THE PA SCIENCE DMZ

Wayne Figurelle
Assistant Director
Institute for Computational and Data Sciences (ICDS)
Penn State

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

PA-DMZ PARTNERSHIP

- Partnership
  - Growing research networking connectivity needs in smaller institutions
  - KeystoneREN/KINBER planning grant
  - Penn State land grant mission and ICDS strategic goals

- New Proposal
  - NSF Program (CC): Campus Cyberinfrastructure
  - New Institutions - New Use cases
  - New capabilities
  - Novel approach
  - PI: Wayne Figurelle, Penn State University

Wayne Figurelle
Asst. Director for Innovation and Outreach, Penn State University

Grant Dull
Executive Director, KeystoneREN

Dr. Rick Adkins
Professor and IUP STEAMSHOP Director,
Indiana University of Pennsylvania

Dr. Jason Simms
Research Computing Manager,
Swarthmore College

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

PA-DMZ PROJECT OVERVIEW

- Frictionless Network (DMZ)
  - Goal - establish the foundation for a statewide Pennsylvania Regional Science DMZ (PA-DMZ) that enables and enhances access for under resourced PA institutions of higher education to cyberinfrastructure-based resources and services in support of science driven research and education applications.
- Grant supports
  - Networking hardware and connectivity
  - Installation and support for 2+ years (organizations to provide support years 3-5)
  - Broader Impacts and Research Enablement
- Grant Link: https://new.nsf.gov/funding/opportunities/campus-cyberinfrastructure-cc

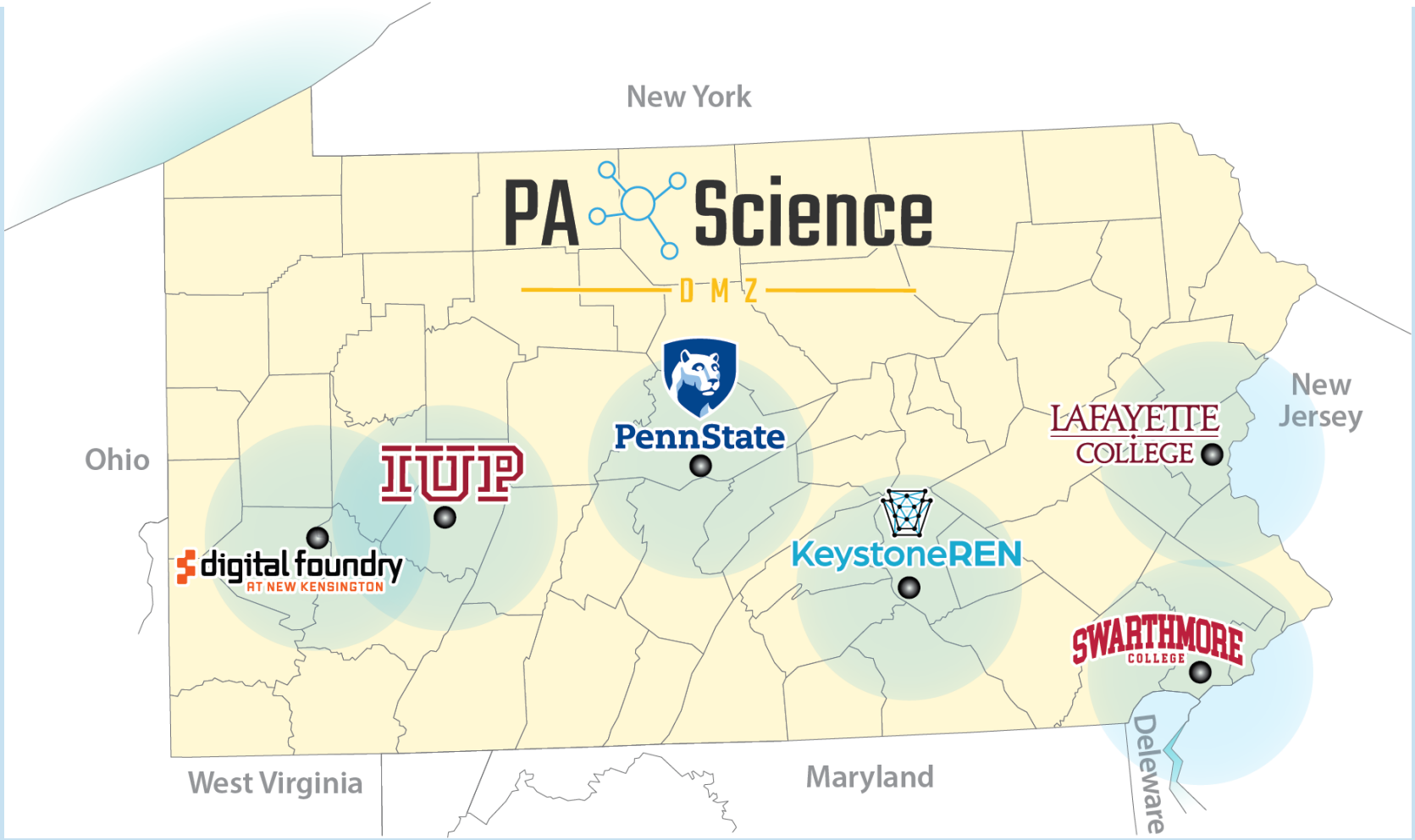
This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

PA-DMZ PROJECT OVERVIEW

- $1.1M funding - NSF Award #2346589
- 6 partners
- Under-resourced Institution Current-Future State
  - Existing 1-2Gb/s Internet only
  - Adding 10/25Gb/s router, 10Gb/s Internet2, with 10G perfSONAR and 10G DTNs
- 2-year award PLUS 1-year NCE
  - 2024 - Team Formation, Site Visits, Requirements, Design, Procurement
  - 2025 - Install and Operational, Research Enablement
  - 2026 - Expansion and Sustainment Plan

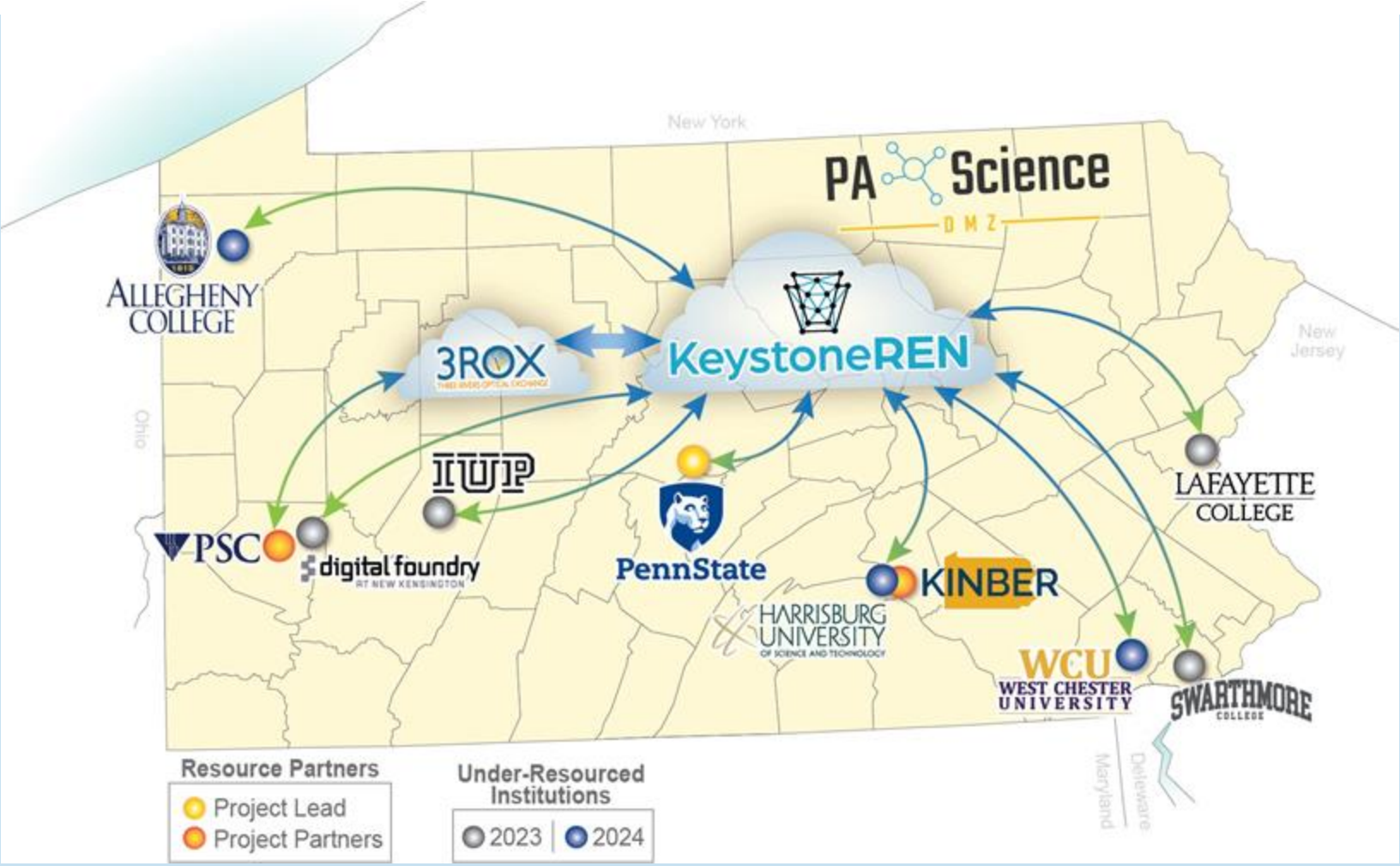
This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

PA-DMZ CURRENT STATE



This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

PA-DMZ NEXT STATE



This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

USE CASE SUMMARY

- Current - 10 Initial Use Cases
  - Biology
  - Chemistry
  - Math and Computer Science
  - Psychology
  - Visual Neuroscience
- Others include:
  - CyberSecurity, Linguistics, Smart Manufacturing

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

SCIENCE DRIVER IMPACT

PRE:

Establish a Baseline by gathering existing data transfer bottleneck or limitations

IMPLEMENTATION:

Measure Total Science Data Transferred

POST:

How has Science Improved both quantitatively and qualitatively?

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

MATH & COMP. SCI. USE CASE

Science Driver Description and Needs

- AI machine learning - training large learning models for classification or prediction.
- 10 TB data sets stored in an old electronics factory - not well-connected to the campus
- Fast connect back to HPC and storage

Key Data/Computational Challenges

- High speed campus bandwidth but access ports in offices/classrooms limited to 1Gig.
- RAID backup option for large datasets, more remote compute options - cloud based storage and remote compute options.

Projected Value

- Data sets will be securely shared outside of IUP to support remote access by IUP students for research and remote training activities with other colleges and local businesses

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

BIOLOGY USE CASE

Science Driver Description and Needs

- Images from a confocal microscope located on campus.
- Using portable hard drives to move data -- Slow and Error Prone
- Generating hundreds of images of 100GB each
- Archive such images automatically after a set time.

Key Data/Computational Challenges

- Use Globus client on that system to watch a directory and move files automatically to HPC for analysis.

Projected Value

- Capture machines and internet sources; currently going to student laptops, etc. but project provides ability to centralize on the HPC

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

CHEMISTRY USE CASE

Science Driver Description and Needs

- Multiple TB of CryoEM data stored on a servers at other institutions
- Transfer to local campus is always failing.
- Project continues to scale making problem worse.

Key Data/Computational Challenges

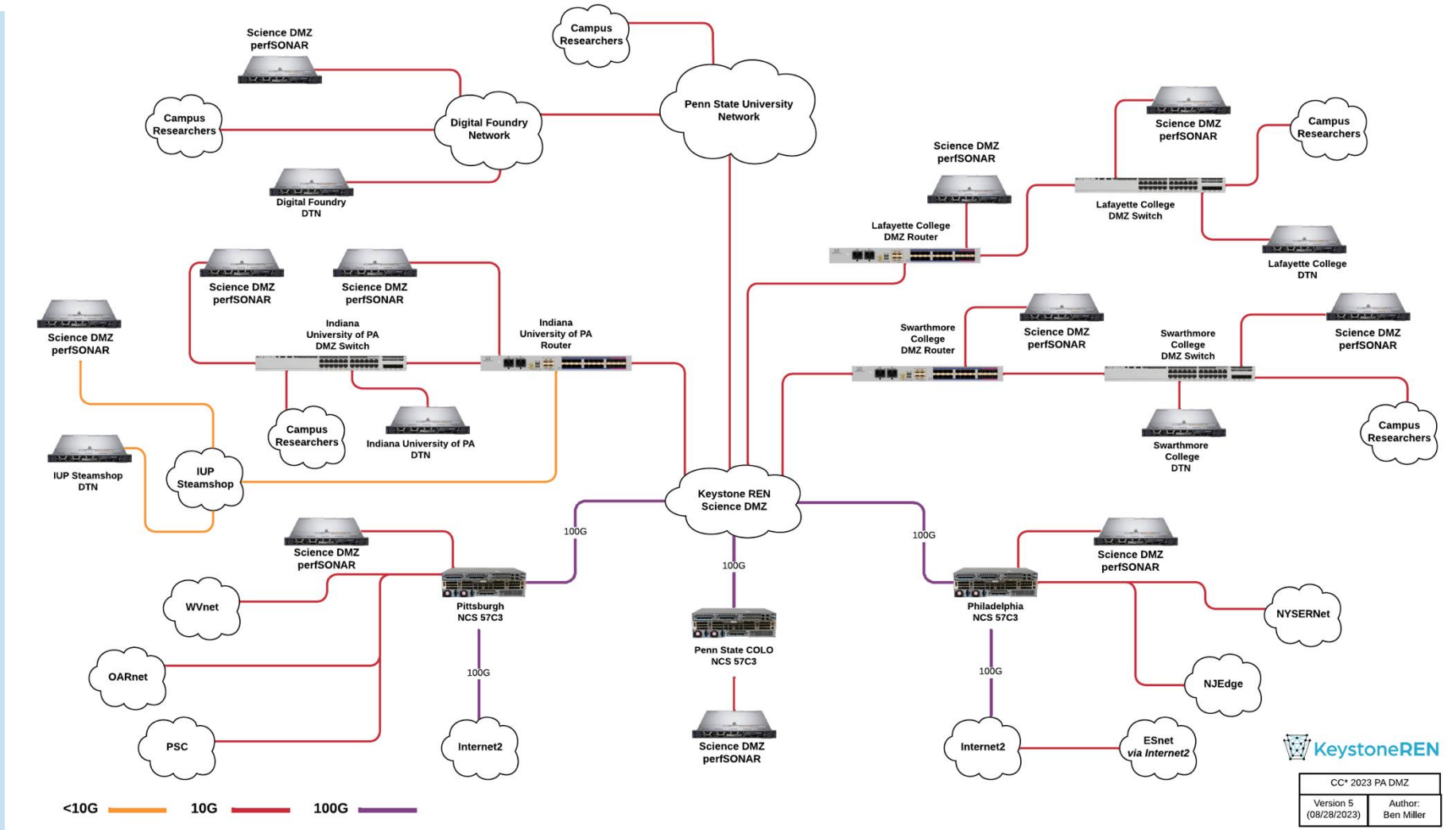
- Need to transfer to local servers to process data on CryoSPARC
- Need local storage for results and raw data

Projected Value

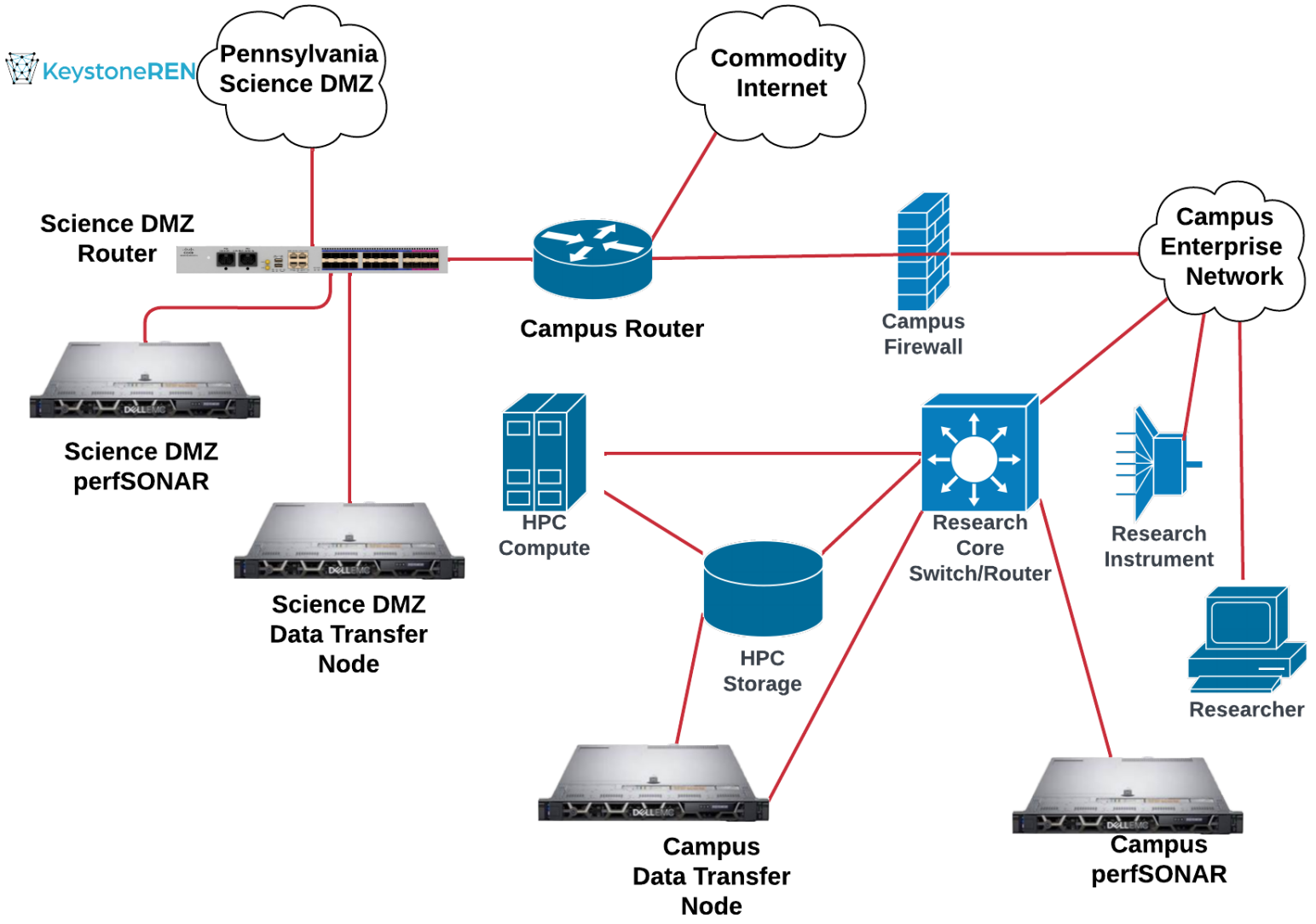
- Enables projects and research to scale enabling larger grant submissions and collaborations

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NETWORK DIAGRAM



NETWORK DIAGRAM



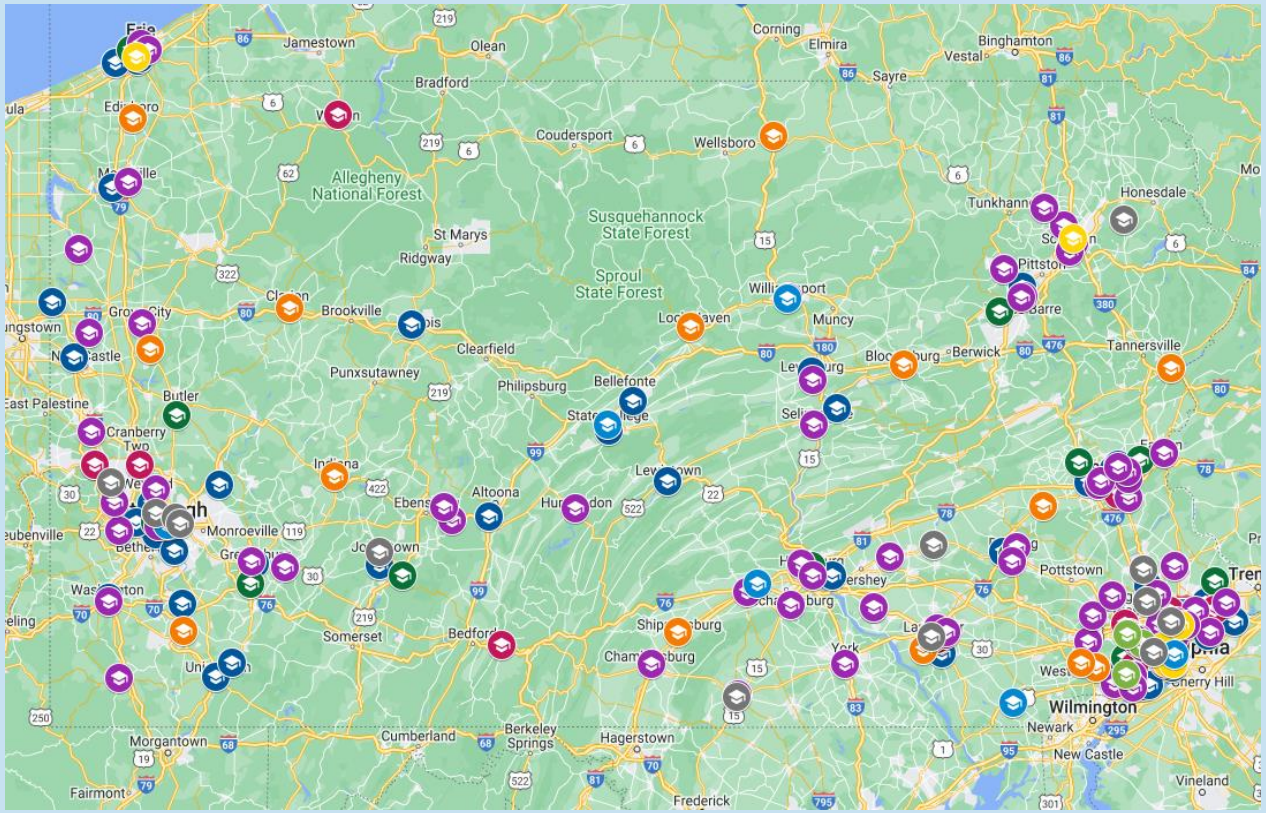
PA-DMZ TECHNICAL OVERVIEW

- CyberSecurity
- perfSONAR testing IPv4 and IPv6
- Data Transfer Node testing
  - Data Transfer Scorecard
- Globus Transfers

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

PA-DMZ PROJECT EXPANSION

- PA has received over 20 CC* awards as of Jan 2024
- Most of the higher ed institutions are under-resourced
- Chance to make a change
  - Conference Presentations
  - CyberAccelerate Workshops – over 80 attendees
  - CyberAccelerate Roadshow
  - IT Training
  - >30 additional institutions contacted



This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

THANK YOU

Questions

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 2346596. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.