Science Data and the NDN paradigm

Inder Monga

CTO, ESnet

Division Deputy of Technology, Scientific Networking Division

Lawrence Berkeley National Lab
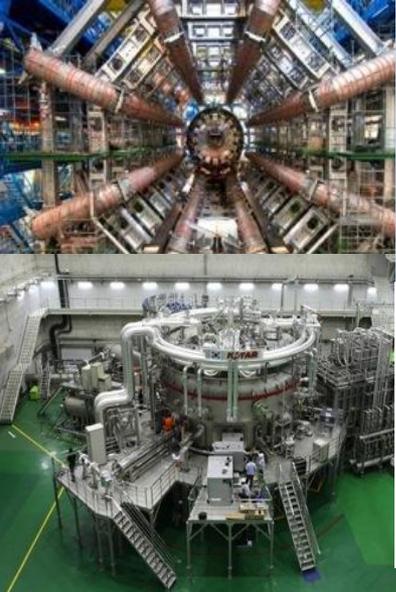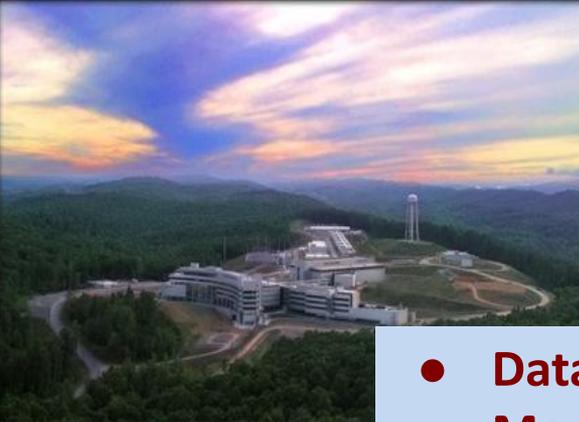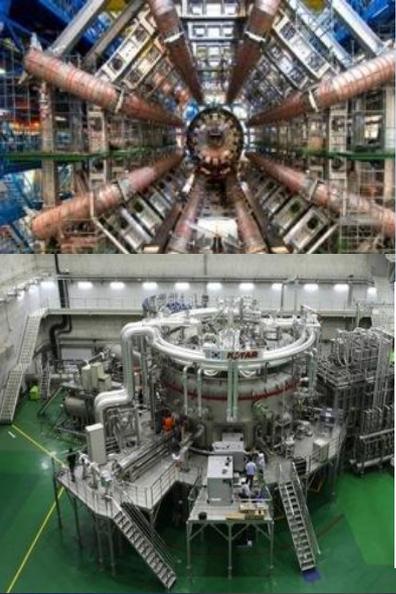
NDN Comm 2015

U.S. DEPARTMENT OF **ENERGY**
Office of Science

BERKELEY LAB

Experimental and observational science deals with big and small

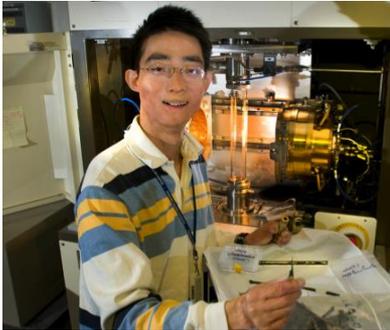# Experimental and observational science deals with big and small

- **Data volumes are increasing faster than Moore's Law**

- **New algorithms and methods for analyzing data**

- **Infeasible to put a supercomputing center at every experimental facility**
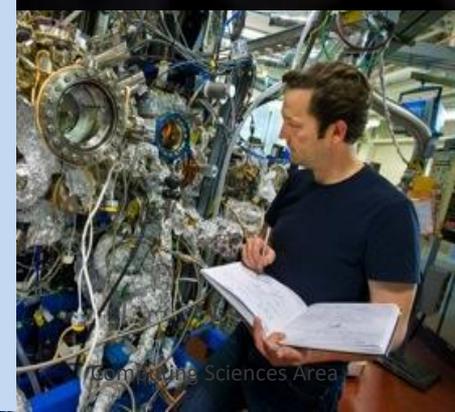
Computing Sciences Area

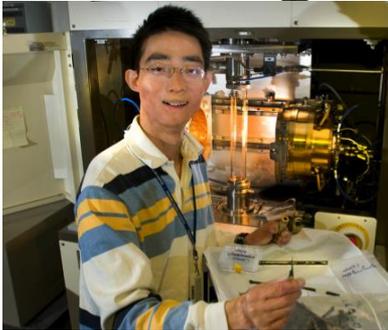# All too common process of discovery

ESnet

All too common process of discovery

All too common process of discovery

All too common process of discovery

All too common process of discovery
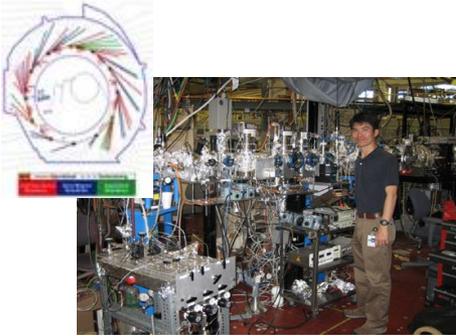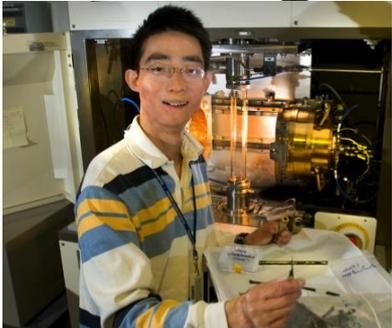
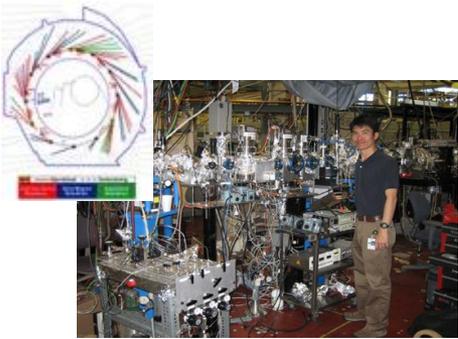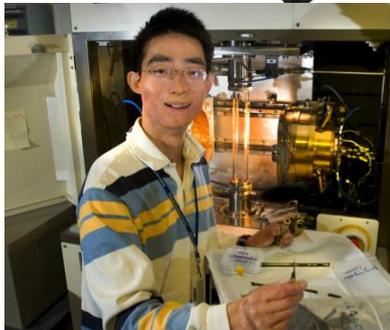# All too common process of discovery

All too common process of discovery

All too common process of discovery
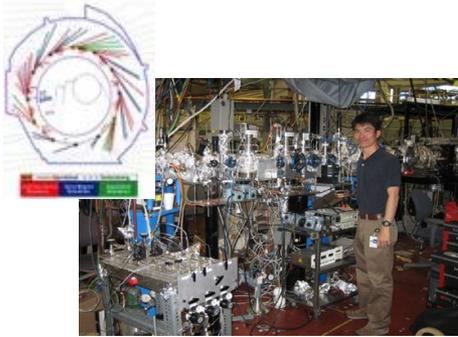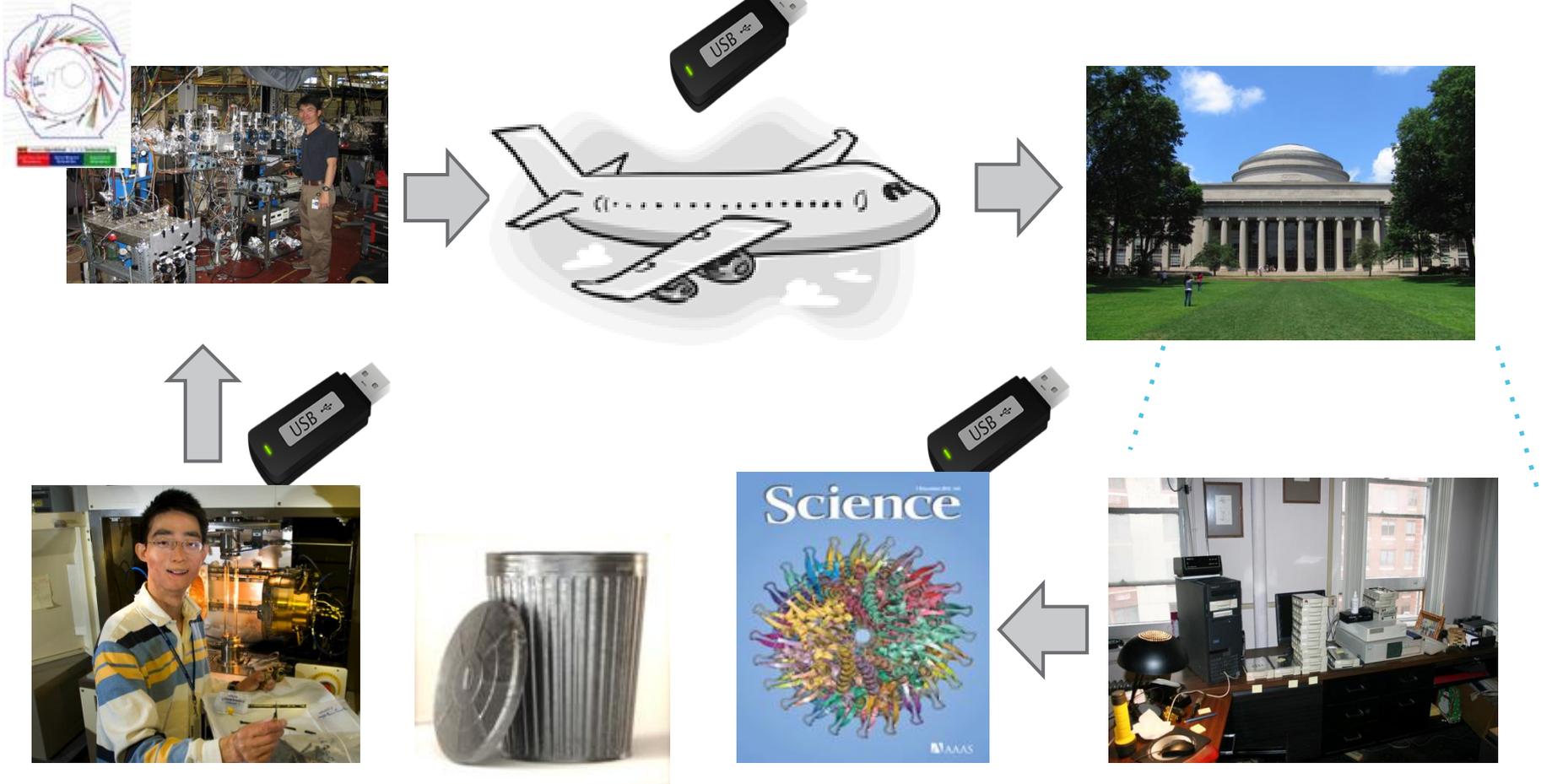
All too common process of discovery

All too common process of discovery

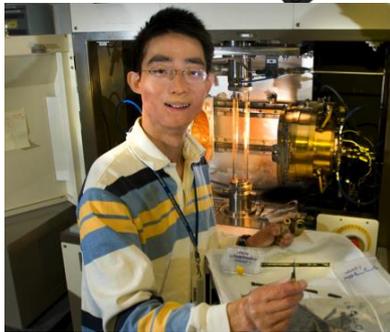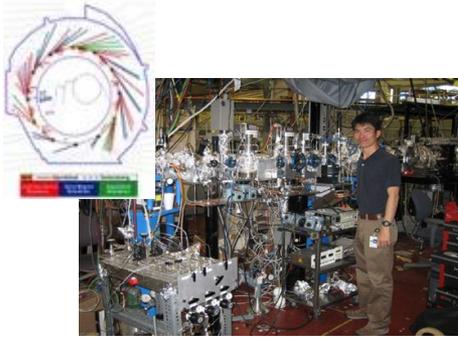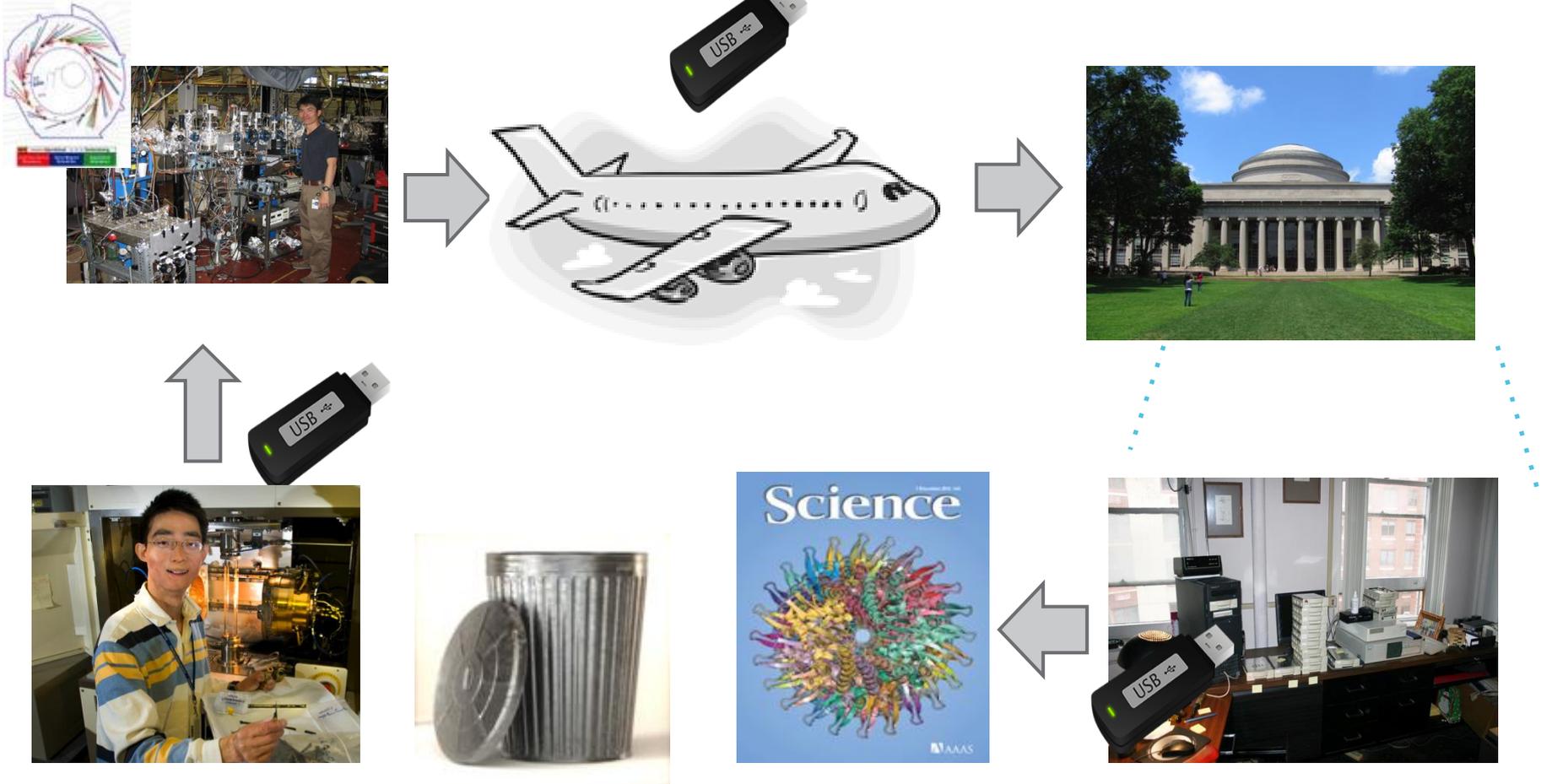# All too common process of discovery

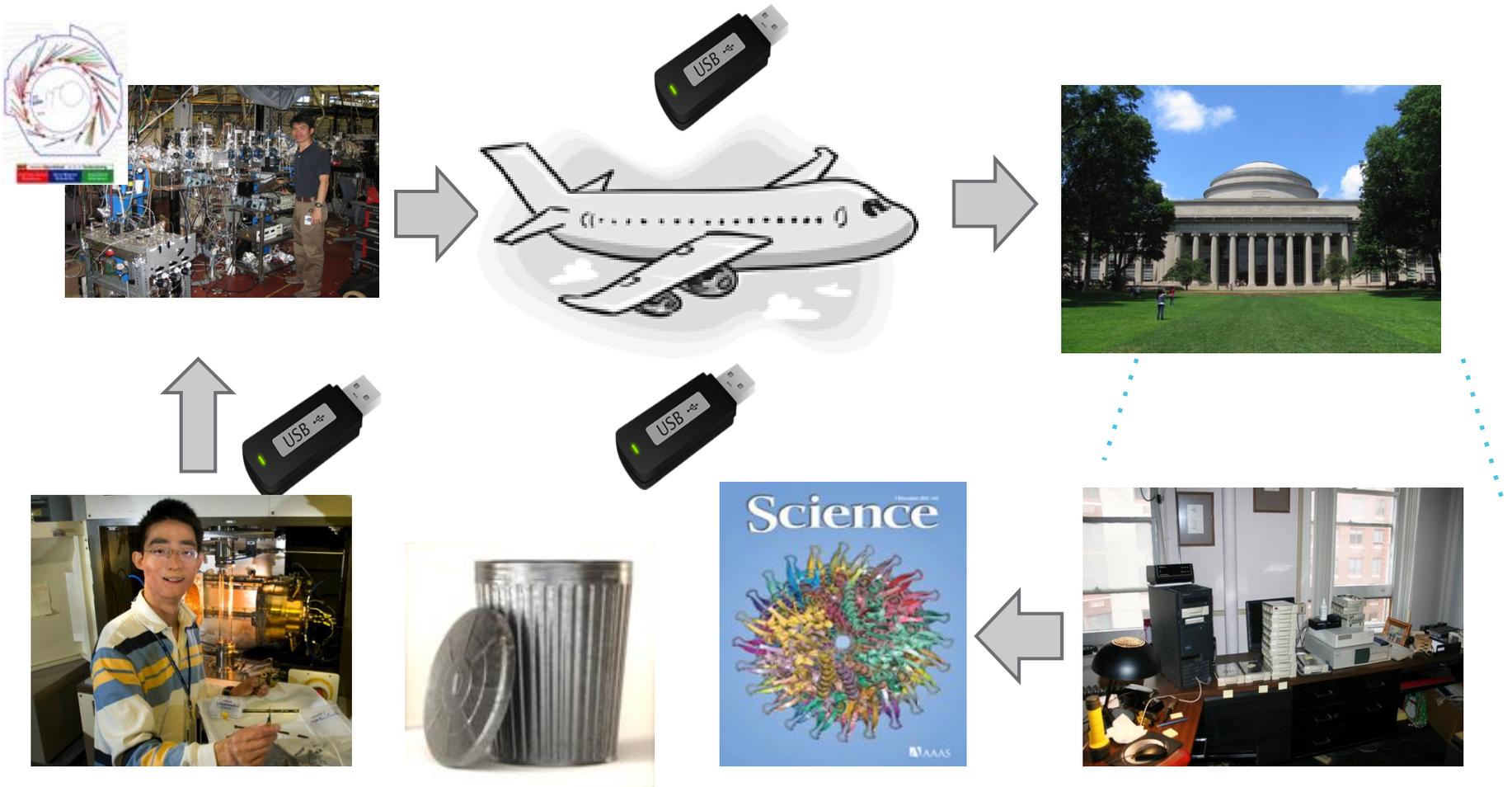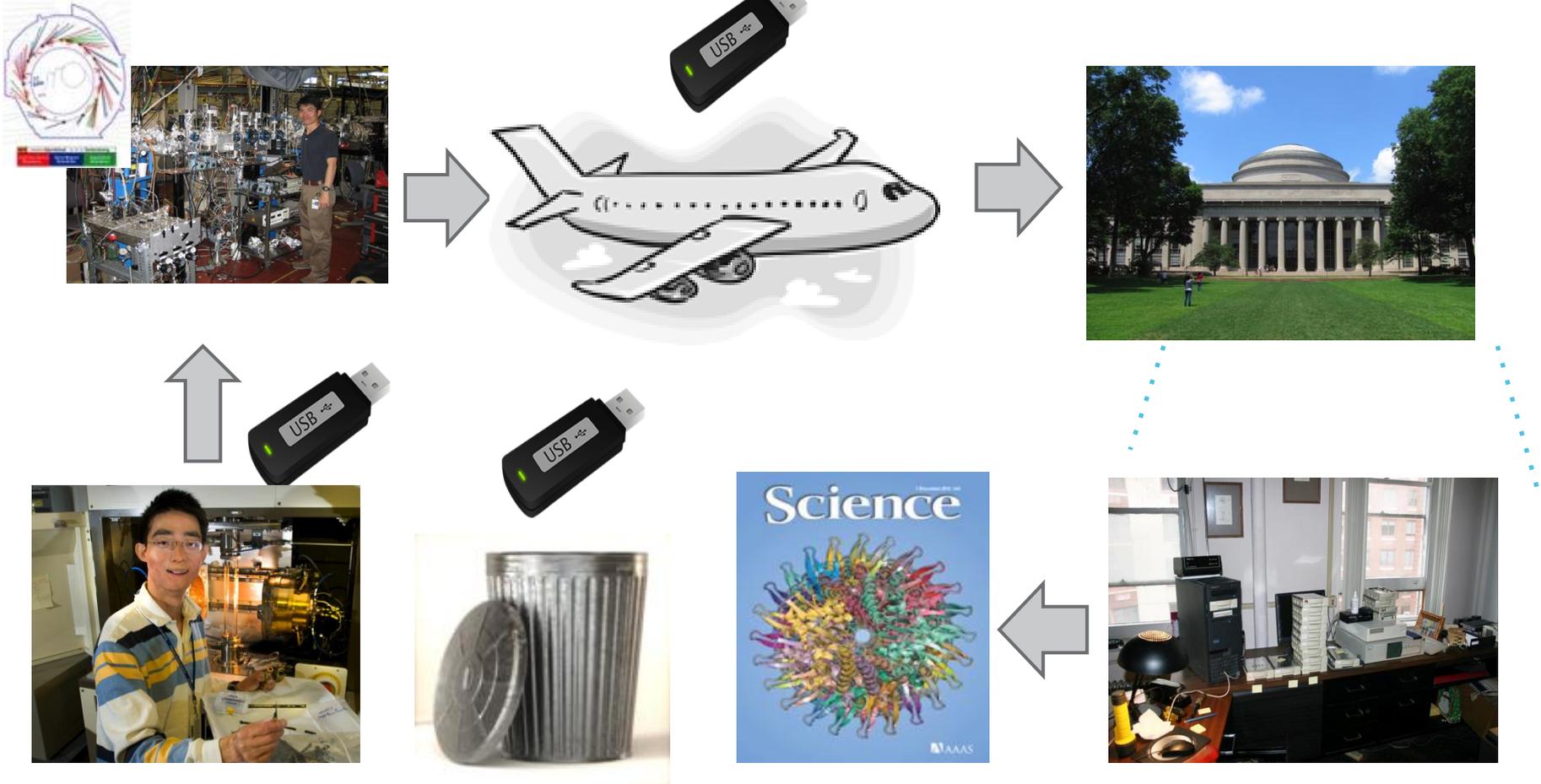# 'Superfacility' Vision: A network of connected facilities, software and expertise to enable new modes of discovery

'Superfacility' Vision: A network of connected facilities, software and expertise to enable new modes of discovery

ESnet is a dedicated mission network engineered to accelerate a broad range of science outcomes.
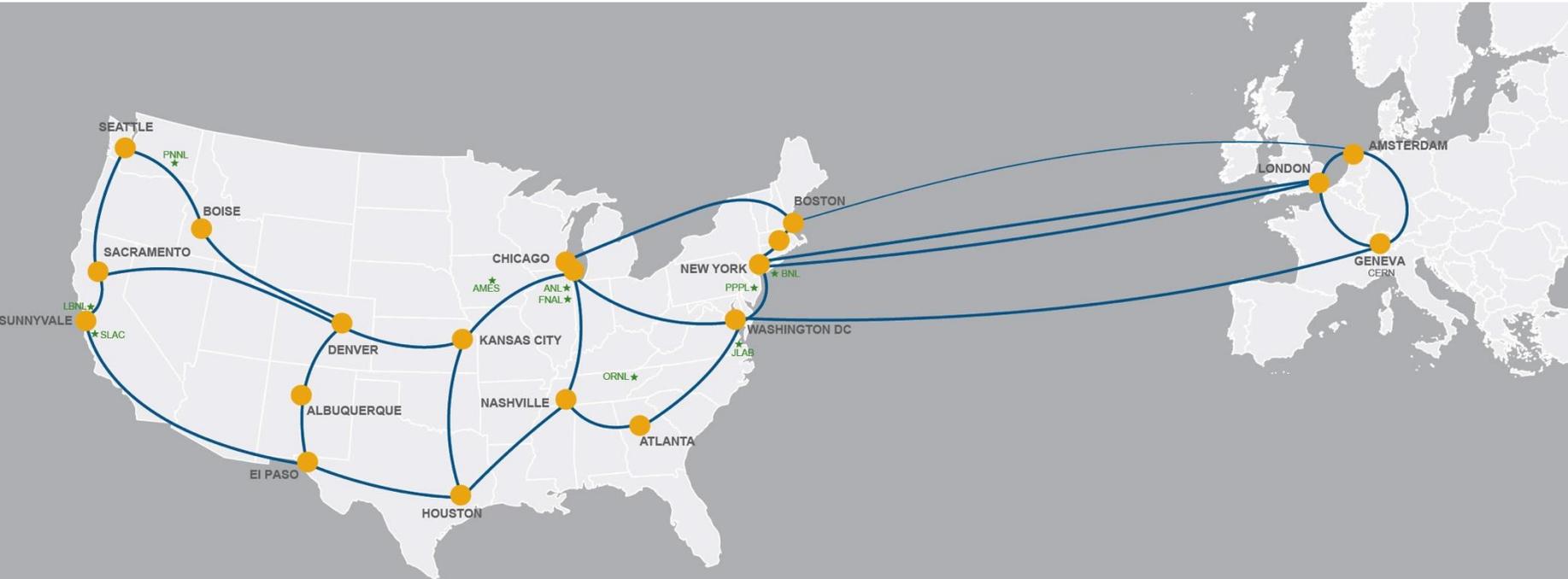
ESnet is a dedicated mission network engineered to accelerate a broad range of science outcomes.
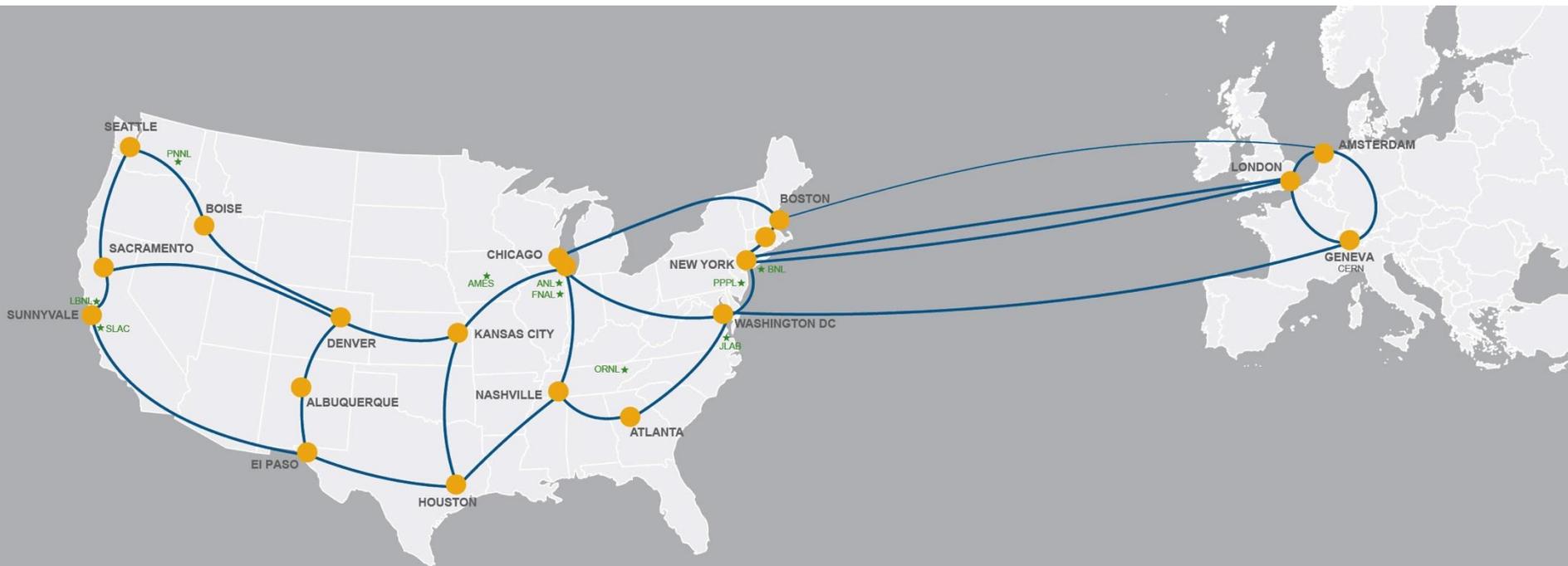
We do this by offering unique capabilities, and optimizing the network for data acquisition, data placement, data sharing, data mobility.

ESnet is designed for different goals than general Internet.
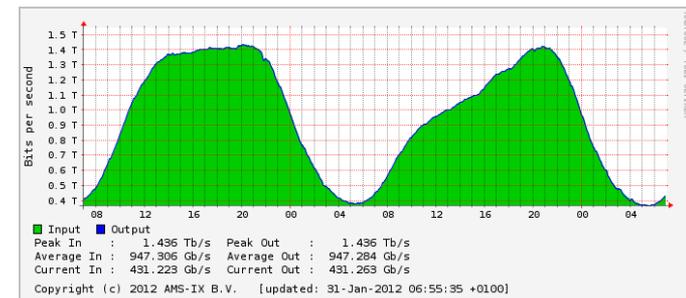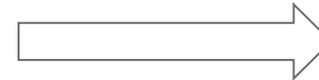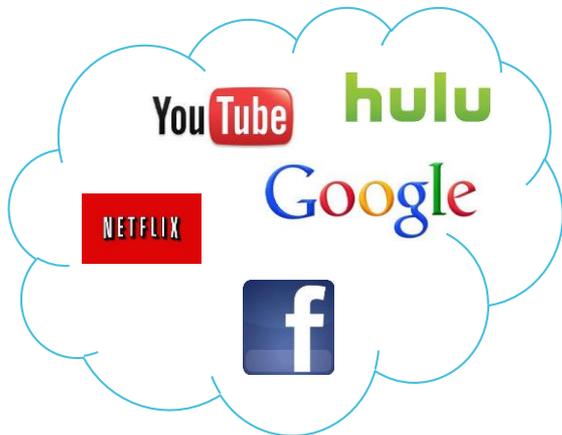
ESnet is designed for different goals than general Internet.
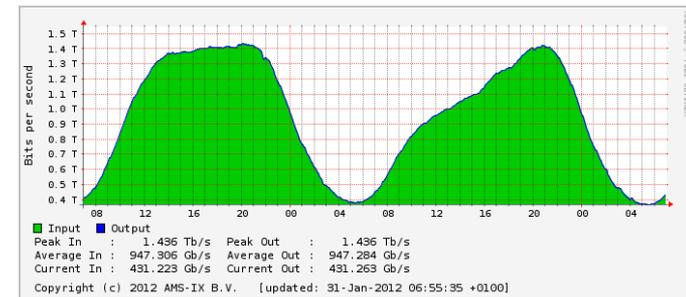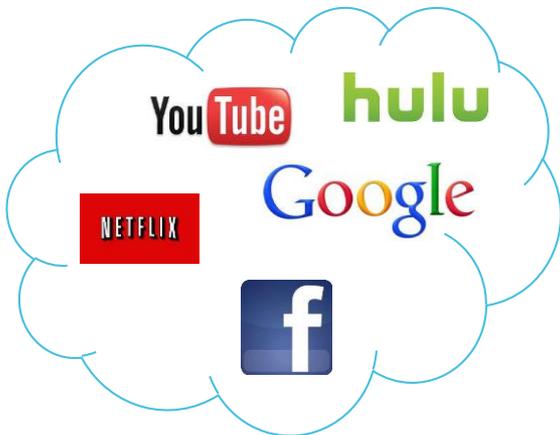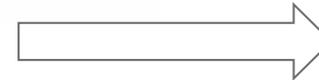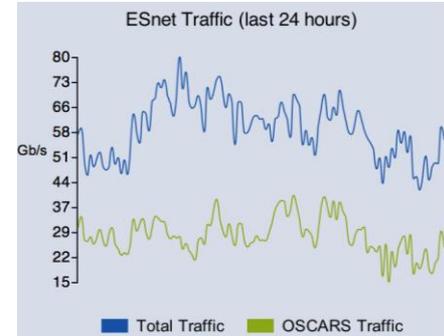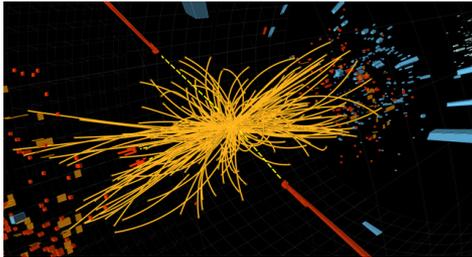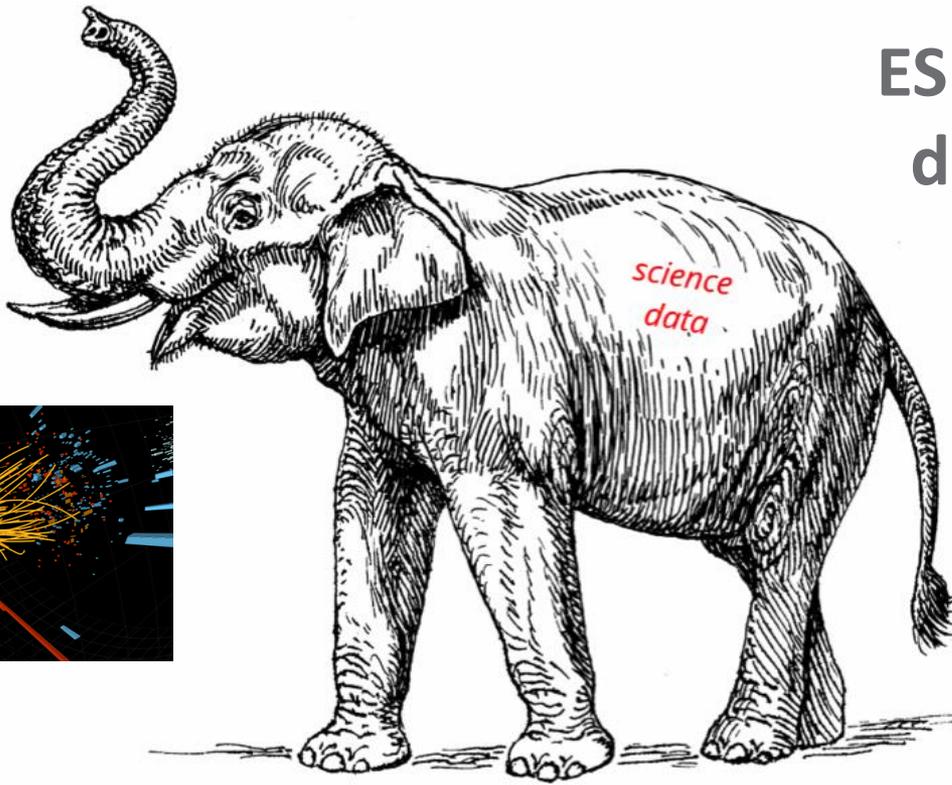
Lots of data to move around

Bytes of Science Data Transferred Each Month
by the Energy Sciences Network

August 2015:
29.13 PB

Total
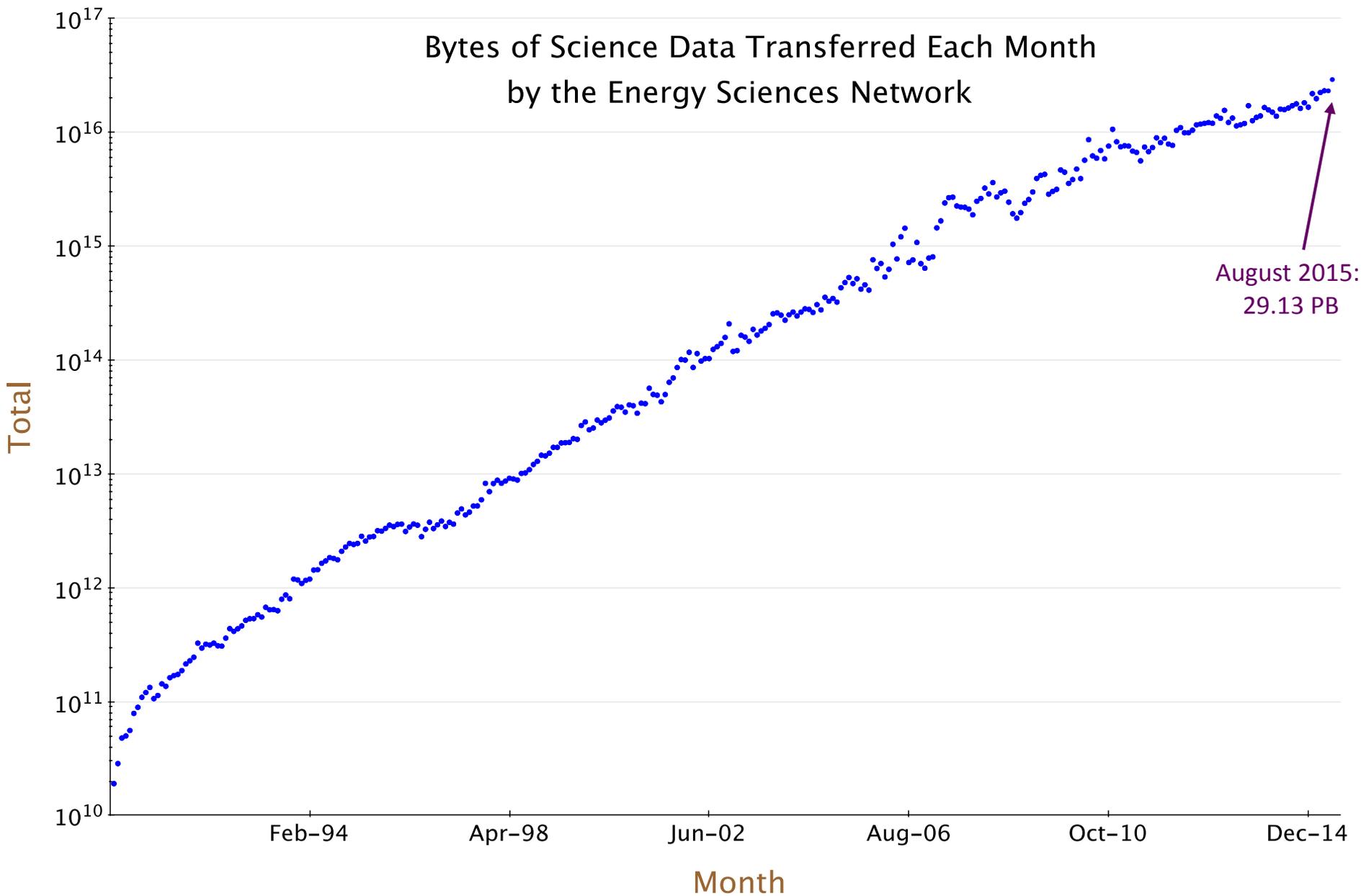
Month

Lots of data to move around (contd.)

Bytes of Science Data Transferred Each Month
by the Energy Sciences Network

August 2015:
29.13 PB
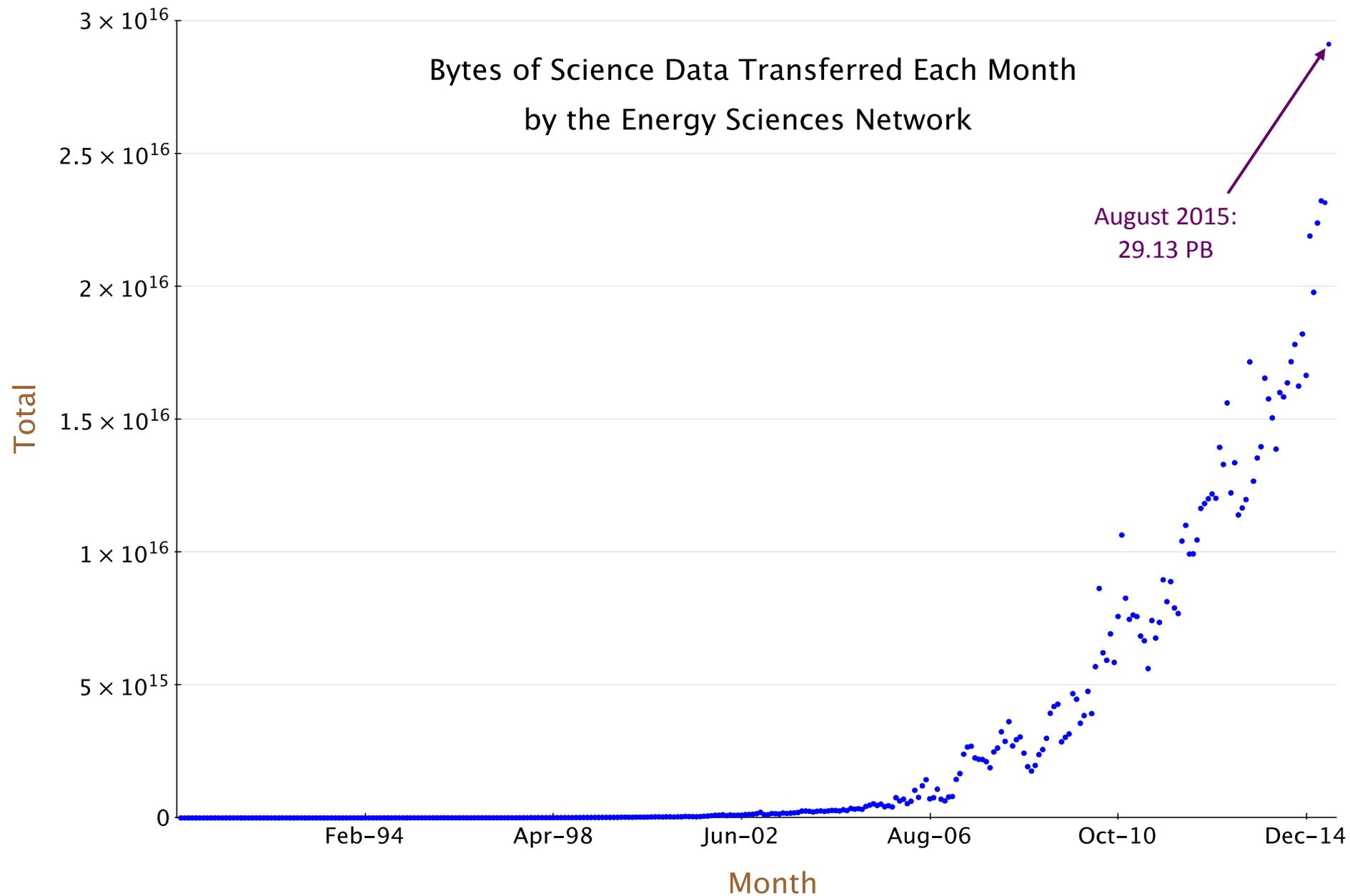
High-level objectives for scientific data: alignment with NDN approach

- Radically simplify how scientific users manage, move and manipulate large, distributed, science data repositories, but with **high-throughput end2end**

- Abstract the storage and network capability and location dependence from the user-data interaction

- Enable the ability for users to specify and retrieve portions of data the workflow needs

- Create a secure, scalable framework based on integrated data management and network transport

ESnet

Use Case #1

Researchers from Berkeley Lab and SLAC conducted protein crystallography experiments at LCLS to investigate photoexcited states of PSII, with near-real-time computational analysis at NERSC.
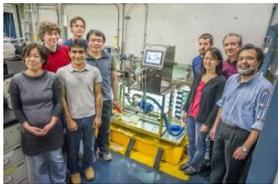
"Taking snapshots of photosynthetic water oxidation using femtosecond X-ray diffraction and spectroscopy," *Nature Communications* 5, 4371 (9 July 2014)
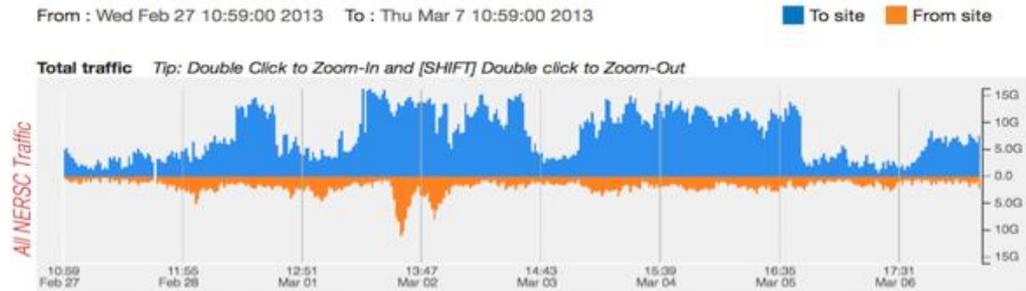
50TB moved a night

ESnet

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

15 September 2015 – WEJohnston, ESnet, wej@es.net

See http://lhcone.net for more detail.

Use Case #3: International Climate Data

ESnet
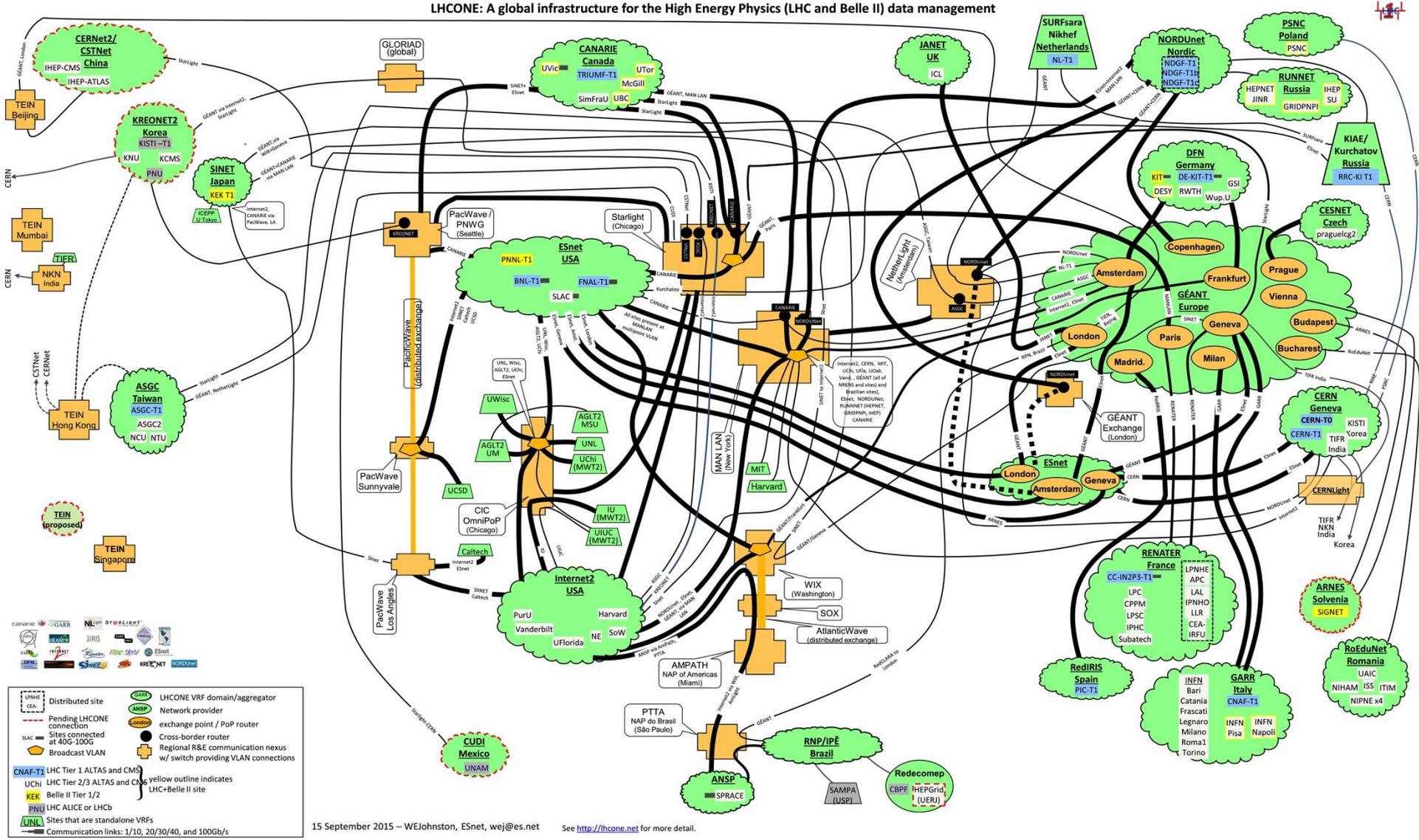
Use Case #3: International Climate Data
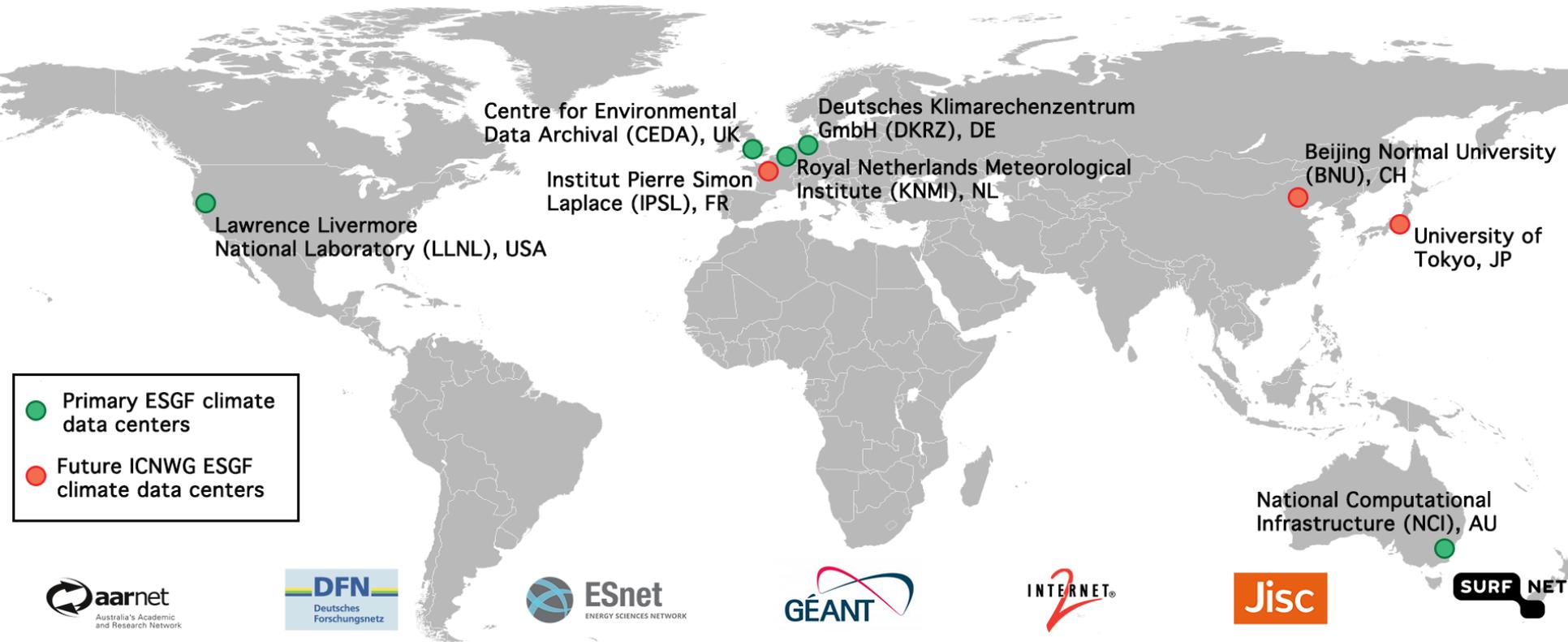
Use Case #3: International Climate Data
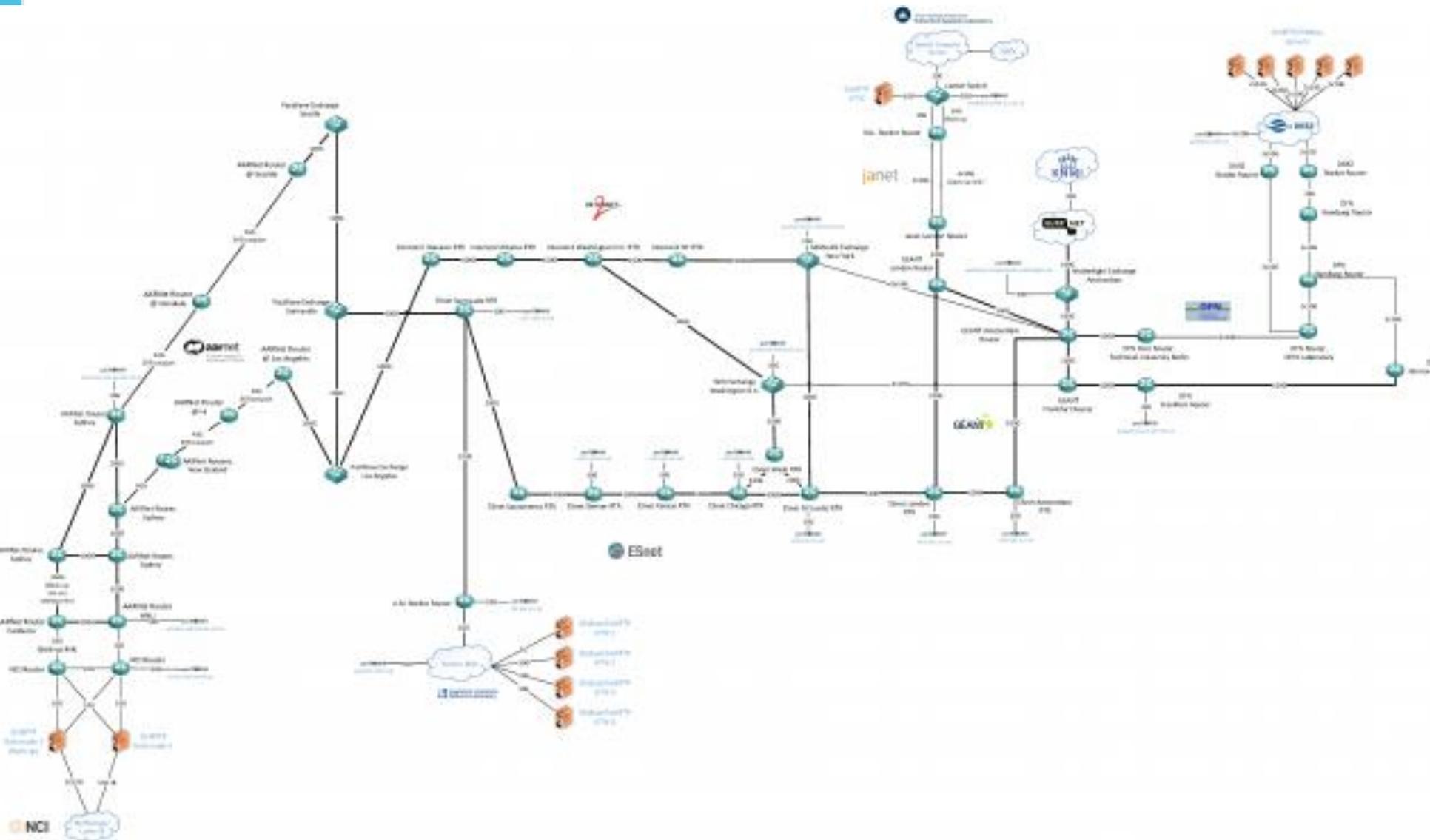
Perception of limitations of NDN motivating research questions

1. If I am moving 50TB of data through a single path, from an experiment to a storage facility, I really do not want to cache it at every intermediate NDN node

    – What is the right strategy for allocating disk resources to caching? What if one data transfer consumes all cache resources or there is not enough space?

2. What is the performance of the end-to-end data transfer? How can I get line rate **throughput**?

3. How do I leverage the knowledge of network capability in choosing the transfer path? How do I build in the knowledge of underlay into the NDN overlay?

4. How do I leverage network programmability to do the above?

5. And many other questions….

ESnet

Where are we at?

- Collaboration with Christos and Colorado State – high-powered NDN devices between three representative climate sites as a testbed
  - Susmit working on answering some of the high-level objectives as described

- HEP and ASCR interest in NDN from a research perspective – paper earlier this year @ CHEP, and Phil will talk about next-steps right after

- Interest in expanding a federation of high-powered NDN devices with the right strategy for caching and data management

- Combining NDN with SDN – we have a next-gen SDN testbed across US and Europe – can we combine that to provide the right primitives for high-performance NDN?
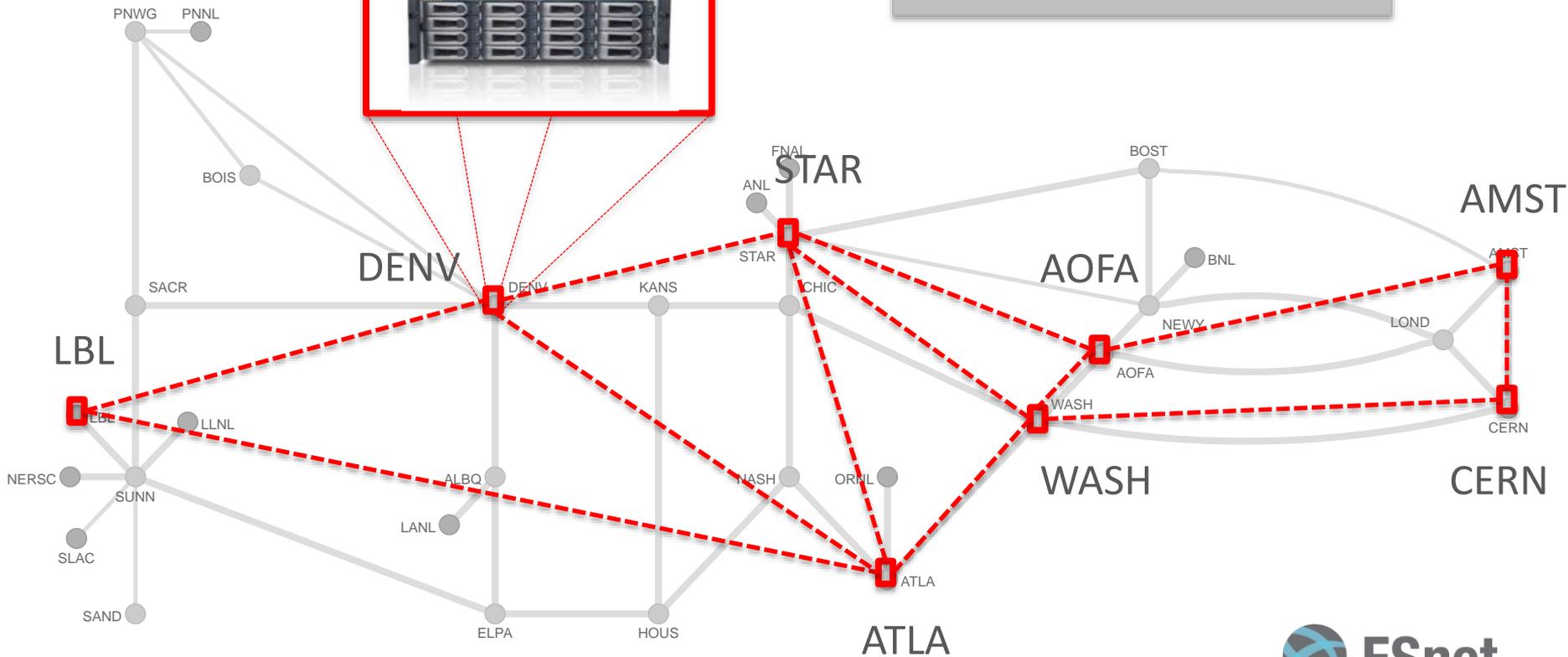 - Lets do iterative experimentation and improvement!!!!!!!

ESnet

ESnet SDN Testbed

ESnet PE Router

(2+)x10GE

(n)x10GE

Testbed Host
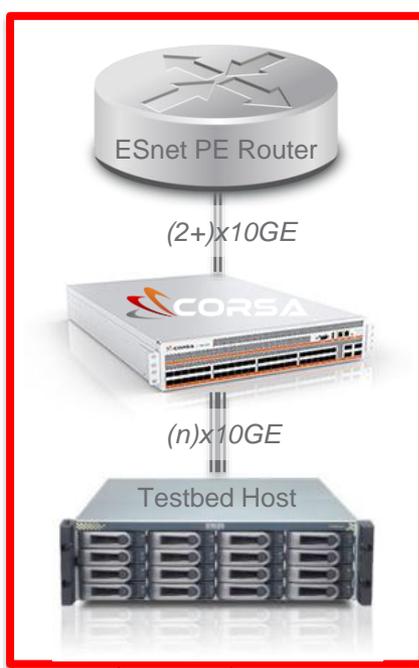
Status Update:

- Testbed deployed at all locations
- QoS support verified, press release next week
- ENOS demo on Testbed @ SC

Deployed SDN Testbed node locations

Deployed SDN Testbed connectivity overlay (using OSCARS circuits)

PNWG
PNNL
BOIS
DENV
SACR
LBL
NERSC
SUNN
SLAC
SAND
LLNL
ALBQ
LANL
ELPA
HOUS
FNAL
ANL
STAR
KANS
CHIC
NASH
ORNL
ATLA
BOST
AOFA
BNL
NEWY
AMST
LOND
WASH
CERN

ESnet

Thank you!

- Please feel free to email me with questions, comments or arrows at

imonga at es dot net

ESnet