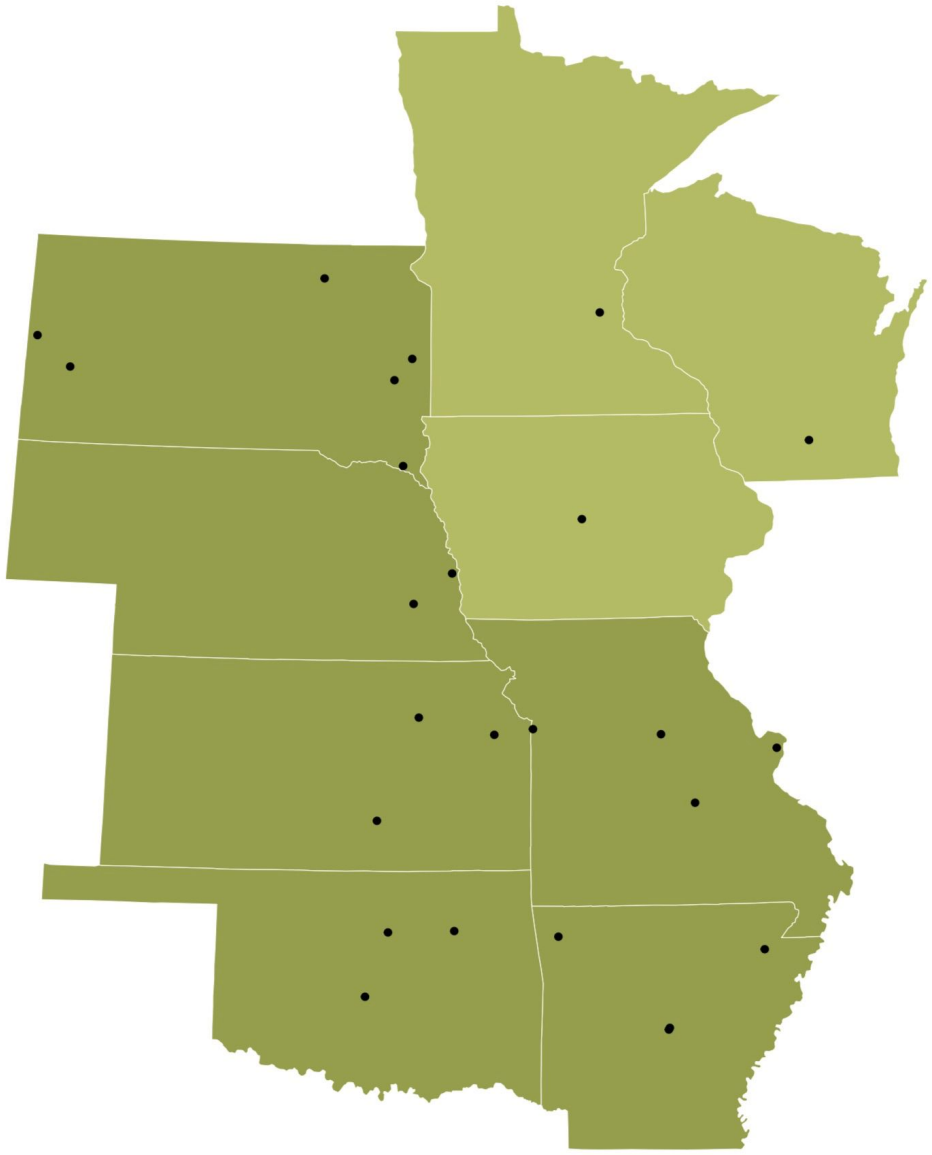
Lessons Learned (So Far) from Developing a Research Platform in the Great Plains

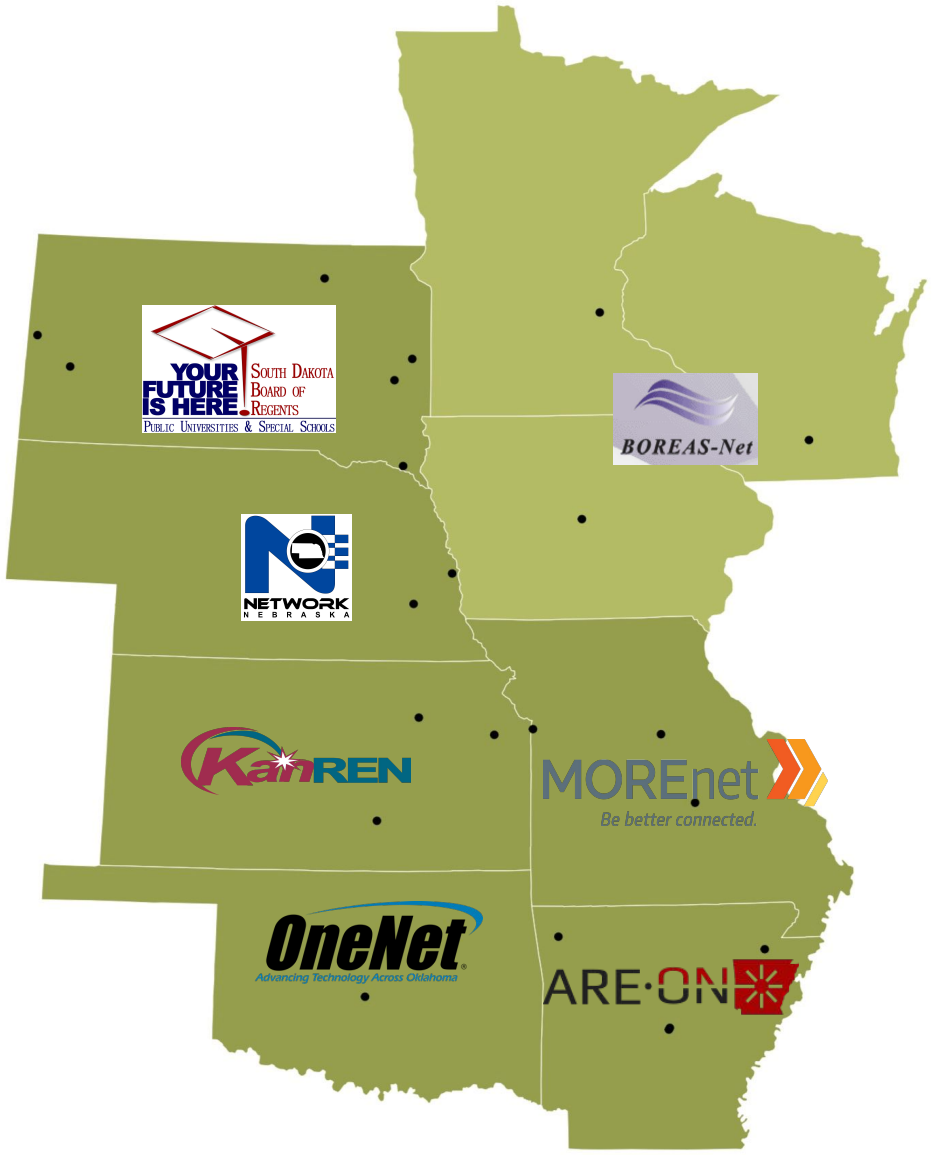
Kate Adams
kate@greatplains.net

Who is the Great Plains Network?



- The Great Plains Network (GPN) is a non-profit consortium aggregating networks through GigaPoP connections while advocating research on behalf of universities and community innovators across the Midwest and Great Plains who seek collaboration, cyberinfrastructure and support for big data and big ideas, at the speed of the modern Internet.

- Over two dozen universities
  - Across 9 states
    - More than 20 years of collaborating in research and education networking



Connected via 6 state networks and a RON

- ARE-ON
- KanREN
- MOREnet
- Network Nebraska
- OneNet
- SD-REED

- BOREAS

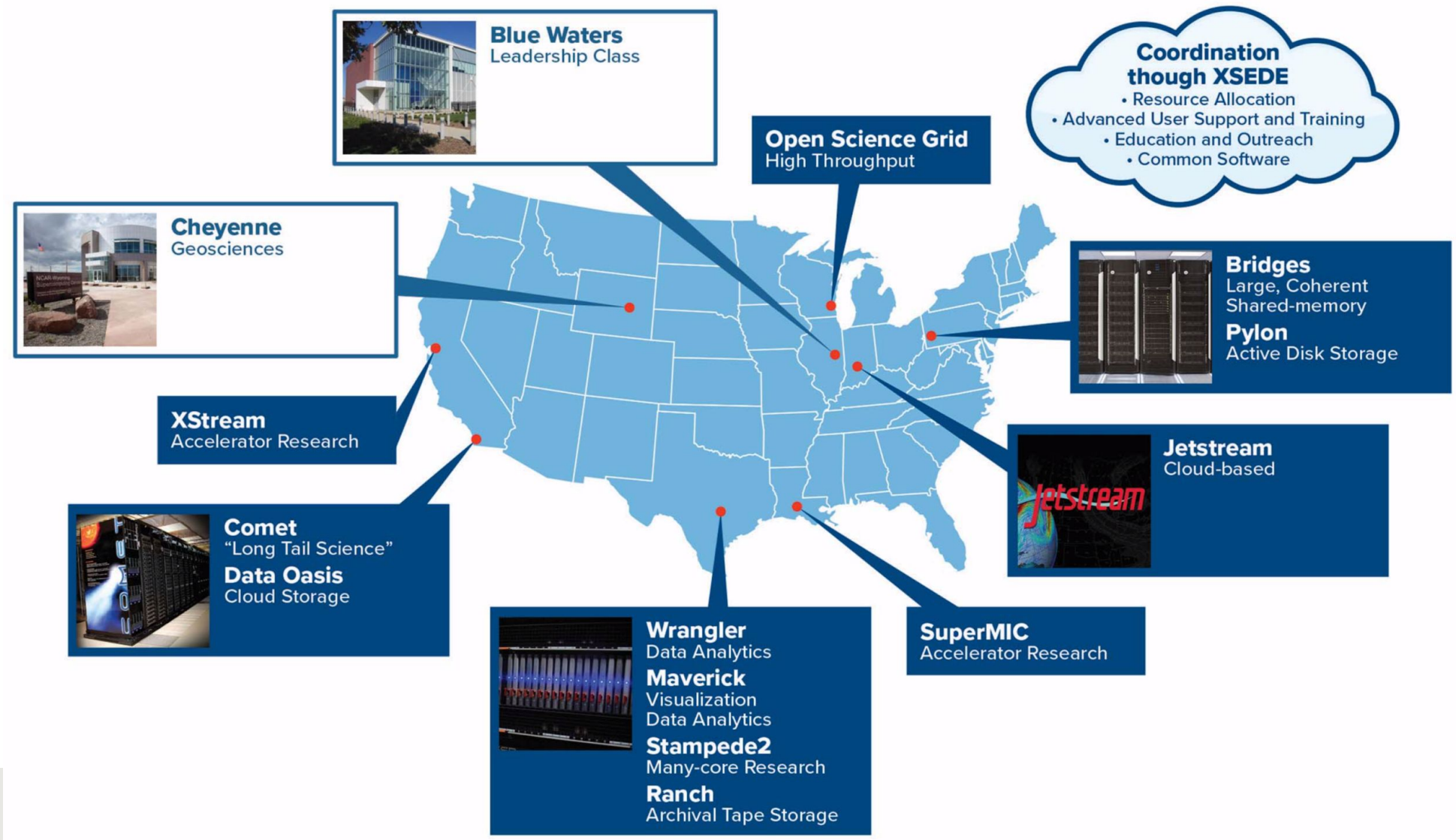
What is a Research Platform?

No one single definition. (Sorry, folks.)

Let's break down the two words, starting with Research.

The following slides show large scale research facilities.

XSEDE (eXtreme Science and Engineering Discovery Environment)



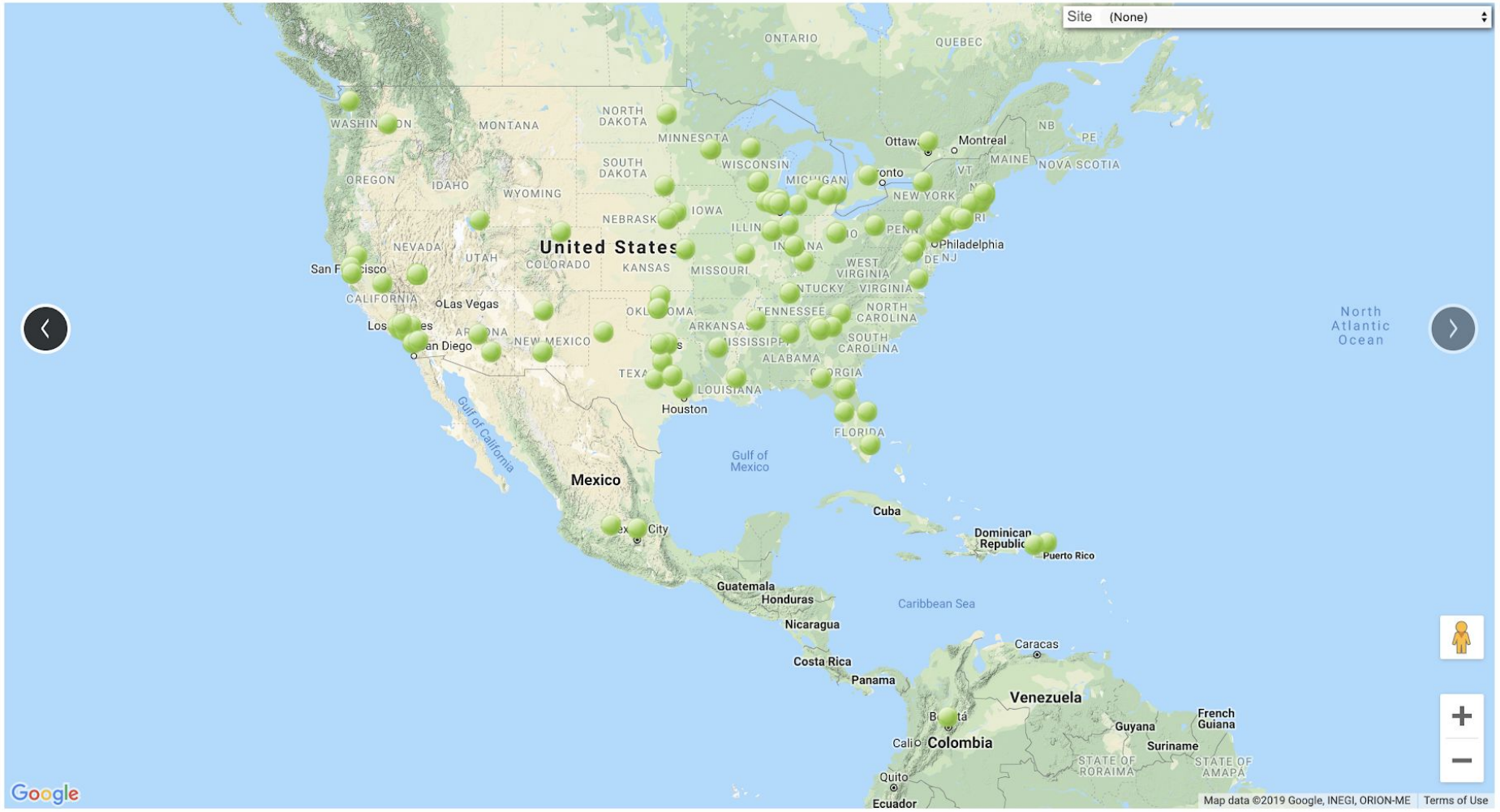
GREAT PLAINS NETWORK

CONNECTING NETWORKS • ADVOCATING RESEARCH • CULTIVATING COMMUNITY

Open Science Grid

A national, distributed computing partnership for data-intensive research

Status Map | Jobs | CPU Hours | Transfers | TB Transferred



In the last 24 Hours	
177,000	Jobs
4,436,000	CPU Hours
3,344,000	Transfers
796	TB Transfers
In the last 30 Days	
6,730,000	Jobs
124,494,000	CPU Hours
131,561,000	Transfers
25,847	TB Transfers
In the last 12 Months	
104,528,000	Jobs
1,635,467,000	CPU Hours
1,767,450,000	Transfers
256,000	TB Transfers

OSG delivered across 114 sites

https://display.opensciencegrid.org/

We need to move the data to these off-site computing centers.

Enter the Science DMZ. Science DMZs protect payroll and student data from research data.

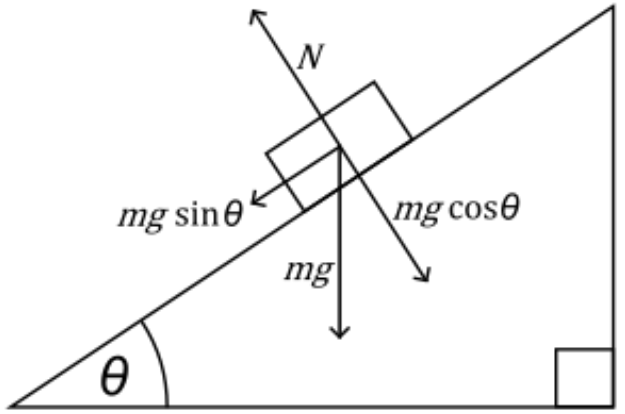
Payroll and student data: secure,
Research: fast
Never the two shall meet.

The Science DMZ* in 1 Slide

Consists of **three key components**, all required:

- "Friction free" network path
  - Highly capable network devices (wire-speed, deep queues)
  - Virtual circuit connectivity option
  - Security policy and enforcement specific to science workflows
  - Located at or near site perimeter if possible
- Dedicated, high-performance Data Transfer Nodes (DTNs)
  - Hardware, operating system, libraries all optimized for transfer
  - Includes optimized data transfer tools such as Globus Online and GridFTP
- Performance measurement/test node
  - perfSONAR
- Engagement with end users

Details at http://fasterdata.es.net/science-dmz/



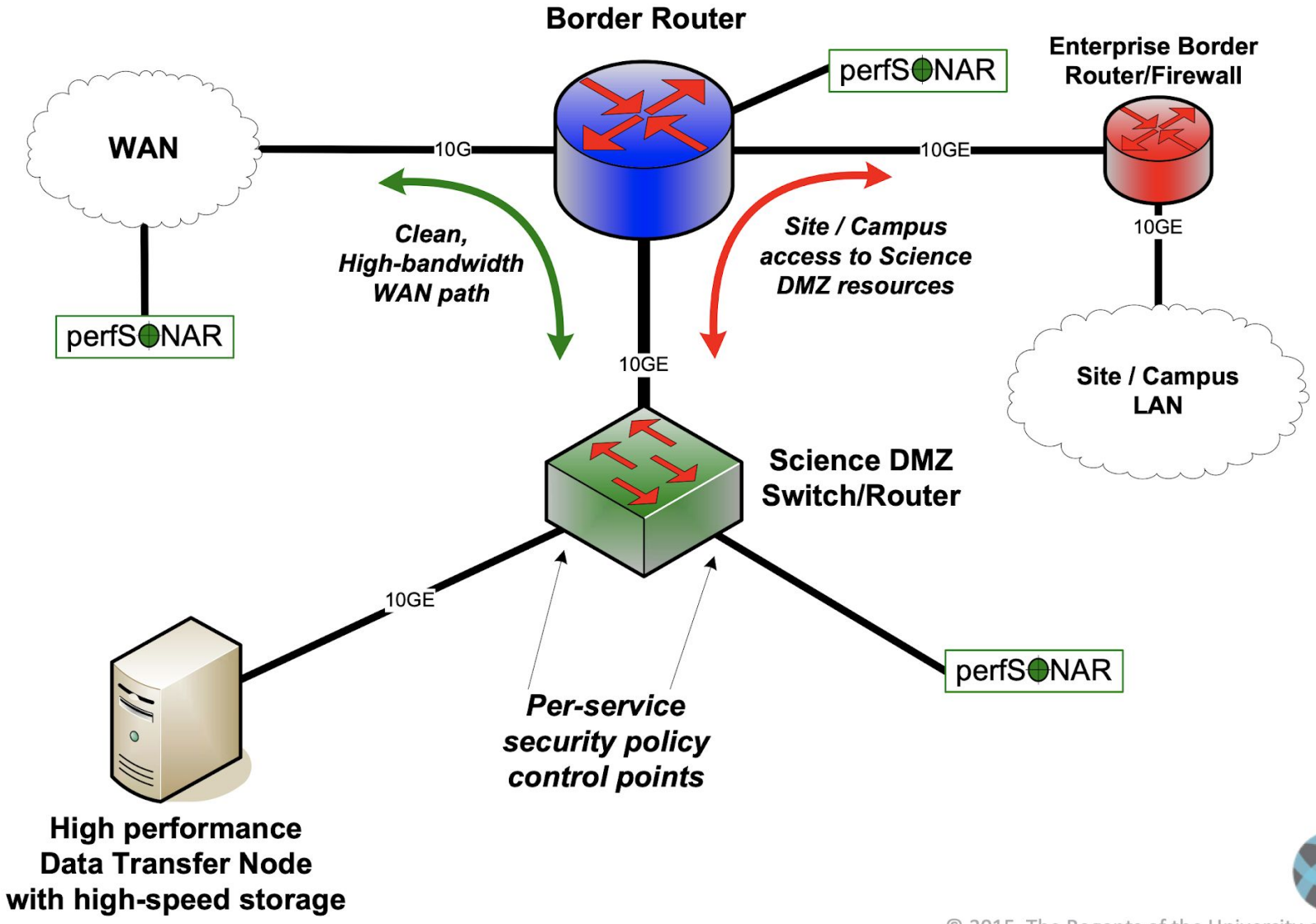
globus

perfSONAR

ESnet

* ***Science DMZ*** is a trademark of The Energy Sciences Network (ESnet)

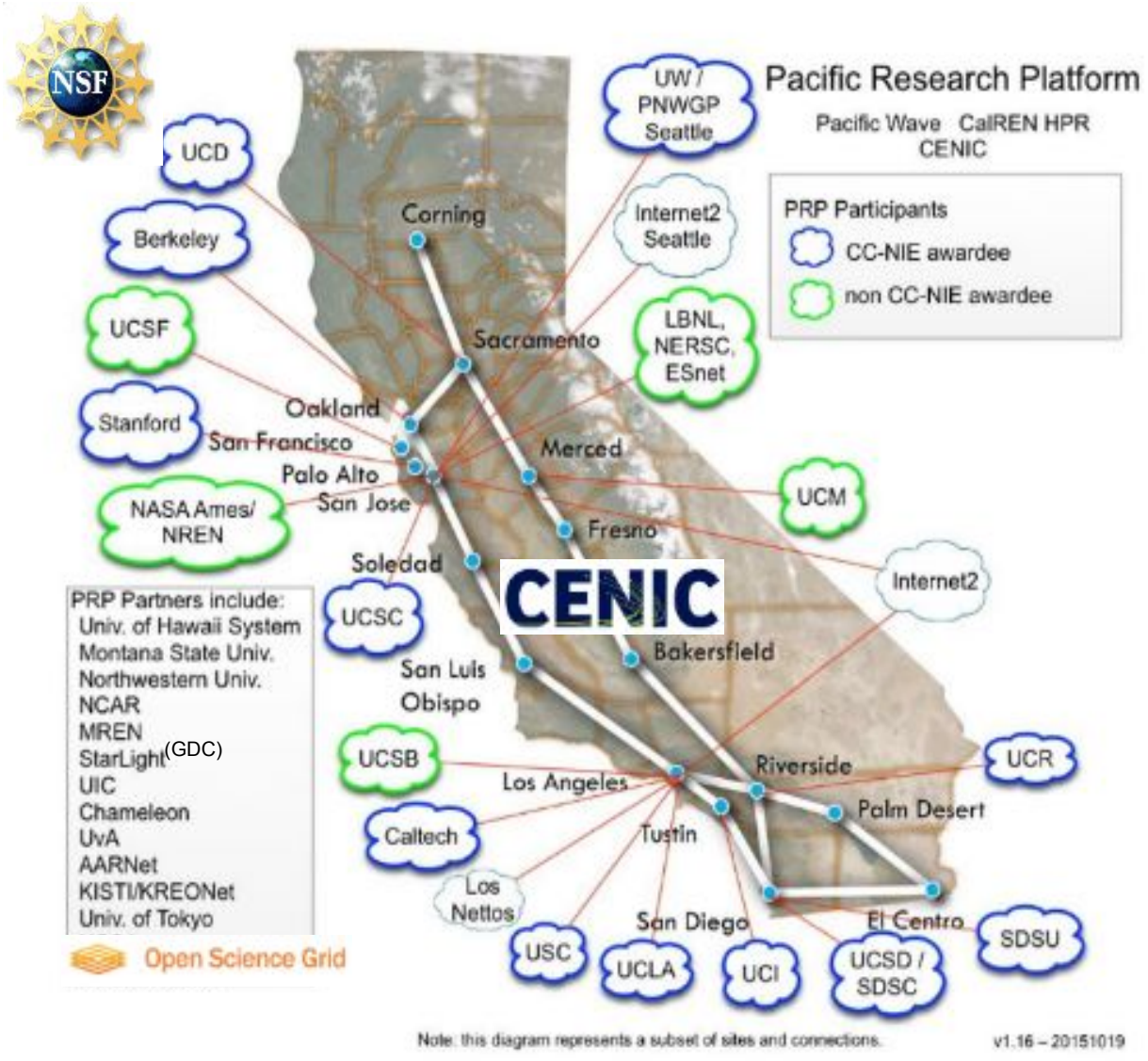
Science DMZ Design Pattern (Abstract)



ESnet

What is the goal with the Science DMZ?

Logical Next Step: The Pacific Research Platform Networks Campus DMZs to Create a Regional End-to-End Science-Driven "Big Data Superhighway" System



Source: John Hess, CENIC, Larry Smarr, Calit2

NSF CC*DNI Grant
$5M 10/2015-10/2020

PI: Larry Smarr, UC San Diego Calit2

Co-PIs:

- **Camille Crittenden, UC Berkeley CITRIS,**
- **Tom DeFanti, UC San Diego Calit2/QI,**
- **Philip Papadopoulos, UCSD SDSC,**
- **Frank Wuerthwein, UCSD Physics and SDSC**

Letters of Commitment from:

- 50 Researchers from 15 Campuses
- 32 IT/Network Organization Leaders

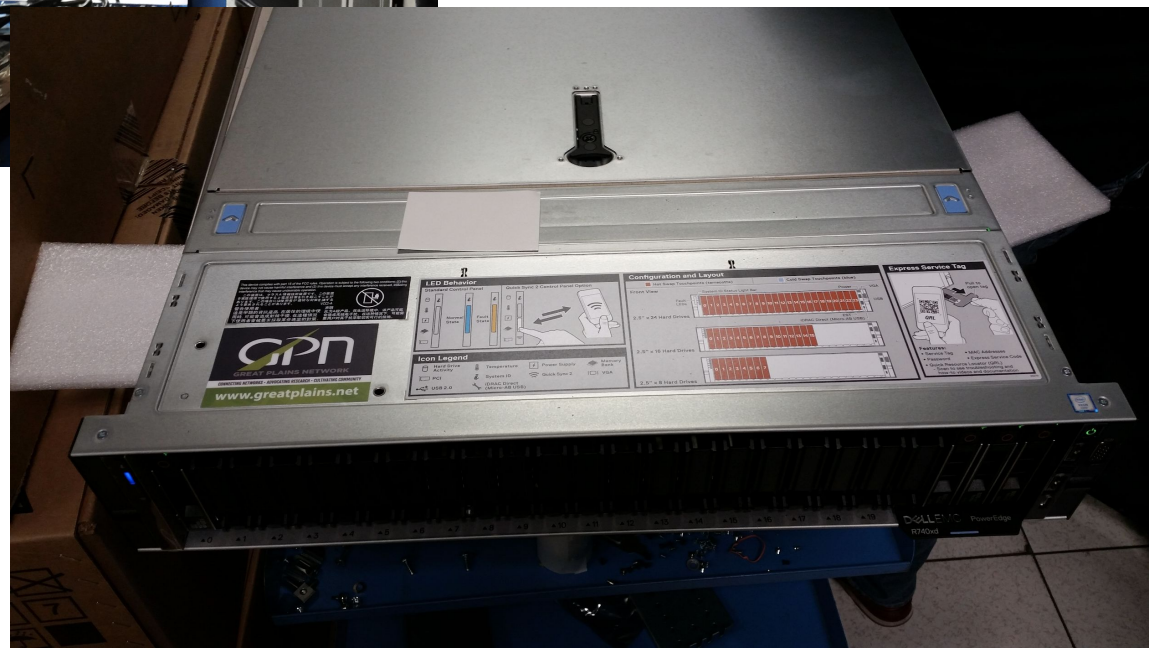
NSF Program Officer: Amy Walton

SDSC

CITRIS BANATAO INSTITUTE

Let’s look at the second word now: Platform


DELL EMC


GPN
GREAT PLAINS NETWORK
www.greatplains.net
LED Behavior
Configuration and Layout
Express Service Tag
Icon Legend


MD14
GPN
DELL EMC

The DMZ is new. Everything else is old.

The network has been around for two decades. XSEDE and OSG predate the idea of a research platform.

So what is a research platform*? There are examples on the next slide. These are platforms built around research...

***lower case**

History of CI Sharing and Facilitation in the Region

ONEOCII
OneOklahoma Cyberinfrastructure Initiative
http://www.oneocii.okepscor.org/

ACI-REF Virtual Residency

SHOW ME CI .org/

GPN Cyberinfrastructure Program Committee (CIP)

THE CARPENTRIES

ENCITE:

Enhancing Cyberinfrastructure by Training and Engagement

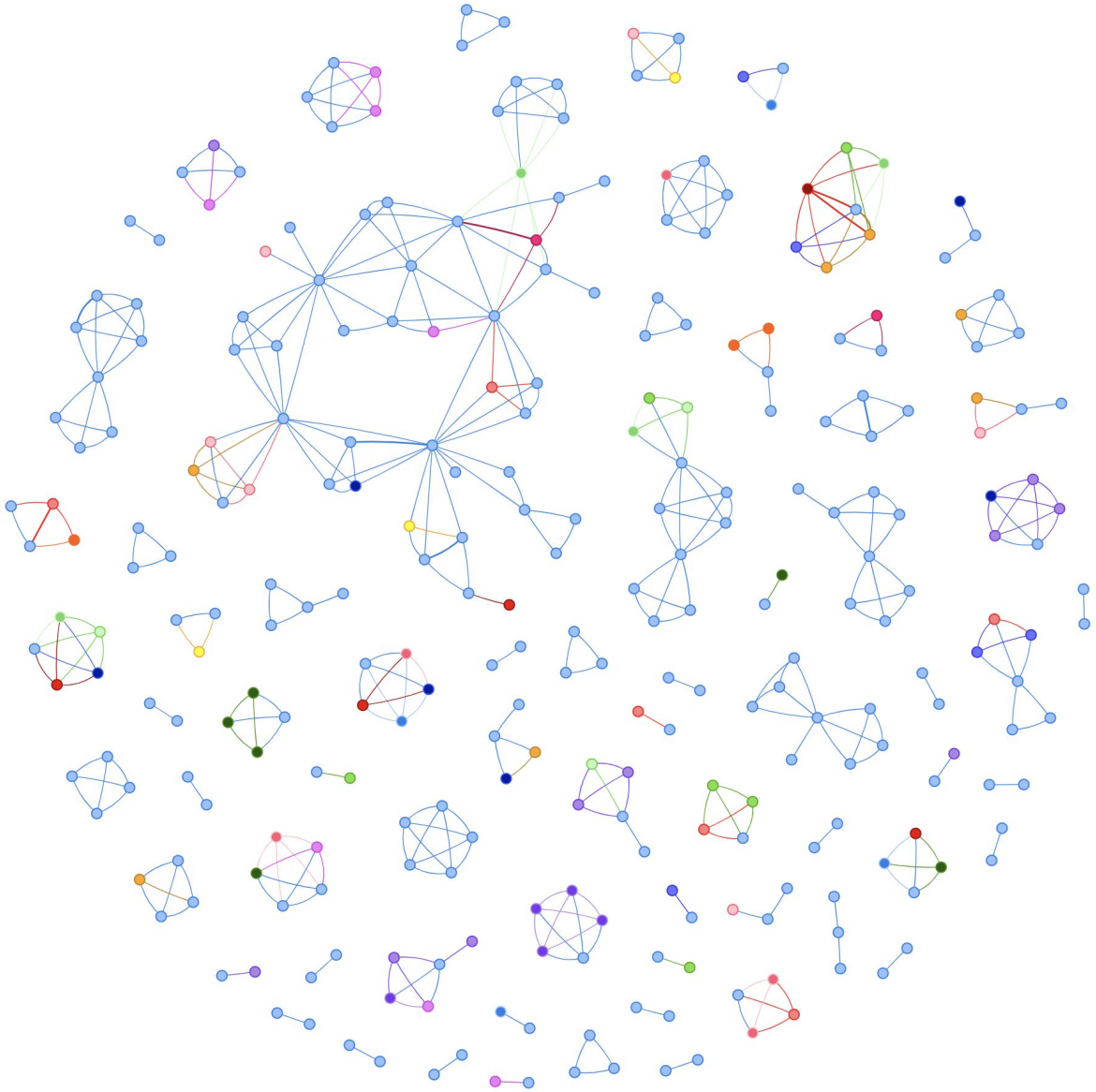
https://www.greatplains.net/archives/presentations/

What is this new Research Platform?

The “R P” in Research Platform
is
Really People

Social Network Analysis of Awards

- **Each dot is a person!** The internet and compute resources augment, not replace, them.
- Reviewing data from NSF and NIH currently
- Text analysis to increase focus on projects with potential greater amounts of data movement
- Color of dots represent the state the PI or co-PI resides
- Lines represent common NSF awards between PIs, thickness represents number of awards between PIs
- As familiarity with the graphs grow, can visually identify interesting inter-institutional and/or interstate collaborations facilitating engagement opportunities



How to leverage/grow a Really People Research Platform

- Regular meetings are essential.
- System administrators, network engineers, HPC directors, research facilitators should attend.
- You can base these meetings around a project, a popular person, or anything else that draws an audience.

How to leverage Really People, continued

- Not every topic is relevant to everyone every week. This is great! The best discussions happen when a sys admin asks a network engineer a question (or vice versa).
- People will surprise you, really!

How do you collaborate?