather, Store, Distribute and Analyse ~200 Petabytes of LHC Data Each Year**
- **WLCG – Partnership of EGI (European Grid Infrastructure), OSG (Open Science Grid), and NeIC (Nordic e-Infrastructure Collaboration).**

LHCOPN map



LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NoVA, XENON)



Non-LHC Scientific Communities Using LHCONE

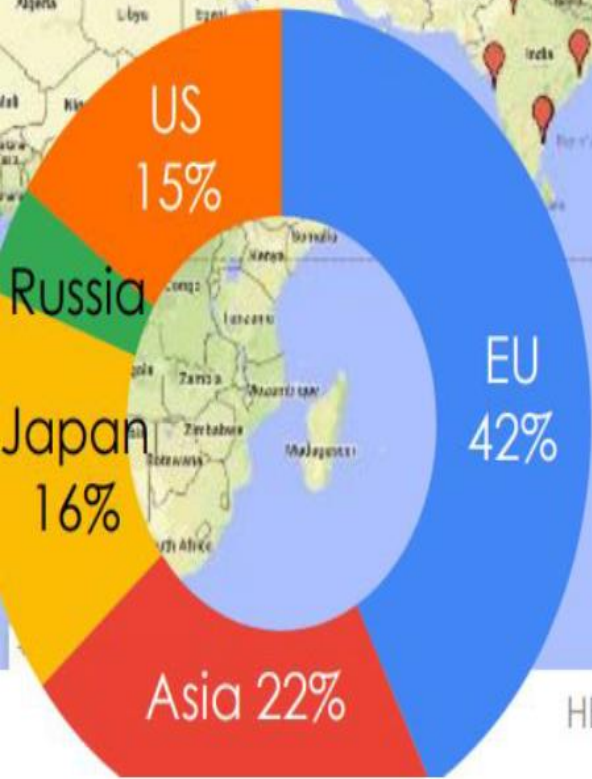
- **Belle II Experiment, Particle Physics Experiment Designed To Study Properties of B Mesons (Heavy Particles Containing a Bottom Quark)**
- **Pierre Auger Observatory, Studying Ultra-High Energy Cosmic Rays, the Most energetic and Rarest Particles in The Universe**
- **LIGO and Virgo (In August 2027 This Collaboration Measured a Gravitational Wave Originating From a Binary Neutron Star Merger.)**
- **NOvA Experiment: Designed To Answer Fundamental Questions In Neutrino Physics**
- **XEON Dark Matter Project: Global Collaboration Investigating Fundamental Properties of Dark Matter, Largest Component of the Universe**
- **DUNE/ProtoDUNE – Deep Underground Neutrino Experiment**





Belle II Collaboration

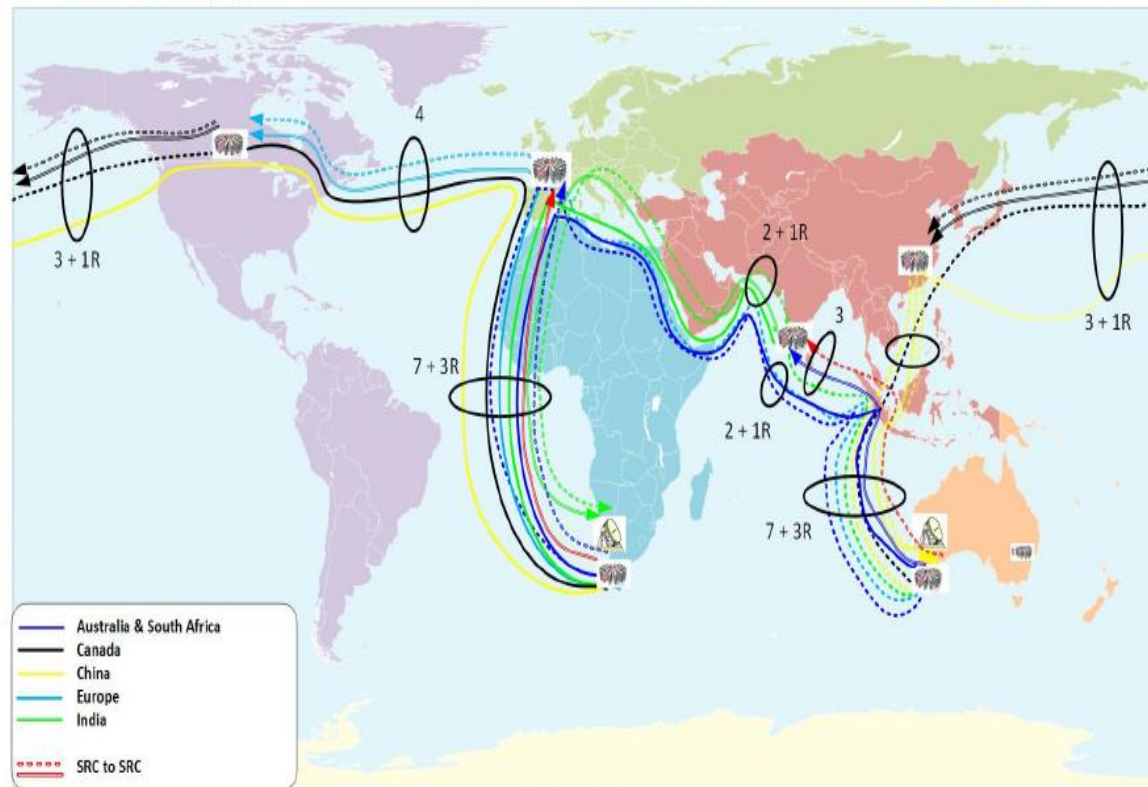
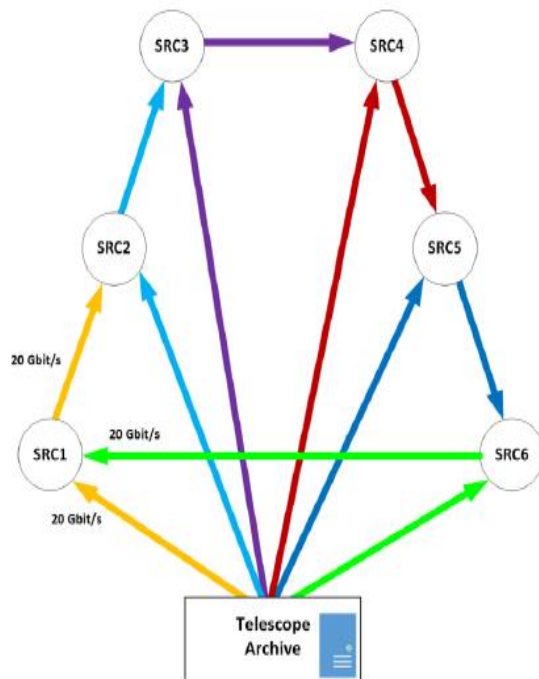
A Global Collaboration
as wide as an LHC experiment



26 countries/regions
123 institutes
1,075 researchers

Global Data Flows if the SRC Re-distribute data – 2 Replicas

- Each SRC accepts its fraction of the Observatory Data Products and re-distributes to another SRC.
- SRC has 20 Gbit/s flow from the telescope & a second continuous 20 Gbit/s flow from another SRC.
- Each SRC sends out a 20 Gbit/s flow.
- Makes substantial use of the shared academic network which would imply charges to the SKA community.
- **Probable cost to SKA community Very approx. ~ 0.8 M USD/year not allowing for the extra BW from the telescopes**



Emerging Next Generation Research Platforms

- *“a comprehensive, scalable, cyberinfrastructure that bridges diverse scientific communities and integrates high--
-performance computing, data, software, and facilities in a manner that brings theoretical, computational, experimental, and observational approaches together to advance the frontier”*
- **Large Scale Science DMZs**
- **Super Facilities**
- **National Research Platforms**
- **Continental Research Platforms**
- **MultiONE Prototype**



Global Research Platform: Prior GRP WSs

- **Science DMZ in GRP: Scaling Up, Eli Dart, ESnet**
- **Prototype National Research Platform, Frank Wuerthwein, UC San Diego/SDSC**
- **Evolution of OSG Resources and Operations, Derek Weitzel, U Nebraska**
- **AutoGOLE/SENSE Thomas Lehman**
- **The Superfacility – Bjoern Enders, NERSC/Lawrence Berkeley National Laboratory**
- **Introduction to Asi@Connect and APAN APRP WG Activities – Jeonghoon Moon, KISTI**
- **GÉANT Project Network Technologies and Services For Science – Ivana Golub, GÉANT**
- **Introduction to Korea Research Platform: R&E Together Project – Ki-Hyeon Kim, KISTI**
- **SINET/NII Hidehisa Nagano, NII**
- **Advanced Infrastructure for Science Susumu Date, Univ Osaka**



Orchestration Among Multiple Domains

- **Instrumentation and Analytic, Storage Resources Are Highly Distributed Among Multiple Domains Interconnected With High Performance Networks**
- **A Key Issues Is Discovering Resources, Claiming Them, Integrating Them, Utilizing Them and Releasing Them**
- **Increasingly, New Software Defined Infrastructure Architecture, Services, Techniques And Technologies Are Addressing These Issues**



Large-Scale High Capacity Data WAN Transport

- Large-Scale High Capacity Data WAN Transport Has Always Been And Remains A Major Challenge, Especially Over Global Paths
- This Issue Is Emphasized By A Next Generation Of Instrumentation That Will Generate Exponentially Large Volumes Of Data That Has To Be Distributed Across the Globe
- Often, This Issue Is Considered Reductively Only In Terms Of Network Capacity
- However, Actually It Is More An E2E Issue, Especially Given Advances In Core Optical Networking Technologies



High-Fidelity Data Flow Monitoring, Visualization, Analytics, Diagnostic Algorithms, Event Correlation AI/ML/DL

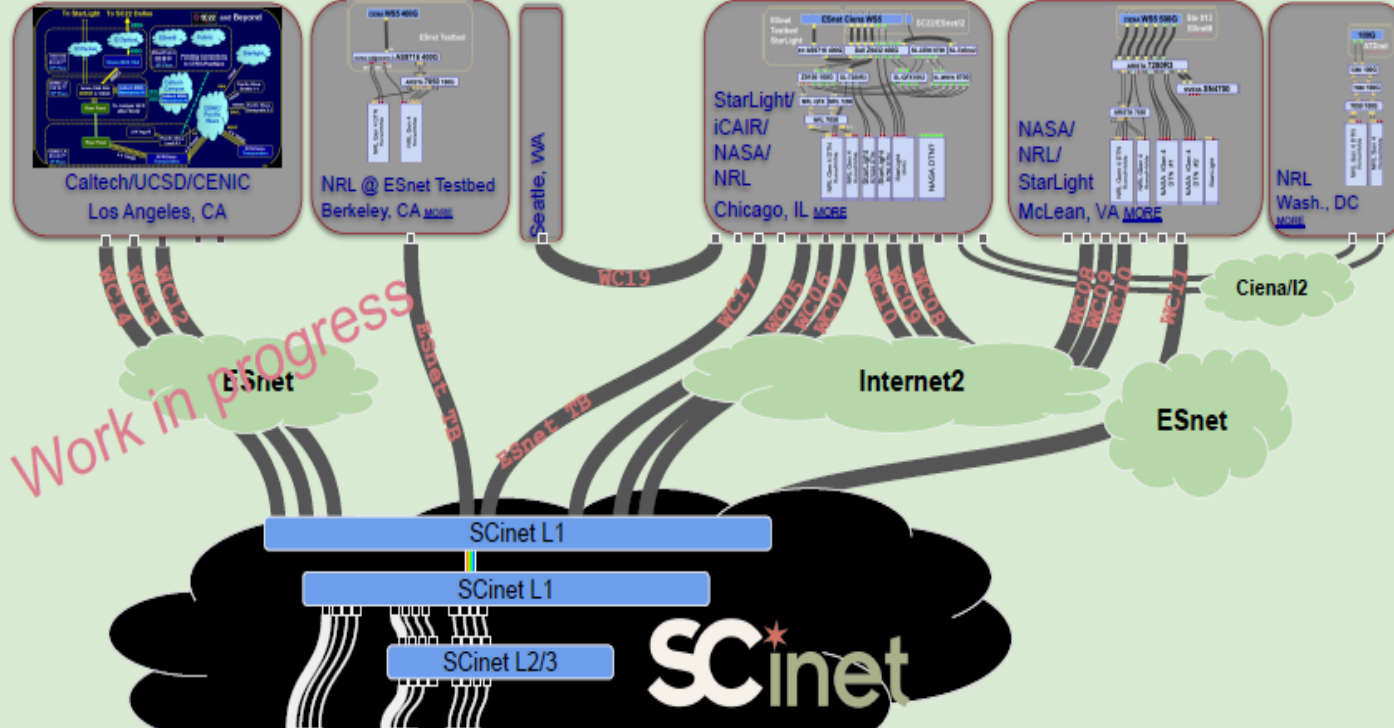
- **A Major Opportunity For Data Transport Optimization Is Being Provided By New Methods For Directly Detecting And Analyzing All Data Flows And Their Characteristics**
- **Because These Techniques Enable High-Fidelity Views Of All Flows, Real Time, Dynamic Traffic Engineering Is Possible With Much More Sophistication Than Traditional Approaches**
- **These Techniques Can Be Significant Enhanced Using AI/ML/DL, Which (Although Still Emerging) Are Becoming Critically Important Tools In The Near Term**



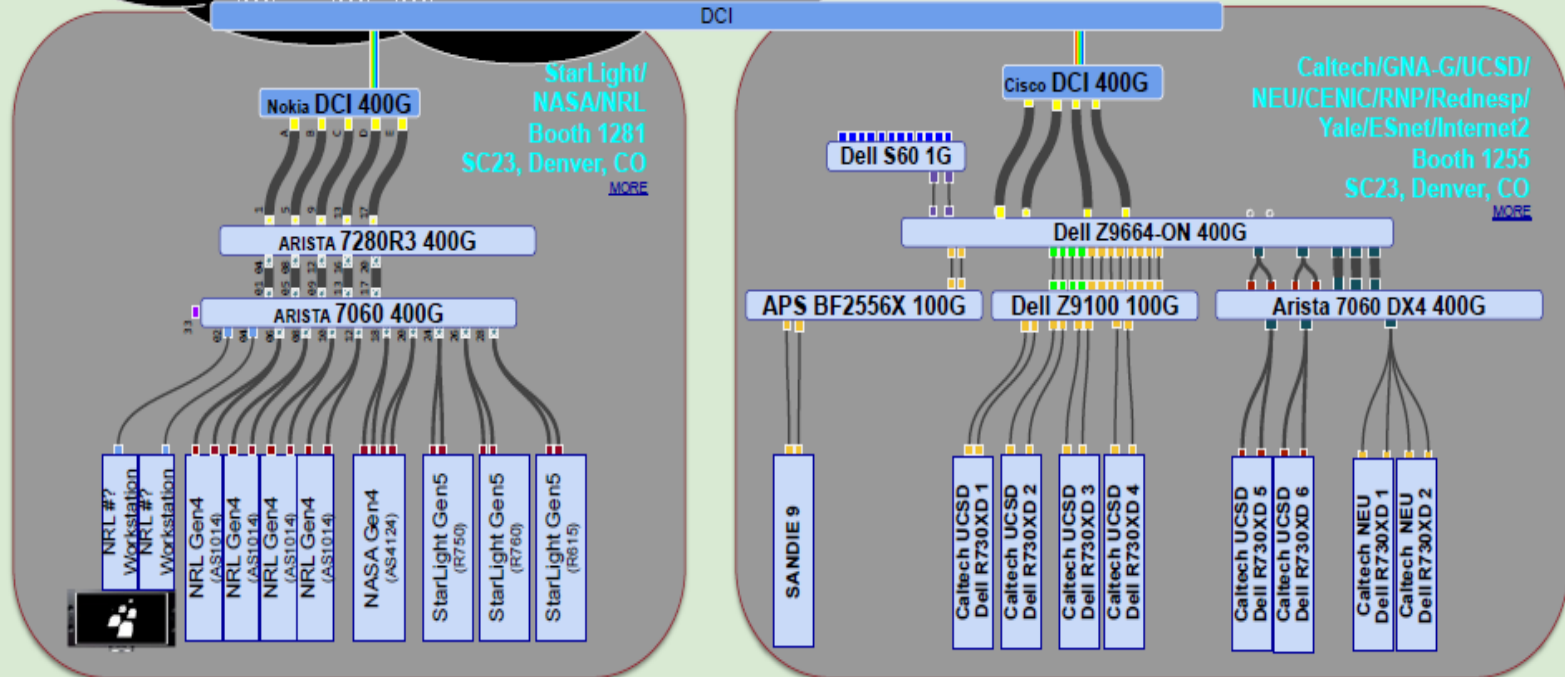
International Testbeds for Data-Intensive Science

- **Challenging Requirements Of Anticipated Large Scale Science Projects Along With Accelerated Rates Of Innovation Require International Testbeds For Pre-Production Investigations And Prototyping Of New Technologies And Techniques Specifically Related To Data Intensive Science, e.g., Tbps E2E WAN Services Among Sites**
- **Such Global Experimental Research Testbeds Exist Today, And They Are Being Developed With Enhanced Capacities, Sites, And Capabilities**





JOINT BIG DATA TESTBED



- 400G - LR4
- 400G - FR4
- 400G - DAC
- 200G - SR4 or DAC
- 100G - CLR4
- 100G - LR4
- 100G - SR4
- 100G - DAC
- 40G - SR4
- 40G - DAC
- 10G
- 1G

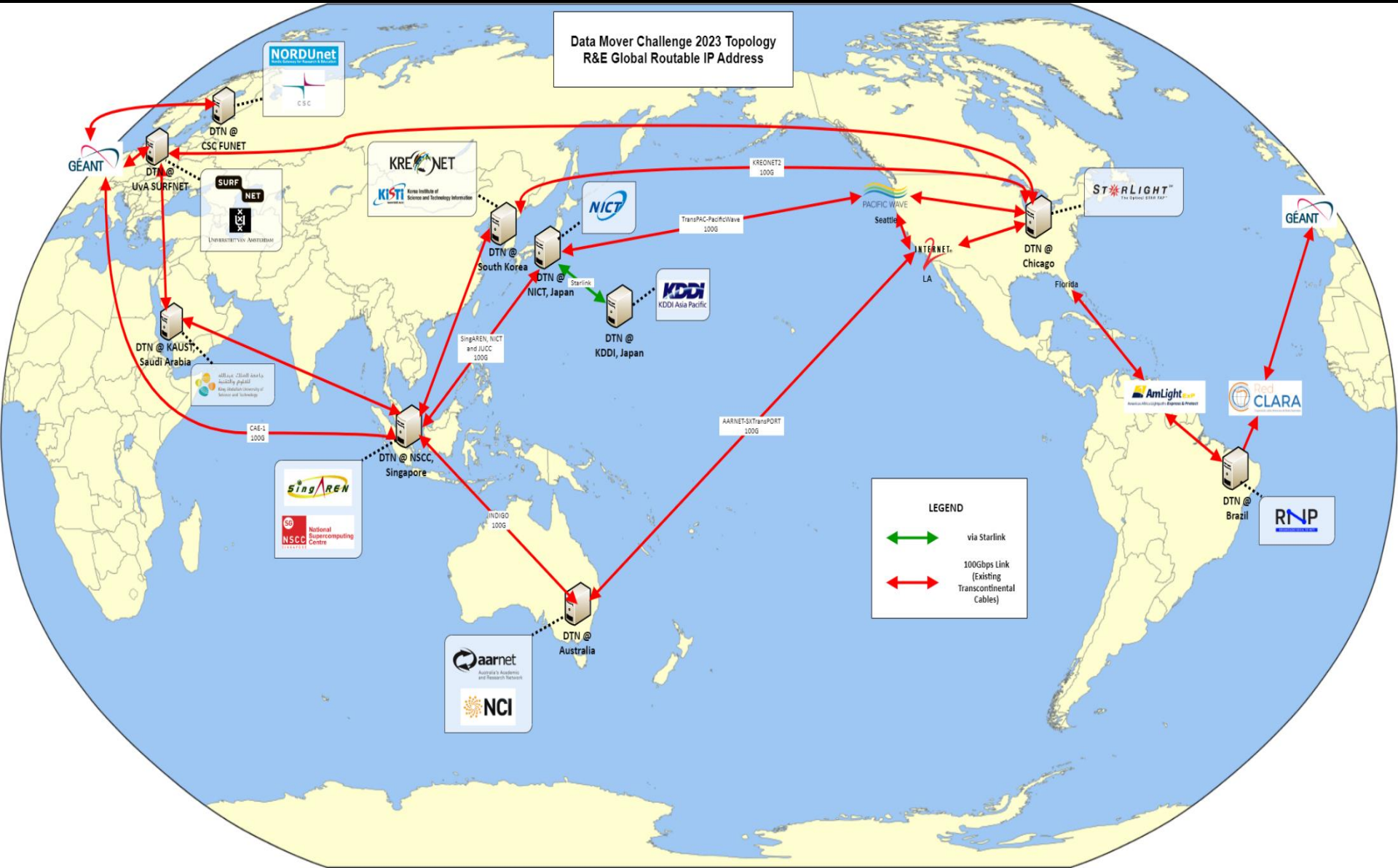
09/19/2023

Latest Version at:
<https://tinynurl.com/SC23-JBDT>
 To request changes, please leave a comment

SC23 floorplan SC22

SC19

Data Mover Challenge 2023 Topology
R&E Global Routable IP Address

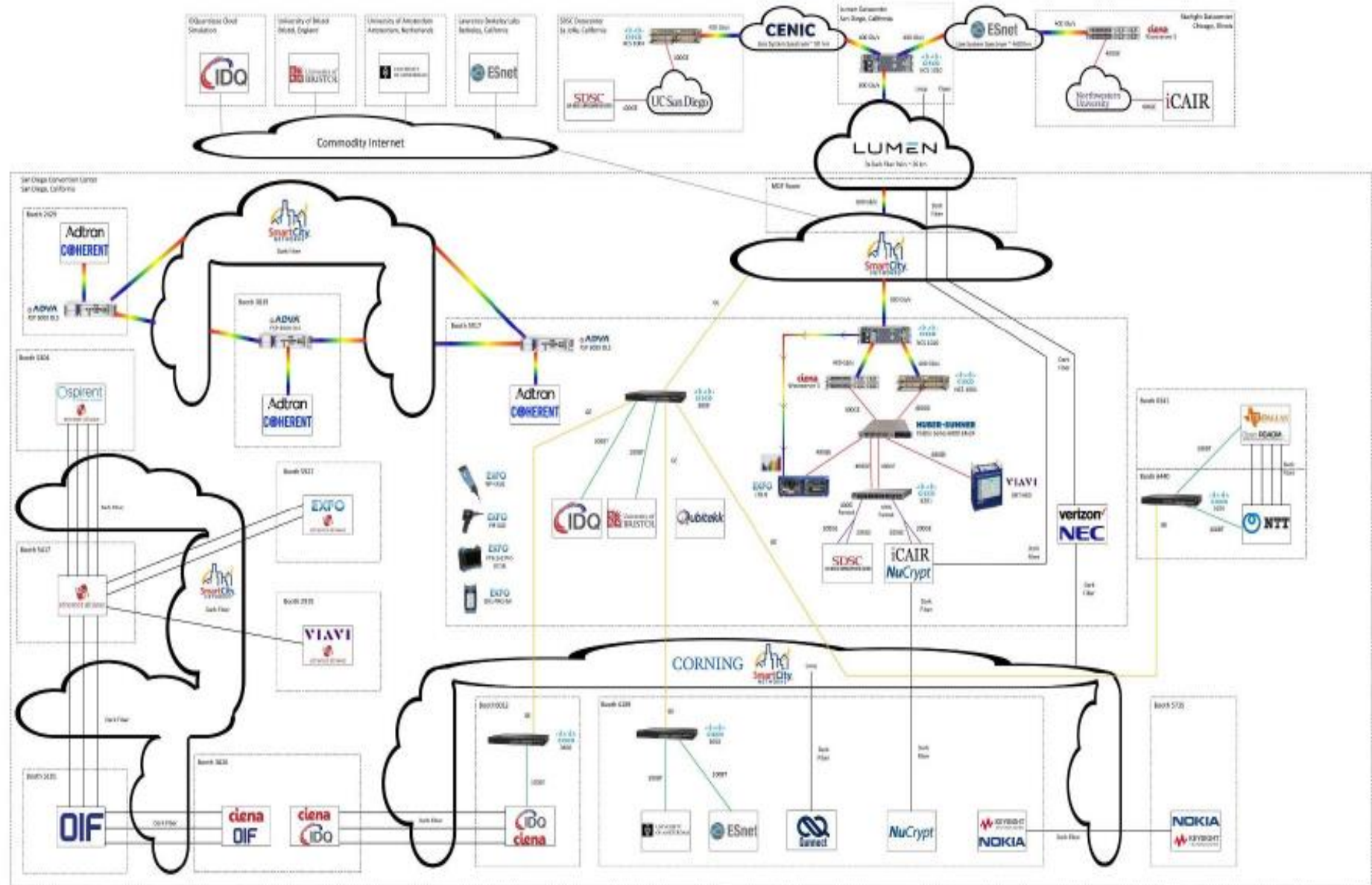


LEGEND

- ↔ via Starlink
- ↔ 100Gbps Link (Existing Transcontinental Cables)



**Support For Large Scale Demonstrations At OFC
San Diego California
March 6-9, 2023**



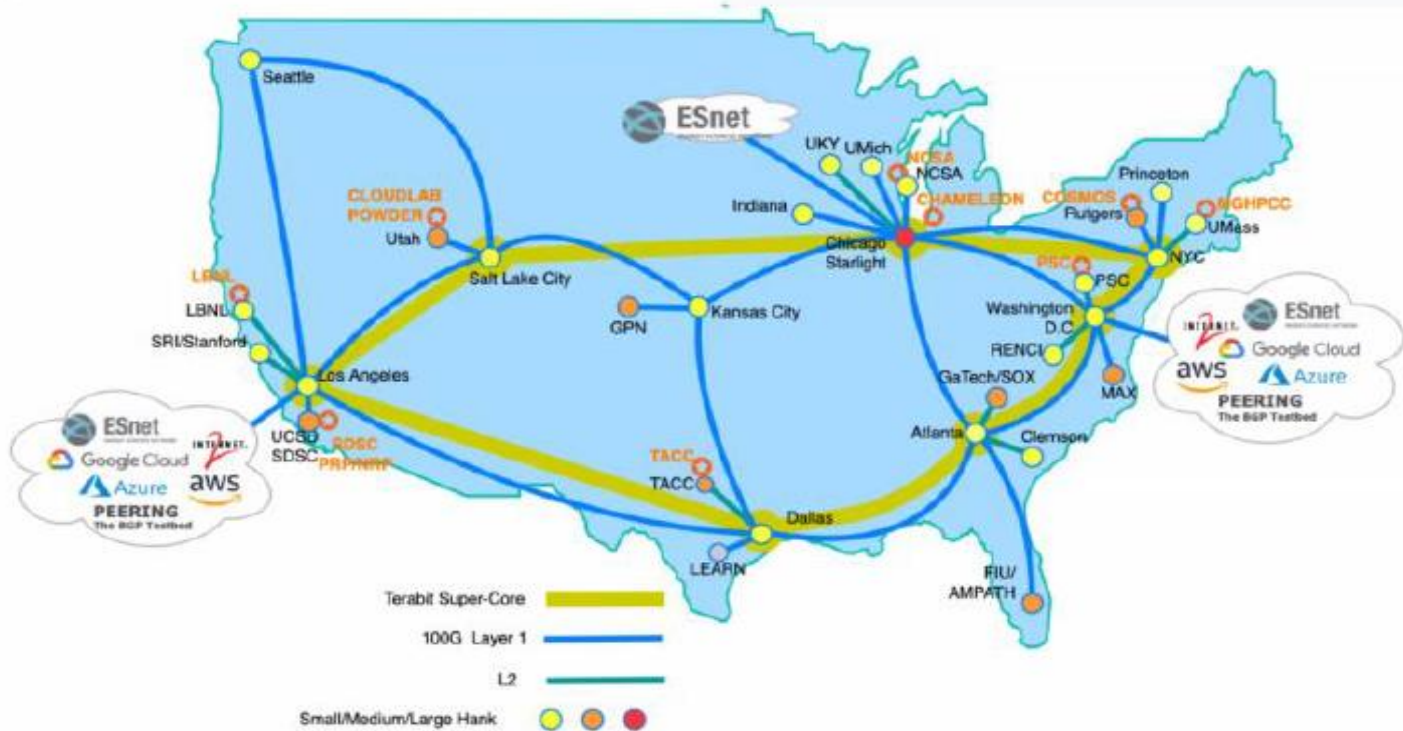
OFC 2023 – OFCnet Architecture Diagram

- Dark Fiber
- 100Base-T
- Gigabit Ethernet
- 300 Gigabit Ethernet
- 400 Gigabit Ethernet
- CWDM
- OFCnet Demonstration

Ilya Baldine PI, RENCI: FABRIC

FABRIC Topology Evolution

FABRIC Phase 2



Core = 3*400 Gbps

STARLIGHTSM



www.chameleoncloud.org

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

Kate Keahey

Joe Mambretti, Pierre Riteau, Paul Ruth, Dan Stanzione

SEPTEMBER 28, 2017

1



www.startup.net/starlight

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

STARLIGHTSM