

# LHC Networking SC23 NRE Demonstrations



Edoardo Martelli, Carmen Misa Moreira, Joe Mambretti,  
Bruno Hoefl, Tom Lehman, Shawn McKee, Marian Babik,  
Vitaliy Kondratenko, Tristan Sullivan, Phil Demar, Syed Asid  
Shah, et al,

LHCOPN-LHCONE MEETING #51

UNIVERSITY OF VICTORIA

BRITISH COLUMBIA, CANADA

OCTOBER 18-19, 2023

# Planning for SC23

- ▶ **IEEE/ACM International Conference On High Performance Computing, Networking, Storage, and Analytics, Nov 12-16, 2023 (SC23), Denver, Colorado**
- ▶ **SCinet Sponsored Network Research Exhibition (NRE) Descriptions (Submitted June 1, 2023)**
- ▶ **NRE Submissions Define Demonstrations and SCinet Requirements**
- ▶ **Prelude To Assessment of Required Resources, Including WANs, Edge Devices, Etc**
- ▶ **Results In Design, Configuration and Implementation of Services/Resources**
- ▶ **Process Also Assists With Pre-Conference Staging Facilities**



# NREs: Verifying/Authenticating New Advanced Concepts

- ▶ **Formulating New Architecture, Services, Techniques, Technologies Through Large Scale, WAN Demonstrations**
- ▶ **Proving Concepts With Empirical, Reproducible Experiments**
- ▶ **Creating Prototypes**
- ▶ **Communicating Results To Wide Audiences**
- ▶ **Leveraging Large Scale Testbeds, e.g., Scinet, Other Testbeds**
- ▶ **Contributing To The Design and Implementation of Testbeds**

# Example SC23 SCinet Network Research Exhibitions

- ▶ **Global Research Platform (GRP)**
- ▶ **SDX 1.2 Tbps WAN Services**
- ▶ **SDX E2E 400 Gbps WAN Services**
- ▶ **400 Gbps DTNs & Smart NICs**
- ▶ **Network Optimized Transport for Experimental Data (NOTED) – With AI/ML Driven WAN Network Orchestration**
- ▶ **SDX International Testbed Integration**
- ▶ **StarLight SDX for Petascale Science**
- ▶ **DTN-as-a-Service For Data Intensive Science**
- ▶ **P4 Integration With Kubernetes**
- ▶ **PetaTrans Services Based on NVMe-Over-Fabric**
- ▶ **NASA Goddard Space Flight Center HP WAN Transport Services**
- ▶ **Resilient Distributed Processing & Rapid Data Transfer**
- ▶ **PRP/NRP Demonstrations**
- ▶ **Open Science Grid Demonstrations**
- ▶ **N-DISE Named Data Networking for Data Intensive Science**
- ▶ **Orchestration With Packet Marking (SciTags)**
- ▶ **Data Tsunami**

# The GRP: A Platform For Global Science



## GLOBAL RESEARCH PLATFORM

*A Next Generation, Software Defined,  
Globally Distributed, Multi-Domain  
Computational Science Environment*

*"The global advancement of science by realizing a multiresource infrastructure through international collaboration."*

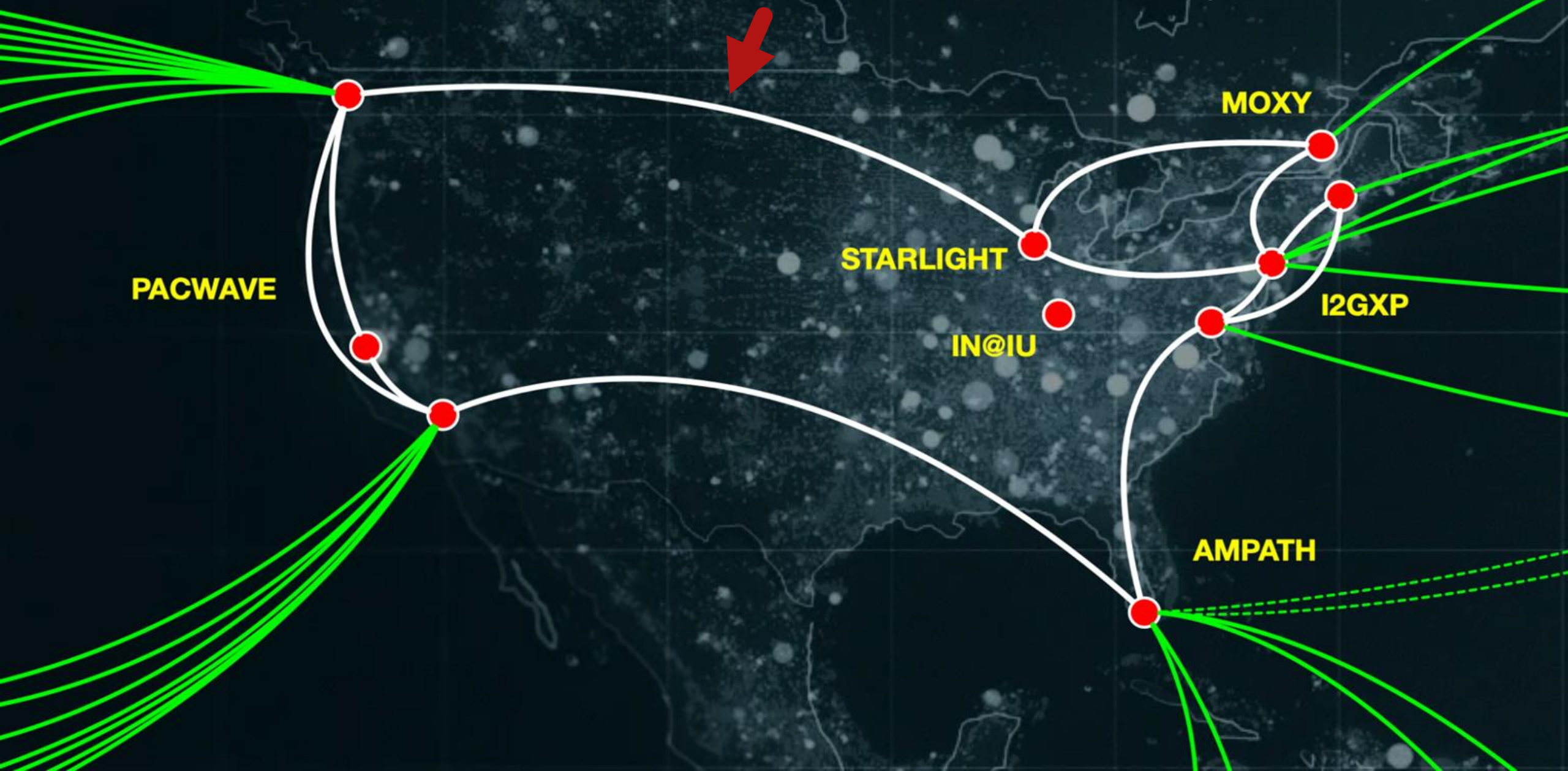


Schematic overview of the GNA-G AutoGOLE

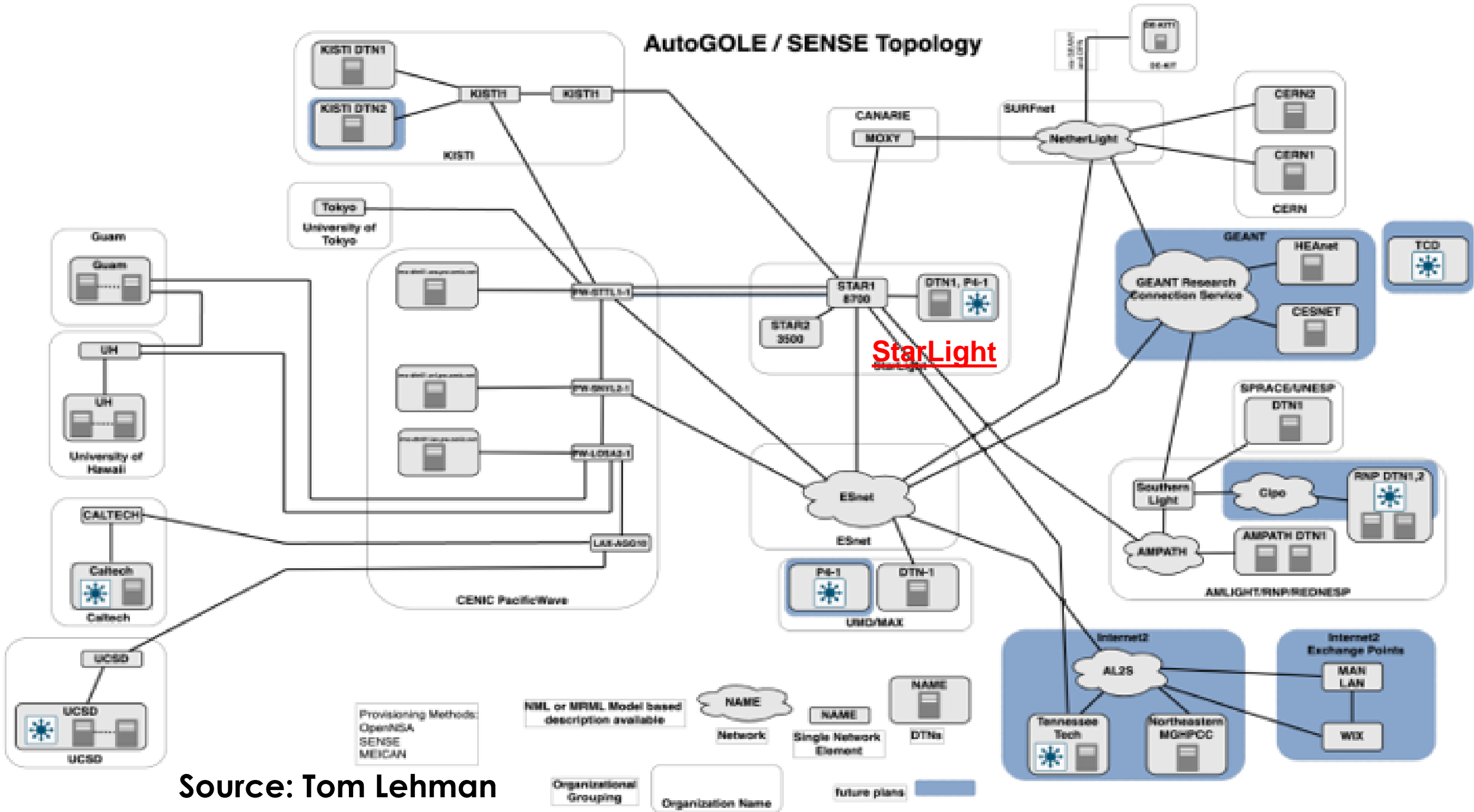


# Global Research Platform/AutoGOLE Open R&E Exchanges

# NA-REX – 400 Gbps WAN Prototype = SC23 NRE, Supporting NOTED



# AutoGOLE / SENSE Topology



Source: Tom Lehman



# MEICAN: AutoGOLE front-end UI

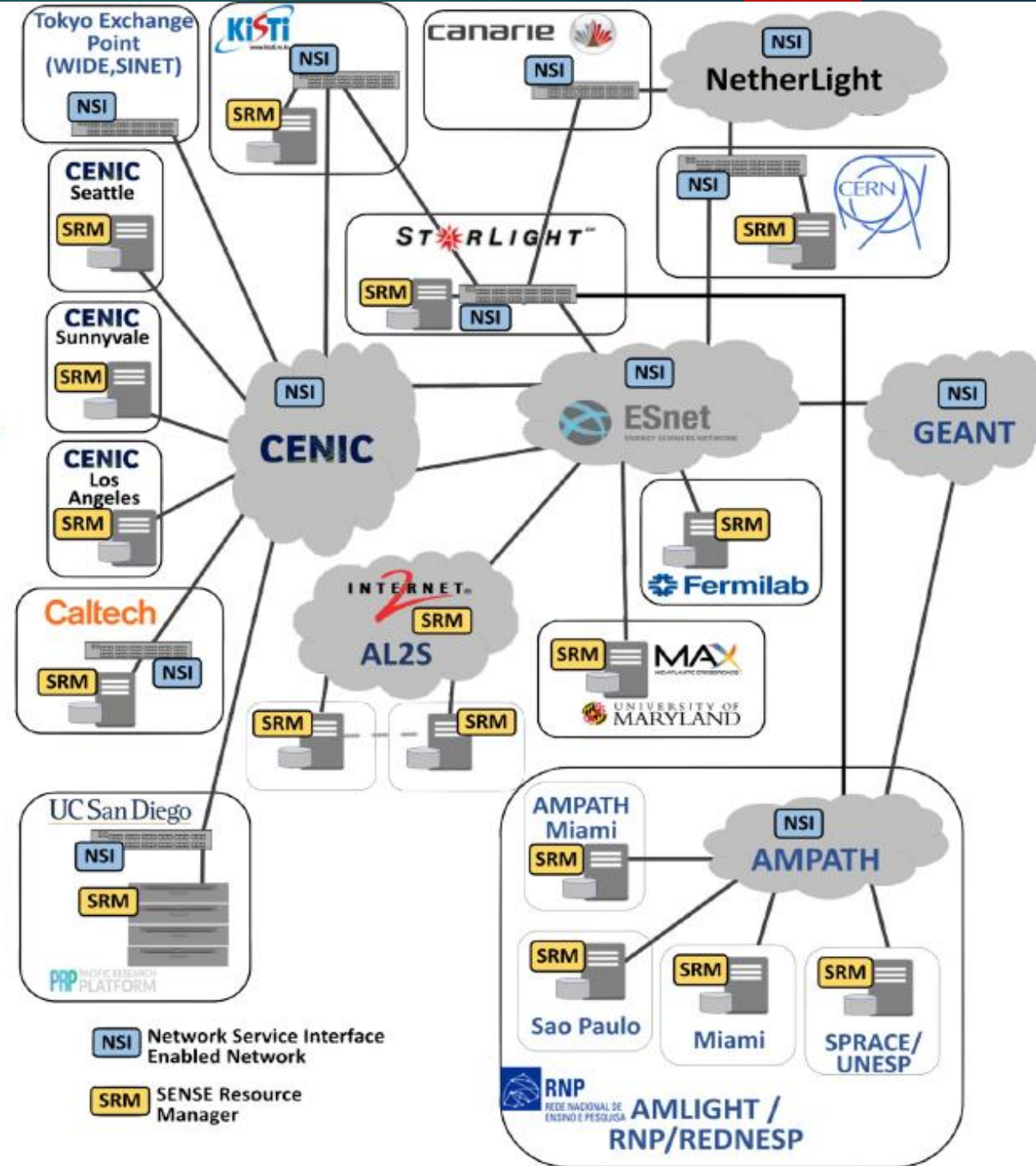


Management Environment of Inter-Domain Circuits for Advanced Networks

# SENSE/AutoGole

- AutoGOLE, NSI, and SENSE working together provide the mechanisms for complete end-to-end services which includes the network and the attached End Systems (DTNs).

Source: Tom Lehman



# SENSE provisioning system

SENSE (SDN for E2E Networked Science at the Exascale): provision system that dynamically builds end-to-end virtual guaranteed networks across administrative domains without manual intervention.

- ❑ Provisioning automation: bring-up and management of services without human involvement.
- ❑ Multi-domain: multiple administrative domains, independent policies and AUP (Acceptable Use Policy).
- ❑ Resource orchestration: allocation and reservation of resources including compute, storage and network.
- ❑ End-to-end: DTN NIC to DTN NIC, across Science DMZ (Demilitarized zone), WANs, Open exchange points...

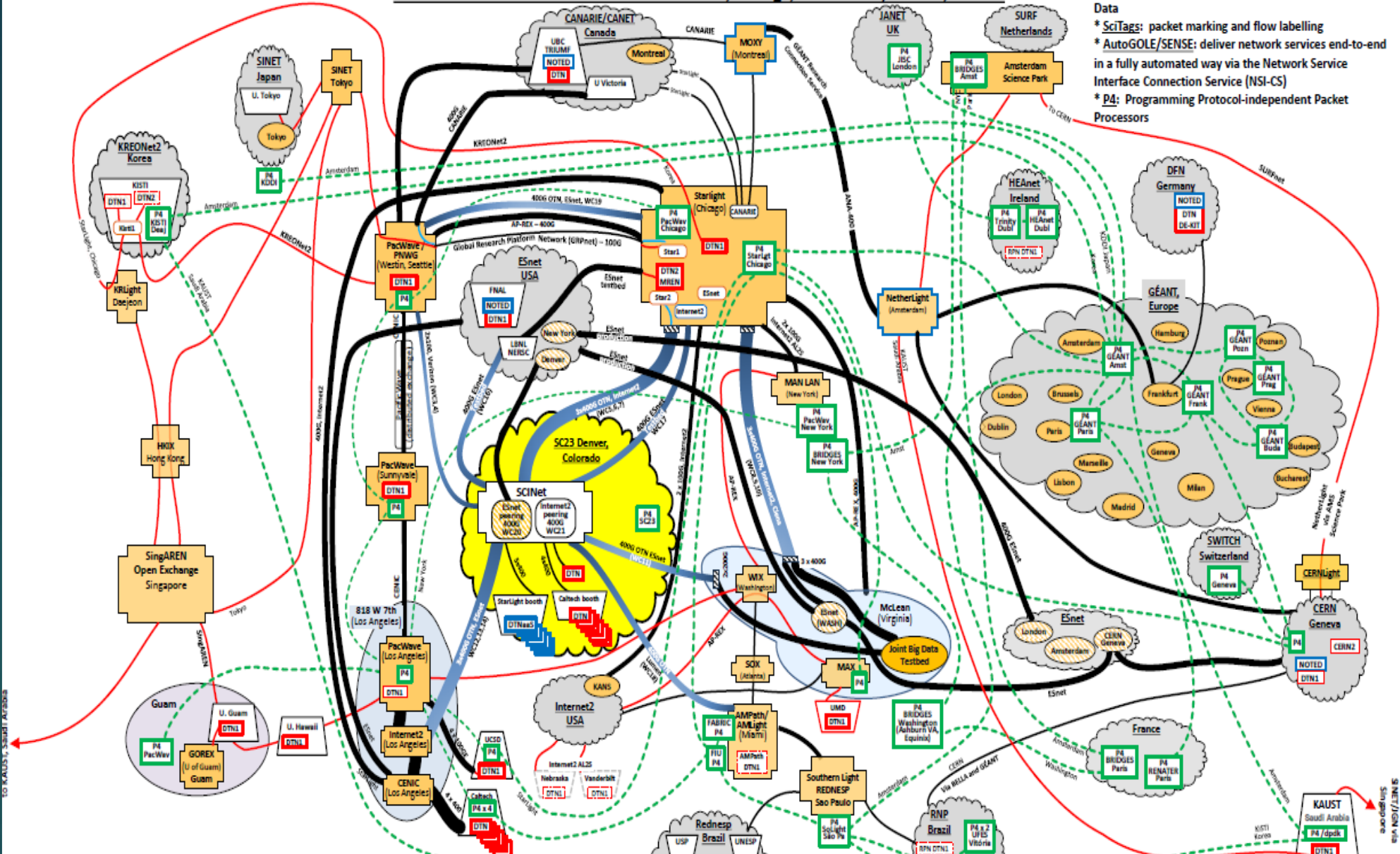


**ESnet**

ENERGY SCIENCES NETWORK

# SC23 Network Research Exhibitions NOTED, SciTags, AutoGOLE / SENSE, and P4

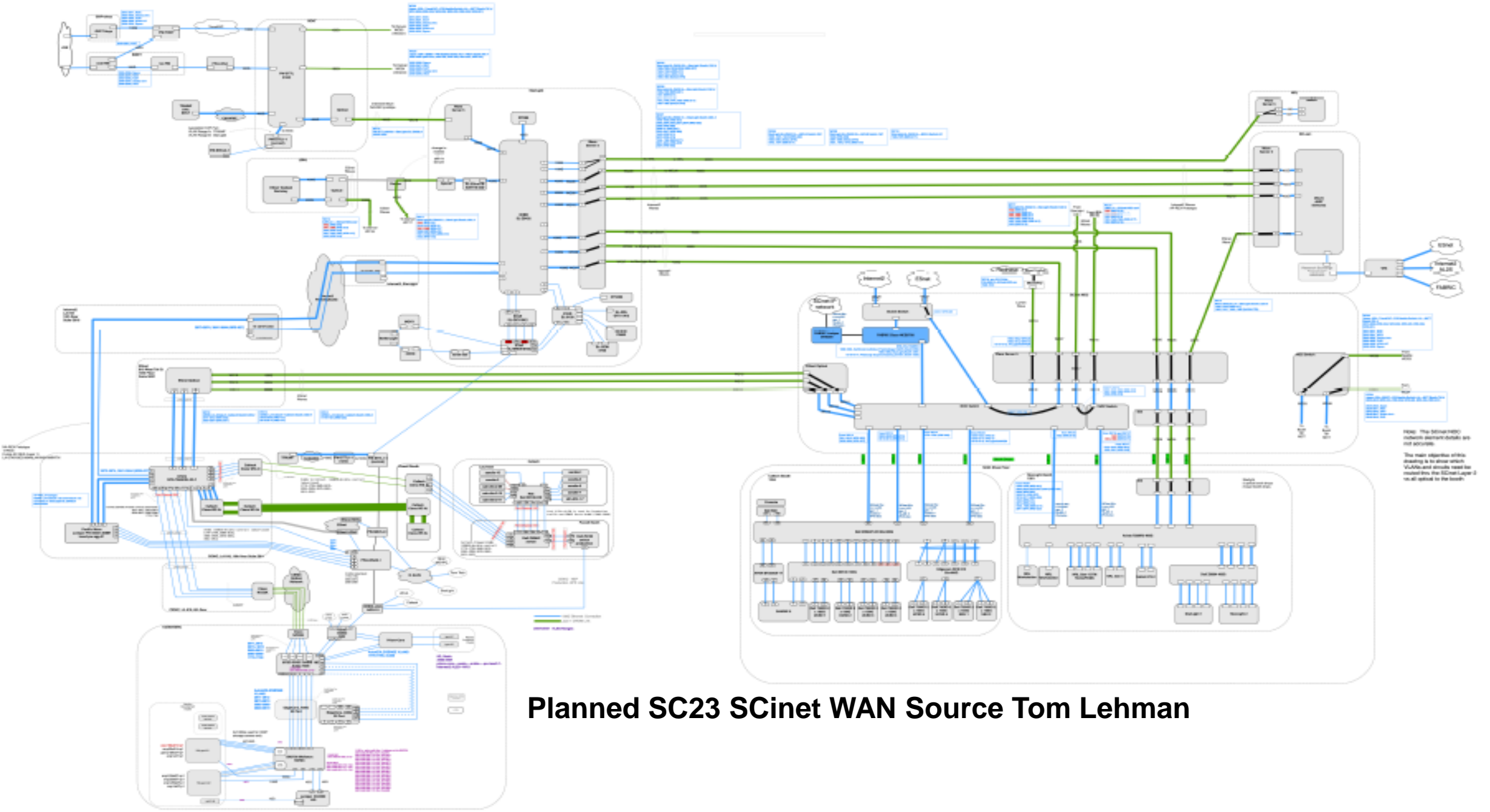
\* **NOTED**: Network-Optimized Transfer of Experiment Data  
 \* **SciTags**: packet marking and flow labelling  
 \* **AutoGOLE/SENSE**: deliver network services end-to-end in a fully automated way via the Network Service Interface Connection Service (NSI-CS)  
 \* **P4**: Programming Protocol-independent Packet Processors



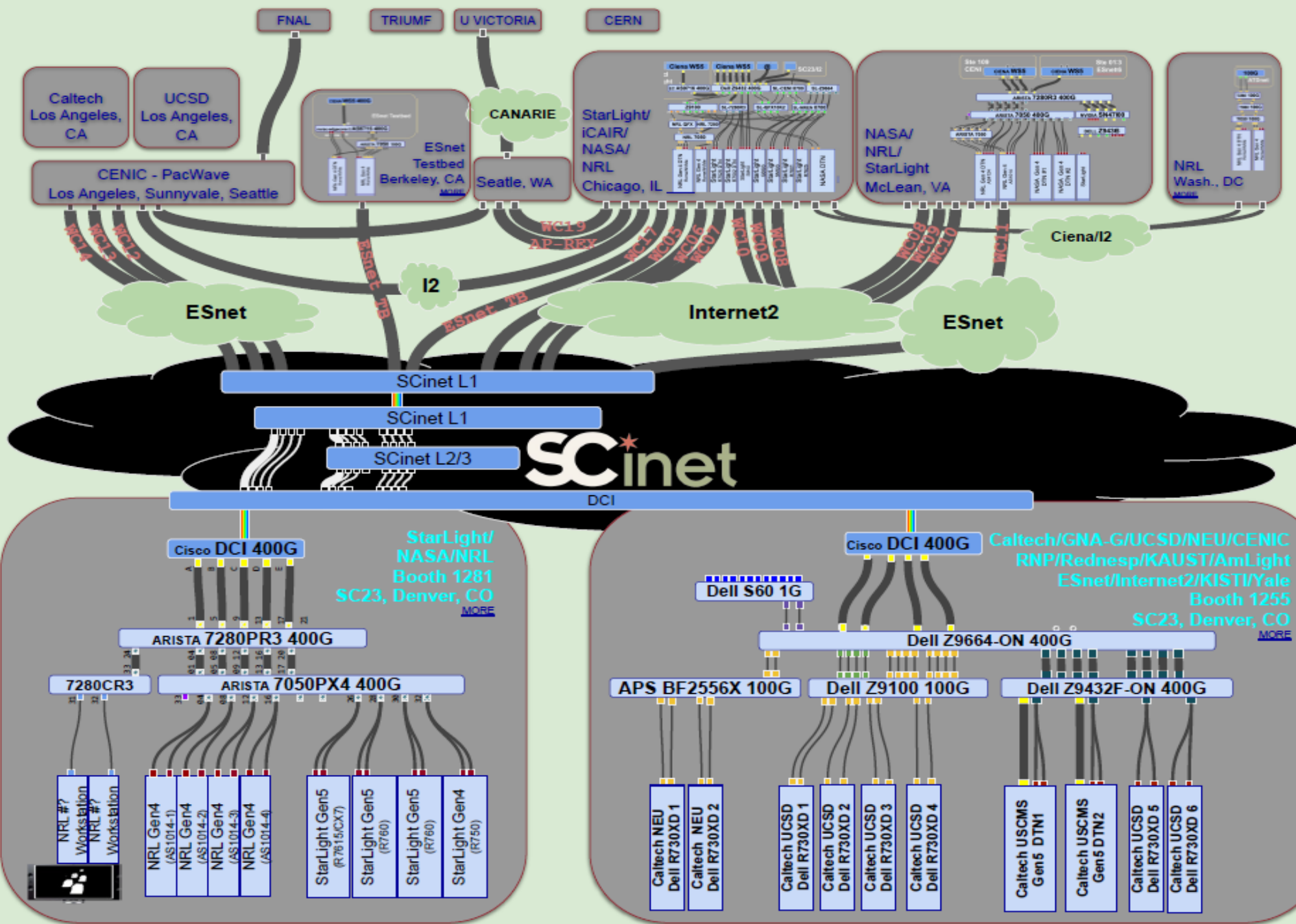
SC23 NRE map v. 11, 2023-10-10 – WEJohnston, ESnet, [wej@es.net](mailto:wej@es.net)

<b>NOTED</b>	SC21 NOTED infrastructure is in blue		ESnet PoPs with High-Touch (line-rate, per packet) monitoring		McLean	Carrier hotels, etc		Regional/national networks: Internal connectivity is assumed to be hi-bw, full mesh.
<b>AG/SENSE</b>	AutoGOLE / SENSE infrastructure is in red		Shared circuits supporting demonstrations		Paris	Dials are points of presence in regional infrastructure		100G
<b>P4</b>	P4 infrastructure is in green		Circuits supporting AutoGOLE / SENSE		ESnet	Rounded rectangles are individual switch/router		200G
<b>general</b>	Shared or general infrastructure is in black		P4 connectivity (not particular circuit infrastructure or bandwidth)		Calltech	Sites		400G
					SOX	Exchange points (external or internal to a site)		800G
					UCSD			1 Td/s
					Calltech			SCINet managed

- NOTES**
- 1) Within exchange points, etc, line width does not usually indicate bandwidth
  - 2) Map files (JPEG, PDF, and PPTX) are at <https://www.dropbox.com/sh/pzwoyypvubek17q/AAAMGFS08xUfQsp3pR1xLta?dl=0>
  - 3) P4 connections are only topological and are not associated with particular network links



Planned SC23 SCinet WAN Source Tom Lehman



**SC23**  
Denver, CO | i am hpc.

**JOINT  
BIG  
DATA  
TESTBED**

- 400G - LR4
- 400G - FR4
- 400G - DAC
- 200G - SR4 or DAC
- 100G - CWDM4
- 100G - LR4
- 100G - SR4
- 100G - DAC
- 40G - SR4
- 40G - DAC
- 10G
- 1G

10/16/2023

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

SC23 floorplan

SC22

SC21

SC19

OPEN items for booth in callouts:

This network diagram in draw.io:  
[https://app.diagrams.net/#G1-6JMIhZM MY1u-rCUC2Ey9W2Hs1O1i\\_A](https://app.diagrams.net/#G1-6JMIhZM MY1u-rCUC2Ey9W2Hs1O1i_A)

Based on a guess for booth drops - need to get actual assignments

- A WC05
- B WC06
- C WC07
- D WC11
- E WC16/17

Need an Optics/Cables count for the booth - here's a start:  
 5 - 400G FR4 OSFP  
 5 - 400G DAC OSFP-OSFP  
 9 - 400/2x200G DAC OSFP-QSFP  
 2 - 100G SR4 to Workstations?  
 2 - 10G LR

We need this 7280SE to start down to 10G  
 s like the 7060 has a couple P ports so I removed the SE

Please go here to see VLAN info, etc

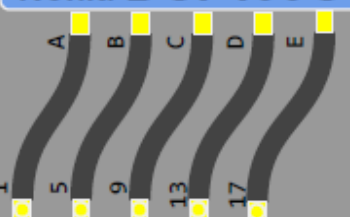
[JBDT team VLAN spreadsheet](#)

[Tom's folder](#)

[Tom's diagram](#)

[Tom's VLAN spreadsheet](#)

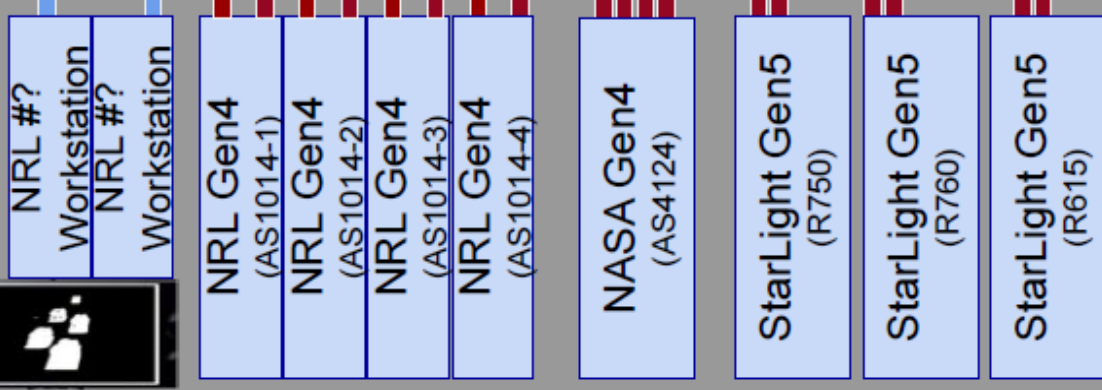
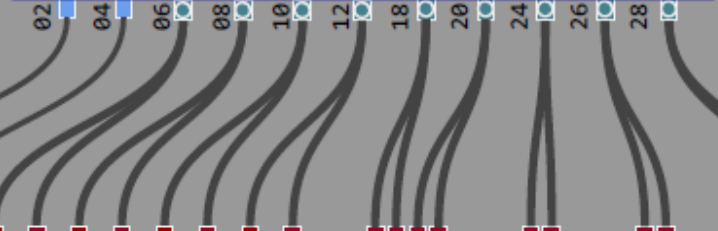
Nokia DCI 400G



ARISTA 7280R3 400G



ARISTA 7060 400G



ARISTA 7010T 1G

9/21/23	
24	Arista 7010T MGMNT
23	NRL AS1014
22	NRL AS1014
21	NRL AS1014
20	NRL AS1014
19	
18	
17	7280PR3-24
16	7060PX4-32
15	
14	Z9664
13	
12	SL-R750
11	
10	SL-R760
9	
8	SL-R615
7	
6	
5	
4	
3	
2	
1	
24U rack	

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

StarLight/  
NASA/  
NRL

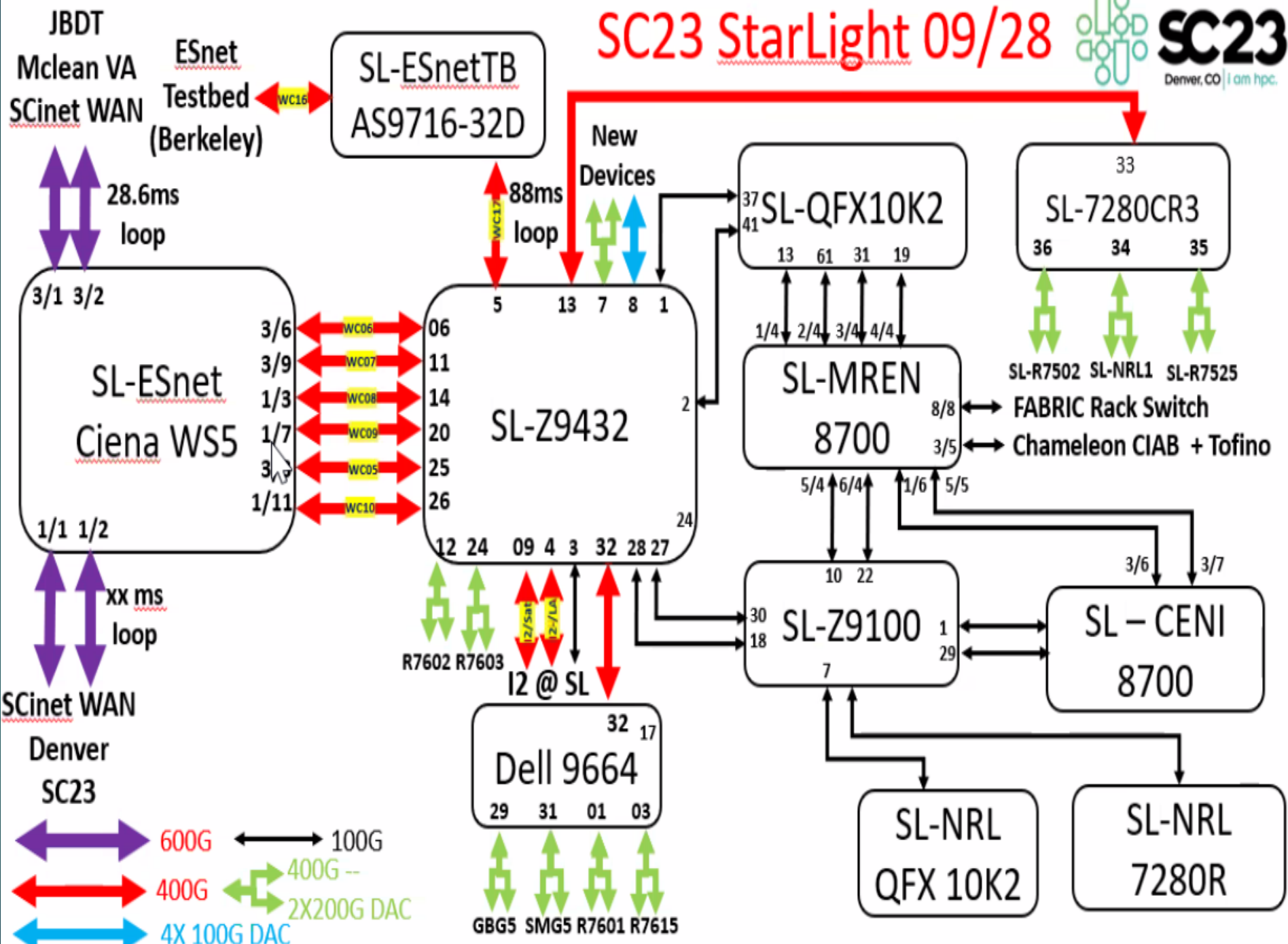
Booth  
1281  
SC23

7010T ports xx & xy for internet connections

09/19/2023

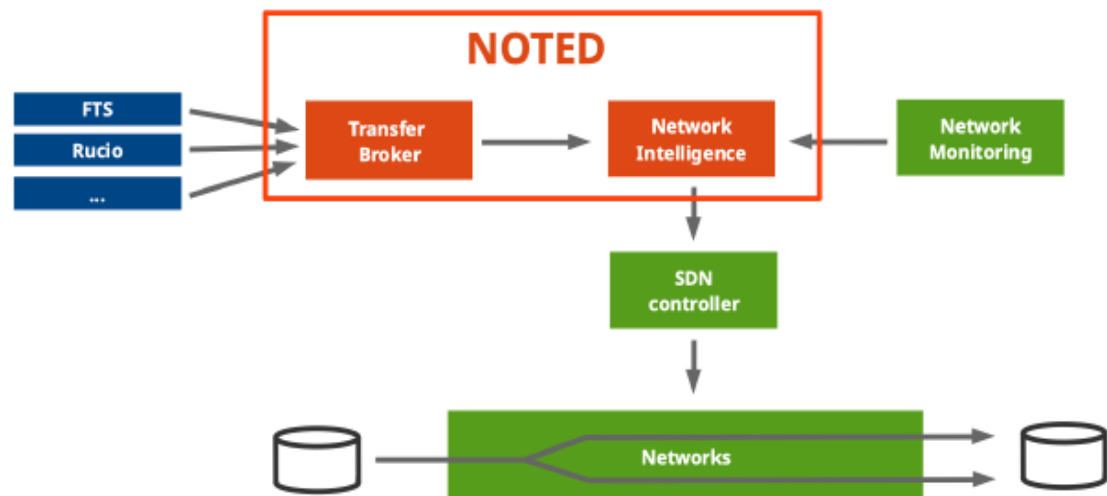
SC22-JBDT

# SC23 StarLight 09/28





## SKELETON AND ELEMENTS OF NOTED



FTS (File Transfer Service):

- Inspect and analyse data transfers to estimate if an action can be applied to optimise the network utilization → get on-going and queued transfers.

CRIC (Computing Resource Information Catalog):

- Enrichment to get an overview and knowledge of the network topology → get IPv4/IPv6 addresses, endpoints, rcsite and federation.

## FLOWCHART AND DATASET STRUCTURE

- Input parameters: configuration given by the user
  - In noted/config/config.yaml → define a list of {src\_rcsite, dst\_rcsite}, maximum and minimum throughput threshold, SENSE/AutoGOLE VLANs UUID and user-defined email notification among others.
- Enrich NOTED with the topology of the network:
  - Query CRIC database → get endpoints that could be involved in the data transfers for the given {src\_rcsite, dst\_rcsite} pairs.
- Analyse on-going and upcoming data transfers:
  - Query FTS recursively → get on-going data transfers for each set of source and destination endpoints.
  - The total utilization of the network is the sum of on-going and upcoming individual data transfers for each source and destination endpoints for the given {src\_rcsite, dst\_rcsite} pairs.
- Network decision:
  - If NOTED interprets that the link will be congested → provides a dynamic circuit via SENSE/AutoGOLE.
  - If NOTED interprets that the link will not be congested anymore → cancel the dynamic circuit and the traffic is routed back.

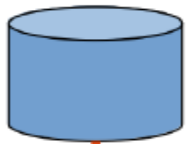
# Components and participants

## Components:

- ❑ NOTED controller and FTS at CERN.
- ❑ NOTED controller at KIT.
- ❑ Data storage at CERN, TRIUMF, KIT.
- ❑ AutoGOLE/SENSE circuits between CERN-TRIUMF and KIT-TRIUMF SENSE circuits are provided by ESnet, CANARIE, STARLIGHT, SURF.

## Participants:





Rucio

FTS

NOTED at KIT

NOTED at CERN

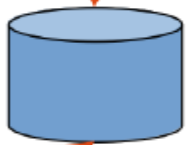
AutoGOLE SENSE

Direct Dynamic Circuit

LHCOPN default path via CERN



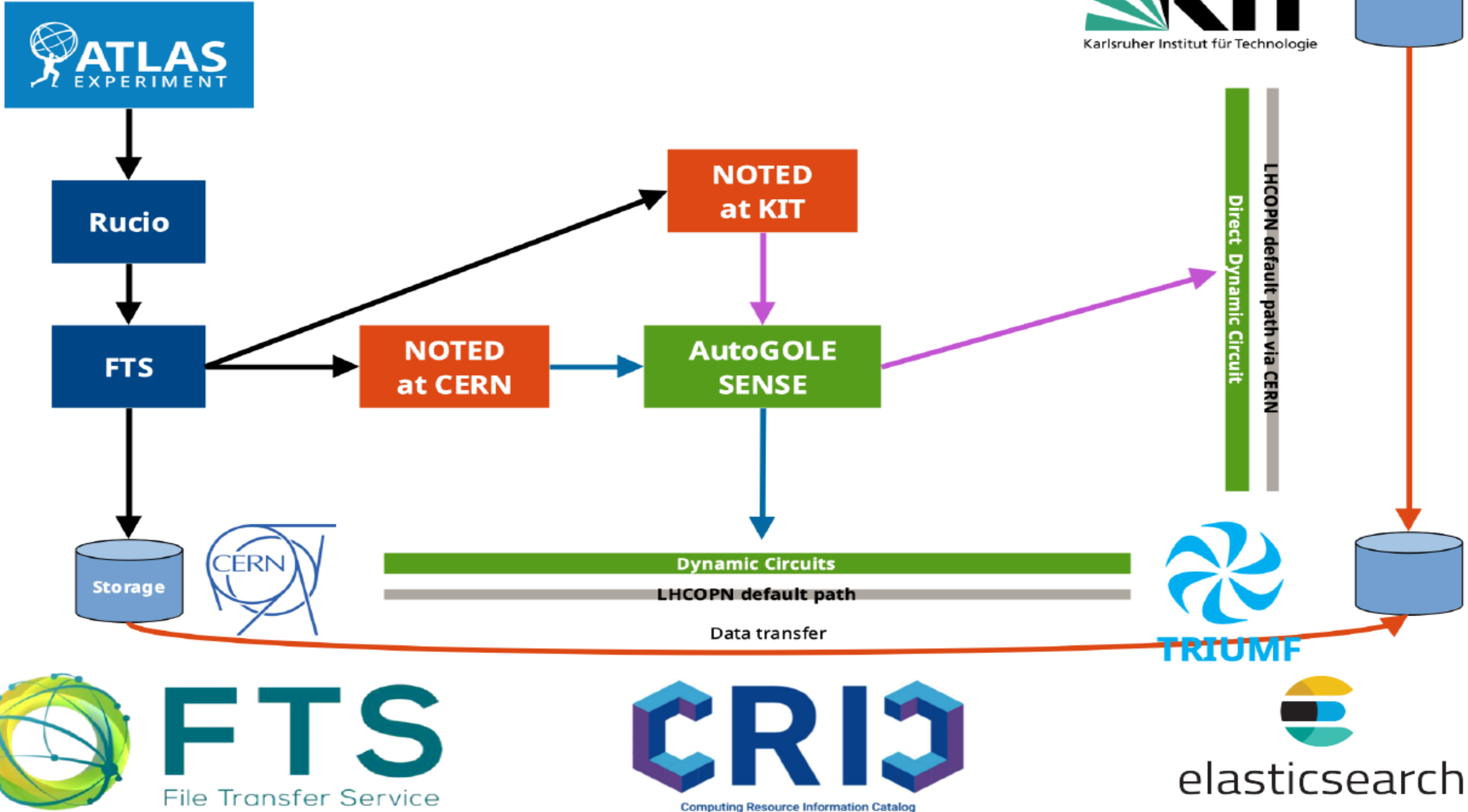
Data transfer



FTS File Transfer Service

CRIQ Computing Resource Information Catalog

elasticsearch

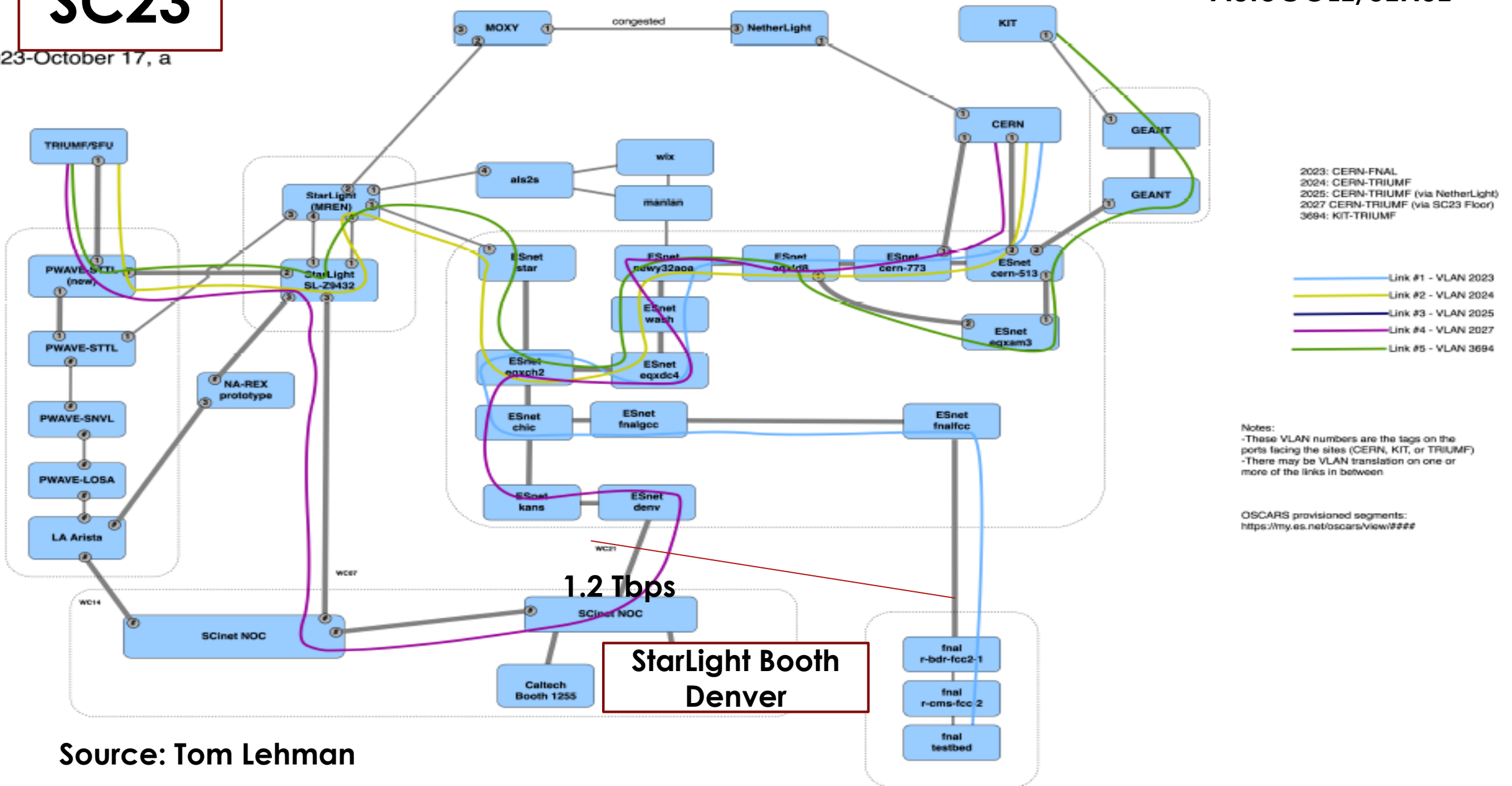


# SC23

2023-October 17, a

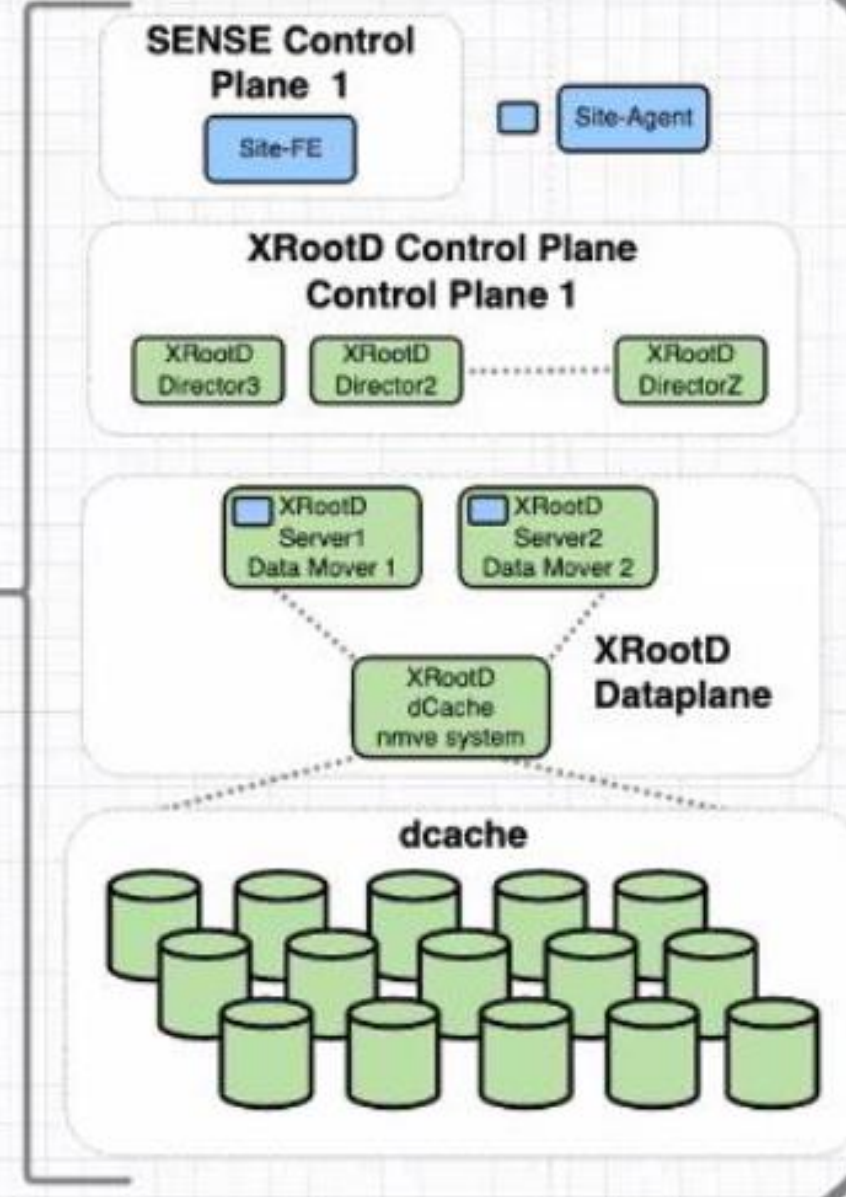
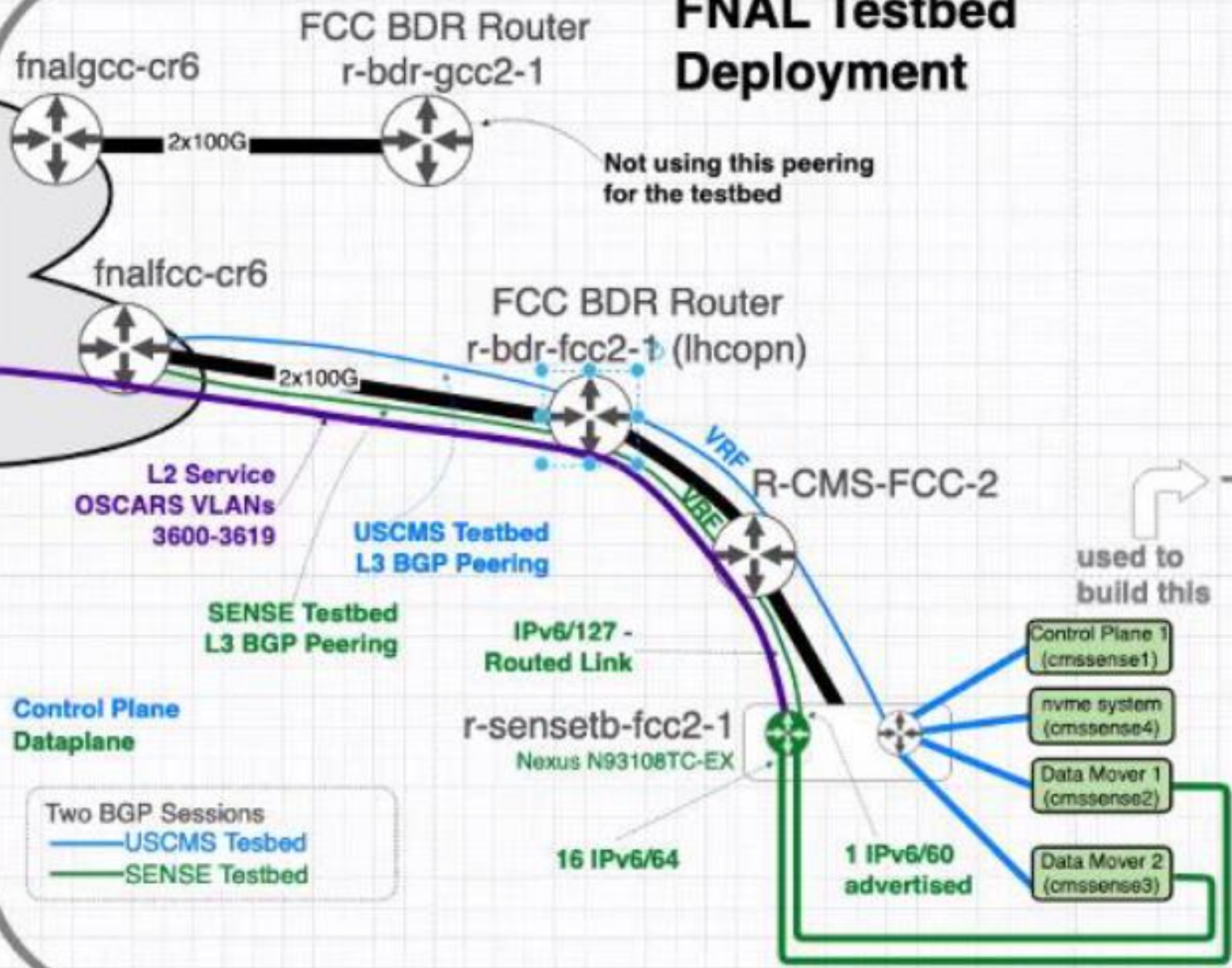
NRE-005, LHC Networking And NOTED

NOTED  
AutoGOLE/SENSE



Source: Tom Lehman

# FNAL Testbed Deployment



# Scitags Initiative

Leads= Shawn McKee, Marian Babik

- **Scientific Network Tags** (scitags) is an initiative promoting identification of the science domains and their high-level activities at the network level.




- Enable tracking and correlation of our transfers with Research and Education Network Providers (R&Es) network flow monitoring
- Experiments can better understand how their network flows perform along the path
  - Improve visibility into how network flows perform (per activity) within R&E segments
  - Get insights into how experiment is using the networks, get additional data from R&Es on behaviour of our transfers (traffic, paths, etc.)
- Sites can get visibility into how different network flows perform
  - Network monitoring per flow (with experiment/activity information)
    - E.g. RTT, retransmits, segment size, congestion window, [etc.](#) all per flow

# SC23 Packet/Flow Marking NRE

- ▶ **Concept: The Goals of the SC23 Packet and Flow Marking NRE Demonstrations Will Be Building On the SC22 Demonstrations To Showcase The Capabilities of The Scitags Architecture And Methods For Optimizing Data Intensive Science**
- ▶ **Five Demonstrations Will Be Staged**
  - ▶ IPv6 Packet Marking With eBPF-TC (100 Gbps)
  - ▶ XRootD Packet Marking with Flowd+eBPF-TC
  - ▶ Accounting For Flow Labeled Packets Using a P4 Programmable Switch
  - ▶ Measurements via Esnet High-Touch Processes
  - ▶ Scitags Integration With DTN-as-a-Service.
- ▶ **Participants:**
  - ▶ CERN, University of Victoria, KIT, ESnet, StarLight, CANARIE, Fermi National Accelerator Laboratory, SCInet, Digital Alliance, etc

# Booth Posters Being Made As With SC22 Shown Here




## NOTED: AN INTELLIGENT NETWORK CONTROLLER TO IMPROVE THE THROUGHPUT OF LARGE DATA TRANSFERS IN FILE TRANSFER SERVICES BY HANDLING DYNAMIC CIRCUITS

Carmen Misa Moreira<sup>1</sup>, Edoardo Martelli<sup>1</sup>, Bruno Hoelt<sup>2</sup>, Vitaliy Kondratenko<sup>3</sup>, Joel J. Mambretti<sup>4</sup>

<sup>1</sup>CERN (Conseil Européen pour la Recherche Nucléaire), IT department CS-NE section, email: firstname.lastname@cern.ch  
<sup>2</sup>KIT Karlsruhe Institute of Technology, Steinbuch Center for Computing, email: Bruno.Hoelt@kit.edu  
<sup>3</sup>TRIUMF ATLAS Tier-1 Computing Center, email: vitaliy.k@triumf.ca  
<sup>4</sup>International Center for Advanced Internet Research, Northwestern University, email: j-mambretti@northwestern.edu

### SKELETON AND ELEMENTS OF NOTED



FTS (File Transfer Services):  
 - Inspect and analyze data transfers to estimate if an action can be applied to optimise the network utilisation → get on-going and queued transfers.

CRIC (Computing Resource Information Catalog):  
 - Enrichment to get an overview and knowledge of the network topology → get IPv4/IPv6 addresses, endpoints, route and utilization.

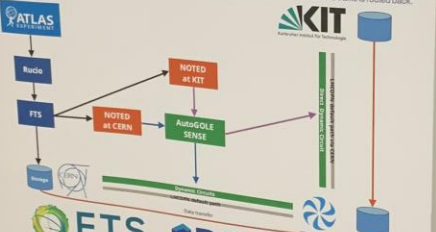
### FLOWCHART AND DATASET STRUCTURE

Input parameters: configuration given by the user  
 In: noted.config.yaml → define a list of (src\_rscale, dst\_rscale), maximum and minimum throughput, threshold, SENSE/AutoGOLE VLANs UUID and user-defined email notification among others.

Enrich NOTED with the topology of the network:  
 Query CRIC database → get endpoints that could be involved in the data transfers for the given (src\_rscale, dst\_rscale) pairs.

Analyse on-going and upcoming data transfers:  
 Query FTS recursively → get on-going data transfers for each set of source and destination endpoints.  
 The total utilization of the network is the sum of on-going and upcoming individual data transfers for each source and destination endpoints for the given (src\_rscale, dst\_rscale) pairs.


Network decision:  
 If NOTED interprets that the link will be congested → provides a dynamic circuit via SENSE/AutoGOLE.  
 If NOTED interprets that the link will not be congested anymore → cancel the dynamic circuit and the traffic is routed back.



### NETWORK UTILIZATION REPORTED BY LHONELHCOOP PRODUCTION ROUTERS

Router	Max	Min	Std
lhonelhcoop-01	20.000	0.000	10.000
lhonelhcoop-02	20.000	0.000	10.000
lhonelhcoop-03	20.000	0.000	10.000
lhonelhcoop-04	20.000	0.000	10.000
lhonelhcoop-05	20.000	0.000	10.000
lhonelhcoop-06	20.000	0.000	10.000
lhonelhcoop-07	20.000	0.000	10.000
lhonelhcoop-08	20.000	0.000	10.000
lhonelhcoop-09	20.000	0.000	10.000
lhonelhcoop-10	20.000	0.000	10.000
lhonelhcoop-11	20.000	0.000	10.000
lhonelhcoop-12	20.000	0.000	10.000
lhonelhcoop-13	20.000	0.000	10.000
lhonelhcoop-14	20.000	0.000	10.000
lhonelhcoop-15	20.000	0.000	10.000
lhonelhcoop-16	20.000	0.000	10.000
lhonelhcoop-17	20.000	0.000	10.000
lhonelhcoop-18	20.000	0.000	10.000
lhonelhcoop-19	20.000	0.000	10.000
lhonelhcoop-20	20.000	0.000	10.000

### NOTED ALERTS CONCERNING NETWORK UTILIZATION OF WLCG SITES IN LHONELHCOOP



NOTED is aware of the network utilization of the WLCG sites in LHONELHCOOP. In orange are shown the alerts generated by NOTED concerning the network utilization, where, a dynamic circuit is produced by SENSE/AutoGOLE as an alternative link to increase the bandwidth and improve the throughput of large data transfers in FTS.

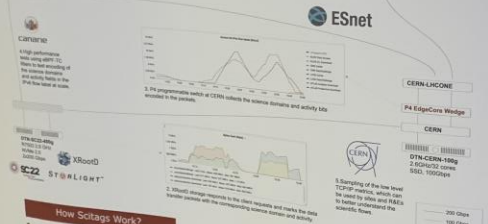


# Scitags Poster SC22

## Scientific Network Tags Packet Marking for Data Intensive Scientific Workflows



Managing large scale scientific workflows over networks is becoming increasingly complex, especially as multiple science projects share the same foundation resources simultaneously yet are governed by multiple divergent variables: requirements, constraints, configurations, technologies, etc. A key method to address this issue is to employ techniques that provide high fidelity visibility into exactly how science flows utilize network resources end-to-end. This demonstration showcases one such method, Scientific Network Tags (Scitags), an initiative that is promoting identification of the science domains and their high-level activities at the network level. This open system initiative provides open source technologies to help Research and Education Networks (REN) understand resource utilization while providing information to scientific communities on the behavior of their network flows.



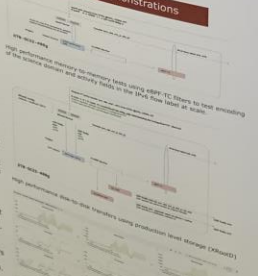
### How Scitags Work?



**Scientific Network Tags (Scitags)** is an initiative promoting identification of the science domains and their high-level activities at the network level.

- Enables tracking and correlation of the data transfers with the REN flows.
- Helps science domains better understand how their network flows perform along the path.
- Improves visibility into network flows performance (per activity) within REN segments.
- Provides insights into how science domains are using behaviour of the transfers (traffic, paths, etc.)
- Since get visibility into how different network flows science domain and activity.

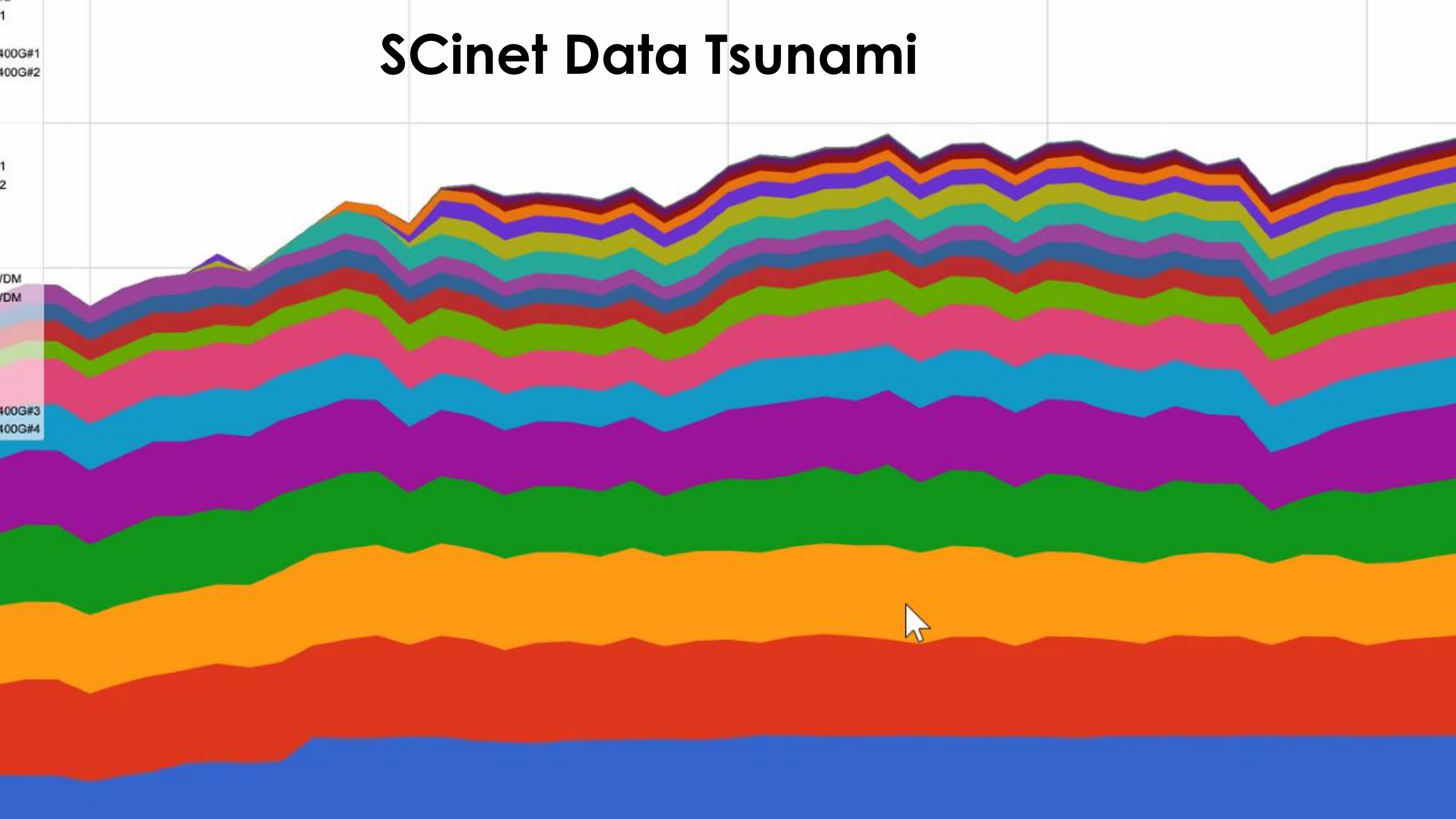
### SC22 Demonstrations



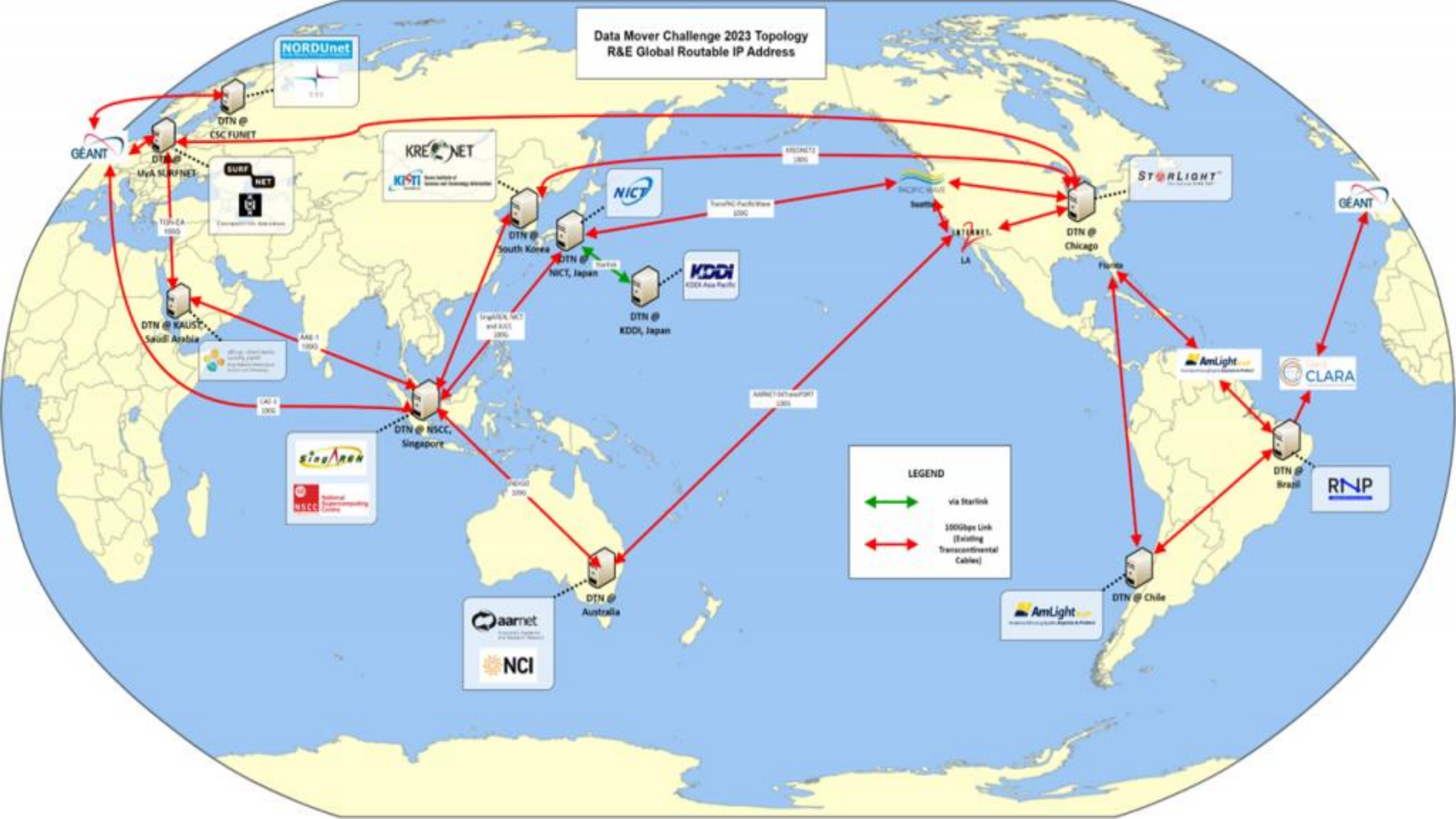
### What's next?

- Deploying and testing XRootD and iCache implementations.
- Organizing validation in collaboration with the WLCG experimental network of functional nodes across a cross-domain network and more networks (Container/Federated).
- Develop functional nodes and storage systems, understand their requirements and work on Scitags implementation.
- Develop available technologies with new sciences that have scientific workflows.
- Expanding available technologies for collecting Scitags at network level and work on SC22.
- Testing and validating ways to generate flow identifiers for network flows and their management systems.
- Work on solving REN informational RFC.

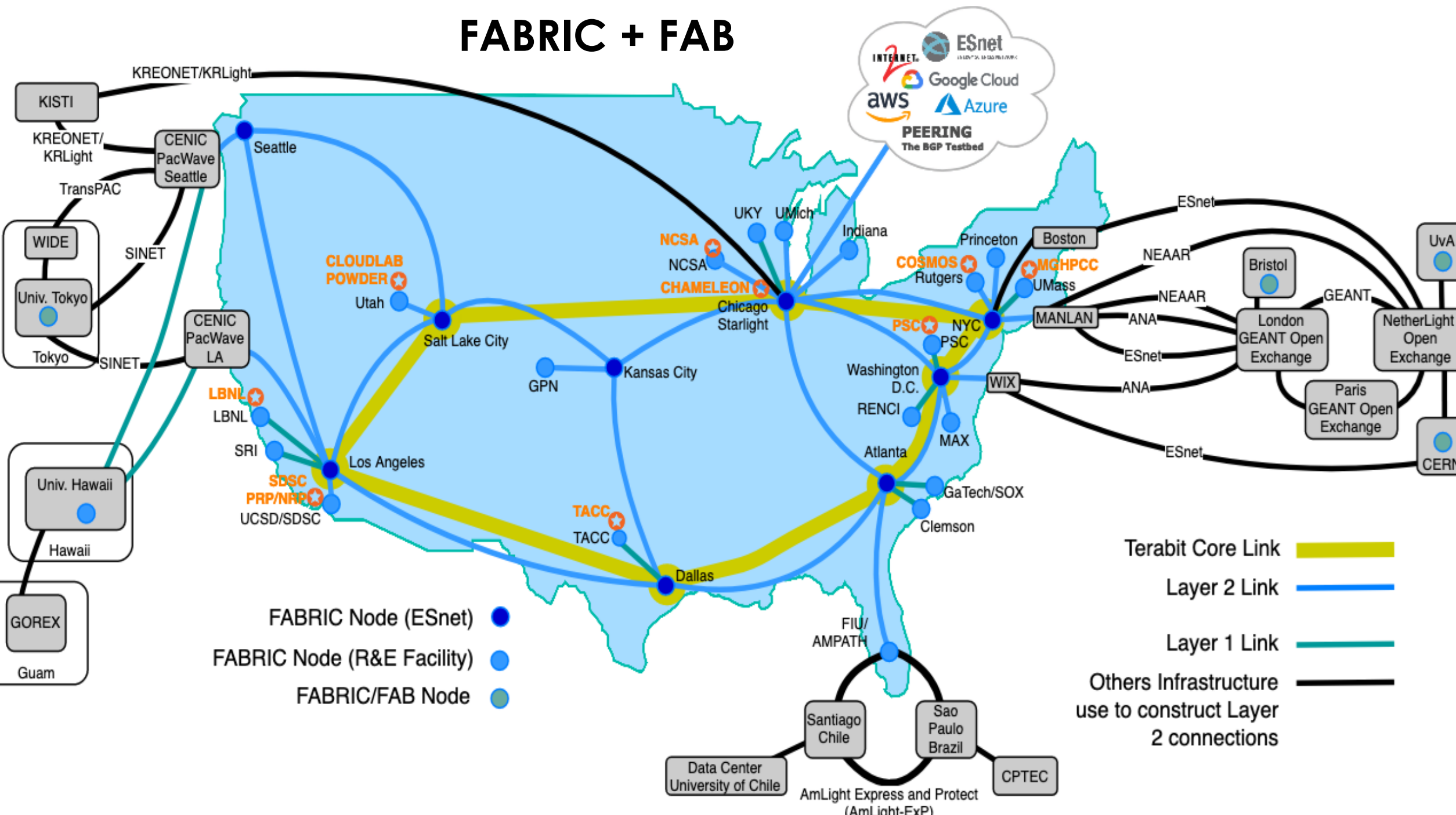
# SCinet Data Tsunami

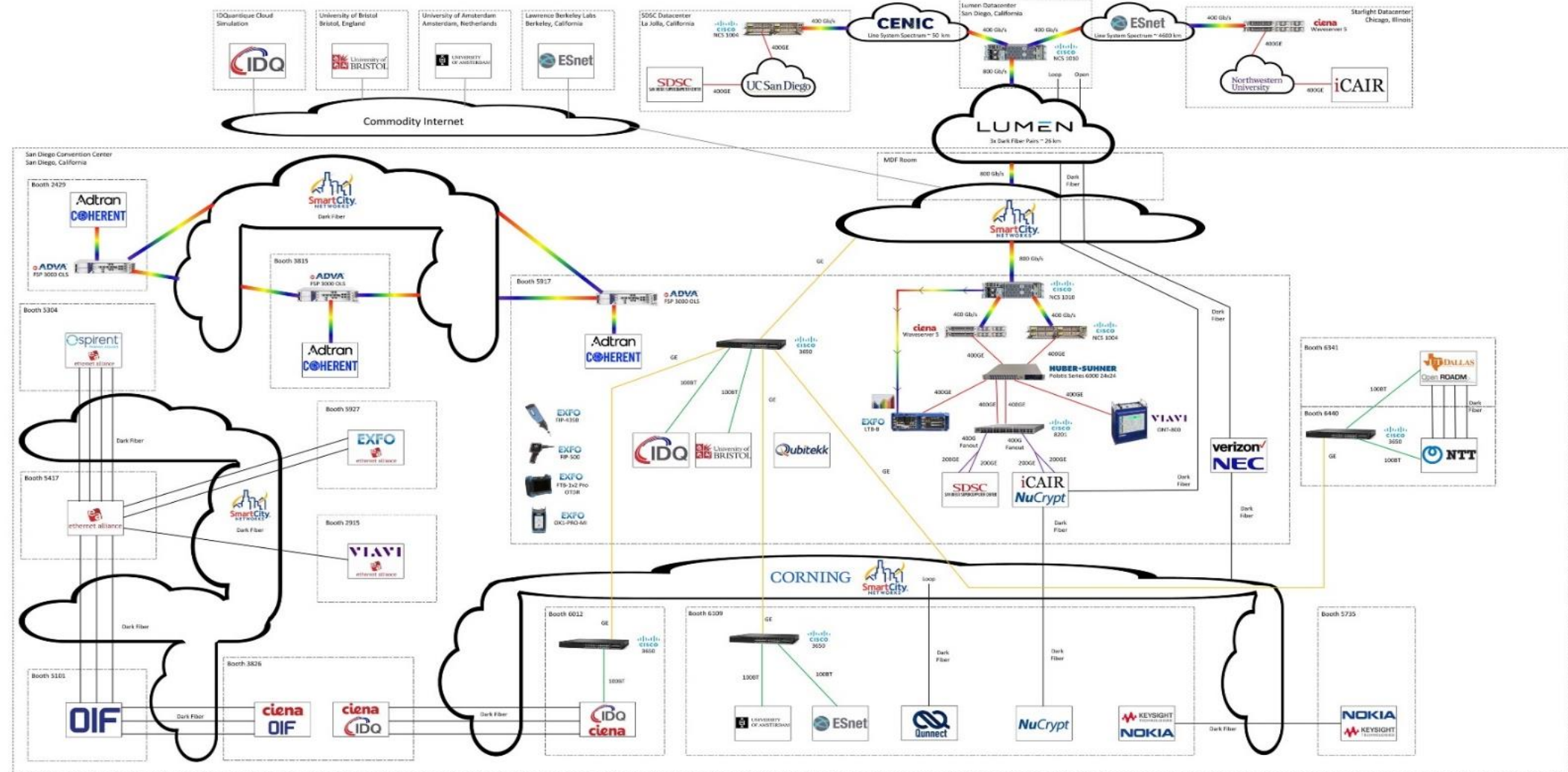


# Data Mover Challenge 2023 Topology R&E Global Routable IP Address



# FABRIC + FAB





## OFC 2023 – OFCnet Architecture Diagram

# Annual Global Research Platform Workshop – Co-Located With IEEE International Conference On eScience Oct 9-10, 2023



CALLS - PROGRAM - TRAVEL

## '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



# Futures

- ▶ Data Challenge 2024
- ▶ Quasi Permanent SCinet Facility Proxy (e.g, Shippable Rack)
- ▶ SC24
- ▶ OFCnet 2024
- ▶ MultiONE
- ▶ Etc

*Thanks!*

▶ Questions?