Towards a Research Data Zone

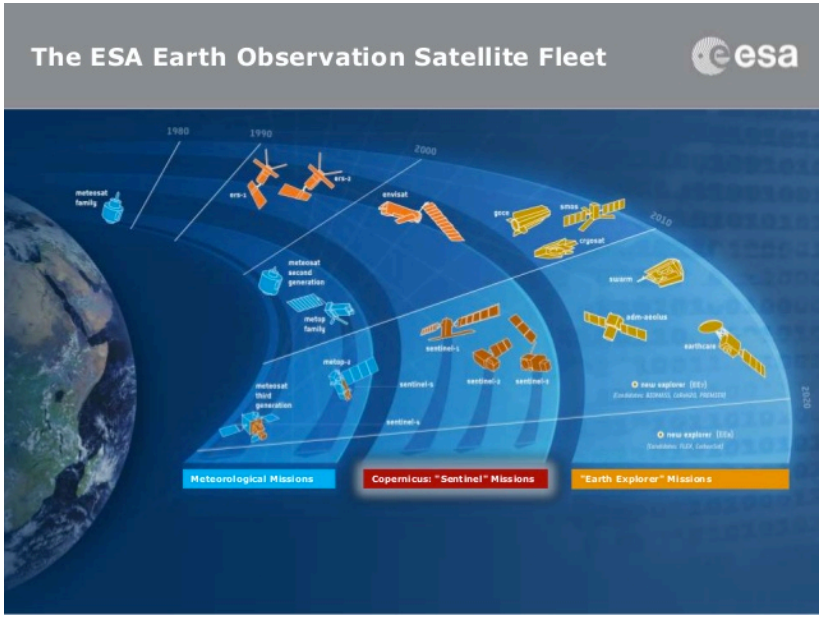
OPTIMIZING CAMPUS INFRASTRUCTURE FOR RESEARCH

Mary Hester, SURFnet (mary.hester@surfnet.nl)
Peter Hinrich, SURFnet (peter.hinrich@surfnet.nl)

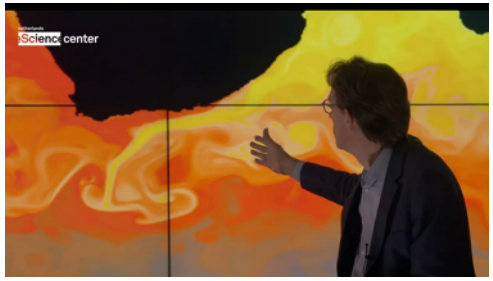
Motivation:
Data size gets bigger, networks become faster

Data size gets bigger and becomes problem in more disciplines

growing number of international archives for communities



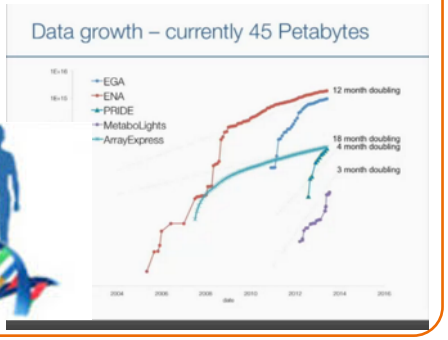
Satellite data



Climate data

Radio astronomy

DNA sequencing



Hybrid network (IP and lightpaths)

- High capacity Internet and lightpaths for heavy use
- Lightpaths offer high, guaranteed bandwidth, low latency and exclusive use
- Large uptake for business operations
- Uptake in research slow (cost, funding model, complexity etc.)

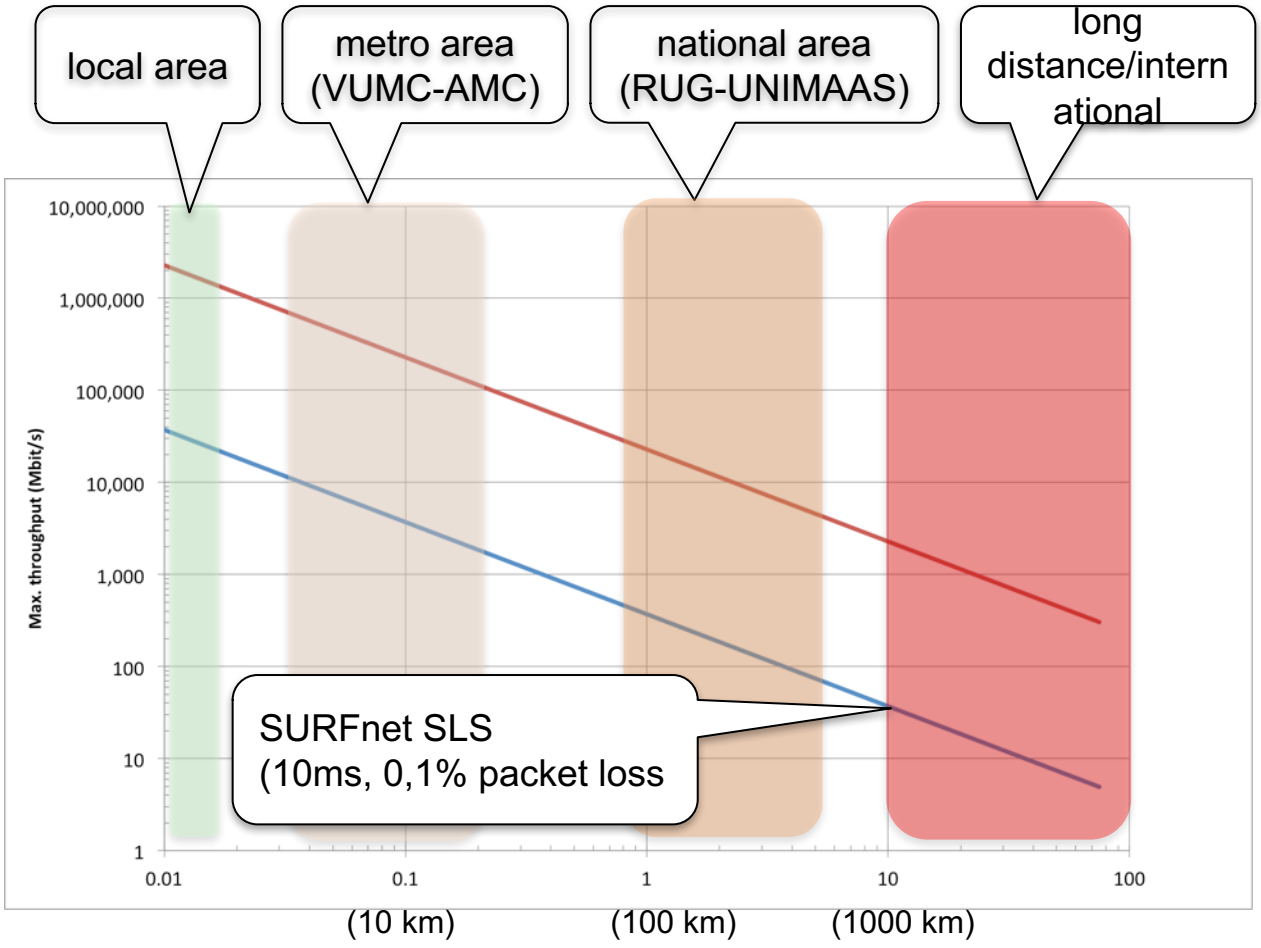
Current situation:

- Research traffic (large data transfers) mixed with commodity Internet:

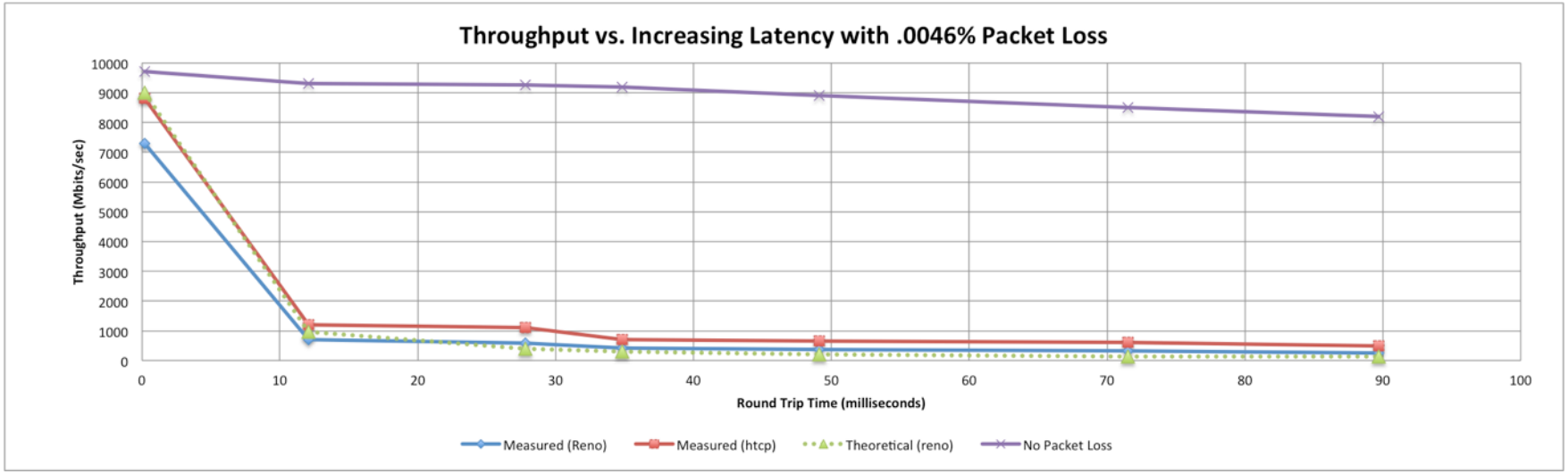
  This is an increasing problem

Limitations of TCP

- The theoretical maximum of a single TCP stream is limited
- Maximum throughput decreases with increasing packet loss and delay (distance)
- Exponential decrease
- Many researchers use TCP based transfers
- SURFnet SLS is 10ms and 0,1% packet loss (worst case scenario)
- Graph shows throughput with 0,1% and 0,00001% packet loss



Limitations of TCP

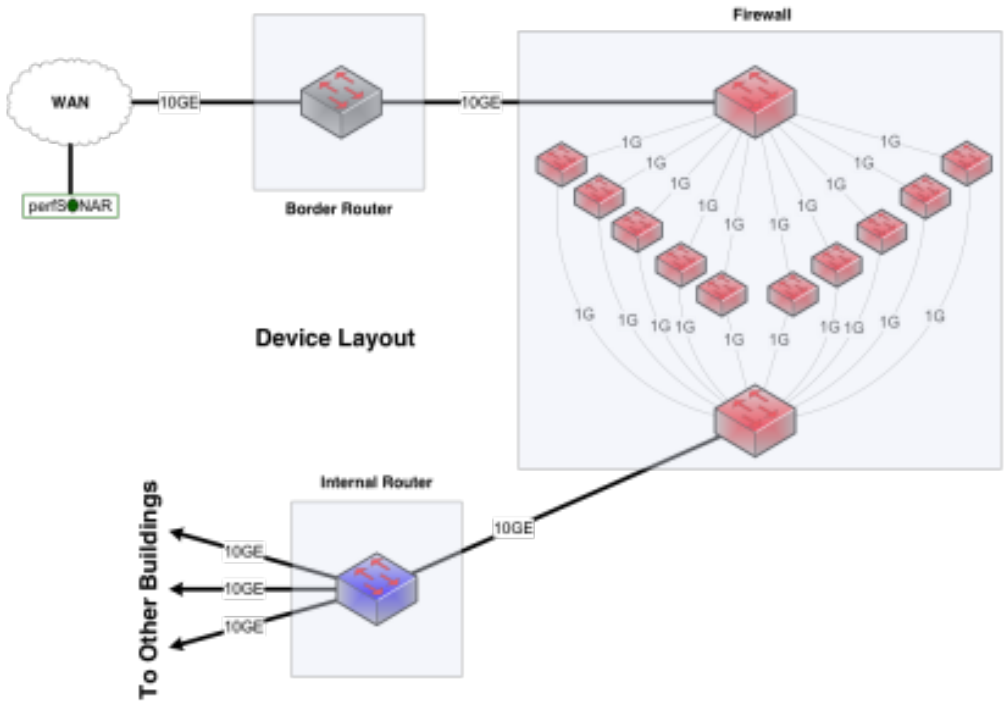


- Research data transfers often based on TCP
- Maximum throughput is limited, packet loss has a major impact
- Packet loss is seen as congestion, and decreases throughput

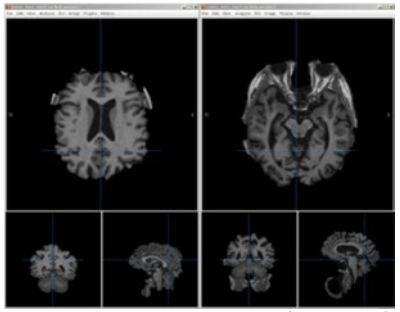
With courtesy to Brian Tierney (Esnet)

Firewall issues

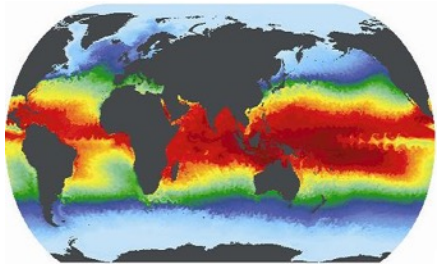
- Limited throughput
- Large number of small flows
- Small buffers
- Increased latency
- Not suitable for ‘exotic’ protocols



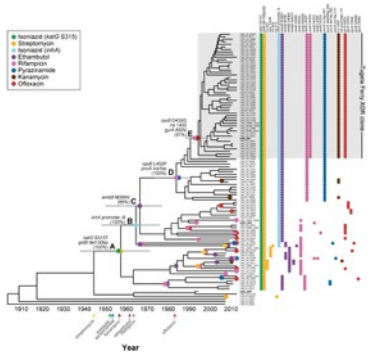
Current approach:
support for specific projects



Population Imaging (LUMC)



Climatology (UU)



Bacterial drug resistance discovery (TUDelft)

Optimization:

- Larger packet size (jumboframes)
- Other network protocols (UDP)
- Specialized data transfer software (GridFTP)
- Access control in stead of firewall

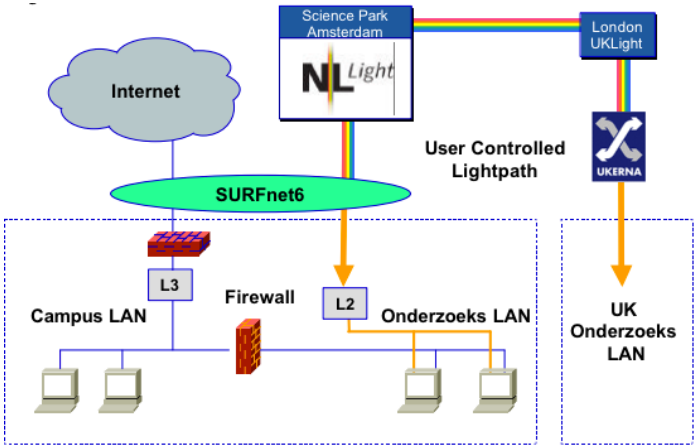
And many more...

Current approach with ad hoc support not efficient

Need for a generic approach:
research data zone

Drivers

- Increased need for collaboration
- Collaboration with commercial research
- Increased datasets in various disciplines
- Emerge of European infrastructures
- Sharing of confidential data
- Urge for outscaling to (HPC) cloud
- Emerging of science DMZ concept in US and Research LAN between UMC's



From "Evolutie en Revolutie in de Campusinfrastructuur".

Responsibility for SURF

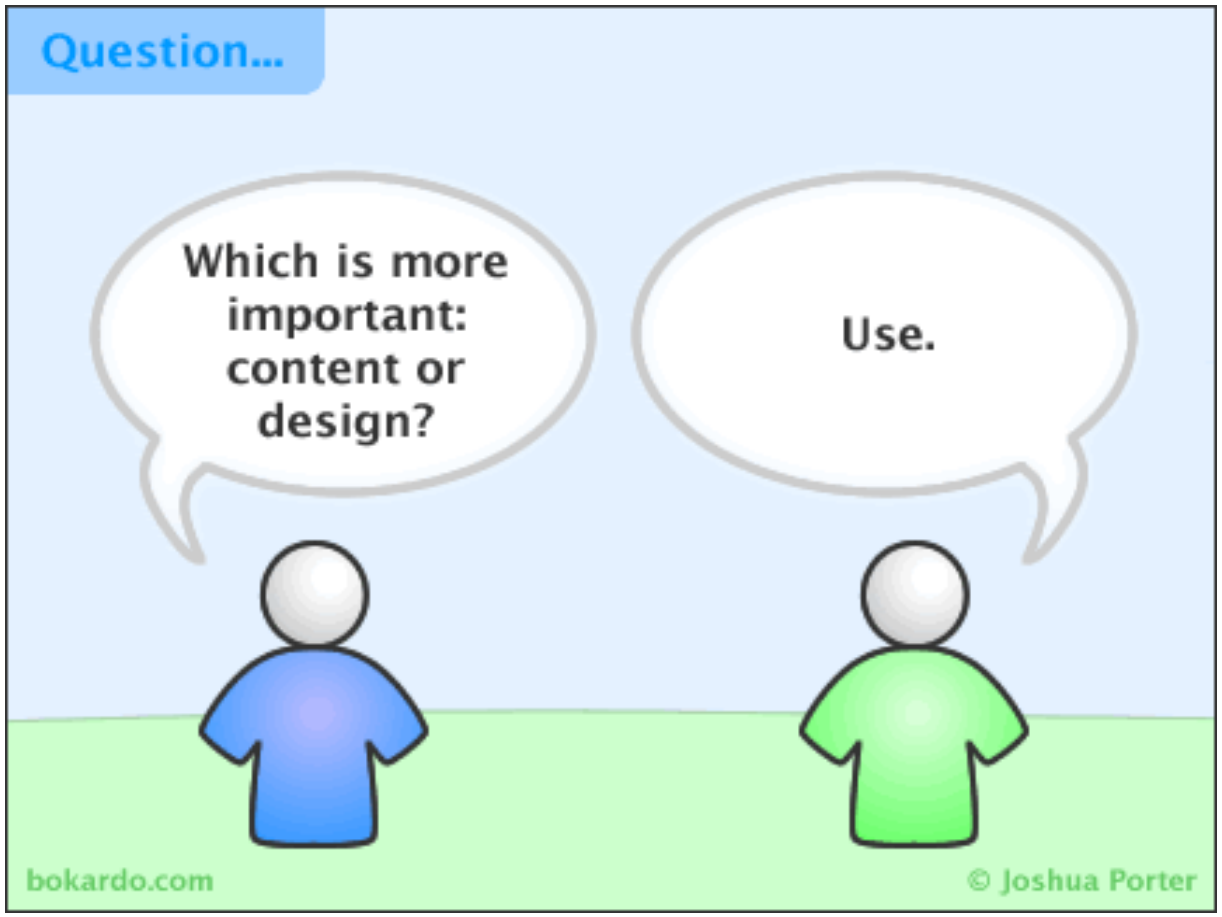
In general:

- Optimal services and future proof solutions
- Ensuring that local architectures and solutions are compatible
- Coordination and knowledge

Objectives (ao) the UvA/SURF project:

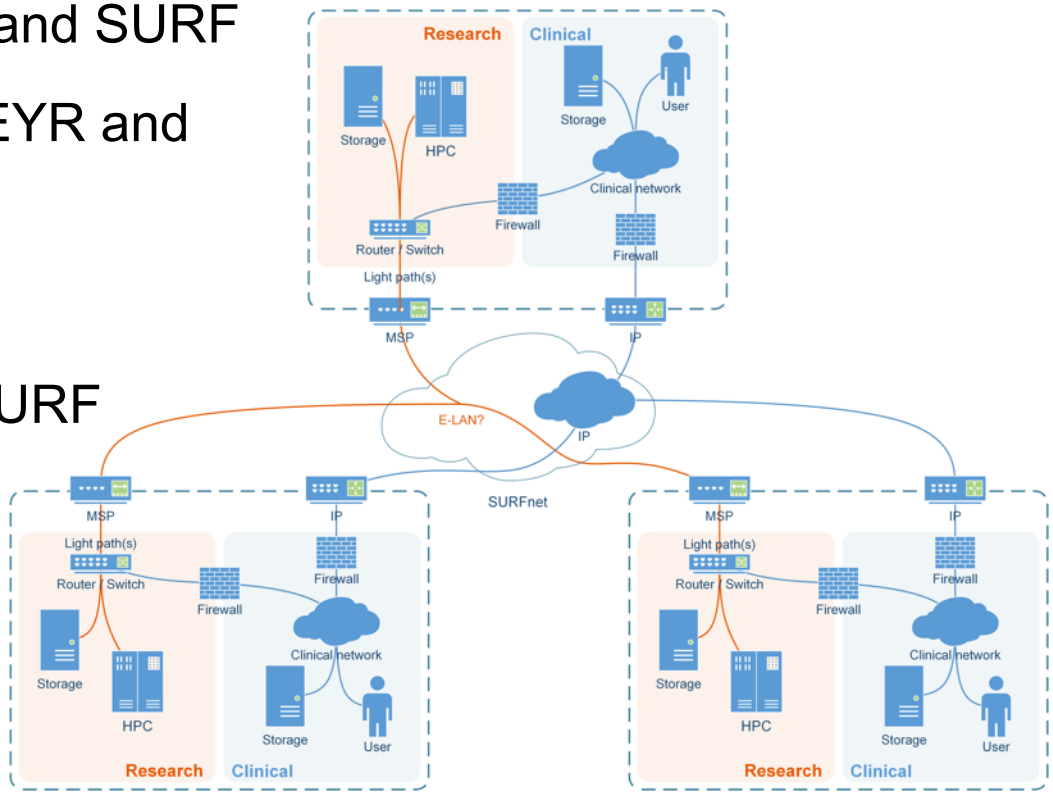
- Blueprint
- Supporting larger numbers of researchers
- Increase scientific output

Responsibility for SURF



Research LAN UMC's

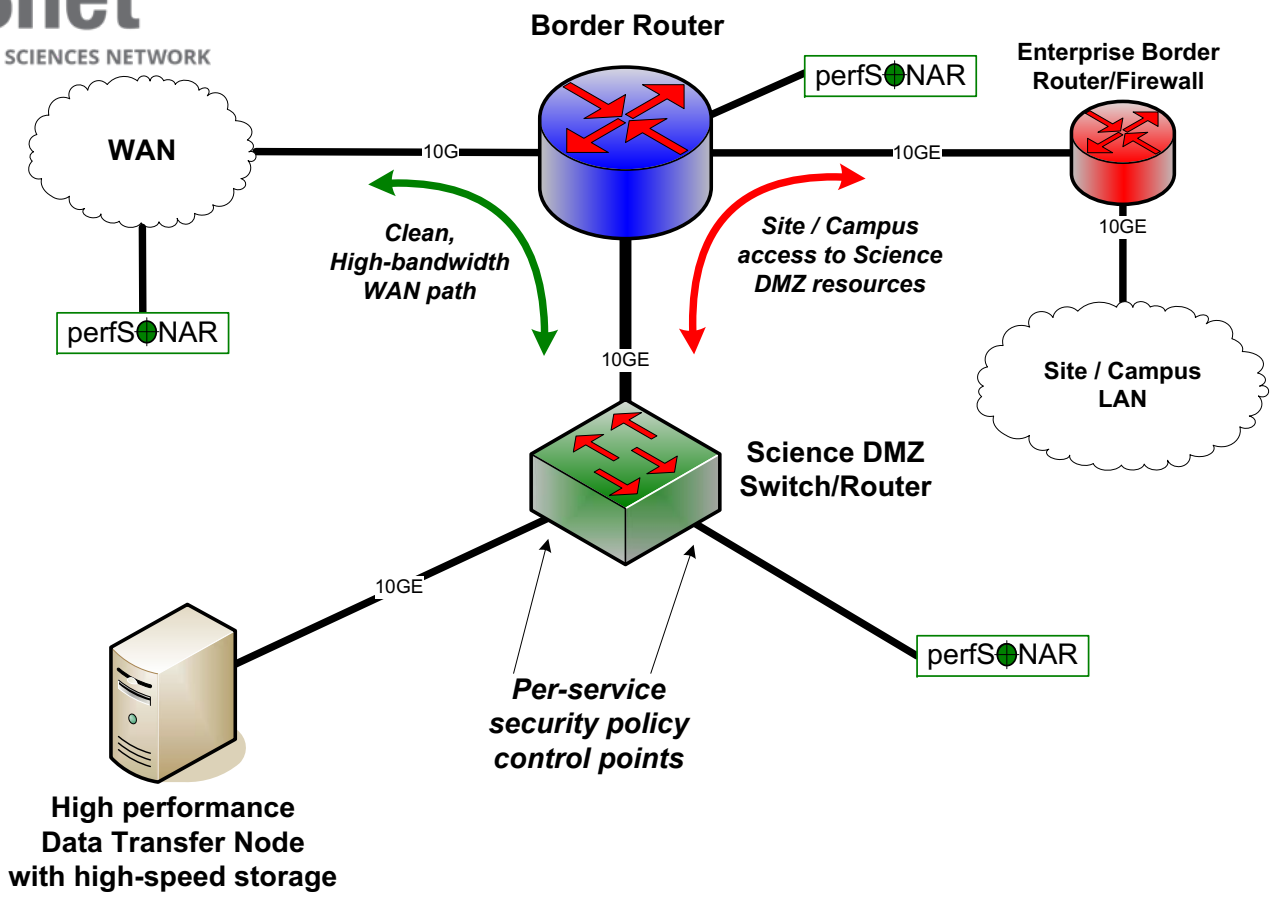
- Specific architecture for UMC's
- Started with LUMC, UMCU and SURF
- Based on experience from EYR and Campus Challenge
- Extension to other UMC's
- Coordination by NFU and SURF



International

ESnet
ENERGY SCIENCES NETWORK

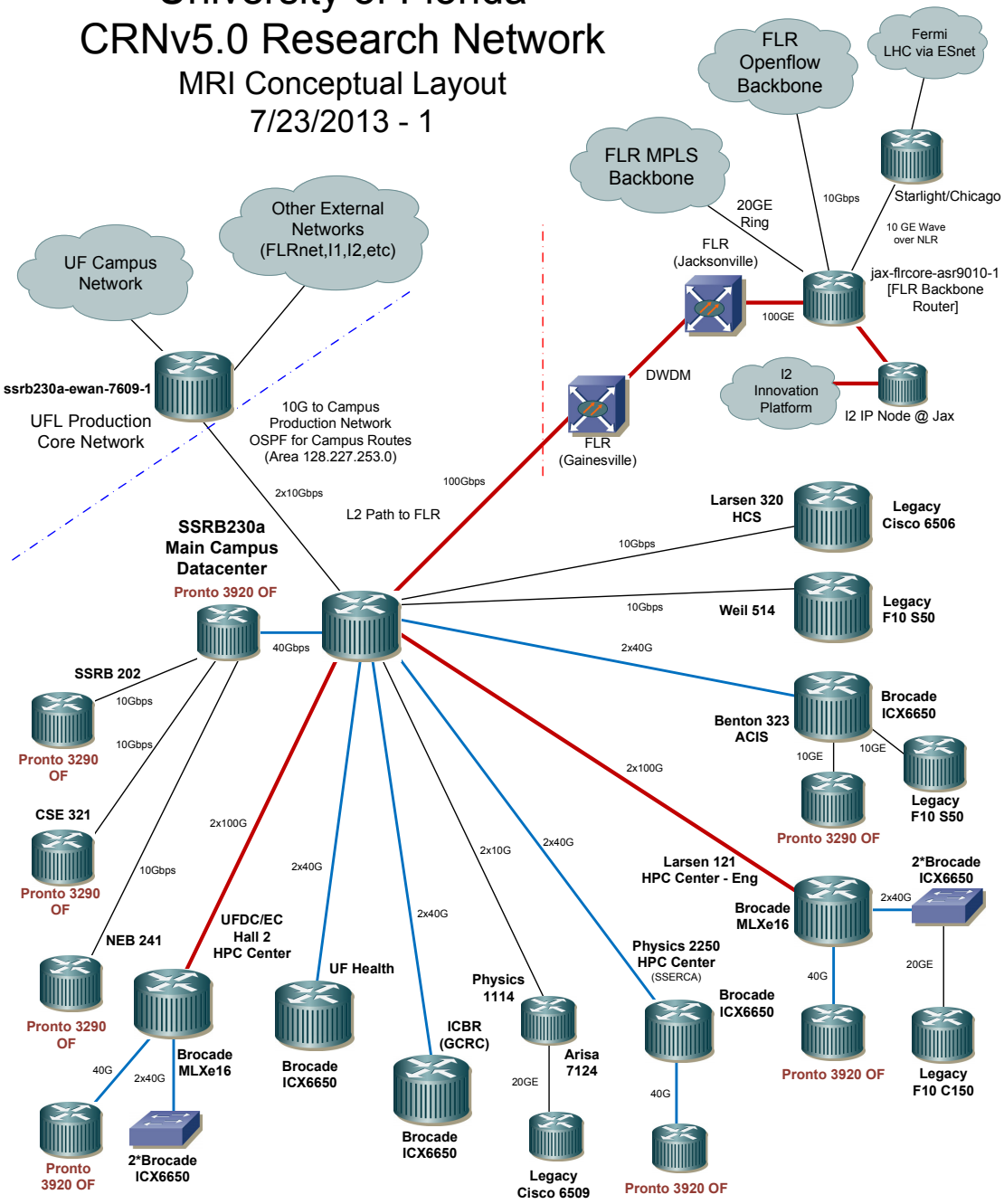
Science & Technology Facilities Council

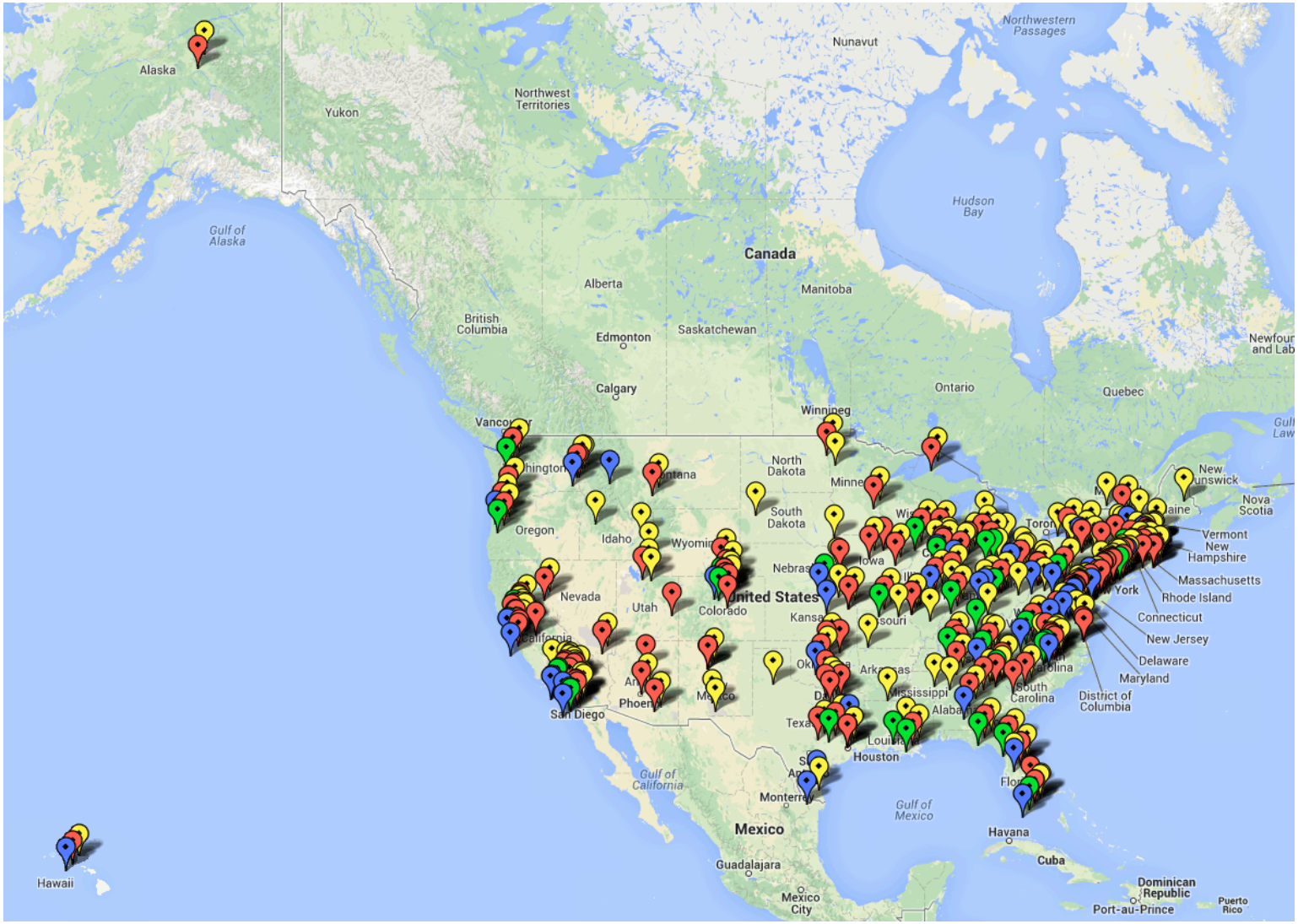


*Fasterdata knowledgebase: http://fasterdata.es.net/science-dmz/

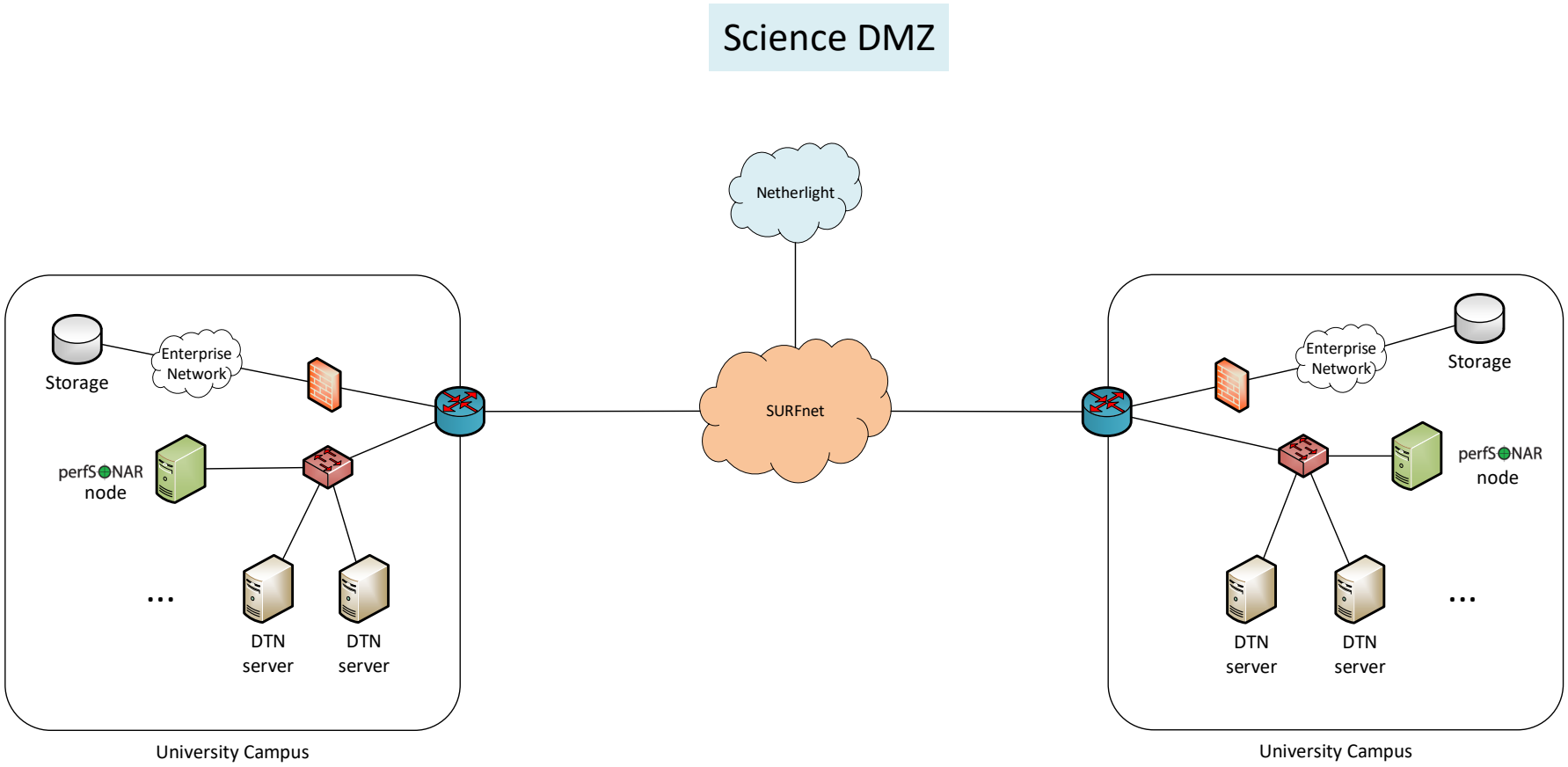
© 2015, The Regents of the University of California, through Lawrence Berkeley National Laboratory and is licensed under CC BY-NC-ND 4.0

University of Florida
CRNv5.0 Research Network
MRI Conceptual Layout
7/23/2013 - 1




Alaska
Yukon
Northwest Territories
Nunavut
Northwestern Passages
Hudson Bay
Gulf of Alaska
Canada
Alberta
Manitoba
British Columbia
Edmonton
Saskatchewan
Calgary
Ontario
Quebec
Winnipeg
Vancouver
North Dakota
South Dakota
Oregon
Idaho
Nevada
Utah
Colorado
United States
Nebraska
Iowa
Kansas
Missouri
Arkansas
Mississippi
Louisiana
Texas
Houston
San Diego
Phoenix
Gulf of California
Monterrey
Mexico
Gulf of Mexico
Havana
Cuba
Guadalajara
Mexico City
Dominican Republic
Port-au-Prince
Puerto Rico
Hawaii
Toronto
New Brunswick
Nova Scotia
Vermont
New Hampshire
Massachusetts
Rhode Island
Connecticut
New Jersey
Delaware
Maryland
District of Columbia
South Carolina

Proposed schematic with SURFnet and universities



References

- Gigaport rapport “Evolutie en Revolutie in de Campusinfrastructuur” SURFnet/Stratix, 2005 http://www.stratix.com/documents/Campusinfrastructuur.pdf
- Greenpaper Science DMZ, SURFnet, 2015
- "Science DMZ", ESnet. URL: http://fasterdata.es.net/science-dmz/
- “Lichtpaden bij het ErasmusMC”, Arnout Terpstra SURFnet, 2015
- LUMC Strategiestudie Research ICT 2018 https://www.lumc.nl/res/att/1364304

Questions?

Thanks for your attention!