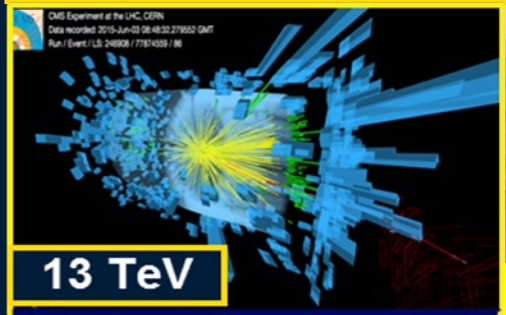
Global Petascale to Exascale Workflows

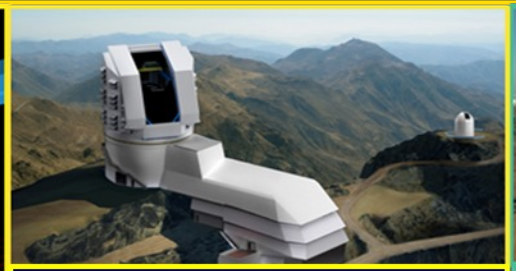
Next Generation Network-Integrated System for Data Intensive Sciences

SC22 Dallas, TX | hpc accelerates.

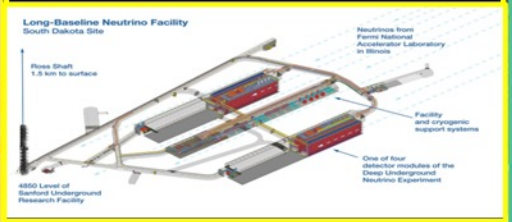
Booth 2820

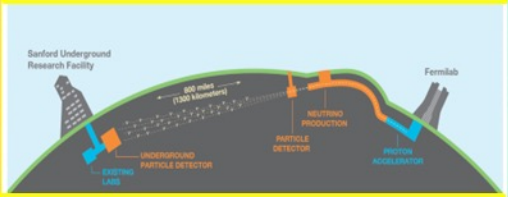


LHC



LSST





LBNF/DUNE

SKA

LHC Run3 and HL-LHC

DUNE

VRO SKA

BioInformatics

Earth Observation

Gateways to a New Era

GNA-G Global Network Advancement Group

SC22 Network Research Exhibition

NRE-19 and Partner NREs: Booth 2820

California Institute of Technology 1891

See https://www.dropbox.com/s/1opcg4vjlhjk6g5/NextGenDISSystems_hbn111222.pptx?dl=0

Global Petascale to Exascale Workflows for Data Intensive Sciences Accelerated by Next Generation Programmable Network Architectures and Machine Learning Applications

Network Research Exhibition NRE-19. Abstract:
https://www.dropbox.com/s/qcm41g7f7etjxvy/SC22_NRE_GlobalPetascaleWorkflows_V8072022.docx?dl=0

- A Vast *Partnership* of Science and Computer Science Teams, R&E Networks and R&D Projects; Convened by the GNA-G DIS WG; with GRP, AmRP, NRP
- Mission: Demonstrate the road ahead
  - To meet the challenges faced by leading-edge data intensive programs in high energy physics, astrophysics, genomics and other fields of data intensive science;
    - ✸ *Compatible with other use*
  - Clearing the path to the next round of discoveries
- Demonstrating a wide range of latest advances in:
  - Software defined and Terabit/sec networks
  - Intelligent global operations and monitoring systems
  - Workflow optimization methodologies with real time analytics
  - State of the art long distance data transfer methods and tools, local and metro optical networks and server designs
  - Emerging technologies and concepts in programmable networks and global-scale distributed systems
- Hallmarks: Progressive multidomain integration; compatibility internal + external; A comprehensive systems-level approach

GNA-G DIS WG: Worldwide Partnerships at SC22

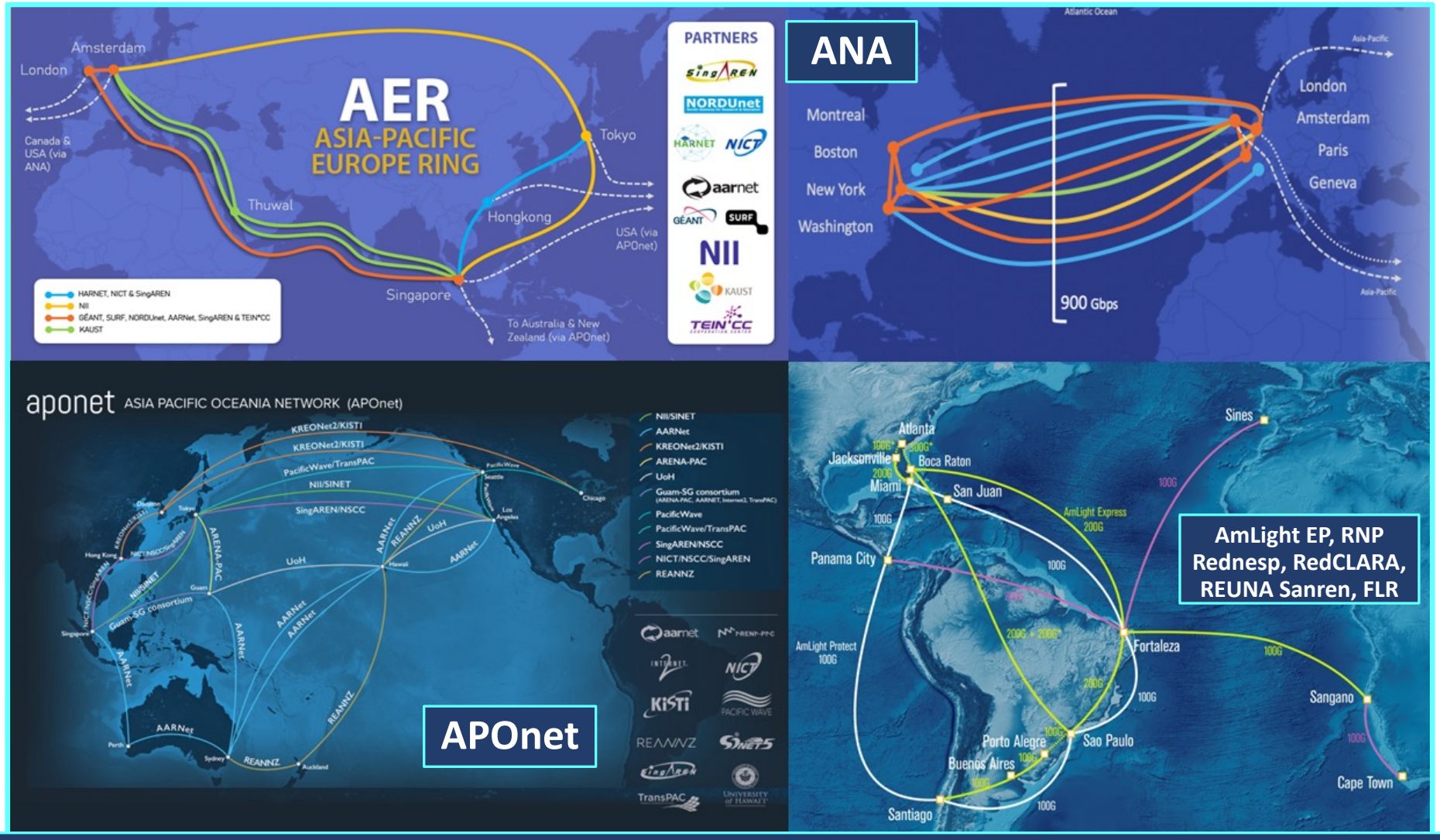
Partners: **Group Leads and Participants, by Team**

- **Caltech HEP:** Harvey Newman (newman@hep.caltech.edu), Justas Balcas (jbalcas@caltech.edu), Raimondas Sirvinskas (raimis.sirvis@gmail.com), Catalin Iordache, Preeti Bhat, Andres Moya, Sravya Uppalapati
- **Caltech IMSS:** Jin Chang (jin.chang@caltech.edu), Azher Mughal (azher@caltech.edu), Dawn Boyd, Larry Watanabe, Don S. Williams
- **UCSD/SDSC/NRP:** Frank Wuerthwein (fkw888@gmail.com), Tom deFanti (tdefanti@eng.ucsd.edu), Larry Smarr, John Graham, Tom Hutton (hutton@ucsd.edu), Dima Mishin, Jonathan Guiang, Diego Davila, Igor Sfiligoi, Aashay Arora,
- **Yale:** Richard Yang (yry@cs.yale.edu), Jensen Zhang
- **Northeastern University:** Edmund Yeh (eyeh@ece.neu.edu), Yuanhao Wu, Volkan Mutlu, Yuezhou Liu
- **Tennessee Tech:** Susmit Shannigrahi (sshannigrahi@tntech.edu), Sankalpa Timilsina
- **UCLA:** Lixia Zhang (lixia@cs.ucla.edu), Jason Cong (cong@cs.ucla.edu), Michael Lo, Sichen Song
- **Fermilab:** Oliver Gutsche (gutsche@fnal.gov), Phil Demar (demar@fnal.gov)
- **ESnet:** Inder Monga (imonga@es.net), Chin Guok (chin@es.net), Tom Lehman (tlehman@es.net), John MacAuley, Xi Yang, Justas Balcas, Mariam Kiran
- **LBNL/NERSC:** Alex Sim (asim@lbl.gov)
- **Nebraska/UNL:** Garhan Attenbury (garhan.attebury@unl.edu)
- **Vanderbilt:** Andrew Melo, andrew.m.melo@accre.vanderbilt.edu
- **CERN:** Edoardo Martelli (edoardo.martelli@cern.ch), Carmen Misa (carmen.misa@cern.ch)
- **Qualcomm Gradient Graph:** Jordi Ros-Giralt (jros@qti.qualcomm.com), Sruthi Yellamraju
- **UFES:** Magnos Martinello, Moises R.N. Ribeiro (moises@ele.ufes.br), Christina Dominicini (cristina.dominicini@ifes.edu.br), Everson Borges (everson@ifes.edu.br), Rafael Guimaraes
- **RNP:** Marcos Schwarz (marcos.schwarz@rnp.br), Leandro Ciuffo (leandro.ciuffo@rnp.br)
- **RENATER/GEANT/RARE:** Frédéric LOUI (frederic.loui@renater.fr)
- **UNESP (SPRACE NCC UNESP):** Sergio Novaes (Sergio.Novaes@cern.ch),Rogerio Iope (rogerio.iope@unesp.br)
- **Rednesp:** Antonio J F Francisco, Ney Lemke **(UNESP)** (ney.lemke@unesp.br), Carlos Antonio Ruggiero **(USP)** (toto@ifsc.usp.br), Jorge Marcos de Almeida (USP) (jorge@usp.br)
- **UERJ:** Alberto Santoro (Alberto.Santoro@cern.ch)
- **George Mason/BRIDGES:** Bijan Jabbari (bjabbari@gmu.edu), Jerry Sobieski, Liang Zhang
- **Xiamen:** Qiao Xiang (xiangq27@gmail.com), Chenyang Huang, Ridi Wen, Yuxin Wang, Jiwu shu
- **Colorado State:** Chengyu Fan (chengy.fan@gmail.com)
- **CENIC:** Louis Fox (lfox@cenic.org), Sana Bellamine (sbellamine@cenic.org), Tony Nguyen
- **Pacific Wave/USC:** Celeste Anderson (celestea@usc.edu)
- **Starlight/MREN/iCAIR:** Joe Mambretti (j-mambretti@northwestern.edu), Jim Chen, Fei Yeh
- **Internet2:** Christian Todorov, (ctodorov@internet2.edu), Rob Vietzke (rvietzke@internet2.edu)
- **AmLight/FIU:** Julio Ibarra (Julio@fiu.edu), Jeronimo Bezerra, Vasilka Chergarova
- **Amlight/ISI:** Heidi Morgan (hlmorgan@isi.edu)
- **Ciena:** Scott Kohlert (skohlert@ciena.com), Rod Wilson
- **KISTI/KREONET:** Buseung Cho (bscho@kisti.re.kr), Mazahir Hussain, Tergel Munkhbat
- **CANARIE:** Thomas Tam (Thomas.Tam@canarie.ca)
- **KAUST:** Alex Moura (alex.moura@kaust.edu.sa), Kevin Sale
- **DE-KIT:** Bruno Hoeft (bruno.hoeft@kit.edu)
- **JPL:** Lee, Carlyn-Ann (Carlyn-Ann.Lee@jpl.nasa.gov)
- **NIST:** Davide Pasavento (davide.pasavento@nist.gov)
- **Hawaii:** Chris Zane (czane@hawaii.edu)
- **SURFNet:** Hans Trompert (hans.trompert@surfnet.nl)
- **CESNET:** Michal Hažlinský, (hazlinsky@cesnet.cz)
- **Clemson:** Cole McKnight (cbmckni@g.clemson.edu)
- **NCHC/TAWREN:** Li-Chi Ku, lku@narlabs.org.tw
- **GNA-G/AArNet:** David Wilde (David.Wilde@aarnet.edu.au)
- **GNA-G AutoGOLE / SENSE WG Members**: https://www.gna-g.net/join-working-group/autogole-sense
- **GNA-G Data Intense Science WG Members:** https://www.gna-g.net/join-working-group/data-intensive-science/
- **STORDIS:** Waldemar.Scheck@stordis.com

SC22 and Beyond: Persistent Development on a Global Testbed; Trajectory to Production

GNA-G
Global Network Advancement Group
https://www.gna-g.net/

Mission: To Support global research and education using the technology, infrastructures and investments of its participants



The GNA-G exists to bring together researchers, National Research and Education Networks (NRENs), Global eXchange Point (GXP) operators, regionals and other R&E providers, in developing a common global infrastructure to support the needs.

Global Petascale to Exascale Workflows for Data Intensive Sciences

- **Advances Embedded and Interoperate within a 'composable' architecture of subsystems, components and interfaces, organized into several areas:**
  - **Visibility: Monitoring and information tracking and management including IETF ALTO/OpenALTO, BGP-LS, sFlow/NetFlow, Perfsonar, Traceroute, Qualcomm Gradient Graph congestion information, Kubernetes statistics, LibreNMS, P4/Inband telemetry**
  - **Intelligence: Stateful decisions using composable metrics (policy, priority, network- and site-state, SLA constraints, responses to 'events' at sites and in the networks, ...), using NetPredict, Hecate, G2, Yale Bilevel optimization, Coral, Elastiflow/Elastic Stack**
  - **Controllability: SENSE/OpenNSA/AutoGOLE, P4/PINS, segment routing with SRv6 and/or PolKA, BGP/PCEP**
  - **Network OSes and Tools: GEANT RARE/freeRtr, SONIC, Calico VPP, Bstruct-Mininet environment, ...**
  - **Orchestration: SENSE, Kubernetes (+k8s namespace), dedicated code and APIs for interoperation and progressive integration**

Next Generation Network-Integrated System

- **Top Line Message: In order to address the challenges and meet the needs, we need a new dynamic and adaptive software-driven system, which**
  - **Coordinates worldwide networks as a first class resource along with computing and storage, across multiple domains**
  - **Simultaneously supports the LHC experiments, other major DIS programs and the larger worldwide academic and research community**
- **Systems design approach: A virtualized global dynamic fabric that flexibly allocates, balances and conserves the available network resources**
  - **Negotiating with site systems that aim to accelerate workflow; Use of ML**
- **Builds on ongoing R&D projects: from regional caches/data lakes to intelligent control and data planes to ML-based optimization**
  [E.g. SENSE/AutoGOLE, NOTED, ESNet HT, GEANT/RARE, AmLight, Fabric, Bridges; NetPredict, DeepRoute, Hecate, ALTO, PolKA ...]
- **A key milestone: integration of SENSE + network services with FTS & Rucio**
- **We are also leveraging the worldwide move towards a fully programmable ecosystem of networks and end-systems (P4, PINS; SRv6; PolKA), plus operations platforms (OSG, NRP; global SENSE Testbed; BRIDGES)**
- **The LHC experiments together with the WLCG, the GNA-G and its Working Groups, and the worldwide R&E network community, are the key players**
  - **Directions also taken up by other programs: LBNF/DUNE, VRO, SKA**

Partner NREs at SC22

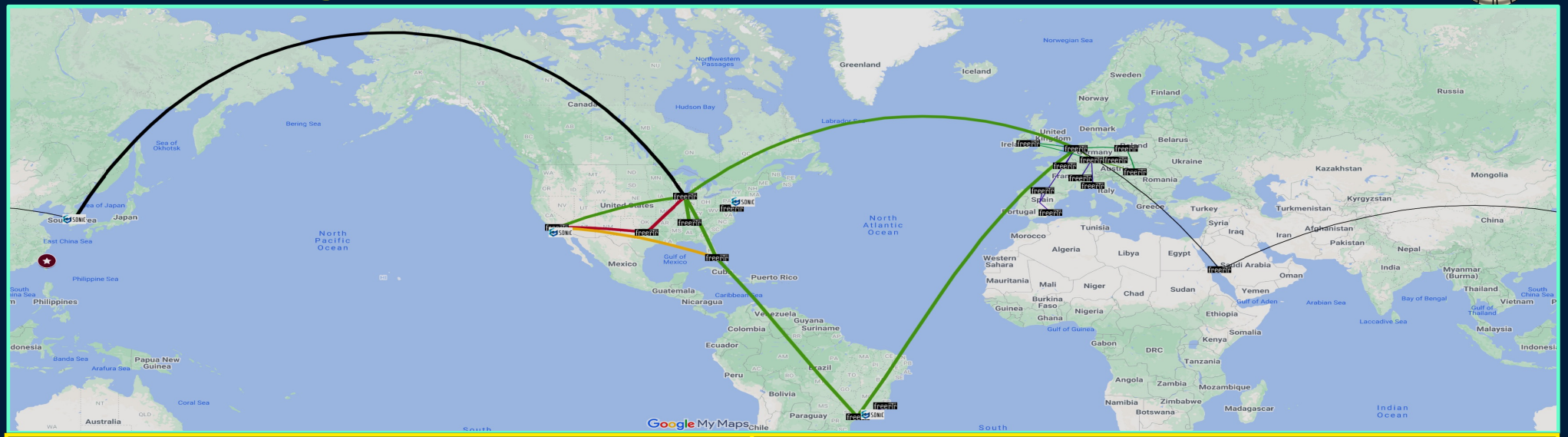
The NRE demonstrations hosted at or partnering with the Caltech Booth 2820:

NRE-001	Edmund Yeh	eyeh@ece.neu.edu	N-DISE: NDN for Data Intensive Science Experiments
NRE-004	Joe Mambretti	j-mambretti@northwestern.edu	1.2 Tbps Services WAN Services: Architecture, Technology and Control Systems
NRE-005	Joe Mambretti	j-mambretti@northwestern.edu	400 Gbps E2E WAN Services: Architecture, Technology and Control Systems
NRE-007	Edoardo Martelli	edoardo.martelli@cern.ch	LHC Networking And NOTED
NRE-008	Joe Mambretti	j-mambretti@northwestern.edu	IRNC Software Defined Exchange (SDX) Multi-Services for Petascale Science
NRE-009	Jim Chen	jim-chen@northwestern.edu	High Speed Network with International P4 Experimental Networks for The Global Research Platform and Other Research Platforms
NRE-010	Magnos Martinello	magnos.martinello@ufes.br	Demonstrating PolKA Routing Approach to Support Traffic Engineering for Data-intensive Science
NRE-011	Qiao Xiang	xiangq27@gmail.com	Coral: Fast Data Plane Verification for Large-Scale Science Networks via Distributed, On-Device Verification
NRE-013	Tom Lehman	tlehman@es.net	AutoGOLE/SENSE: End-to-End Network Services and Workflow Integration
NRE-015	Tom Lehman	tlehman@es.net	SENSE and Rucio/FTS/XRootD Interoperation
NRE-016	Marcos Schwarz	marcos.schwarz@rnp.br	Programmable Networking with P4, GEANT RARE/freeRtr and SONIC/PINS

Global Petascale to Exascale Workflows for Data Intensive Sciences

- **Development Trajectory:** Parallel developments + mission-driven progressive interfacing and system-level integration
- **Overarching Concept: Consistent Network Operations:**
  - Stable load balanced high throughput workflows cross optimally chosen network paths
  - Provided by autonomous site-resident services dynamically interacting with network-resident services
  - Up to preset or fliexible *high water marks* to accommodate other traffic
  - Responding to (or negotiating with) site demands from the science programs' principal data distribution and management systems
- **Data Center Analogue for Networks**
  - Classes of "Work" (work = transfers, or overall workflow), defined by task parameters and/or priority and policy
  - Adjust rate of progress in each class to respond to network or site state changes, and "events"
  - Moderate/balance the rates among the classes to optimize a multivariate objective function with constraints

P4 Tofino + Tofino2 + SONIC
Programmable Global Persistent Testbed



22 Active GNA-G/RARE P4 Testbed Sites/Devices

- Caltech, Pasadena-US: 4 x FreeRtr/P4+SONIC
- CERN, Geneva-CH: FreeRtr/P4
- FIU, Miami-US: FreeRtr/P4
- GEANT, Amsterdam-NL: FreeRtr/P4
- GEANT, Budapest-HU: FreeRtr/P4
- GEANT, Frankfurt-DE: FreeRtr/P4
- GEANT, Paris-FR: FreeRtr/DPDK
- GEANT, Poznan-PL: FreeRtr/P4
- GEANT, Prague-CZ: FreeRtr/DPDK
- HEAnet, Dublin-IE: FreeRtr/P4
- RENATER, Paris-FR: FreeRtr/P4
- RNP, Rio de Janeiro-BR; FreeRtr/P4
- SouthernLight (FIU/RedClara/Rednesp/RNP), São Paulo-BR: FreeRtr/P4
- StarLight, Chicago-US: FreeRtr/P4
- SWITCH, Geneva-CH: FreeRtr/P4
- TCD, Dublin-IE: FreeRtr/P4
- Tennessee Tech, Cookeville-US: FreeRtr/P4
- UFES, Vitória-BR: FreeRtr/P4
- UMd, College Park, Maryland-US: FreeRtr/P4

+ 7 Sites in October – November (by SC22):

- JISC, London-UK: FreeRtr/P4
- KAUST, Saudi Arabia: FreeRtr/DPDK
- KISTI, South Korea: SONiC/P4
- RNP, Rio de Janeiro-BR+1 FreeRtr/P4
- SC22 Caltech Booth, Dallas-US: FreeRtr/P4
- UCSD, San Diego-US: SONiC/P4
- UFES, Vitória-BR: +1 FreeRtr/P4

Global Petascale to Exascale Workflows
Cornerstone system components and concepts (III)

- **PolKA: Polynomial Key-based Architecture for Source Routing**
  Creation of an overlay network with PolKA tunnels forming virtual circuits, integrating persistent resources from the GNA-G AutoGOLE/SENSE and GEANT RARE testbeds, validated using 100G+ transfers of science data.
  - **Underlay congestion will be detected by tunnel monitoring and signaled to the overlay so that the traffic is steered away from congested tunnels to other paths.**
  - **Comparisons between SRv6 segment routing and PolKA regarding controllability and performance metrics.**
  - **PolKA full deployment enables extreme traffic engineering demands of data-intensive sciences to be met, through a new range of network functionalities such as: multipath routing, in-network telemetry and proof-of-transit with path attributes to support higher level stateful traffic engineering decisions.**
- **Network traffic prediction and engineering optimizations using the latest graph neural network and other emerging deep learning methods, developed by ESnet's Hecate /DeepRoute project.**