The First National Research Platform Workshop:
Toward A National Big Data Superhighway

August 7-8, 2017
Chairs: Jim Bottum and Larry Smarr
Program Chair: Tom DeFanti
Report by Richard Moore
Editing by Tom DeFanti and Camille Crittenden and Molly Wofford

Held at Montana State University
Bozeman, Montana

This workshop was supported in part by NSF Cooperative Agreement, ACI-1541349 to UC San Diego and UC Berkeley. Additional support was received from Montana State University (MSU), the Corporation for Education Network Initiatives in California (CENIC), the Center for Information Technology Research in the Interest of Society (CITRIS), the California Institute for Telecommunications and Information Technology (Calit2), Calit2’s Qualcomm Institute, and Pacific Interface, Inc. Support for the webcasting was made available from MSU’s CC*DNI Award ACI-1541252.
Table of Contents

1 Executive Summary and Recommendations .................................................. 5
2 Workshop Context and Objectives ..................................................................... 6
3 Workshop Sessions and Discussions .................................................................. 8
   3.1 Agenda Overview .......................................................................................... 8
   3.2 Keynote: NSF Cyberinfrastructure Investments and Opportunities: An Update to the First National Research Platform Workshop (Irene Qualters) .................. 9
   3.3 Keynote: The Pacific Research Platform: Leading up to the National Research Platform (Larry Smarr) ........................................................................ 10
   3.4 Keynote: Next-Generation Integrated Architecture: An SDN Ecosystem for LHC and Exascale Science (Harvey Newman) .................................................. 12
   3.5 Session One: Scaling Science Communities: Lessons Learned and Future Plans (Chair, Ruth Marinshaw) ........................................................................ 13
      3.5.1 Advanced Cyberinfrastructure – Research and Educational Facilitation (ACI-REF): Campus-Based Computational Research Support (Tom Cheatham) .......... 14
      3.5.2 Scaling Science Communities: Lessons Learned by and Future Plans of the Open Science Grid (Frank Würthwein) ................................................................... 14
      3.5.3 Strategies for Scaling Science: an ENet View (Inder Monga) ..................... 15
      3.5.4 Software infrastructure for a National Research Platform (Ian Foster) ............. 16
      3.5.5 SLATE: Services Layer at the Edge (Rob Gardner) ...................................... 17
      3.5.6 Session Q&A .......................................................................................... 18
   3.6 Deep Dive Sessions 1–4 ................................................................................. 19
      3.6.1 Deep Dive One: Roles of Regionals, QUILT, LEARN, NYSENET, FLR, MREN, PNWGP & Others (Moderators: John Moore & Ronald Hutchins) ....................... 19
      3.6.2 Deep Dive Two: Scaling and Security: Not at the App Level, Please! (Moderators: Philip Papadopoulos & Von Welch) ...................................................... 19
      3.6.3 Deep Dive Three: Opportunities in Research Cyberinfrastructure at NSF (Moderator: Amy Walton) ................................................................................. 20
      3.6.4 Deep Dive Four: How Does PRPv1 Work? – An Explanation of DMZs, DTNs... (Moderator: Celeste Anderson) ........................................................................ 20
   3.7 Session Two: What Scaling Means to the Researcher (Chair: Patrick Schmitz) 21
      3.7.1 Cyberarchaeology, Immersive Visualization, and the UC Catalyst Project (Chris Hoffman) ................................................................................................. 21
      3.7.2 OsIRIS: Scalable, Computable, Collaborative Storage (Shawn McKee) .......... 22
      3.7.3 Necessity Is the Mother of Invention: An End User’s Quest to Scale up the Cyberinfrastructure Needed to Move & Process Exabytes of Complex Genetic Systems Data (Alex Feltus) .......................................................... 23
      3.7.4 SciDAS Underneath (Claris Castillo) .......................................................... 24
      3.7.5 Scaling Up Water Science Research (Ray Idaszak) ...................................... 24
      3.7.6 Session Q&A .......................................................................................... 25
   3.8 Session Three: Engaging Communities of Researchers (Chair: Tom DeFanti) 27
      3.8.1 ENet Science Engagement; Eli Dart .......................................................... 27
      3.8.2 NCAR Climate Data Creation, Storage, Analysis, and Distribution Engagement Strategies (Marla Meehl) ................................................................. 28
      3.8.3 Engaging Communities of Researchers: Experience from the PRP; Camille Crittenden .................................................................
CHASE-CI: Building a Community of Machine Learning Researchers on the Pacific Research Platform (Tom DeFanti) ................................................................. 29
Session Q&A ........................................................................................................ 29

3.9 Keynote: The Need for Big Data, Networking, Storage, and Machine Learning (Exascale Numerical Laboratories) (Alex Szalay) .......................................................... 30

3.10 Special Session: High-Speed Equipment (Chair: John Hess) ................................................................. 32
3.10.1 Science DMZs Behind Firewalls: Heresy? (J.J. Jamison) ................................................................. 32
3.10.2 Arista Network Equipment Deployment at USC (Azher Mughal) ......................................................... 33
3.10.3 Software-Defined Access: Enabling the Digital Transformation (Tim Martin) .......................... 34
3.10.4 Session Q&A .................................................................................................................. 34

Session Four: NRP Impact on Big and Small Campuses (Chair: Gil Gonzales) ........................................ 35
3.10.5 Supporting the NRP with a Lean Cyberinfrastructure Staff (Jerry Sheehan) 35
3.10.6 NRP Medical School Challenges (Tracy Futhey) ..................................................................... 36
3.10.7 A Science DMZ in Every Pot? (Von Welch) ............................................................................. 36
3.10.8 What Measurements Can We Report to Show Engagement? (John Hess) ............................. 37
3.10.9 Engaging Tribal Colleges and Universities in R&E Networking (Jason Arviso) ...................... 37
3.10.10 Session Q&A .................................................................................................................. 38

3.11 Deep Dive Sessions Five–Eight ....................................................................................... 39
3.11.1 Deep Dive Five: Creation of Community Data Resources (Moderators: Jerry Sheehan and Alexander Szalay) .................................................................................. 39
3.11.2 Deep Dive Six: What's Worked and What Hasn't for High-Performance Networking (Moderator: Richard Moore) ................................................................. 40
3.11.3 Deep Dive Seven: Strategies for Moving Forward: How to Build A Network of Regional DMZs (Moderators: Jim Bottum and Larry Smarr) ........................... 42
3.11.4 Deep Dive Eight: SC17 Demos (Moderators: John Graham and Azher Mughal) 45

3.12 Session Five: Towards a Global Research Platform (GRP) (Chair: Joe Mambretti) ................................................................. 46
3.12.1 Toward a Global Research Platform (GRP) – DMZs in Korea (Jeonghoon Moon) 46
3.12.2 LSST Scaling Issues and Network Needs (Heidi Morgan) .................................................. 47
3.12.3 Pacific Rim Application and Grid Middleware Assembly (PRAGMA) (Phil Papadopoulos) .................................................................................................................. 48
3.12.4 International Options: How can we scale? (Jennifer Schofield) ............................................ 48
3.12.5 Session Q&A .................................................................................................................. 49

3.13 Session Six: Democratizing Collaboration (Chair: Maxine Brown) ................................................. 50
3.13.1 Network Startup Resource Center (Steve Huter) ....................................................................... 50
3.13.2 Regional Research Platforms: Regional Example – The Virtual Data Collaboratory; Wendy Huntoon ........................................................................................................ 51
3.13.3 Democratizing Collaborations: Equity and Access (Gil Gonzales) ........................................... 52
3.13.4 Role of the Regional Networks in Scaling a National Research Platform (Jen Leasure) 52
3.13.5 Global Collaborative Research Groups (CRGs) (Cees de Laat) ............................................... 53
3.13.6 Session Q&A .................................................................................................................. 54

3.14 Closing Session: What's Next (Jim Bottum and Larry Smarr) ................................................. 55

4 Findings ................................................................................................................................. 56
5 Acknowledgements ............................................................................................................... 61
6 Appendices......................................................................................................................... 63
  6.1 Workshop Presentation Materials and Videos......................................................... 63
  6.2 Selected Attributed Opinions of the Participants.................................................... 63
  6.3 Workshop Agenda (Original).................................................................................... 72
  6.4 Workshop Registrants ............................................................................................... 76
1 Executive Summary and Recommendations

Scientific exploration and discovery are enabled by increasingly specialized information technology infrastructure. The requirements of a wide array of scientific research domains provide challenges that regularly exceed the capabilities of existing infrastructure, necessitating a continuous renewal of existing services, architecture, and technologies. This process must not simply provide incremental improvement but truly transform current capabilities. The workshop to launch the National Research Platform focused on examining opportunities for designing and implementing future-oriented capabilities for multiple science research communities.

Leading up to this initiative is the Pacific Research Platform (PRP), an ambitious project driven by the high-speed networking needs of collaborative, big-data science. Funded in 2015 by a $5-million grant from the National Science Foundation, the PRP is creating a researcher-defined and data-focused network whose requirements are driven by direct engagements with sophisticated, cyberinfrastructure-knowledgeable science teams chosen from several data-intensive fields, including particle physics, astronomy, biomedical sciences, earth sciences, and scalable data visualization. The PRP is a partnership of more than 25 institutions, including four National Science Foundation, Department of Energy, and NASA supercomputer centers. The purpose of this workshop was to bring together representatives from PRP partners and outside institutions, including domain scientists, network and system administrators, campus CIOs, regional network leaders, and representatives of ESnet, Internet2, the Quilt, XSEDE, and the National Science Foundation (NSF) to discuss expanding the PRP and address the potential challenges and benefits of scaling the Science DMZ model to a national level and creating a National Research Platform (NRP).

The NRP workshop was held August 7–8, 2017, at Montana State University in Bozeman, Montana. Day one of the workshop began with a welcome address from Irene Qualters, Director of the Office of Advanced Cyberinfrastructure at the NSF who stressed the need for bold ideas for advancing the nation’s cyberinfrastructure. Qualters’ welcome was followed by a brief history of the PRP presented by its Principal Investigator Larry Smarr, and a keynote address from Harvey Newman at Caltech, a scientist who pushes the boundaries of advanced networking and cyberinfrastructure. These three introductory presentations set the tone for the rest of the day’s discussions, which focused on scaling strategies and challenges to creating a national platform, as well as the need for science engagement with communities of researchers. Day two of the event included a special session on high speed equipment and capabilities followed by discussions on the impact of an NRP on campuses both big and small, the creation of a Global Research Platform (GRP) with reports from international partners, and the need to cultivate collaborations by providing broad access to data. All sessions included time for questions and comments from the audience.

In addition to the six main sessions that made up the workshop, each day included several optional Deep Dive sessions held during the lunch hour. The less formal Deep Dives provided opportunities for participants to learn and discuss the finer details of creating and scaling a regional research platform. Topics included: explanations and demos of software and hardware, security issues, an introduction to regional platforms, and how to create community data resources. The final day of the workshop concluded with a lively discussion about the next steps needed to establish a functioning National Research Platform.
This first NRP workshop successfully gathered key stakeholders, including administrators, scientists, network engineers and researchers, interested in creating such an infrastructure. Participants were deeply engaged and committed to ascertaining the science-driven requirements of the project and discussing ideas and priorities its progress. (See extensive text in Section 4 and Appendix.)

Based on the extensive Findings laid out in Section 4, the workshop led to the following recommendations:

- The NSF should not pursue just incremental changes over the next 3-5 years, but rather transformative changes. U.S. campuses cyberinfrastructure capabilities have advanced greatly under NSF’s Campus Cyberinfrastructure (CC*) programs, laying the foundation for an emerging NRP.

- Specifically, we recommend that NSF:
  
  - Continue funding its Campus Cyberinfrastructure program, including more opportunities for campus cyber-engineers and cyber teams.
  
  - Issue a call for proposals to address the tough technical and sociopolitical issues that will need to be addressed in an NRP; in fact, this could be an excellent opportunity for DOE and NSF to work together on a joint program.
  
  - Engage science engagement facilitators on campuses to collaborate with each other across campuses in support of building the NRP.
  
  - Support more regional Science DMZs to be formed from existing CC* grants to campuses. Multi-campus networking organizations should be encouraged to take the initiative to create regional DMZ proposals, including campuses that have not previously received NSF CC* grants. Parallelism is the key to scaling rather than central control.

2 Workshop Context and Objectives

For the last five years, the National Science Foundation (NSF) has made a series of competitive grants to over 100 U.S. universities to aggressively upgrade their campus network capacity for greatly enhanced science data access. NSF is building on that distributed investment by funding a $5-million, five-year award to UC San Diego and UC Berkeley to establish the Pacific Research Platform (PRP), a science-driven high-capacity data-centric “freeway system” on a large regional scale. Its goal is to give data-intensive researchers at participating universities and related institutions the ability to move data 1,000 times faster compared to speeds on today’s inter-
campus shared Internet. The PRP does this by federating campus DMZs into a regional DMZ, a concept developed by the Department of Energy in 2010 and adopted by the NSF in its CC-* solicitations.\(^2\) Documentation, including reports and materials resulting from all PRP workshops are archived.\(^3\)

When the PRP was proposed and awarded, it was anticipated that successes and lessons learned on a regional scale could eventually inform extensions (under other future awards) to a national scale. In this regard, the PRP Cooperative Agreement provided that

> “After approximately 18 (or TBD) months, a site visit and comprehensive review of progress towards meeting project milestones and goals and overall performance and management processes will take place, including user community relationships, scientific impacts, and the status of the project as a model for potential future national-scale, network-aware, data-focused cyberinfrastructure attributes, approaches, and capabilities.” [Emphasis added]

This report documents the First National Research Platform (NRP) workshop held at Montana State University in Bozeman, Montana August 7–8, 2017. The purpose of this workshop was to bring together representatives from interested institutions, both national and international, to discuss implementation strategies for scaling the PRP to a national scale in order to realize a vision for a national research platform to advance data-intensive science and engineering research. The workshop involved diverse stakeholders, including domain scientists, network and system administrators, campus CIOs, regional network leaders, and representatives of ESnet, Internet2, the Quilt, XSEDE, and the National Science Foundation, as well as international networks and universities. Sessions of the NRP workshop were devoted to science-driver application researchers, describing their needs for high-speed data transfer, including their successes and frustrations. Discussions focused on requirements from the domain scientists and the networking architecture, policies, tools, and security necessary to scale up from the current regional-focused PRP that connects ~25 institutions to a National Research Platform that connects ~200-institutions.

Workshop objectives included:

- Engage potential NRP teams and collaborators and facilitate in-person interactions (the “other kind of networking”)
- Gain a common understanding of the science engagement process
- Explore and capture data storage, computing, and networking requirements across scientific domains
- Identify common science-driven technical requirements for the NRP
- Exchange technical ideas for scaling the NRP, including
  - Network architecture
  - Data transfer nodes
  - Software and tools
  - Security


Near-term and longer-term NRP capabilities

It is clear from the PRP and many historical efforts that a successful NRP will require a successful intersection of two networks – not only high-performance data networks, but also productive human networks among diverse stakeholders. Scientists need to pursue publishable science, not be mired in networking/IT support. At the same time, scientists must provide the requirements for the NRP and test and evaluate the networking capabilities that are developed. An effective national partnership will require cyberinfrastructure experts to work with scientists at their interface and understand the desired scientific outcomes, rather than viewing the technology as an end to itself. By creating an opportunity for constructive interactions among a diverse group of domain scientists, network engineers, administrators, campus IT leaders, and vendors, this NRP workshop strived to address both the human and technological networks necessary for success.

The workshop attracted a large number of participants from beyond the initial PRP partnership. Approximately 155 people registered for the workshop and 135 ultimately attended. Fewer than 25% of registrants were from the original PRP partnership. More than 30 states and five foreign countries (Brazil, Korea, Australia, Netherlands, and Singapore) were represented. A significant number of participants were from EPSCoR states, smaller research universities, and minority-serving institutions/organizations. Most registrants were from universities, but there were more than 30 people from regional/national networking organizations, ten from government, nine from industry, and seven from other non-profit organizations. A complete list of registrants with affiliations and job titles is provided in the Appendix to this report.

3 Workshop Sessions and Discussions

3.1 Agenda Overview

The PRP experience has identified two key areas within its scope that will continue in the vision of an NRP: science engagement and technical approach. Science engagement is the bridge to domain scientists, understanding their science objectives and workflows, translating those to technical requirements for networking and supporting tools, and working directly with science teams to apply the technical tools to solve their specific problems and enable collaborative research. The technical approach encompasses more traditional networking and IT efforts: establishing and configuring networks, designing and deploying necessary hardware and software, and working through the technical steps necessary to achieve end-to-end high-performance networking for science groups. In addition to these two broad areas, scaling was a pervasive theme throughout the workshop agenda. How do we scale up from the regional PRP to a much larger NRP – across more domains, more scientists, more use cases, more institutions, and more networks?

The first day of the workshop focused on science engagement, including examples of specific science projects that could utilize a NRP, and scaling up the engagement with broader scientific communities. The second day focused on scaling to include more institutions and a broader geographical reach. The technical approach elements of the PRP and NRP were largely addressed in a few of the Deep Dive parallel sessions rather than plenary sessions.
The agenda included several keynote talks, seven panel sessions with multiple speakers on common themes, and eight Deep Dive sessions that offered opportunities to discuss key topics. Only the Deep Dive sessions were held in parallel (two parallel sessions in each of four one-hour time slots); all others were plenary sessions.

High-level summaries of the presentations and discussions are provided in the following subsections. These notes are not intended to replicate the presentation material and should be read in conjunction with the presentation slides and the videos of most presentations (available at pacificresearchplatform.org).

In each section, quoted comments come from the speaker in that session.

3.2 **Keynote: NSF Cyberinfrastructure Investments and Opportunities: An Update to the First National Research Platform Workshop (Irene Qualters)**

Irene Qualters is the Office Director of the Office of Advanced Cyberinfrastructure (OAC) at the National Science Foundation (NSF). OAC funds the CC* program, which has made major investments in campus networking (see Figure 1, with >$100M across >210 awards), the Data Infrastructure Building Blocks (DIBBS) program (>126M, in collaboration with other NSF directorates), and many other programs that would contribute to or leverage a National Research Platform. The PRP award is one of the current DIBBS awards.

![Figure 1: CC* 210+ awards across 44 states 2012-2017](https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504730)

Since 2012, NSF cyberinfrastructure investments have been guided by an expansive framework defined in the “Cyberinfrastructure Framework for 21st Century Science and Engineering Vision and Strategic Plan” (CIF21). Now five years later, as CIF21 sunsets, NSF is accepting community input and developing new plans to guide future investments. These include a 2016 National Academies Report on NSF Advanced Computing, a Request for Information on Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020,” 2016, National Academies Press.
Needs for Advanced Cyberinfrastructure to Support Science and Engineering Research (NSF CI 2030),

NSF and joint agency assessments of the National Strategic Computing Initiative Exascale Requests for Information, and numerous NSF-sponsored workshops. As Qualters stated, “this is the time for big, bold ideas.”

OAC is developing a set of guiding principles:

- **Promote Science Excellence**
  - Enable fundamentally new scientific advances
  - Attend to current trends in research
    - Multidisciplinary, geo/institutionally-agnostic research and research teams
    - Complex problems; dynamic workflows; data-rich
    - Robust and reliable science

- **Focus on Unique NSF contributions to CI**
  - Holistic view
    - Build capability, capacity, and cohesiveness of national CI ecosystem
    - Consider both human and technical aspects of CI
  - “Lean forward” to new approaches and technologies
  - Encourage reuse of investments in CI from industry, federal agencies, academic institutions, etc.
  - Foster partnerships and community development
  - Incentivize measurement and sharing of results

3.3 **Keynote: The Pacific Research Platform: Leading up to the National Research Platform (Larry Smarr)**

Larry Smarr (UCSD) is the principal investigator of the Pacific Research Platform (PRP) project and a co-chair of this workshop.

There is a long history of NSF adopting and deploying nationally DOE mission-developed technical concepts. This productive partnership can be traced back to the origin of the NSF supercomputer centers 30 years ago, when the National Center for Supercomputing Applications (NCSA) and the San Diego Supercomputing Center (SDSC) modeled their systems on the Lawrence Livermore National Laboratory (LLNL) and the Magnetic Fusion Energy Network (MFEnet). In recent years, a similar pattern has emerged with NSF adopting and funding DMZs. Similarly, the NSF has developed prototypes of the NRP in prior programs, notably twenty years ago, when the Partnerships for Advanced Computational Infrastructure (PACI) program was essentially a national research platform, tied together by vBNS, with standard approaches and specific requirements for campus users (a precursor to elements of NSF’s current CC* program).

Today’s PRP emerged at UCSD from research carried out over the last 15 years, with PI Smarr and co-PIs Phil Papadopoulos and Tom DeFanti collaborating on the OptIPuter, Quartzite, and PRISM awards, essentially using the UCSD campus as a model for developing what has become the PRP. Much of the foundation for the current PRP comes from work led by ESNet over the

---


past decade on Science DMZs, Data Transfer Nodes, perfSONAR and other measurement tools, and security policies.\(^7\)

Since 2012, NSF has had a large-scale program for investments in campus networking cyberinfrastructure, collectively referred to as the CC* program, an outgrowth of an NSF Campus Bridging Task Force Report\(^8\) and the Science DMZ concept. According to Smarr, the CC* awards are “the biggest change in campus networking capability in the history of NSF. It’s changed the entire national landscape.”

The current PRP program is funded by a $5M, five-year cooperative agreement from NSF with the goal of developing a regional high-performance data-sharing capability across ~25 PRP partner institutions. “This is not primarily a networking project; it’s primarily a community socio-technological problem,” explained Smarr. Well before the PRP award, the concept needed to be sold to UC CIOs and Vice Chancellors for Research, to CENIC members, to networking technical staff across many institutions, and to key science groups that would use the proposed capability. Smarr emphasized that “one should not underestimate” the extensive people-to-people interaction required to establish this collaboration on a regional level.

One of the key PRP innovations has been to utilize Flash I/O Network Appliances (FIONAs) as uniform endpoint optical network Data Transfer Nodes (DTNs) across the PRP, each of which can be tailored for 1/10/40/100 Gbps connectivity. The FIONAs, developed under the NSF-funded CC* PRISM grant (PI: Papadopoulos), use Layer 3 networking and ubiquitous SSDs to enable high-bandwidth data flows to be captured by the FIONAs without TCP/IP backing off and degrading network performance. By having (affordable) similar FIONAs at all endpoints, as well as the same data transfer software (e.g., GridFTP) and tools (e.g., perfSONAR), has facilitated collaboration and assurance of continuous performance. The use of the ESnet-developed MaDDash visualization matrix of throughput for all pairs of endpoints (green/yellow/orange/red)

---

\(^7\) For further information, see [http://fasterdata.es.net/](http://fasterdata.es.net/).

has worked remarkably well in not only monitoring the system, but in incentivizing partners through peer-pressure to “go green,” i.e., exceed 5Gb/s for large files every day. As discussed extensively in this workshop, the PRP also has a science engagement track to work directly with domain scientists who can utilize the PRP. “The hard part is to get application people working together with the networking community,” said Smarr.

A number of planned science applications for the PRP were reviewed, including the Large Hadron Collider (LHC) high-energy physics data analysis, astronomy survey data (e.g., Dark Energy Sky Survey and the Palomar Transient Factory), earthquake engineering (e.g., Pacific Earthquake Engineering Research Center), biomedical ‘omics (e.g., the Cancer Genomics Hub, initially hosted by UC Santa Cruz at SDSC, now at the University of Chicago), visualization/collaboration (e.g., sharing CAVEcam data between UCSD and UC Merced). There is growing interest from additional scientific applications, such as linking Cryo-EM facilities with computational resources, expansion of the High Performance Wireless Research Education Network (HPWREN) for real-time support of firefighters, a cyber-archaeology program (see talk by Chris Hoffman in Session 2), and collaborative research on atmospheric water in the western region. For the atmospheric water team, by linking researchers to better computational resources, the workflow time has been reduced from 20 days to overnight, a transformative change for the team’s productivity and scientific impact.

PRP’s cooperative agreement contained a requirement to look at extending the regional PRP concept to a national level (see excerpt in Section 2 above). This extension makes sense from a number of perspectives, and Larry Smarr and Jim Bottum (representing the ACI-REF and CaRC programs) have teamed up to explore the extension of the PRP from a regional to a national or global platform, the motivation for this workshop.

3.4 Keynote: Next-Generation Integrated Architecture: An SDN Ecosystem for LHC and Exascale Science (Harvey Newman)

Harvey Newman (Caltech) is a high-energy physicist who has pushed the boundaries of advanced networking and cyberinfrastructure for decades in support of LHC data analysis and other scientific projects. This talk focused on the increasing challenges posed by major science programs (e.g., LHC, SKA, LSST, genomics) and new technical approaches to address those challenges. The summary below is excerpted directly from the presentation.

**NGenIA Summary**

- The decadal challenges of globally distributed Exascale data and computation faced by major science programs must be addressed
- New approaches: Deeply programmable software driven networked systems. We have just scratched the surface
  - A new “Consistent Operations” paradigm: SDN-driven goal-oriented end-to-end operations, founded on
    - Stable, resilient high throughput flows (e.g., FDT)
    - Controls at the network edges, and in the core
    - Real-time dynamic, adaptive operations among sites, the switched and optical network layers
    - Increasing negotiation, adaptation, built-in intelligence
    - Coordination among VO and Network Orchestrators
- Bringing Exascale HPC and Web-scale cloud facilities, into the data intensive ecosystems of global science programs
- Petabyte transactions a driver of future network and server technology generations

**Figure 3: Summary points**

The Large Hadron Collider (LHC) has been a highly successful scientific instrument and a major driver for advanced cyberinfrastructure. High-performance networks – e.g., LHCP0N, LHCONE, Internet2, GEANT, ESnet (with EEX), CENIC, national networks in Latin America and Australia/New Zealand, and US state/regional networks – move LHC data worldwide at all scales to a tiered network of analysis sites. For example, ESnet often carries ~200 Gbps, with a large fraction of that
long-term ESnet traffic has grown by \(~72\%\)/year and was \(~64\text{ PB/month}\) during 2016. (Newman noted that while technology growth can enable higher growth rates, a constant financial budget can sustain network growth of \(~25\%\)/year, so there will be cost-driven constraints on growth.) LHC-driven demands will increase significantly during “Run 2” and later, the “HL” phase of LHC. In addition, other instruments such as the large synoptic survey telescope (LSST), the square kilometer array (SKA), next-generation light sources, and the genomics community will add exponentially to the demands, increasing data volumes and traffic. The AmLight network between North and South America is being built up to support LSST, and additional networks will be developed for SKA (instruments in Africa and Australia).

The network capacity will be overwhelmed by raw instrument data if we are not thoughtful about architecture and transfer. To address this challenge and others, Newman envisions a major opportunity to exploit the synergy between global operations data and workflow management systems, deeply programmable agile software-defined networks (SDN), and machine learning/modeling/game theory. “We have in our hands the ability to destroy networks,” said Newman, and a new consistent operations paradigm is being posited to manage networks/storage/compute resources for exascale science. Several current programs represent pathfinders towards this paradigm. The Next-Generation Exascale Network Integrated Architecture (NGenIA-ES) program is designed to accomplish new levels of network and computing capabilities through the development of a new class of intelligent, agile networked systems. The SENSE program (SDN for End-to-end Networked Science at the Exascale) is a collaboration between ESnet, Caltech, Fermilab, Argonne, and the University of Maryland to develop SDN to improve end-to-end workflow performance and enable new paradigms. The CHASE-CI program spearheaded by UCSD (see talk by Tom DeFanti in Session 3) will provide a cyberinfrastructure targeted to machine learning. CHASE-CI is developing reasonable-cost NVMe-based data transfer nodes (DTNs) that will try to sustain \(~100\text{ Gbps}\). The group continues to target demonstrations of new capabilities, including both high bandwidth transfers (22 \(100\text{-Gbps}\) links available) and initial examples of consistent operations at the SC17 conference (see Deep Dive Session #8).

Looking ahead to exascale ecosystems in the early/mid-2020s, one would like to readily conduct petabyte transactions, e.g., a PB data transfer on today’s 100 Gbps networks would take 24 hours, but this could be reduced to 90 minutes on a 1.6 Tbps link. These time scales provide the agility and flexibility that will be required for productive scientific workflows.

In his address, Newman stated that the National Research Platform can have profound benefits to
- Expand the reach and shorten the path to the next rounds of scientific discovery
- Reimagine data intensive networks with real-time analytics in science and education and integrate them more deeply in daily life
- Spark a new generation of students and developers and drive the emergence of AI

### 3.5 Session One: Scaling Science Communities: Lessons Learned and Future Plans (Chair, Ruth Marinshaw)

Ruth Marinshaw (Stanford) referred to Irene Qualters’ earlier presentation and pointed out that there are science communities behind all of the inter-operating technical components mentioned in her presentation. The focus of this session was to illustrate some of those community efforts and to discuss challenges those communities have faced in scaling up.
3.5.1 Advanced Cyberinfrastructure – Research and Educational Facilitation (ACI-REF): Campus-Based Computational Research Support (Tom Cheatham)

Tom Cheatham (University of Utah) described the NSF-funded ACI-REF program and its follow-on, the Campus Research Computing Consortium (CaRC). The focus of both programs is about the human networking process (not moving data).

The ACI-REF program was a two-year award starting in March 2014, with six partners. The primary purpose was to address the campuses’ limited user-facing technical support by funding facilitators who could work with users across the partnership according to their expertise; this level of support filled a gap between the deep support of programs like XSEDE ECSS\(^9\) and Campus Champions.\(^10\) In addition, the partnership established a collaborative network of expertise and developed/shared best practices across partners. Challenges included sustainability of the partnership and facilitator positions, providing a career path/identity/visibility for the facilitator positions, and growing to include additional institutions that wanted to join.

The original ACI-REF team developed a new proposal for NSF’s Research Coordination Network (RCN) solicitation that resulted in the current program for Campus Research Computing Infrastructures – The CaRC Consortium. There are currently ~30 members in the consortium, and the costs of efforts within the consortium are primarily institutionally-funded. CaRC is in its early stages and is open to ideas from its members and prospective members as to what the group should prioritize and execute. Committees are currently being formed on workforce development, stakeholder/value proposition, facilitator network, expertise/resource sharing, administrative structure, and marketing/communications. Figure 4 contains two informative “wordles,” graphics that illustrate the words and their frequency (the bigger the word, the more often it occurs) in responses to two questions across a survey of members.

![Figure 4: Wordles (frequency count) for responses to the following two questions: Left: “If CaRC Consortium could deliver one thing to you, “a must have,” what would it be? (Something that you personally value or that is professionally useful to you. It would motivate you to want this to move forward.)” And Right: “What is the biggest barrier preventing or limiting your ‘must have’?”](image)

3.5.2 Scaling Science Communities: Lessons Learned by and Future Plans of the Open Science Grid (Frank Würthwein)

---

\(^9\) See [www.xsede.org/ecss](http://www.xsede.org/ecss)

\(^10\) See [www.xsede.org/campus-champions](http://www.xsede.org/campus-champions)
Frank Würthwein (UC San Diego) is a PRP co-PI and the Executive Director of the Open Science Grid (OSG), an at-scale cyberinfrastructure resource created in 2004. OSG integrates hardware across 100 institutions, supports scientists at more than 100 institutions, and supports science across more than 30 domains (while focused primarily on high-energy physics). From the beginning, OSG has been a PRP partner. Over the last year, OSG delivered >1.5B CPU-hours of time and executed >2B data transfers.11 As a long-term at-scale cyberinfrastructure organization, OSG provided valuable lessons learned, such as the following:

- Be open to all – resource providers and user communities at all scales, and accommodate institutions with differing business models for their services.
- One tool does not fit all – services must be flexible; seamless integration is key.
- Open source is mandatory – need multiple sources and need to survive transience of software providers (or their licenses).
- Stay engaged with software providers and IT shops.
- Funding agencies are fickle – science timelines often exceed agency program timelines and need structures that can survive changing winds in Washington, D.C.

OSG has an ambitious long-term vision to integrate all compute/storage/networking resources at U.S. research institutions and commercial cloud providers, for all science domains.

3.5.3 Strategies for Scaling Science: an ESnet View (Inder Monga)

Inder Monga (ESnet) described ESnet’s ambitious vision: “Scientific progress will be completely unconstrained by the physical location of instruments, people, computational resources, or data.” Scaling up the network is implicit to this vision – in performance, geography and the number of institutions/endpoints.

11 See http://display.opensciencegrid.org/.
A general strategy for scaling was proposed to “infer and codify the underlying design patterns.” Using this construct, seven design patterns were discussed, the latter three designated as emerging design patterns:

- Protect your *large* science (a.k.a. elephant) flows which require almost lossless networks.
- Unclog your data taps – i.e., the Science DMZ architecture is designed to accomplish, simplify, and effectively on-ramp science data to a capable WAN.
- Prepare your data cannons – i.e., DTNs are designed for high-performance data transfers (and should be limited to that function).

- Keep flossing the network – i.e., use tools such as perfSONAR to monitor performance and identify issues.
- Build a bridge between science and infrastructure: *science engagement* is a human-intensive endeavor and includes partnerships, education, consulting, documentation, and knowledge bases.
- Share, train, listen, learn, and share – e.g., training provided by the Operating Innovative Networks model. Again, in-person training is human-intensive; virtual training has been initiated in some places and can scale to broader audiences with comparable effort.
- Well-tuned end-to-end science infrastructure is critical for next-generation science.

The SENSE program (SDN for End-to-end Networked Science at the Exascale) is halfway through its three-year duration, with the vision to “enable national labs and universities to request and provision end-to-end intelligent network services for their application workflows.”

The construct of design patterns can be an effective tool for scaling up. A systems approach is required to identify these patterns, crossing boundaries of compute/storage/networking, as well as observation and persistence. Finally, it is critical to engage end users – positive impact on their science is the goal.

### 3.5.4 Software infrastructure for a National Research Platform (Ian Foster)

Ian Foster (University of Chicago) described eight principles that should drive any strong software infrastructure and provided examples of how the Globus data transfer and management software matches up against these principles.
Globus plays a key role in high-performance data transfers among many institutions. The platform is hosted on cloud services, which minimizes replication of hardware and expertise across sites, and has had 99.5% uptime. There are more than 70,000 registered users (12,000 active/year), with 10,000 active endpoints. More than 290PB and 50 billion files have been transferred using Globus.

3.5.5 SLATE: Services Layer at the Edge (Rob Gardner)

Rob Gardner (University of Chicago) described a new DIBBSs program to (a) equip the Science DMZ with a service orchestration platform, potentially federated to create scalable, multi-campus science platforms, and (b) develop underlayment-as-a-service (UaaS) for platform builders and science gateway developers. The objectives are to

- Reduce barriers to supporting collaborative science
- Give science platform developers a ubiquitous "CI substrate"
- Change distributed cyberinfrastructure operational practice by mobilizing capabilities in the edge
- Achieve scale through multi-site service meshes: ensembles of federated, orchestrated services

The program comes out of experience in distributed platforms (e.g., IVDGL, Grid3, Open Science Grid, WLCG) that identified friction in the use of distributed services. For example, latency in updates across sites makes it difficult to deploy and innovate platform services; distributed expertise in specialized software stacks is required, and resource providers need to become experts in a range of software services. SLATE is designed to move services to the edge of campus resources, located in the Science DMZ (including, as an example, the PRP FIONA boxes), and to enable those services to be centrally managed and updated.

Security is a potential challenge and will be undertaken in collaboration with campus network security officers. The program would like to work with partners to inform technology choices, align with emerging Science DMZ standards, address security issues, promote usability and integration with campus resources, and facilitate science gateway developers and distributed service deployment teams.
### 3.5.6 Session Q&A

John Graham (UCSD) asked Rob Gardner: Have you chosen a container orchestration engine for SLATE? Gardner: it hasn’t been decided, but they will probably start with Kubernetes.

Harvey Newman (Caltech) asked Rob Gardner: What do you expect from the network to orchestrate services? Gardner: They would like to essentially take any capability (e.g., SDN-assisted transfers) and lower the barrier.

A participant asked all the panelists to comment on the biggest challenges, both in general and for people specifically, in sustaining the PRP/NRP.

- Frank Würthwein: Sustainability boils down to funding. There are lots of technical and organizational approaches to sustainability, but sustained funding is hard.
- Ian Foster recommended moving to cloud services where possible.
- Tom Cheatham stressed the importance of institutional buy-in and a willingness to sustain investments.
- Inder Monga commented that it takes a lot of energy to get people to adopt something – then once it’s accomplished, scientists expect capability to be sustained. Unfortunately, funding agencies only take sustainability so far then often transition to other efforts.
- Rob Gardner recommended focusing on reducing operational costs (e.g., cloud services).

Richard Moore asked the panelists if they could identify common threads across their experiences for scaling, including addressing scaling components that require (expensive) people.

- Rob Gardner highlighted Inder Monga’s concept of identifying design patterns, and finding out what users want to do.
- Monga commented that an organization must set uncomfortable design goals for itself. Some design patterns are derived from seeing how people do things and looking for replication. Scaling people is the hardest, but sharing technology, lessons learned and best practices helps.
- Cheatham reiterated that scaling is hard, and people/organizations need to work together. For example, ACI-REF did better facilitation on campuses by borrowing expertise from elsewhere when appropriate.
- Foster recommended finding common elements and design patterns then building further on those. Capture things you see people doing all the time, then communicate the patterns and the tools, so others know about them.
- Würthwein noted that all the speakers see the need for integration, for example, the need for a national platform in the first place. Next steps will be to identify the need for experts, develop a model for the general community to access those experts (so you don’t have to replicate expertise), and minimize operational effort. The only way to get buy-in from many organizations/teams is to minimize the incremental efforts necessary for adoption.

Someone commented that the university research community is fractured. While the infrastructure exists, sustained engagement with researchers is often missing, and face-to-face engagement is hard to scale. How do you sustain the engagement with researchers, which is critical to gaining sustained institutional support? Würthwein replied that “you can’t force someone to eat. If a domain researcher doesn’t see the problem that you can help solve, you have to just live with that. There should be a mix of proactive and responsive measures.” The campus network community is in a good position to be proactive, since they can measure who uses the
network and how much. In a reactive mode, it can help to have meetings on a regular schedule where people can hear what’s going on and perhaps see opportunities.

3.6 Deep Dive Sessions 1–4

3.6.1 Deep Dive One: Roles of Regionals, QUILT, LEARN, NYSERNET, FLR, MREN, PNWGP & Others (Moderators: John Moore & Ronald Hutchins)

The first Deep Dive session was intended to begin the conversation about what role the regional networks should play in rolling out a national research platform.\(^{12}\) It was structured as a conversation rather than a presentation, and the group attending was engaged and enthusiastic. Regional networks face a number of challenges. Finding or developing the skills needed to effectively engage with researchers, managing the technical requirements that a national research platform would require, and finding the funding for both are at the top of the list. Suggestions were made during the conversation on how to organize and make progress.

From an organizational perspective, regionals should start working from existing science collaborations and let those drive the effort to extend beyond the west coast. The more these efforts can then leverage existing regional network collaborations and other national projects, such as CASC, ACI-REF, CaRC, etc., the better. From a technical perspective, having a list of the minimal standard infrastructure and systems needed to get started would prove very helpful. Training is essential, and the effort should take advantage of OIN workshops, the ESnet Science Engagement Team, and other outlets to help with the skills shortage. It was suggested that focusing on building MadDash dashboards for specific collaborations and focusing on “getting to green” is a great way to motivate a team. Several regional networks have already set up MadDash dashboards and exhibited them during the workshop.

Building trust among the technical team, between the researchers and support folks, and generally creating a team with a clear set of common objectives was highlighted as essential to success. Strategies such as developing a cadence of regular technical calls to discuss status and help each other work through issues and utilizing social media tools to foster collaboration are key.

The group agreed to share the conversation with those who are interested but were not in attendance and convene again at existing national meetings, such as the Quilt, CC\(^*\) workshops, etc., to continue discussions and planning.

3.6.2 Deep Dive Two: Scaling and Security: Not at the App Level, Please! (Moderators: Philip Papadopoulos & Von Welch)

The moderators led a group discussion about how to approach security at scale as the number of institutions and systems increases significantly from the limited PRP collaboration to a much broader national or international research platform. This increase brings issues of scale both in terms of number of participants and in the heterogeneity of approaches and level of resources at availability.

\(^{12}\) Many U.S. regional networks were represented at the workshop (see Section 5.3), including the Arkansas Regional Educational Optical Network (AREON), Corporation for Education Network Initiatives in California (CENIC), Florida LambdaRail, Great Plains Network, I-Light Network, Keystone Initiative for Network Based Education and Research (KINBER), MCNC, New York State Education and Research Network (NYSERNET), Ocean State Higher Education Economic Development and Administrative Network (OSHEAN), Ohio Academic Resource Network, Pacific Northwest GigaPoP, Pacific Wave, the Northern Tier Network, and Southern Crossroads.
the institutions. One key issue is the appropriate balance between central standardized security procedures across all partners and autonomous decision-making authority at each participating site. Standard procedures must be sufficient to establish not just interoperability but trust among participants, while preserving the autonomy that may be required locally in order to participate. A related issue is the distribution of security responsibilities between the NRP and the sites, with different sites seeking different distributions and levels of autonomy based on their available resources to handle security themselves. As an example, differences of opinion may arise on a permissive versus restrictive access policy between NRP institutions.

3.6.3 Deep Dive Three: Opportunities in Research Cyberinfrastructure at NSF (Moderator: Amy Walton)

Amy Walton (NSF/OAC) reviewed cyberinfrastructure-related activities and plans within the National Science Foundation. The discussion focused upon the following:

- The array of current and recent cyberinfrastructure funding opportunities at the National Science Foundation. Slide 7 in the presentation provided an overview of data-focused research solicitations, including BIGDATA, DIBBS, CAREER, and Big Data Hubs/Spokes.
- Future directions for cyberinfrastructure now that the Cyberinfrastructure in the 21st Century (CIF21) initiative is coming to an end. FY18 will be a year emphasizing strategic planning and community input. Slides 11 and 12 in the presentation provided summary information on a recent NSF Dear Colleague Letter requesting input from the community.¹³

3.6.4 Deep Dive Four: How Does PRPv1 Work? – An Explanation of DMZs, DTNs... (Moderator: Celeste Anderson)

This session discussed the design considerations and choices for PRPv1 and the new features that will be provided in PRPv2, a parallel environment to the existing testbed. There was also a brief discussion on what a typical science DMZ looks like and an update on the FIONA DTN.

Eli Dart explained why the PRP chose to use a routed Layer 3 approach instead of a Layer 2 approach in PRPv1 after some experience with Layer 2 issues in PRPv0 and how this moves us closer to self-service for researchers and scientists. Reducing the complexity of the inter-domain connections allows scientists to focus on the data that needs to flow over the network and not how long it will take to coordinate the many networks for the data flow needed. Scalability is also at the forefront of the design as people do not scale well, so by simplifying and standardizing processes, this allows networks to automate tasks, reduce errors, focus on fixing performance issues, introduce new tools, and, ultimately, create that illusive friction-free science network.

John Hess and Tom Hutton presented the plans for PRPv2, which will be a parallel environment (testbed) to explore security models, IPv6, Software Defined Network (SDN)/Software Defined Exchange (SDX), and cooperating research groups. Unlike PRPv1, which uses the CENIC CALREN HPR control plane, PRPv2 will have its own control plane. It is using autonomous system number (ASN) 395889, which could be a problem for those networks that cannot handle the 4-byte ASN. There will be a subset of six current PRP participants at the outset: UCSD, SDSC, Caltech, University of Chicago, NCSA, and UCSC. A Cooperating Research Group (CRG) is formed by scientists and resources at a subset of campuses and resource centers, and

such a group should be able to shield access (via firewalls and authorization mechanisms) to its resources even from other CRGs. This will be explored on the PRPv2 as will the possibility that IPv6 be the default. The authentication tools will be compatible with Globus (CI).

Also mentioned in this talk were the AutoGOLE-NSI-orchestrated circuit services available to participants traversing each of the Seattle, Sunnyvale, and Los Angeles GOLEs. More information about these services will be available after the September 2017 Global LambdaGrid Workshop, the Global Lambda Integrated Facility (GLIF) annual meeting in Sydney, Australia where some of the features will be demonstrated, including the Management Environment of Inter-domain Circuits for Advanced Networks (MEICAN) application. Joe Mambretti gave a more detailed explanation of what a Software Defined Exchange (SDX) is and how it can be used to slice networks. With Layer 2 at the core and Layer 3 at the edge, network operators can easily access and use DTNs (NSI, MEICAN, etc.). A simpler security implementation limits software running on a DTN. Using the SDX, you can slice off a piece of a DTN and make a policy container if you need specific software and security. People who maintain DTNs on campus should consult with security folks; many don’t understand high performance computing (HPC) or the performance implications of legacy firewalls.

John Graham's talk focused on an update of the current FIONA DTNs, including the less expensive FIONette model. Some in the audience mentioned that they had downloaded the basic configuration and created their own DTN using the FIONA as a model, and that it is very valuable to build FIONA boxes and deal with simple security implementations that can reach a broad audience easily. Also included in this short presentation was a description of the JupyterHub FIONA and plans for a Supercomputing 2017 (SC17) demonstration.

### 3.7 Session Two: What Scaling Means to the Researcher (Chair: Patrick Schmitz)

Patrick Schmitz (UC Berkeley) opened the session by commenting on the various dimensions of scaling. One needs to scale networks to move data among researchers, instruments, and data processing facilities, and scale up computing and storage resources. Science engagement and consulting resources must also keep pace to make sure researchers have what they need and can make productive use of physical resources. Such people are at the front lines of scaling, not just to help individuals but also to bridge groups and cross scientific domains.

#### 3.7.1 Cyberarchaeology, Immersive Visualization, and the UC Catalyst Project (Chris Hoffman)

Chris Hoffman (UC Berkeley) described a cyberarchaeology project (PI: Tom Levy, UCSD) funded by the University of California Catalyst program to enable the capture, preservation, analysis, and visualization of data from archaeological sites, especially those at-risk from erosion/pollution or destruction/looting. Participants in the project include four UC campuses and the Hearst Museum of Anthropology. The PRP has played a key role in this project, particularly in connecting visualization facilities at UC Berkeley, UCLA, UC Merced and UCSD for displaying archaeological imagery.
Scaling challenges with the project are significant. For example, there are 3.8M objects that could potentially be digitized, managed, and preserved. Much of this work is labor-intensive with high skill levels and close interaction with the archaeologists throughout the workflow required. They are exploring various avenues, both from an IT and domain perspective, to scale up this work, and the Research IT group at UC Berkeley has a broader commitment for science engagement and building relationships – and trust – with researchers across campus.

3.7.2 OSiRIS: Scalable, Computable, Collaborative Storage (Shawn McKee)

Shawn McKee (University of Michigan) described the Open Storage Research InfraStructure (OSiRIS) project whose goals are to (a) make data access and sharing easy across research groups and institutions (including direct access to storage across sites and support for block, object, and file system storage in the same cloud), (b) contain costs, (c) enable pre-existing authentication/authorization tools, and (d) simplify extension and replication of the infrastructure. The meta-goal is to “enable scientists to collaboratively pursue their research goals without requiring significant infrastructure expertise.”

The infrastructure builds on distributed Ceph disk pools, with a software-defined storage layer as the interface between users and those pools. The system is readily scalable in capacity and performance (by adding storage nodes) and across institutions by providing a common storage...
platform. The OSiRIS team is working with researchers in physical ocean modeling and high-energy physics as early adopters, because OSiRIS addresses their emerging requirements and they are willing collaborators; additional adopters will be added through the course of the program.

Important to the program’s sustainability is that each participating institution agreed to continue to maintain and evolve their components of the infrastructure at the end of the project, assuming science users at their institutions are benefiting from OSiRIS.

3.7.3 Necessity Is the Mother of Invention: An End User’s Quest to Scale Up the Cyberinfrastructure Needed to Move & Process Exabytes of Complex Genetic Systems Data (Alex Feltus)

Alex Feltus (Clemson University), a domain scientist in agricultural genomics, works with Clemson software engineer Claris Castillo; their talks provided their perspectives on the collaboration. The initial point was that the research network must have comparable reliability and robustness to the campus enterprise network in order to be adopted by scientists. It cannot be an experiment.

Feltus cited familiar examples of the exponential growth in biology data, and may have been the first to characterize this growth as “ginormous.” With multi-terabyte data transfers, high-performance networking has a significant impact on scientific productivity.

Feltus reviewed a recent DIBBS award titled “Tripal Gateway, a Platform for Next-Generation Data Analysis and Sharing.” Tripal is a toolkit for construction of online biological community databases (for genetics, genomics, breeding, etc.). He would like to extend the PRP to include the partners on this project and understands the difficulty of achieving “green dashboards” for the matrix of end-point connections.

Feltus is also PI of a CC* Data award for “National Cyberinfrastructure for Scientific Data Analysis at Scale (SciDAS).” The premise of the program is that domain scientists face challenges scaling up their data-intensive scientific applications to take advantage of advanced cyberinfrastructure. He believes it is better to embed active end users moving and processing large amounts of data within agile cyberinfrastructure developer teams. This framework promises to integrate existing subsystems that discover domain data, manage fluid data movement across advanced networks, launch comfort-zone scientific applications, and improve flexibility and accessibility to national and global resources, a framework likely to be more successful than using an “if you build it they (the users) will come” approach. Instead, the program advocates “they will help build it while using it.”
3.7.4 SciDAS Underneath (Claris Castillo)

Claris Castillo (Clemson University) is a collaborator and co-PI with Alex Feltus on “Scientific Data Analysis as Scale.” SciDAS builds on various awards made under NSF’s CC* program since 2013 and continues their process of building a data-centric infrastructure step-by-step.

Building Data-centric Infrastructure Step by Step

They take a fluid approach to data analytics at scale with four primary objectives, each addressed with some degree of independence: compute-centric, data-centric, network-centric, and cost-efficiency. The guiding principles of the project include Virtualization, Abstraction, Policy-driven, and Federation. A number of implementations of an integrated infrastructure have already been deployed, and more will be developed.

3.7.5 Scaling Up Water Science Research (Ray Idaszak)

Ray Idaszak (University of North Carolina) described several programs in water science research. In 2021 NASA will launch the Surface Water and Ocean Topography (SWOT) satellite, which will measure water levels globally to within 1 cm, as well as inundation extent in oceans, large lakes, rivers, and reservoirs. The data production will be ~10 TB/week for the northern hemisphere. In preparation for this mission, the SWOT science community plans to produce ~3 years of simulated data (~1.5 PB). SWOT is just one of many water science programs by NASA, NOAA, USGS and...
NSF that represent heterogeneous data products and workflows requiring efficient cost-effective data transfers and computing and a flexible cyberinfrastructure ecosystem.

The Hydroshare program (www.hydroshare.org/) began in 2012 and was designed to provide cyberinfrastructure for hydrologic researchers to solve problems of size and scope not otherwise solvable using desktop computing. A user can share data and models with colleagues; manage access to content; share, access, visualize, and manipulate a broad set of hydrologic data types and models; use the web services API to program automated and client access; publish data and models to document research findings supporting open data, reproducibility, transparency, and trust in results (as well as meet the requirements of your data management plan and receive a citable digital object identifier [DOI] to get credit for your work); discover and access data and models published by others; and use web apps to visualize, analyze, and run models on data in HydroShare. The data management software iRODS is a key element of the back-end of HydroShare in managing the data and providing community access to the data. HydroShare also leverages capabilities already available and being developed in the SciDAS program.

3.7.6 Session Q&A

Someone asked a question of Chris Hoffman about ownership and privacy of the digital archaeological data that’s being collected. Specifically, given past issues with museums repatriating physical artifacts to the place of origin, will similar issues arise with digital data? Hoffman: He knows the Hearst Museum takes these ownership issues seriously, certainly with physical artifacts. In terms of digital data, he does not know the answer to this question, but expects there would be an active discussion between the museum and originator.

Another person asked Hoffman whether they are looking beyond the academic archaeological community for collaborations in digitizing/visualizing sites. Hoffman: They are and, for example, the Native American community has expressed interest in documenting sacred sites, so they could be known to and appreciated by the public without being physically accessible.

Jim Bottum: It has been a “grand challenge” to get various collaborating institutions connected to Alex Fletus at Clemson, making progress “one step at a time.”

In response to a question from Eli Dart (ESnet), Shawn McKee commented that within a small group of ocean modelers, file organization is understandable and not a problem. However, as you go beyond that limited group, the file system may seem difficult to navigate, and they are considering ways to organize it better to make information more discoverable and easier for others to use, but it’s still a work in progress.

Patrick Schmitz asked the panel what they considered the biggest challenge from a researcher’s perspective.

- Alex Feltus – the volume of data storage required for genomics.
- Shawn McKee – researchers have a hard time figuring out how to scale up what they do. It helps to share best practices. OSIRIS is trying to provide access to data from many locations.
- Chris Hoffman – having skilled people who understand what researchers need to accomplish and help them interface to the cyberinfrastructure.
- Claris Castillo – it’s important to let scientists step back from scaling challenges and for them to be able to delegate them to infrastructure experts.
- Ray Idaszak – the science community wants access to existing workflows and models, and they want it to be easy rather than on the cutting edge.

Laurin Herr (Pacific Interface): As datasets grow, is there concern about long-term preservation of data? For example, iRODS has management capabilities, but who’s going to fund long-term storage and preservation?

- Shawn McKee: Part of the challenge is to figure out what data can be discarded. For data that must be preserved, it is a challenge to ensure it is documented and archived and that analysis tools remain accessible (e.g., even on a ten-year timeframe, let alone "long-term").
- Ray Idaszak said NSF-funded cyberinfrastructures are important because one can’t initially anticipate the longer-term value of data. Communities can seek funding for the data that emerge as very important.

Frank Würthwein (UCSD) asked Alex Feltus if the genomics community has had a discussion of the merits of generating new data versus retaining old data. Feltus replied that NCBI has been depositing sequence data since the 1980s. He can dump raw data easily, but he has had problems managing the analysis data over the years, and he could really use help from a university library/librarian. Chris Hoffman commented that a domain scientist can usually manage data during active research, but needs to engage broader discussions with librarians regarding curation over the long-term.

Ruth Marinshaw (Stanford) referred back to Frank Würthwein’s comment about being able to withstand the shifting winds of funding agencies, and asked Ray Idaszak what the water science community would do if agencies stopped funding tomorrow? Idaszak replied that the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) is a consortium and actually has different funding streams. In addition, the Renaissance Computing Institute (RENCI) has committed to preserving the data, even if funding went away.

Larry Smarr (UCSD) asked McKee about collaborations across DIBBS awards on the Ceph work. McKee replied that this was a good idea. Harvey Newman (Caltech) expressed interest in this work and commented that he’s heard both that Ceph is a great file system and, other times, that it is still a work in progress. McKee replied that working in a DevOps environment is a challenge and that versions of Ceph have had quirks.14

---

14 At the closing of the workshop, Mark Berman (GENI project office) expressed his enthusiasm for the "coming attractions" described in the talks about scaling in Sessions 1 and 2. These talks highlighted some of the capabilities that may well be in the next wave of NRP functionality, such as Inder Monga's new "instrument-to-supercomputer phone call" twist on his previous superfacility concept; Rob Gardner’s talk on SLATE; Shawn McKee’s talk on OSiRIS; and Alex Feltus and Claris Castillo’s talks on SciDAS. Some concepts presented were conceived and/or demonstrated in the context of the GENI project, as well as some other testbed environments. While all of these talks were exciting, it is likely that only a subset will prove successful in terms of meeting a clear demand with a high-quality implementation. For that subset, there will be all the usual integration and interopration questions of how best to combine these new capabilities with an existing platform and with each other.
3.8 Session Three: Engaging Communities of Researchers (Chair: Tom DeFanti)

3.8.1 ESnet Science Engagement; Eli Dart

Eli Dart (ESnet) opened his talk with a live demo of moving a TB-scale dataset (1.5TB in size) from a National Center for Atmospheric Research (NCAR) climate data repository in Boulder to the Argonne Leadership Computing Facility (ALCF) in Chicago. The data transfer, using Globus, was initiated by Eli in about a minute with a few clicks. By the end of the 15-minute talk, the dataset had been successfully transferred and was ready for computation at Argonne. This demonstrated the ease, speed, and utility of current high-performance networking capabilities.

A key perspective is that “scientists want to do science, not IT” and “if they [scientists] have to look under the hood, most won’t bother.” According to Dart, the role of science engagement is for “technology professionals working with scientists to help make technology a productive scientific tool.” Instead of technologists providing a “bag of Swiss Army knives” to the scientist to understand and integrate, he recommends a collaborative relationship between the technologists and scientists with each contributing expertise.

The vision of the ESnet science engagement team is that “Collaborations at every scale, in every domain, will have the information and tools they need to achieve maximum benefit from global networks through the creation of scalable, community-driven strategies and approaches.”

Science engagement includes capturing requirements based on scientific needs and user experiences, working collaboratively between scientists and cyberinfrastructure experts to develop solutions using available technologies, and training/outreach to both cyberinfrastructure experts and scientists. Throughout this process, building trust is key to success. ESnet has developed various techniques for engagement, including the CrossConnects workshops, the Operating Innovative Networks (OIN) training series, and the Fasterdata Knowledgebase. In the context of the scaling theme of this workshop, the latter two are specific examples of how ESnet and its partners are effectively scaling out.

Dart pointed out that it is important to engage key facilities, such as supercomputer centers and data repositories, that are hubs for major data traffic. ESnet currently has the Petascale DTN project for routine 15 Gbps data transfers amongst ALCF, LBNL, Oak Ridge National Laboratory (ORNL), and NCSA – this allows users of those supercomputer centers to download data at high speed from modern data portals such as the NCAR RDA in support of climate and earth science (as Dart demonstrated during his talk). This capability is an example of the synergistic benefits of improving key cyberinfrastructure and of the PRP/NRP model.
Since engagement has a broad mission scope and domain knowledge, it is crucial for participants to work together. Many groups are doing this in various forums (e.g., Indiana University, ACI-REF, OSG, XSEDE, etc.) and can share best practices and knowledge and develop trusting collaborations across these efforts.

### 3.8.2 NCAR Climate Data Creation, Storage, Analysis, and Distribution Engagement Strategies (Marla Meehl)

Marla Meehl (NCAR) described NCAR’s mission in climate science and the data-related infrastructure at NCAR for supporting that mission.

One element of NCAR’s mission is to develop and run the Community Earth System Model (CESM); the infrastructure demands are rapidly increasing as the complexity, resolution and ensemble size increase for CESM. Similarly, the Coupled Model Intercomparison Project (CMIP) is a crucial, worldwide modeling effort, and each iteration (CMIP 6 is in progress) increases substantially in scope and requirements. NCAR maintains production compute systems and data repositories that are widely used nationally and internationally.

NCAR recently deployed the 5PF Cheyenne supercomputer at the Wyoming datacenter. It is expanding the GLADE disk storage system (~40 PB) and HPSS archival system (current capacity 320 PB) and operates additional data analysis/visualization clusters. Because of the challenges with moving high-volume data from NCAR repositories to distributed researchers, NCAR is looking at moving the computing capability to the data, and developing flexible compute platforms for remote researchers to use at NCAR.

NCAR has 100 Gbps connectivity from the Wyoming datacenter to the Front Range GigaPOP and is a current PRP node with a FIONA box connected at 40 Gbps, with access to data in the GLADE storage system. NCAR is also participating in other collaborations for high-speed networking to key cyberinfrastructure facilities (e.g., Eli Dart’s real-time demonstration mentioned in Section 3.8.1 above). Meehl also strongly endorsed the ESnet science engagement reports and the OIN training.

### 3.8.3 Engaging Communities of Researchers: Experience from the PRP; Camille Crittenden

Camille Crittenden (CITRIS, UC Berkeley), PRP co-PI, reviewed science engagement efforts within the PRP program. The PRP team has established a bi-weekly teleconference for science engagement leaders across its network, and has held several cross-disciplinary workshops for domain scientists and network engineers, and initiated training efforts. In addition to the general PRP workshop in October 2015, which included significant participation by domain scientists, four specific science engagement workshops have been held in the last 18 months: at UC Davis,
UC Berkeley, UC Merced and UC San Diego. The PRP science engagement team has also initiated two projects with students, who worked on the cyberarchaeology project and developed tools for monitoring PRP network performance.

Crittenden admitted there have been challenges, including the fact that science engagement is a relatively new field/career. She explained that bridging the gap between scientists and CI engineers is often harder than expected, hiring people to do science engagement has been difficult (particularly in the Bay Area), and there is a need to develop better methods to scale up the human infrastructure for cross-institution efforts. As solutions to some of these challenges, Crittenden advocated continuing to work with campus research IT, ESnet, and related groups, attending conferences and sharing best practices across various efforts, documenting and sharing case studies for successful projects, and growing a workforce in engagement, particularly starting with students and other tactics for workforce development.

### 3.8.4 CHASE-CI: Building a Community of Machine Learning Researchers on the Pacific Research Platform (Tom DeFanti)

Tom DeFanti (UCSD) described a recent NSF award for the Cognitive Hardware and Software Ecosystem, Community Infrastructure (CHASE-CI) program to add a machine learning (ML) layer to the PRP, with the goal of enabling ~30 ML researchers at ten of the PRP campuses. Machine learning is a major emerging area of computer science research with a wide variety of algorithmic approaches and is a foundational basis for artificial intelligence. GPUs are well-suited to machine learning, but most systems available to researchers (e.g., Amazon or XSEDE resources) use double precision (64-bit) GPUs with error-correcting memory that are much more expensive than the commodity “gamer” single precision (32-bit) GPUs usable for machine-learning. Thus, this program will provide a hybrid cloud of at least 32 FIONA nodes, each with eight commodity GPUs, specifically for machine learning researchers and students. In addition, non-von Neumann architectures may be evaluated, optimal speed/energy trade-off studies will be conducted, and additional software infrastructure for workflows will be developed.

#### Table: Single vs. Double Precision GPUs: Gaming vs. Supercomputing

<table>
<thead>
<tr>
<th>Nvidia Card</th>
<th>~Cost</th>
<th>32-bit GF</th>
<th>GB per GF</th>
<th>per GB</th>
<th>cores</th>
<th>8-GPU PC</th>
<th>256 GPU CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTX 1080 Ti 11GB</td>
<td>$736</td>
<td>10609</td>
<td>11</td>
<td>$0.07</td>
<td>$67</td>
<td>$14,888</td>
<td>$476,401</td>
</tr>
<tr>
<td>P100 16GB</td>
<td>$6,747</td>
<td>10609</td>
<td>16</td>
<td>$0.64</td>
<td>$422</td>
<td>$62,976</td>
<td>$2,015,232</td>
</tr>
<tr>
<td>AWS p2.xlarge EC2 (8) K-80 GPUs+disk for 3 years</td>
<td>$239,040</td>
<td>$7,649,280</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWS p2.xlarge EC2 (8) K-80 GPUs+disk for 3 years +5% ICR</td>
<td>$370,512</td>
<td>$11,856,384</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8 x 1080 Ti: >1 million GPU core hours every 2 days: $21

$72/week, based on 4-yr life
$100/week with 2-yr GPU refresh

Figure 16: Comparison of single- and double-precision GPUs, mid 2017.

### 3.8.5 Session Q&A

Someone asked Tom DeFanti how the CHASE-CI infrastructure will couple with data sources. He replied that the GPU servers are built around FIONA boxes, so they are designed to handle large datasets drawn from across the PRP.
Harvey Newman (Caltech) had a general question for the panel about technology issues with storage, such as file systems, formats, compression, etc., that need to be addressed in any cyberinfrastructure.

- Eli Dart: File formats really matter and standardizing helps. For example, the climate science community converged on NetCDF and that has provided big leverage. The genomics community has the opposite problem—things are changing so fast, it’s hard to keep up.
- John Graham (UCSD): The CHASE-CI GPU deployment will use Kubernetes and rook/Ceph, and they will collaborate with SDSU, which has a BeeGFS file system.

Jeff Weekly (UC Merced) asked when the PRP will engage graduate students.

- Tom DeFanti: Perhaps we could invite graduate students to the Kubernetes workshop in the fall.
- Alex Szalay (JHU): “Most good ideas for his systems came from end users and from grad students.”

John Towns (UIUC/NCSA) asked if there has been formal documentation of use cases and research needs in order to develop requirements. Camille Crittenden replied that the PRP is probably at a good point now to start doing that. Towns noted that XSEDE has documented various science use cases that could be reviewed.

Claris Castillo (Clemson) asked the panel about security of data. The panel replied as follows:

- Marla Meehl: The security she was talking about referred to the provenance and integrity of data, i.e., that the data is assured to be the original published data.
- Eli Dart: The Earth System Grid included checksums that could be used to validate that you’re getting the original data published. In terms of network security, most scientists bump into this issue when things just aren’t working well (e.g., firewalls, bottlenecks). Current processes to ensure that systems are not hacked are not perfect, but they are better than they used to be.

Tom DeFanti commented on the science engagement staffing challenges. He said that it is easier for universities to hire/retain people in the domain sciences than it is in IT. Maybe recruitment for science engagement should emphasize domain scientists rather than IT experts.

### 3.9 Keynote: The Need for Big Data, Networking, Storage, and Machine Learning (Exascale Numerical Laboratories) (Alex Szalay)

Alex Szalay (Johns Hopkins University) presented a keynote talk in which he described the Sloan Digital Sky Survey (SDSS) data infrastructure and lessons learned from that project. He discussed applications of the SDSS-related tools to other science domains and projected ahead to requirements and approaches to future “exascale numerical laboratories.”

The SDSS project started in 1992 and finished in 2008; the survey data have been and continue to be available to the public and represent a tremendous success story. For example, the SkyServer website has had 4 million distinct users with 2.4 billion web hits over 16 years, with 7,000 papers published using the data and 350,000 citations. The scale and accessibility of the SDSS data transformed the astronomy community, and it serves as a model for other science communities.

But important lessons have also been learned—the hard way. The initial project plan for development and operations from 1992–2000 called for a total budget of $25 million, with 8% of...
that allocated to software/data. The actual price was $100 million, with ~33% in software/data costs, so the overall budget increased 4X, and the software/data budget increased 13X. The cost—and significance—of the software/data efforts were drastically underestimated at the inception of the project in the early 1990s.

Another lesson learned, according to Szalay, is that “one cannot emphasize enough how important trust is—it’s hard to build and easy to lose.”

Third, the project realized that instead of just providing a way for users to submit queries against the survey database, it is more effective to empower the users and provide a workbench for them to build and share their own databases, analysis tools, and archival capabilities.

In order to continue a robust, modern SDSS SkyServer capability and to broaden the application of the tools to other science domains, JHU received an NSF DIBBS award for SciServer. SciServer is currently being used not only to service SDSS data but also in turbulence, cosmology simulations, oceanography, materials science, and cancer immunotherapy.

Through the course of this work, especially evangelizing to new science domains, Szalay has experienced “the valley of death”—the period of time when research resources decline but before commercialization resources increase. While they are getting adoption in some areas, such as materials science, other areas are still in the valley of death. “The best collaborators are the ones that are desperate,” said Szalay, “and are willing to change their ways and are happy to take advice and collaborate.”

Looking ahead to large instruments and exascale HPC systems that are prodigious data generators and whose data should be made widely accessible, new approaches to data are required. According to Szalay, it is “too easy” to collect more data, and no scientist will say they want less data than is available. It is important to look at trade-offs and decide in advance how to collect the most relevant data. The LHC uses “in-situ triggers” to achieve enormous reductions in data upfront (and is still left with large quantities of data). Similar approaches are planned for instruments like the SKA (pre/post-correlator data), and triggers need to be explored for exascale HPC simulation data. This includes analysis hardware and new algorithms tailored to information content as opposed to just pixels. They are conducting tests now on burst buffer triggers on Trinity and are already able to illustrate the concept with a 5:1 data reduction.

The summary slide is copied verbatim below:

- Science is increasingly driven by data (big and small)
- Data explosion everywhere, changes non-incremental!
  - Machine Learning appearing everywhere
  - Need to move Petabytes with ease
- Changing sociology – surveys analyzed by individuals
- Convergence of physical and life sciences
- New instruments: “microscopes” and “telescopes” for data
- A new, Fourth Paradigm of Science is emerging…
- SDSS has been at the cusp of this transition
- But: successful conversion of a community takes a long time (LHC, Human Genome, SDSS…)
- We need to make the technologies more shareable!
Eli Dart (ESnet) pointed out that while the SDSS software/data budget grew significantly from the initial projections, the end capability was transformative for astronomy, both for professionals and a new generation of Internet scientists, in ways that were not initially anticipated and was curious about the best way to tell this story. Szalay replied that when he recently recounted the story to a prostate cancer group, it met with a sober silence in the room as people realized how dramatically costs could be underestimated. He thinks some projects, such as the Dark Energy Survey (DES) and LSST, have learned from the SDSS experience, but other projects have not.

Miron Livny (University Wisconsin) put a different spin on the budget topic and asked “if you put more money into the network or software or computing, you have to take it from somewhere else – the instrument, the sensor, etc. How do you work within a finite budget to be transformative?” Szalay agreed that this is the trade-off that needs to be made. Larry Smarr referred to this as the “silicon-for-steel trade-off.”

Harvey Newman (Caltech) recalled that at the beginning of LHC planning within DOE, the first step was to convince people that the cyberinfrastructure budget would not be zero. He sees numbers around 20% for many projects. He also suggested that, based on his experience, it would be valuable to form scenarios of network needs, and derive examples of science-driven requirements. Szalay replied that LSST is a good example. While SDSS was small enough to deliver end-to-end data to users, the LSST is too big and will need a multiple hierarchy, similar to LHC’s hierarchy. It remains to be seen if LSST will build its own infrastructure for distributing data or if the community can build something (e.g., an NRP) that can move PBs on demand for distributed users. (As an example, during a breakout later it was suggested that storage repositories could be NRP-facilitated.)

3.10 Special Session: High-Speed Equipment (Chair: John Hess)

3.10.1 Science DMZs Behind Firewalls: Heresy? (J.J. Jamison)

J. J. Jamison (Juniper) talked about new Juniper hardware that could provide the security of a firewall but still enable the performance required for Science DMZ traffic.

It is a common misperception among some IT security staff that Science DMZs are not secure. There are typically router/switch security controls, often other security systems are in place (e.g., Bro intrusion detection). In fact, Science DMZs are often the most observed parts of the network. Firewalls are still used because they are easier to implement – both technically and bureaucratically (e.g., no special waivers, especially for HIPAA and FISMA environments).

High-performance firewalls can be put in front of a Science DMZ and provide both enhanced security and performance. “Elephant flows” bound for the Science DMZ from trusted content providers can take an Express Path with full

![Express Path = “TSA Precheck” for Elephant Flows](image)

- Low latency security screening for known travelers flows
- Full Stateful Firewall Service
- Security screens for: IP address sweeps; port scans; denial of service (DOS) attacks; ICMP, UDP, and SYN floods
- Application Layer Gateways for: FTP, H.323, MGCP, MSRPC, PPTP, RSTP, SCCP, SIP, SunRPC, Telnet, TFTP
- SPU based Fast Path
- Supports all the services of Express Path plus:
  - Content Security: IPS; antivirus; antispam; Web, content filtering
  - Application Security: AppID; AppFW; AppQoS; AppTrack; AppRouting

Figure 17: Functions of Express Path and Fast Path
state-full firewall services, security screens for IP address sweeps, port scans and DOS attacks, and application layer gateways for transfer protocols. All other traffic (e.g., campus enterprise traffic and “mice flows” for the Science DMZ) follows the higher-latency Fast Path with additional content and application security screening. (Furthermore, a campus could choose to use Express Path for non-Science DMZ content from trusted cloud/web providers.)

NERSC and SDSC (as well as Goddard) are testing this equipment. Recent results indicate little bandwidth degradation along the Express Path with the firewall in place, although there is 0.3–0.4 msec additional latency.

3.10.2 Arista Network Equipment Deployment at USC (Azher Mughal)

Azher Mughal described the ITS Transformation Project, a major overhaul of the University of Southern California (USC) campus networking built around Arista equipment. He described key features of the Arista equipment that supported their technical requirements. The requirements going into the project included

- Future-proof platform
- Operationally easy to understand and maintain (industry standard commands)
- Network core with lowest latency
- Deep buffers at the edges to absorb traffic bursts
- Port density and multiple speed options
- Supports software defined networking
- Allows custom programming and integration with Northbound applications
- Supports VXLAN and BGP EVPN
- Vendor (as partner) support for faster deployment

The conclusions from the project to date are shown below verbatim from the presentation:

- Our ITS transformation project is steady on its path to uplift the campus infrastructure
- With the help of Arista and their state of the art equipment, we are actively deploying a highly flexible, future-proof infrastructure that accommodates significant increases in capacity and at the same time reducing the cost.
- With software-defined technologies from leaders such as Openstack and Vmware, we are hoping to deploy an agile and converged infrastructure to respond to business dynamics in a multi-tenant services area.
- We are hoping to soon deploy full scale intelligent monitoring and telemetry support across the entire infrastructure for efficient fault isolation and faster response to users.
3.10.3 Software-Defined Access: Enabling the Digital Transformation (Tim Martin)

Tim Martin (Cisco) described current challenges in network performance and management and how Cisco is working to address those challenges. Traditional networks are difficult to segment, complex to configure, and have inconsistent policies for various types of devices. Software-defined access can provide end-to-end segmentation, network automation, and data assurance, with common policies across the enterprise and research components of a campus network. Cisco has developed the Digital Network Architecture (DNA) Center as a centralized management application for the network. The DNA Center runs on Cisco’s software-defined networking (SDN) controller, the Application Policy Infrastructure Controller Enterprise Module (APIC-EM).

3.10.4 Session Q&A

Jeff Weekly (UC Merced) asked J.J. Jamison how much effort would be required to configure networks for the firewall. Jamison replied that they are working to establish baseline configurations and may post those configurations to the fasterdata ESnet website.

Eli Dart commented on Jamison’s talk and said that he is glad the marketplace is responding to the need for elephant flows in both the research and enterprise community. This technology will help sites that are required to have a firewall. He thinks it is important to come up with best practices and is anxious to see how this equipment performs in SCinet 2017 in November.

Tim Lance (NYSERNet) said that this may be a time to look at these network security issues again, and he wants to talk with researchers at his institutions.

Ian Foster (University of Chicago) noted that the Globus group is working with Los Alamos National Laboratory (LANL) to establish a call-out to signal a firewall to open up a specific port.

Michal Krsek (CESNET) commented that, in his experience working with audio/video data transport, non-permanent data streams are fragile with respect to delay/jitter. He said it is okay to bypass a firewall, but he needs signaling, e.g., there are six firewalls to bypass to get data from the Czech Republic to San Diego. He encouraged hardware developers to think about multi-domain environments and signaling. Jamison replied that signaling gets complicated and doesn’t scale well, and he would like to keep things simple.
Jerry Sheehan (Montana State University, Vice President & CIO) described his efforts and strategies to support research cyberinfrastructure in a smaller campus with constrained resources. As examples, Montana State is a community of ~20,000 students/faculty/staff, and his IT budget for campus is ~$14M (75 FTEs). Virtually all of his decisions are economically driven. His total networking staff for all campus support services is three full-time employees (FTEs), and only two FTEs are allocated to research cyberinfrastructure.

In the technology adoption curve, MSU doesn’t have the resources to be an innovator, but they can be an early adopter if there is a research-driven need (e.g., a climate scientist was the driver for building their Science DMZ) and existing community resources are leveraged. Indeed, it is important for smaller institutions to be early adopters and add invaluable use cases from a more resource-constrained environment. “If not,” warned Sheehan, “the NRP will be built for the 1% by default (not by intent).”

MSU’s campus science network would not have been possible without NSF’s CC-DNI program. The NSF funding represented ~40% of MSU’s annual IT capital budget, a cost that could not otherwise have been justified relative to other priorities.

A smaller campus must keep things simple to be an early adopter with lean staff. More complex efforts (e.g., IPv6) just have to wait. Sheehan noted that NSF’s CC-DNI calls have an inherent tension with this approach – in order to show the value of federal investment, one must add more science drivers, but more use cases lead to more complexity. For an early adopter with lean budgets to be successful they must leverage resources from the national CI community, such as the PRP, FIONA designs, Globus, perfSONAR, and technical mailing lists. MSU has also enjoyed a successful partnership with their networking vendor Cisco.

The bottom line is that research networking is challenging for smaller research-intensive universities (both why and how), but by leveraging NSF programs, keeping designs simple, collaborating with the national community, and having private sector partners, such universities can still be early adopters.
3.10.6 NRP Medical School Challenges (Tracy Futhey)

Tracy Futhey (Duke University) is the Duke University CIO, and her responsibilities overlap with the Duke Medical Center’s CIO. The Duke Medical School is “the big gorilla” on campus, and there is substantial overlap with faculty that also do clinical work, with associated additional security requirements.

Futhey presented a number of examples of rewards (collaboration, discovery, impact, translation to industry) and risks (security, privacy, regulations, data provenance, compliance). She cautioned however, “the beneficiary of the rewards is not (by and large) responsible for the risks.” The rewards generally go to researchers, but CIOs and other IT/administrative staff are responsible for the risks. Futhey stated the NRP concept can clearly increase the rewards to researchers via collaborations and access to data, but the “risk people have veto power.” And arguments must be convincing for them not to use that veto power. “The NRP is unlikely to succeed in the high reward medical arena (except anecdotally) unless it reduces the risk as well.”

There are a number of techniques the PRP can use to reduce risk, including federation/virtualization, moving computation to where data resides (local protection), secure multi-party computation, honest third-party brokers, simplifying and scaling consent, data de-identification, and synthesizing data. Medical center administrators are very risk-averse and for the NRP to have impact in medical research, it must substantively address their risk concerns.

3.10.7 A Science DMZ in Every Pot? (Von Welch)

Von Welch (Indiana University) directs the NSF Center for Trustworthy Scientific Cyberinfrastructure (the NSF Cybersecurity Center of Excellence). He began his talk with two main points:

- Networks to support science on campuses exist: Science DMZs
- Scaling is blocked by social problems

The social problem is illustrated by the conflict between a hypothetical research computing advocate saying “Science DMZs are great! They optimize the network for large science flows, removing the friction, and make data movement fast!” while the Security Officer hears “we get rid of the firewall … (and nothing else matters).”

He proposed a shift in that conversation: Start with the premise that the goal of cybersecurity is to support a mission by managing risks to IT, and research and science are part of the mission in the same way that enterprise functions are part of the mission. In this context, an enterprise network and a science network, both with appropriate cybersecurity for their uses, makes sense. For
scaling the NRP to a large number of institutions, the NRP needs to get more CIOs and CISOs behind this different perspective. To do so, it would help for the CIOs and CISOs from early adopter institutions to communicate early success stories to their peers.

### 3.10.8 What Measurements Can We Report to Show Engagement? (John Hess)

John Hess (CENIC) gave a talk spanning measurement tools, engagement efforts, and a recent example working with the University of Guam.

A large set of measurement tools and visualizations are valuable for operating and monitoring high-performance networks, including perfSONAR, the MaDDash matrix display, NetSage (focused on international links), a new prototype trace-route visualization tool (by D. Mishin at SDSC), Globus-stats for transfer statistics, and Tstat for DTN traffic characterization.

Hess summarized a number of ongoing engagement efforts:

- **ESnet** – science requirements reviews/reports, CrossConnects workshops, CI Engineering Brown Bag
- **Internet2** – Global NOC, workshops (e.g., OIN with ESnet, TechEx, Global Summit), working groups: Peering and Routing, IPV6, Security – 25 total!
- **NSRC** (IRNS-funded) – international workshops and direct engineering assistance (DEA). Development of indigenous CI resources. Online resource library.
- **PRP** – Science engagement calls and workshops, engineering calls
- **ACI-REF** – facilitators and collaborations

A recent effort by the Network Startup Resource Center (NSRC, see talk by Steve Huter in Session 6) working with the University of Guam provides an interesting example of creating a new capability at a “non-1%” institution. The University identified several initial science drivers to provide use cases for the network. The University of Guam hosted a campus design workshop with the NSRC and a number of other Pacific island universities. Follow-up steps included direct engineering assistance by NSRC, the University of Oregon, and CENIC staff to successfully establish a basic Science DMZ capability, including a perfSONAR node and a FIONette DTN.

The talk concluded with a reminder that low-cost FIONAs and perfSONAR nodes are available. The cost of FIONAs ranges from ~$1K for a 1Gbps/2TB storage capability to ~$8K for 10/40Gbps systems and ~$15K for 100Gbps FIONAs. In addition, boards as low as $100 can serve as perfSONAR nodes.

### 3.10.9 Engaging Tribal Colleges and Universities in R&E Networking (Jason Arviso)

Jason Arviso (Navajo Technical University and the American Indian Higher Education Consortium, AIHEC) talked about efforts to engage students in tribal colleges and universities (TCUs), STEM programs within the TCUs, and a proposal submitted to NSF for a cyberinfrastructure initiative within TCUs.
There are 37 TCUs in 16 states, with a total enrollment of ~160,000 students. Many TCUs are community colleges, but some are four-year universities and others offer graduate master’s programs. Their missions are broader than just academic education and employment; they include comprehensive education and social systems, professional workforce preparation, tribal culture sustainability, and economic impact and job creation. The research conducted at TCUs is “native research” addressing tribal land, water, health, and community issues.

Post-secondary and STEM education face various challenges at TCUs. For example, the process of laying fiber has an unusual number of stakeholders that govern land use (e.g., the Bureau of Indian Affairs, Bureau of Land Management, tribal councils, etc.). In some cases, TCUs have installed microwave hauls to overcome the delays and complexity; some telephone companies are also providing more network connectivity.

TCU STEM programs must address K-12 system shortcomings, provide remedial classes, engage students in their community (projects that affect them), build student competence, stimulate job creation, and be responsive to local/regional STEM workforce needs.

Several STEM initiatives are supported by advanced cyberinfrastructure, including the Advanced Manufacturing Initiative, environmental resilience, institutional data analytics and the National Tribal University.

The AIHEC has submitted a proposal to NSF for a TCU Cyberinfrastructure Initiative. This Initiative proposes to examine TCU cyberinfrastructure capabilities, provide a primary planning and evaluation resource for TCU CI capabilities, and collect the information necessary for each TCU to articulate and begin implementing a sustainable campus cyberinfrastructure plan. This effort will offer an opportunity to write reports to the TCUs as well as a comprehensive report to the NSF about the path forward to better assist TCUs with STEM efforts and research, including engaging the TCUs in research-and-education networking.

3.10.10 Session Q&A

Mount Allen (SF Jazz Organization, a cultural organization that is part of CENIC) was excited to see TCUs represented and expressed the desire to see HBCUs and other minority-serving institutions represented as well. Allen advocated equitable treatment across organizations and reminded the networking community not to give the equivalent of “second-hand books” to the communities that have historically gotten them.

Mike Sosnkowski (Virginia Commonwealth University) commented that, along with the transport problem, he sees the NRP as a security fabric, including authentication. Von Welch replied it is a

![37 TCUs - More than 75 Campus in U.S. – 16 States](image)

Figure 21: Tribal Colleges and Universities
tough challenge to get the right security expectations from sites: If the bar is not high enough, they don’t get enough value from the federation, but if the bar is too high, then it is too difficult to join the federation. He acknowledged there may be tiers to that problem and observed it is “easier to pull the Band-Aid once, up front,” and ensure that the difficulties are clear in the beginning and don’t stretch out over time. Finally, he said it helps to build on other available widespread tools, such as InCommon.

Laurin Herr (Pacific Interface) asked Tracy Futhey if the costs she cited for security breaches included compliance penalties or legal costs or both. Futhey thought it includes both costs, but she can check. He then asked if the cost logic needs to be applied to science more generally or just to medical records? Futhey replied she is just focused on medical records. Von Welch commented he does not have quantifiable data for the cost of scientific data breaches outside the medical field.

3.11 Deep Dive Sessions Five–Eight

3.11.1 Deep Dive Five: Creation of Community Data Resources (Moderators: Jerry Sheehan and Alexander Szalay)

This deep dive session was initiated with a presentation by Alex Szalay (Johns Hopkins University), “Building the Open Storage Network,” which explored the premise that although NSF has made substantial investments in an effective national computational infrastructure (resources integrated by the XSEDE program), the same is not true for a national data storage infrastructure. Universities build their own storage systems with associated capital, and operating inefficiencies often make them inadequate for growing petabyte-scale datasets. Meanwhile commercial cloud storage is often over-provisioned and expensive, particularly for data access transactions. The Open Storage Network (OSN) proposal is for NSF to fund (~$30M–$40M) a national, distributed storage system with standard 1–2PB storage racks deployed across ~100 universities with high-speed network connections. A large-scale, uniform storage tier could amplify NSF’s investment in networking, enable big-data projects, link to other ecosystems like XSEDE, National Data Service (NDS), Research Data Alliance (RDA), and provide “meat” for the big data hub communities. Key issues to be discussed before deployment of an OSN include whether to have hierarchical storage tiers, object-store versus POSIX, software stack, trust/security across partner sites, whether to include some computing resources at the storage sites, backup strategies, and management of the distributed resources.

Harvey Newman commented that the OSN is a good idea. However, there are tactical operations where data needs to be redistributed across locations, and this process needs to be worked out.

Tim Lance (NYSERNet) also thought the OSN is a great idea and highlighted the two comments that policy management is hard, and wide buy-in and trust are required. He thinks the big data
hubs are too new and inadequately funded to serve as catalysts for this, but regional collaborations, as illustrated by the PRP, already have a community of trust.

Glenn Ricart (US IGNITE) weighed in with his support for the proposal, commenting that big data is evolving to be at least as important as big computing in scientific research, and in the context of this workshop, high-performance networking is a key enabler of the distributed storage system.

Chris Hoffman (UC Berkeley) thought that in selling this concept to universities, it would be important to highlight not only petabyte-scale datasets, but also terabyte- or even gigabyte-scale datasets. Also, the OSN could work with libraries and other units on campuses that are not involved in extreme-scale science.

Alex Szalay replied that Andrew Moore at Carnegie Mellon University is interested in petabyte-scale datasets for machine learning, not datasets that can be easily downloaded. His experience at Johns Hopkins demonstrates that many researchers on campus need at least terabyte-scale datasets.

Larry Conrad (UC Berkeley) asked what the right funding source is for the OSN. Alex Szalay proposes the NSF seed the OSN with standard storage building blocks and software stacks, with perhaps a planned refresh in five years. Universities could augment the NSF-provided resources with their own funding.

Tim Lance (NYSERNET) is not convinced that NSF should seed the proposal. Perhaps more could be learned with a process that requires distributed funding from stakeholders. Separately, he commented that if campuses are to distribute their data to other sites, there needs to be a believable way to get data quickly back to the original campus.

Laurin Herr (Pacific Interface) commented that cloud providers are heavily marketing the media industry with the argument that their data objects are too large to move, and they can use cloud services for self-contained processing, distribution, and archiving (without large bandwidth transaction costs). Alex Szalay replied with an anecdote about a surprise $70K bill he got for download costs (which was waived one-time). Perhaps a reasonable cost model could be negotiated with cloud providers for scientific data.

Fred Prior (University of Arkansas Medical School) sees governance/allocation issues for a finite resource even within an institution, let alone a national storage system. A few of the largest users/institutions could use up any finite resource. Alex Szalay replied: first, he was not anticipating this system would be used for restricted data (e.g., medical clinical data) and that if NSF funds the project, NSF-sponsored research projects would have priority. Second, a participating university could get a quota and then purchase more storage (in standard building blocks) if their researchers required it. He added that this system could be used as a reliable, centrally managed backup facility, albeit with some fraction of the capacity committed to backup. Many universities now have diverse backup solutions, and the OSN could serve as a cost-efficient backup service, with geographically-distributed sites.

3.11.2 Deep Dive Six: What's Worked and What Hasn't for High-Performance Networking (Moderator: Richard Moore)

The purpose of this Deep Dive session was to discuss successes and lessons learned from past high-performance networking efforts to better inform future efforts, such as the NRP. Session
Barr von Oehson, Director of the Office of Advanced Research Computing at Rutgers, described the history and context of research computing at Rutgers. Faculty engagement is strong on campus, and he regards his dual-reporting to the campus CIO and Vice President for Research as a positive structure. Some challenges are the decentralized decision making (e.g., inconsistent IT environments across and within campuses) and overcoming the inertia of familiar past practices. With respect to networking, rather than question the bottlenecks that limited current utilization and envision potential scientific impacts, many campus staff initially saw no need to upgrade performance because existing networks had low utilization. Fortunately, NSF’s CC* program and funding helped alter this perspective. He sees positive developments with science engagement efforts, the CC* CI-Engineer awards, and the Science DMZ and DTN concepts (including FIONAs) but would like more standardization and best practices to emerge.

Eli Dart (ESnet) provided his personal perspectives of what has worked and what has not in five different areas, condensed below (the full presentation can be found on the PRP website at pacificresearchplatform.org):

- **High Performance Everywhere**
  - Mostly a failure – problem is too big
  - Some good things emerged – kernel auto-tuning, new TCP implementations, Grid/Globus tools

- **Science DMZs**
  - Lots of success – and proof that we can do high-performance TCP, in production, even on routed Layer 3 networks
  - Some limitations – security (including IDS), some institutions won’t take risk beyond firewall, trust issues, need engagement with scientists
  - Need to reduce scope from “high performance everywhere” and make major investments in funding, evangelism, and training

- **Virtual Circuits (e.g., Ethernet VLANs, OSCARS, SDN)**
  - Mixed successes – SDN shows lots of promise, but still needs active research
  - Need to provide services that are not available in routed Layer 3 networks
  - User interfaces, especially at domain scientist level, are not effective

- **perfSONAR**
  - Mostly successful – widely deployed effective tool, stable software
  - Interoperability needs to be seamless
  - Interaction with security is still challenging

- **Engagement**
  - Successful in many environments, in building communities and capturing requirements
  - People don’t scale well (need to collaborate on best practices), and engagement needs a better career path within institutions

Steve Fulkerson (AREON) commented that state/regional networks have done well in working with campuses to set up Science DMZs, but often not as good a job with smaller campuses. He also commented that although perfSONAR is widely installed, it requires better interoperability (Quilt has made efforts here).

Inder Monga (ESnet) opened his comments with “culture eats strategy for lunch.” Institutional cultures, particularly with regard to security and risk management, must be addressed. A uniform
understanding of security requirements and approaches is needed. A cultural transformation is also required with respect to science engagement; more generalists are needed who can work across multiple technical systems and work with scientists to solve their problems. Finally, he noted that when something is working, there is a reluctance to change. How do you keep encouraging new things and change infrastructure in ways that are rational but not risky?

Following the panelists’ comments, substantial discussion ensued about career paths and retention for CI engineers working on campuses and regional/national networking organizations. Many institutions lose excellent staff to positions with better security (especially relative to soft money positions) and better pay. Many advocated getting staff into more stable positions (e.g., permanent funds), with good career prospects for pay and advancement, although one participant expressed concern that stability leads to less motivation, and appreciated the innovative fire that can come with relying on soft money.

One participant highlighted new capabilities in the latest perfSONAR release (e.g., it is easy to add to the framework and use as a distributed information sharing service) and encouraged people to develop additional capabilities beyond just performance monitoring.

Jim Pepin (Clemson) referenced Azher Mughal’s talk about refreshing the USC networking infrastructure (~20 years old) and how campuses often go through irregular investment cycles. Celeste Anderson (USC) commented that consultants had helped convince the administration that the networking infrastructure is critical to all functions of the campus enterprise – it shouldn’t be viewed just as a cost center – and will need regular technology refresh going forward. Steve Fulkerson noted that state/regional networks may be able to play a role in this dialog across their member campuses.

3.11.3 Deep Dive Seven: Strategies for Moving Forward: How to Build A Network of Regional DMZs (Moderators: Jim Bottum and Larry Smarr)

This Deep Dive session largely provided an open mic opportunity for workshop participants to express their ideas for moving forward with the NRP concept. As such, and with no presentation materials for this session, the notes here are more comprehensive than for other Deep Dive sessions.

Larry Smarr and Jim Bottum opened the session by commenting on the positive attitudes and energy that have pervaded the workshop. Bottum noted that the NRP is at a point where “we’re building a community out of communities,” citing related community efforts like ACI-REF and the recent PEARC’17 conference, which have brought together previously independent programs. He also expressed the need to engage and retain staff in this community, including professionalizing the career paths of campus cyberinfrastructure experts and the importance of determining what is required to scale up an NRP community nationally and beyond. “1-800-LARRY is not scalable,” said Smarr.

Smarr noted the diversity of potential directions for moving forward. For example, what could the PRP do over the next three years of its grant in terms of its engineering or measurements/evaluations that would be useful to a broader NRP? Is there a way to join an extended PRP that could evolve to an NRP? Could the PRP effectively leverage knowledge gained and lessons

---

15 see www.pearc.org
learned to enable campuses and regional networks to join the effort within their constraints of limited financial/staff resources?

Jeff Weekly (UC Merced) noted that Merced’s PRP experience has not always been smooth, but problems get solved because he knows the individuals within the team who can help. He wondered if this solution was scalable nationally, and noted it is important that NRP builders think “from a user perspective, with an emphasis on user service, quality of service, and quality of experience.”

Cees de Laat (UvA) took pride in being one of the most “problematic” partners in the current PRP, with trans-Atlantic latencies and multiple network domains to transit. He considers the PRP to be a template for scaling out. According to de Laat, the Science DMZ and DTN concepts are “Lego blocks you can build on.” He is working with the European Science Cloud and mentioned the FAIR (findable, addressable, interoperable, repeatable) principles: being discussed in the data management community.16

John Towns (UIUC/NCSA) commented on a number of existing efforts related to NRP objectives, observing that there may be some “reinventing of the wheel.” XSEDE has been working on high-performance end-to-end data transfers, at least for a limited community, as has OSG. Finally, while this workshop focused on regional networks to support an NRP, Towns asked what role does Internet2 have? Larry Smarr welcomed suggestions from XSEDE/OSG experiences, noted the linkage between OSG and the PRP with the inclusion of Frank Würthwein as a PRP co-PI and OSG Executive Director, and invited Internet2 to comment on their perspectives.

Larry Conrad (UC Berkeley) observed that the NRP “is more of a mesh architecture than a top-down architecture,” and it is necessary for individual universities in that mesh to want to make this happen. It makes sense for the regional networks to step in to support that type of architecture when there’s a modest number of participants who know and trust one another. It’s important for the regionals (and the Quilt, an organization of regionals) to be active participants in the NRP.

Timothy Lance (NYSERNet) remarked that discussions around the NRP provided an “excuse” for him to get back to researchers and talk with them about their requirements, and it offers a common standard for what potentially can be offered to researchers.

Harvey Newman (Caltech) said he “thinks we’ve lost the thread in this conversation.” The NRP should be addressing some challenges that have not been discussed. For example, many people have not been empowered to solve problems in a way that others have been. Newman said the NRP needs to be about new abstractions, new workflows, configuring end systems well, the nature of networks, the intelligence of networks, and how networks are used and how they react. “If we just talk about empowering communities – and that means being able to launch 10 Gbps flows – this is not a success model,” said Newman. Larry Smarr replied that this is a good example where there’s probably new NSF opportunities to explore network scaling and architectures, as well as DOE/ESnet efforts. The PRP program funds three FTEs with a job to connect various researchers across specific partners. If the PRP consortium is serious about an NRP in five years, NSF (and DOE) should be investing in some fundamental questions (e.g., SDN) now.

Jim Pepin (Clemson) described the PRP “as a giant Dropbox in front of my neighborhood.” He said the PRP provides a good dashboard for endpoints on the edge of campus, but asked how do you reach to/from the labs, not just the campus edge? Larry Smarr replied that the PRP tried to address this by requiring letters of commitment from CIOs of participating campuses stating that they would support the effort – and ensure that end-to-end performance reaches into the labs on campus, not just the edge. A similar commitment was encouraged for prospective “extended PRP” members, so that incentives are aligned.

Cees DeLaat (UvA) noted that proposals in the European Union now need research data management plans. He advocated that researchers should be required also to state how they would use national or international cyberinfrastructure, in order to ensure awareness of available capabilities.

Wendy Huntoon (KINBER) commented that while many people have talked about funding, not everyone will be able to get grants, especially at smaller campuses. KINBER encourages its members to integrate ideas into their campuses and plan for additional capabilities, whether they get grants or not. In addition, she stressed the significant role that “people networking” plays. While she knows many people in the national community, not everyone has the benefit of that people network. By building trust relationships regionally and then nationally, we can move forward more quickly.

Howard Pfeffer (recently appointed President and CEO of Internet2) said he’s very interested in engaging with the regionals and the community on networking, trust and identity, and other issues, and welcomes the opportunity to talk further with the participants.

Eli Dart (ESnet) cited examples from Alex Szalay’s keynote and Marla Meehl’s NCAR presentation about the transformative science that can be accomplished when valuable data are made easily accessible. The NRP should include national data resources like these and make it possible for a broad community of scientists to work with these data.

Alex Feltus (Clemson) asked what it really means to be part of the PRP. Does it mean getting on the dashboard? Does he need to get the CIO on his campus to tell people to do this? Smarr responded that first, researchers on campuses need to be the drivers (and he plans to go back to researchers mentioned in the original PRP proposal and see what they are doing). Second, while the PRP grant funds a limited number of FTEs to accomplish specific objectives, he is talking with NSF about additional resources to extend this to additional campuses. Third, just as with the initial PRP partners, any new participant would be required to have a letter of commitment from the campus CIO. Larry reiterated that each campus that wants to really participate needs to have the CIO’s commitment of support.

Inder Monga (ESnet) noted that the first thing the PRP project did was to ask domain scientist users “are we doing the right thing for your science and is the return-on-investment there?” The PRP has brought together science and network experts in a conversation about needs and technology. Rather than just deploying a box and forgetting about it, the PRP needs scientists to use the capability. Monga also observed that as instruments get larger, there are fewer of them – one LHC, one LSST, a couple exascale machines – and it will be crucial for people to get access to data from those instruments to do research. Freeing data so it can be used is a key value-added issue for the PRP and its evolutions. The more facilities/data sources that connect, the greater the value to the scientific community.
Mark Berman (GENI Project Office) stated that the primary goals of this workshop and immediate next steps naturally focus on community development and planning for deploying basic NRP capabilities across a broader, national footprint. However, the pace of change is rapid in research infrastructure, as demonstrated by several promising technologies presented during the workshop’s “Scaling Science Communities” and “What Scaling Means to the Researcher” sessions. Berman stated that a long-lived NRP must include lifecycle planning for new capabilities, including the opportunity for staged and controlled rollout of candidate capabilities. By incorporating a testbed environment within the NRP and/or working with other research testbeds, first adopters within the NRP community can help to better define, validate, integrate, and harden emerging capabilities.

Camille Crittenden (UC Berkeley) commented that, as a champion of science engagement, the PRP needs to do a better job in science communications. She would like to look at good models for how to explain the science that is being done, collect use cases from scientists, develop good materials that could be posted on a website, and converge on coherent messages. She recognized that ESNet does a good job of this and invited others to suggest ideas. In addition, she asked what should we call people who do science engagement? Facilitators? Being able to define that role could benefit that career path and broaden the pool of candidates.

John Towns (UIUC/NCSA) suggested that to “get to what ‘it’ is, ‘it’ should be driven by researcher needs.” While that is probably implicit, it had not been made explicit in the discussions so far. He stated the PRP needs to start building the use cases that can drive this. XSEDE has been building scientist-driven use cases for some time, and some of these existing use cases could probably be used for the NRP. Larry Smarr welcomed these potential inputs and reiterated that he will be going back to all the science teams that were in the PRP proposal and documenting their use cases, including whether they are using the PRP capabilities and what impact it has made on their science.

Yves Poppe (National Supercomputing Center, NSCC) commented that he made a recent proposal for an Asian Research Platform (ARP), similar to the PRP concept. He also noted that in his proposal, he stressed the platform aspect of the PRP over the network aspect and has followed the PRP model of first talking with researchers and also requiring a CIO commitment of support. He is organizing a workshop on the Asian Research Platform to be held March 26–29, 2018.

The session concluded with Larry Smarr inviting people to submit additional written comments to him as private communications or for inclusion in the workshop report.

3.11.4 Deep Dive Eight: SC17 Demos (Moderators: John Graham and Azher Mughal)

This session was essentially a working group meeting for a number of people involved in some high-performance networking demonstrations planned for the Supercomputing 2017 (SC17) conference (see Figure 23). This is a work in progress, so the description here is intended only as a high-level snapshot of current efforts.
The theme of the primary demonstration is to perform ATLAS machine-learning-based simulation and analysis on a distributed platform comprised of shared GPU machines connected by high speed networks that allow transparent delivery of data through a distributed Ceph file system. The Pacific Research Platform provides this infrastructure. The lessons learned will inform the development of models for the next generation of ATLAS tools targeted for the High Luminosity Upgrade of the LHC.

Additional planned demonstrations include a remote microscopy demonstration from the USC campus, a demonstration of the OSiRIS distributed storage system (see Section 3.7.2), and pathfinder LSST data transfers to Korea Institute of Science and Technology Information (KISTI) and via AmPath to South America (see Section 3.13.2).

3.12 Session Five: Towards a Global Research Platform (GRP) (Chair: Joe Mambretti)

Joe Mambretti (Northwestern University, StarLight) opened this session by reviewing international networking resources that could be available for a Global Research Platform (GRP) extending beyond the NRP. These include PacificWave international peering and the Global Lambda Integrated Facility (GLIF) consortium, including GLIF open exchanges such as the StarLight International/National Communications Exchange. In addition, he noted the planned US Software-Defined Exchange (SDX) interoperable fabric.

3.12.1 Toward a Global Research Platform (GRP) – DMZs in Korea (Jeonghoon Moon)

Jeonghoon Moon (KISTI) provided an overview of high-performance networking, including Science DMZs and DTNs within Korea and in collaborations between Korea and the U.S. KISTI
operates two Science DMZs, one with an SDN-based DTN (20 Gbps) for internal Korean researchers and a Layer 3-based DTN (40 Gbps, to be upgraded to 100 Gbps) connecting the KISTI supercomputer’s Lustre filesystem to external collaborations, including the PRP.

Several science use cases were described, including capturing and analyzing data from the Korean very-long-baseline interferometry (VLBI) network of three multi-frequency telescopes within Korea, accessing data from the LSST and other precursor astronomy datasets like the Dark Energy Survey and sharing results of analysis by Korean researchers, and sharing climate and agrometeorology data. KREONET-S is the first international production SDN-WAN connection from Korea to Starlight and should be completed this year.

Larry Smarr (UCSD) commented that within the PRP, UC Santa Cruz is already accessing and analyzing data from the Dark Energy Sky Survey (via LBNL) as a ramp-up to LSST data, and it would be interesting to test this from Korea, with higher latencies.

3.12.2 LSST Scaling Issues and Network Needs (Heidi Morgan)

Heidi Morgan (USC) discussed the network requirements and design for the LSST, which will be located in Chile, with the primary data archive and access center at NCSA, a control center in Tucson, Arizona, and a satellite processing center in Lyon, France. LSST will be a prodigious data instrument and will produce the largest non-proprietary dataset in the world.

The network has two main channels: a low-bandwidth, high-priority control channel (from Tucson to Chile) and a high-bandwidth (~90 Gbps sustained) data channel (from Chile to NCSA and beyond). In order to produce timely alerts of transient events, each image (~13 GB) must be transmitted from Chile to the U.S. within five seconds of image capture.
While the AmLight network connecting South America to the U.S. in South Florida is a homogenous SDN network with the connectivity and bandwidth to meet LSST needs, the paths north from Miami and Boca Raton, Florida cross multiple domains with heterogeneous configurations and services. Potential solutions using Software-Defined Networks and Software-Defined Exchanges are being considered to facilitate multi-domain provisioning and to provide the quality-of-service and programmability required for LSST. As such, the LSST is a primary use case for the AtlanticWave (AWave)-SDX collaboration.

The network to support the LSST will be deployed and tuned from August 2017 until 2020, primarily using simulated data. “First-light” for science data from LSST is in 2021, and the network must be fully operational by then. AmLight-Express will be the primary path, with AmLight-Protect as a backup path.

3.12.3 Pacific Rim Application and Grid Middleware Assembly (PRAGMA) (Phil Papadopoulos)

Phil Papadopoulos (UCSD) described the Pacific Rim Application and Grid Middleware Assembly (PRAGMA) collaboration, which includes ~25 institutions (~20 from Asia, five from the U.S. and one in Australia) and is a community of “willing participants” interested in conducting scientific expeditions and infrastructure experiments.

The science expedition model is designed to enable a small laboratory to use technology to share data, methods, and techniques with other collaborators. These collaborations are not intended as demonstrations, but rather are meant to enable persistent collaborations on time scales of years.

With respect to networking, in contrast to much of the discussion in the workshop, the data volumes/bandwidth needs of PRAGMA research are not particularly challenging. However, the connectivity and ease of use are crucial. PRAGMA has developed a multi-cloud testbed for supporting experiments. This testbed uses overlay networks to provide a trusted environment for sharing of resources.

Yves Poppe (NSCC) asked whether PRAGMA would have any use for the proposed Asian Research Platform (see comment in Deep Dive #7)? Answer: “Absolutely.”

3.12.4 International Options: How can we scale? (Jennifer Schopf)

Jen Schopf (Indiana University) talked about the international networks overseen by her group at IU and lessons learned with respect to networking and engagement.

The IU team oversees the TransPAC network to Asia and the NEAAR network to Europe and Africa. In general, the international adoption of Science DMZs/DTNs is slower in countries outside the U.S. While the group regularly monitors the international links (including utilization),
engagement with scientists is also a big part of the job. This again reflects the dual elements of successful networking – both technological and human. In addition, it is apparent that a lot of groups, particularly in Africa, are still struggling with basic connectivity, such as reliable Ethernet across campuses. As a result, one must accommodate different realities and make progress within them.

Project NetSage provides a common framework with test points to show performance of NSF-funded links using tools such as perfSONAR, SNMP, and Tstat. While NetSage is just beginning, it is important to be able to measure network utilization data analogous to the accounting databases prevalent in HPC communities.\(^\text{17}\)

Many factors impede understanding data movement: many pieces lie along the data transfer path; no single organization controls all those pieces; expectations for performance vary; asymmetric paths make soft failures hard to find; and there are multiple points of coordination. For example, the path from the University of Utah to the University of Cape Town in South Africa or from IU to Thailand each cross seven network domains.

IU is working with ESnet to develop a Joint Center for Engagement and Networks to provide resources and assistance in networking and data transfers to network and campus engineers, application scientists, and people interested in engagement. The focus is on promoting demonstrated technologies (e.g., Science DMZs, DTNs, perfSONAR) and enabling end-to-end performance. They envision two types of engagement: 1) Case Studies, which are a more deliberate approach (analogous to routine car maintenance), and 2) Roadside Assistance to fix specific problems (analogous to fixing a flat tire).

Schopf summed up with the following conclusions regarding scaling:
- Measurement, monitoring, and engagement are linked
- The people problem is as hard or harder than the technical ones
- Need an army of the willing

### 3.12.5 Session Q&A

Ron Hutchins (University of Virginia) commented that when he recently visited Apple headquarters, he asked how they scale user support. The answer was that they have web documentation, FAQs, and help info, then peers help each other, rather than a central support desk. In the context of the NRP and the human-intensive scaling challenges, he asked if this is a model to explore?
- Jen Schopf replied that her team has started to adopt some of that approach – e.g., fasterdata.es.net is a good example. But how do individual researchers know peers they could go to in order to get help?

\(^{17}\) For example, see [https://xdmod.ccr.buffalo.edu/](https://xdmod.ccr.buffalo.edu/) for NSF’s XSEDE HPC resources.
- Ron Hutchins commented that one needs to get peer support, e.g., chemists helping chemists.
- Someone recommended attending domain conferences to get people interested in technology and then they can help connect their peers.

Celeste Anderson (USC) remarked she encountered a group that didn’t want to install perfSONAR because doing so would create more work for them, and they were already stretched thin. She said it is important to make things easy to overcome resistance to new things (and more work).

Phil Papadopoulos commented that if he’s got a great new technology widget, a domain scientist needs to determine whether adopting this widget is worth the time. It is not always possible to simplify the widget to the point where it is automatic and easy. For example, many of the experiments they’ve done in PRAGMA take a long time – and that project involves “willing people.”

Jim Pepin (Clemson University) suggested using a Slack channel¹⁸ to find support. The NRP may need something at a higher bandwidth for scientists. Jen Schopf commented that the earth systems group at IU has created a slack channel for their community.

3.13 Session Six: Democratizing Collaboration (Chair: Maxine Brown)

Maxine Brown (University of Illinois at Chicago) introduced this session by noting the program committee’s preliminary title, NRP Governance, was deemed premature since there is no NRP organization yet. Instead, the program committee decided to focus on extending the reach of the NRP and how it can democratize collaboration via broad access to data.

3.13.1 Network Startup Resource Center (Steve Huter)

Steve Huter (Network Startup Resource Center, NSRC at the University of Oregon) gave an overview of the NSRC organization and examples of its programs. Founded 25 years ago, going back to the original NSFNet, the NSRC’s mission is to “cultivate collaboration among a community of peers to build and improve a global Internet that benefits all parties.”

Huter perceived some overlap of the work NSRC has done over the years with NRP objectives. The PRP builds on an established set of practices (e.g., Science DMZs) within its partnership, and this workshop addressed how it might be scaled out to an NRP or even a Global Research Platform (GRP). But Huter asked “how do we scale the benefits internationally, so the techniques and resource sharing model can reach 100x or 1000x the

¹⁸ See www.slack.com.
campus networks and research labs around the world?” “NSRC focuses on helping improve campus network infrastructure since scientists do not connect to an NREN for access. Scientists connect to the global R&E network fabric via their campus network, which is the transport layer, and hence critical for the foundation of a National or Global Research Platform,” Huter added.

As examples of recent programs, Huter described video-based education programs for scalable outreach.19 They have worked with the TENET education and research network in South Africa and the Kenya Education Network (connecting 160 institutions), as well as the work with University of Guam described by John Hess in Session 4.

Looking to the future, Huter believes that the model of Science DMZ/DTN/Engagement can be done in a systematic way across not only in the U.S. but also internationally and would improve the distributed cyberinfrastructure and accelerate scientific research. He cautioned that it is important to respect your peers – both in the network sense of peering and in the human sense – to cultivate a community of peers that benefits all parties.

### 3.13.2 Regional Research Platforms: Regional Example – The Virtual Data Collaboratory; Wendy Huntoon

Wendy Huntoon (Keystone Initiative for Network Based Education and Research, KINBER) talked about the Virtual Data Collaboratory (VDC) program, a recent DIBBS award from NSF to form collaborations across Rutgers, Penn State, NJEdge and KINBER. The VDC is a regional effort to create a data DMZ across collaborating institutions using the New Jersey and Pennsylvania regional networks. Its objectives are somewhat similar to those of the Pacific Research Platform. The architecture includes data hubs at the Rutgers and Penn State main campuses, with spokes at six additional campuses, using NJEdge and KINBER networks; FIONA DTNs are installed at all hubs and spokes, and initial planned bandwidths are 10 Gbps. The group is conducting performance tests now using existing perfSONAR platforms, which can be used as a “before” baseline compared to eventual performance.

One of the early lessons learned is that implementation is hard and the devil is in the details. Unexpected problems crop up and must be addressed. For example, one of the datacenters is relocating, resulting in unanticipated project costs, other costs were not covered in the grant, and participants are discussing who will pay.

In summary, Huntoon cited some of the conclusions from the Deep Dive 1 discussion on the role of regional networks in the NRP:

---

19 See learn.nsric.org.
• Work on science collaborations
• Get FIONA boxes and start
• Organize communities – e.g., monthly calls, gain traction
• perfSONAR dashboard is effective – peer pressure works, especially when visible

3.13.3 Democratizing Collaborations: Equity and Access (Gil Gonzales)

Gil Gonzales (Gonzales Consulting) opened his talk by saying this would be the first talk at the workshop with no mention of Science DMZs, network maps, or requests for funding. His focus is to reduce the digital divide. He has worked in Arizona, California and New Mexico, and his talk was primarily about his work in New Mexico.

Because broadband costs were extremely high in New Mexico, those costs drove network designs from a perspective of scarcity. Their efforts have helped reduce those costs. Gonzales said it is important to recognize that progress is relative, especially in under-served communities. He has seen institutions where they have improved from minimal connectivity to reasonable connectivity (e.g., 300 Mbps) and that has been transformative for the institution.

There are 23 universities/colleges in New Mexico, each with its own governance/decisions. These universities can be categorized as research institutions (University of New Mexico, New Mexico State University, and New Mexico Tech), comprehensive colleges, and community colleges. Gonzales’s program targets the middle tier of comprehensive colleges.

In the New Mexico Science & Economic Development Network’s (NMSeed) efforts to expand STEM/broadband capabilities, they went first to CIOs to discuss plans, but then met with university presidents, since the CIOs were often disenfranchised from decision-making or had no access to resources. They then received an NSF regional planning award to work with a broad community to advance broadband capabilities for both education and economic development, including municipal government, public health organizations, K-12 education, higher education institutions, state information technologies, and broadband providers. The three research institutions committed funds to support the effort, as they were committed to its success.

Gonzales’s group is now working to support multiple CC-DNI proposals from comprehensive and community colleges and is conducting initial planning for a statewide data-sharing mechanism built on familiar tools, such as the Science DMZ/DTNs and existing authentication mechanisms (e.g., InCommon).

As a take-away from these efforts, Gonzales concluded that “you need continuous engagement, rethinking the environment so that it’s aligned with what institutions need.”

3.13.4 Role of the Regional Networks in Scaling a National Research Platform (Jen Leasure)

Jen Leasure (The Quilt) talked about the role that regionals can play in scaling the NRP.

The first step, she said, is to recognize the significant diversity across regional networks – and to view this as an asset rather than a liability; people get to learn from each other.
In terms of scaling the NRP, regional networks have an inherent advantage as “trusted conveners” and coordinators among the community of member institutions. In addition, there are existing collaborations (including The Quilt) across the regional networks.

One question: what is the right size for an N*N perfSONAR mesh (e.g., MadDash display)? An all-to-all display is not manageable, but scales of a collaborating research group, a regional network (e.g., PRP), or across regional networks (e.g., http://quiltmesh.onenet.net/maddash-webui/) are manageable.

In moving forward with an NRP, Leasure suggested that the PRP capture an operational list of “things to do to get started,” then convene a “coalition of the willing” of additional institutions and regions that can identify science drivers, leverage institutions with knowledge/experience, and build on existing regional partnerships. Leasure noted that the NRP should “accept that coalition efforts may not look the same,” and there will be a process of building trust among that coalition. Existing engagement efforts (ESnet, Internet2, ACI-REF, CaRC, XSEDE, CASC, NSRC, etc.) need to be engaged, as well as training (e.g., ESnet, OIN workshops) and the availability of standardized affordable hardware (e.g., FIONA, FIONette). Leasure concluded that some level of funding will be required, whether institutional or federally provided.

3.13.5 Global Collaborative Research Groups (CRGs) (Cees de Laat)

Cees de Laat (University of Amsterdam, UvA) talked about a number of efforts in Amsterdam and collaborations both within the PRP and with European partners.
UvA is a partner with the PRP, along with KISTI and the University of Tokyo. These international connections within the regional PRP set the stage for the NRP and Global Research Platform (GRP).

UvA collaborates with KLM Royal Dutch Airlines and others on the iSHARE project, designed to allow managed data sharing among groups that both collaborate and compete – in this case, airlines that want to share certain flight/engine data without sharing other data that would compromise their competitiveness. The iSHARE project is exploring various techniques that provide for control by the data owner, along with data integrity and provenance (e.g., an embassy concept). They are developing system architectures that combine Science DMZs and software-defined exchanges (SDXs) and want to deploy these capabilities into a field laboratory environment.

De Laat noted that international research collaborations pose additional challenges compared to regional/national collaborations – from network latency to humans coordinating across time zones. Various initiatives within the European Union (EU), such as the Open Science Data Cloud (PIRE) and the European Open Science Cloud, provide resources for scientific collaborations. An emerging theme, particularly within the EU, is the FAIR standard: data that is Findable, Accessible, Interpretable, and Reusable.

UvA has established a Science DMZ and is developing a campus cyberinfrastructure plan. DTNs are located at UvA, KLM, and SURFnet, and SURFnet is encouraging other universities to deploy systems. Finally, de Laat advocated that the EU adopt a practice already in place at the Dutch NSF and eScience Center – namely, that all proposals identify their cyberinfrastructure requirements. This requirement not only forces campus CIOs to keep science needs in mind, but also keeps the funding agencies and network organizations informed of the needed resources and geographical locations of cyberinfrastructur.

### 3.13.6 Session Q&A

Yves Poppe (NSCC) asked Cees de Laat if the airline project for sharing data among airlines would be extensible to other airlines. Cees expects it would work if they are part of the SkyTeam alliance, but he would have to check. The main point is that this large-scale data sharing project is a commercial example of what is being attempted in the academically-focused PRP.

Glenn Ricart (US Ignite) said he was excited about Cees’s description of the embassy concept for managed cooperation among different entities (e.g., the airlines) and asked the panelists if they have seen other use cases. Cees cited some other examples in his work, and Wendy Huntoon said a similar concept obtains in the VDC hubs that allow organizations to share data with different protocols and sharing permissions.
3.14 Closing Session: What’s Next (Jim Bottum and Larry Smarr)

Workshop chairs Jim Bottum (Internet2) and Larry Smarr (UCSD) made closing comments. There were no presentations or question-and-answer session during this wrap-up.

Larry Smarr thanked all participants, particularly the international attendees, the NSF representatives, and remote online participants. He also thanked the host Jerry Sheehan (MSU) and conference coordinator Michelle Perez-Robles (CENIC). He interpreted it as a very positive sign of the level of interest that nearly everyone remained to the end of the second day of a two-day workshop. (An audience member lamented the limited number of evening flights out of Bozeman, but everyone else was certain this was not a factor in their level of interest or late attendance.) He also invited any participants to submit comments to him for inclusion in the workshop report.20

Smarr said he was “impressed with the sincere level of discussion and community building” and asked “how do we keep that going?” He noted that while everyone talked about a National Research Platform, it remains only a concept, not a program. Upcoming community gatherings present opportunities for further NRP discussions, such as the Supercomputing conference in November, the Quilt meetings in October and February, and various XSEDE and Internet2 meetings. Smarr reminded the audience that anyone can create an NRP session to get more input from the community and to keep the momentum going forward.

Bottum closed by first thanking Smarr for developing the PRP concept. He described himself as “being on the outside looking in” to the PRP, but said he sees people working hard on this, and he knows many people at the workshop want to join this effort. Bottum said he believed there is a strong “coalition of the willing,” people who are passionate about advancing research and scholarship. He appreciated a number of talks about getting participation from campuses with lesser means, and acknowledged that involving these campuses is yet another element of scaling up: how do you capture collective wisdom and make it broadly available to people at all interested institutions?

Smarr thought that Bottum’s and Wendy Huntoon’s points were on the mark with respect to the role of more institutions, regional networks, and Internet2 in the NRP. The PRP project has tried to make information accessible to adopt commodity hardware, and he said it is great to see MadDash displays popping up from other groups. When the PRP first started, the MadDash displays were not pretty – plenty of reds and oranges, instead of green squares. However, measuring and optimizing end-to-end performance across multiple domains has forced collaboration, required people to work together, and has developed trust among the participants.

Smarr noted that identifying scientists to use research network infrastructure is not in the normal job description for network engineers and CIOs. Based on his own experience in assembling the PRP, Smarr described the process as a human-intensive effort, a “shoe leather interface.” However, the process is rewarding, particularly when researchers start using systems and providing feedback, then things start to change for the better for all participants.

Smarr continued, there are opportunities for pre-existing structures to begin their own efforts toward an NRP or GRP. For example, if a federation of regionals is an effective model for the NRP, regionals could take the initiative. Parallelism is the key to scaling rather than central control.

20 Several of these comments have been received and are incorporated into this report.
Smarr also noted Harvey Newman’s comments that the NRP should not pursue just incremental changes over the next 3-5 years but rather transformative changes. For example, ESnet has been engaged in a strategic planning exercise about its future, and there are likely to be major changes.

Smarr made some observations about several current and prospective NSF programs, saying “if we think that the country needs an NRP over the next three to five years, perhaps the NSF should issue a call for proposals to address the tough technical issues that will need to be addressed in an NRP; in fact, this could be a tremendous opportunity for DOE and NSF to work together.” Smarr said he is pleased that NSF is continuing the DIBBS efforts, several of which were discussed in this workshop. He thanked NSF for the cyberengineer awards and said he hoped NSF will continue that program; Smarr explained that beyond the CC* hardware awards, campuses that have those glue people have made a lot of progress. Even if NSF doesn’t continue funding these positions, campuses should consider hiring cyberengineers anyway, and he encouraged people to go to NSF with ideas (e.g., emerging from the regionals). At a lot of institutions (e.g., supercomputer centers, XSEDE, instrument facilities, etc.) the expertise and resources also reach into the campus (not just the supercomputer datacenter) and connect user labs to a Science DMZ. This approach helps to identify campus networking limitations and open up paths within those institutions.

Smarr concluded by saying that he has been “pleasantly overwhelmed” with the level of interest from people wanting to join/emulate the PRP and he is open to further inquiries. Perhaps he will think about a more formal way to initiate an extended PRP. In whatever form the PRP concept is extended, he believes some of the groundwork for the PRP represents key requirements for success – namely that data-intensive scientists on campus must be recruited to use the system and work on its development, and the campus CIO and regional networks should provide letters of commitment to support the project (with staff time, resources, and policy decisions) to make it a success.

Botum said that he was thinking about what got them motivated at Clemson to push on their cyberinfrastructure and recommended that “if you don’t have one already, find an Alex Feltus” on your campus, a scientist who needs advanced cyberinfrastructure and is willing to work with you. Feltus became a “burr under the saddle” at Clemson and now has evolved to a burr under the saddle of his collaborators’ campuses to bring them up to speed. And it’s a success all around!

4 Findings

The two-day workshop was successful in gathering together key stakeholders in the NRP concept – scientists, CIOs, network engineers and researchers, and administrators – for positive and enthusiastic interactions. With her permission, we include a copy of an email sent to all workshop participants by Irene Qualters, who leads NSF’s Office of Advanced Cyberinfrastructure (OAC).

“Thank you for your individual and collective engagement in the workshop! It struck exactly the right tone and conversation. And this was no accident.

“The agenda, the selected topics, the location, and the presenters exposed the rich set of perspectives that contribute to a national platform vision. And it did so in a venue which provoked engagement and thoughtful discussion – both in the meeting and in adjacent conversations. The mix of issues identified over the workshop span short and long horizons. Some topics are actionable, others require further exploration. Importantly,
This workshop extended the intellectual reach of distinct communities and it did so in a way that will stimulate further conversations, guide community development, and inform NSF plans.

“While Jim and Larry summarized the meeting perfectly, I wanted to add a footnote. It was a privilege to be present as an observer. Thanks to each of you.”

The workshop organizers thank Irene Qualters for these comments and believe they reflect the perceptions of most workshop participants – this was an energetic and important workshop that should have long-term transformative impacts for scientific research.

This section summarizes the findings from the workshop. These are organized by bold headings with justification drawn from participant comments made during the workshop. Many participants indicated this workshop was an important step toward a scalable and sustainable refocusing of NSF efforts, after the successful establishment of regional networks and campus Science DMZs, toward using these infrastructures across a diverse range of scientific applications. It was generally agreed that the country needs a coordinated national support initiative and program to help bring an NRP into existence, and it is in the national scientific interest to do so.

The momentum created by NSF to establish improved campus networks and connections to regional networks is ready to move to the next phase: engagement with campus communities (both scientist and humanists) to remove the technological and sociological barriers to cooperation and collaboration. State and regional networks are critical to the NRP effort, yet leaders must remember that not all institutions are part of such networks. In that regard, it was impressive that smaller institutions not traditionally seen as a constituency of advanced networks participated in this meeting. Significant, global challenges will require everyone’s participation to solve, and the inclusion of small universities, community colleges and minority-serving institutions is a step in the right direction.

Many current large-scale science collaborations require high-performance end-to-end networking, and even more demanding science-driven requirements are on the horizon. Even given those demands, clear opportunities are emerging for transformative scientific advances.

Specific key findings from this NRP workshop are summarized as follows:

**The NRP platform needs to be easy for scientists to implement and use.** Building a seamless data platform that makes data mobility as easy as downloading apps on a smartphone is critical for next-generation scientific discovery. The Pacific Research Platform, and potentially the National Research Platform, provide the tools, personnel, knowledge, and active engagement of researchers to build a seamless platform that will be used in day-to-day scientific processes of discovery. As multiple scientific domains bring new applications, this is an opportunity for disciplines in computing (compute, storage, network/software/middleware) to come together, and build a seamless, science-accessible platform.

**Scientists want to do science, not networking or IT.** An effective partnership has cyberinfrastructure experts working with scientists at their interface and understanding the desired scientific outcomes, rather than viewing the technology as an end to itself.

**This initiative is a social engineering project as well as a technical networking/IT project.** Many of the stakeholders often do not work together and almost all have separate funding sources
and management structures. Stakeholders include scientists/users, science team experts who can work with users and solve end-to-end IT ecosystem issues, campus network engineers and IT personnel from campuses and regional/national networks.

The Science DMZ/DTN architecture is an effective means of enabling high-performance end-to-end networking for campuses and institutions, balancing researcher requirements and network security concerns. There are many DTNs, but FIONAs in different flavors have proven cost-effective and adaptable to various needs from 1-100 Gbps. Common tools such as perfSONAR and Globus also have wide adoption in this community.

The Science Engagement process is crucial to scaling up to a national/international research platform. It is important to engage scientists who will be users of the system at the outset, identify their requirements, design and build a system that responds to those requirements, and work with the science teams to the finish line.

Several classes of cyberinfrastructure elements will benefit significantly and synergistically by being adapted to the PRP model. These elements include national HPC facilities, instruments used by experimenters, data repositories and data portals, campus computing centers and storage systems, and commercial cloud services. Each of these is described below in relation to a National Research Platform architecturally modeled on the advances and capabilities demonstrated by the PRP.

- **National HPC facilities** (and scientific users of those facilities) benefit when they are connected to a high-speed network using the Science DMZ model. In particular, the Data Transfer Nodes connected to the large parallel filesystems at HPC facilities provide a high-speed path between the storage accessible by the HPC resources and other HPC facilities, experimental instruments and facilities, and other resources. When deployed with a capable data management platform (e.g., Globus), the Science DMZ at an HPC facility becomes the conduit for science at scale – very large (terabytes to petabytes) data sets can be moved in and out of the HPC resource without requiring the scientist to scale human effort with data set size.

- **The instruments and facilities used by experimenters**, including synchrotrons, cryo-EM instruments, gene sequencers, telescopes, etc., also benefit hugely from DTNs in a Science DMZ. In this case, the PRP model eliminates the friction from data analysis by providing a high-performance path from the experiment to HPC resources and back again. The emerging super-facility model, in which an HPC resource and an experiment instrument are coupled together to permit at-scale data analysis during the experiment, is a perfect example of the kind of capability enabled by connecting the experiment’s data acquisition system to a high-speed network by means of a DTN. The data scale at experimental facilities is rapidly overwhelming the capabilities of computing systems at the experiment. Integrating the experiment with an HPC center (which provides not only computing but significant storage capabilities and expertise) allows the experimental facilities and their scientific users the opportunity to realize the full scientific promise of the next generation of advanced instruments.

- **Data repositories and data portals** contain the fruits of decades of labor in many scientific disciplines, including earth science, astronomy, biology, and material science. In many cases, however, the data is inaccessible to the large-scale analysis capabilities that today’s HPC facilities provide, and is therefore unavailable for analysis using machine learning and other advanced computing techniques. The reason is that the software architecture of most data portals dates from around the year 2000, when data
portals built on web servers replaced command line FTP for data access. If the data in the portals were accessible via a capable data management system (e.g., Globus) and available in a Science DMZ, the data could flow from the portals to the national HPC facilities as scientists develop machine learning and other advanced applications to conduct analyses heretofore impossible. Several examples of this exist already, including the Research Data Archive at NCAR. The ease of data access at these modern portals allows large-scale data placement at computing facilities without the weeks to months of labor required when using legacy portal systems. Connecting the nation’s scientific data repositories to a National Research Platform would allow for rapid advances in multiple fields and enable continued scientific leadership by the United States.

- **Campus computing centers** can be attached to a National Research Platform in the same way that the national HPC facilities are – via DTNs connected to storage systems. By taking this approach and integrating them with a National Research Platform, campus computing systems become several things at once: a bridge to a national capability for their users; a platform for analyzing large-scale data sets produced or stored at national HPC facilities; a means of interacting with experiments and instruments when the scientist does not yet have an allocation at a national center; and a means of analyzing data stored in advanced data portals as described above.

- **Commercial cloud services** are increasingly being integrated into scientific workflows for multiple reasons, including convenience, accessibility, and cost. The high-speed science networks which would make up the backbone of a National Research Platform already connect to cloud providers at high speed. By integrating high-performance data management tools (e.g., Globus) with commercial cloud services as has already begun, scientists can analyze data at campus computing systems, on national HPC resources, or in commercial clouds as appropriate.

All of the above capabilities reinforce each other and together build a next-generation science complex on top of the high-speed science network backbones and the Science DMZs funded by the highly-successful CC* programs. The PRP has demonstrated that the technological basis for scaling up to a National Research Platform is sound. What remains is to integrate the different parts of the science complex using this model. This will involve Science DMZs at HPC and experimental facilities, the deployment of data management tools on those DTNs, and the integration of those systems using workflow tools built on capable data management platforms. The technical basis for this exists, but a coordinated national effort is needed to realize it.

**Voluntary institutional and researcher commitment is vital for success.** The phrase “coalition of the willing” was repeated in many contexts during this workshop. Based on the experience in the PRP, a broad campus commitment is required to overcome the many potential technical/security/policy/time/financial impediments to high-performance end-to-end data transfers. There must be internal motivation by diverse stakeholders on a campus to take the necessary steps, to participate in collaborations across institutions (with their own set of stakeholders and motivations), and to solve problems that will inevitably arise. Domain researchers must be motivated to improve their workflows. In addition, support, incentives and motivation are needed from the CIO and risk management personnel, as well as within the campus IT team to address the technical and security issues associated with Science DMZs. Financial support (explicit or more often implicit in volunteer labor) from the budgets of stakeholders is also needed. (Even an effort like the PRP is probably leveraged ~10:1 with volunteer labor compared to grant-funded labor.) Coordination is critical at the CIO level, and it is beneficial to engage their colleagues through groups like RUCC, Educause, Internet2, CIC, etc.
A corollary of this finding is that scaling the NRP to a large number of organizations is more likely to be successful as a bottom-up effort among willing, interested campuses volunteering to participate, as opposed to a program organized from the top-down (e.g., an organizational structure with mandates that flow down to campuses).

Related to this “coalition of the willing” theme, the PRP initially used a model of requiring all member campuses to submit a letter of commitment, signed by at least the CIO and sometimes by the use-case scientists and research administrators, promising support to the project and a commitment to invest time and resources to work through the anticipated hurdles and resolve potential conflicting interests and priorities. When considering extending the PRP membership, Larry Smarr was adamant that the same requirement would hold for prospective partners. By extension, this requirement is strongly encouraged for other regional or national-scale collaborations. Such extension activity would involve funding more staff than just those maintaining the PRP.

**NRP cyberinfrastructure needs to be more than interconnected Science DMZs.** The PRP has successfully illustrated the concept of interconnecting Science DMZs, a necessary step when creating an NRP but not sufficient for accelerating science nationally. A set of Science DMZs interconnected by a high-speed network fabric is a platform upon which larger structures must be built. The PRP demonstrates this well by mapping entire science collaborations onto the platform. However, to do this nationally, the NRP must consider classes of cyberinfrastructure elements common to many science collaborations. Working collaboration-by-collaboration will be difficult to scale. Campuses may also face internal challenges, such as outdated hardware and a plethora of different IT organizations on campus, that may impede participating in the NRP.

**The NRP should be prepared to evolve at the same time as it is being launched.** A point made by a few introductory speakers was along these lines of “We can build an NRP. We’ve done it before, and we know how to do it again.” On the one hand, this sentiment is true and a sign of confidence that an NRP will be an achievable success. Cyberinfrastructure is a rapidly advancing field, and the NRP should expect new technological advances that need to be rolled out to the research communities that it will support. On the other hand, the fact that such pep talks are needed suggests that these rollouts remain challenging, and the NRP will need to do what it can to prepare for the next capabilities coming down the road, even while it works to establish a baseline for the platform across a larger footprint.

Tools and protocols for scaling up should be developed and shared to facilitate the process. Scaling is a challenge on many fronts. The PRP has been reasonably successful in part because it is a limited set of campuses and science groups, and the people involved know and trust each other. *Those factors cannot be ignored in scaling up the PRP to the NRP.* Working through all the details of end-to-end performance requires debugging at each site (N) and in some cases, each pair-wise connection (N**2). Some of the ideas suggested to facilitate scaling include:

- Define, document, and share best practices
- Share knowledge – open attitude toward sharing, central knowledge bases, ‘knowing how to get answers easily,’ even sharing staff expertise (e.g. ACI-REF facilitators)
- Training – e.g., OIN
- Standardize and adopt hardware/software where possible
- Develop standard network configurations, some top-down, some bottom-up
- Leverage existing organizations for communications and collaborations – like ACI-REF, CaRC, XSEDE, PEARC
• Conduct science engagement – e.g., CrossConnect workshops, participating in domain workshops, peer-to-peer testimonials
• Begin the path toward an NRP by extending/replicating the PRP concept at the regional network scale
• Commit to cooperation and collaboration amongst the “coalition of the willing”
• Identify design patterns that capture recurring workflows and requirements, communicate the patterns, have the community build tools against those design patterns, and disseminate lessons learned
• Lower the barriers and minimize the incremental efforts necessary for adoption of the NRP concept for institutions. For example, develop a minimal standard infrastructure and systems with standard configurations and steps for deployment and testing.

Trust is a key element for success. Trust was mentioned as a key element for success in MANY different contexts during the workshop. There are trust issues associated with security, especially across institutions, but in most contexts, the reference was to human trust, a vital element in the effectiveness of collaborative teams, both within a campus and across different institutions. For example, it took time for PRP participants to work together and get to know each other, to learn individual roles and strengths (and weaknesses), and to learn to rely on/trust their collaborators. Sustained delivery on promises – to collaborators and funding sources – is necessary to build trust. Trust is a human-intensive endeavor, one relationship at a time, and is not readily amenable to scaling. On the other hand, certain methods can be used to scale trust. For example, identifying and documenting successful collaborations (e.g., the PRP), including specific issues related to trust, can assure other potential participants that they can start with a predisposition to trust. In addition, peer-to-peer communication is an effective way to transfer trust, whether it is a domain scientist talking with a colleague about the benefits of the program, or a CIO talking with his or her counterparts about security/risk/costs, or a Vice Chancellor for Research telling his or her counterpart at a competing university that they are missing the boat.

**Need to think about the Science Engagement career path and how to attract and retain good staff.** There is a general challenge to retain the best IT staff on many campuses because of market demands (including from other campuses) and better compensation packages in industry. This is particularly true for staff in the important, but less well-defined, science engagement field. These individuals may find themselves in a no-man’s land between domain science and IT in terms of career paths, and may not get the recognition merited for this key role. Therefore, it is important to think carefully about career paths for people doing science engagement and find ways to offer the salaries, advancement and technical opportunities, and recognition needed to retain them.

**General consensus was reached that improved measurement techniques and metrics should be developed and adopted.** The PRP currently tests with perfSONAR but also with disk-to-disk GridFTP tests since both provide data ingestible by ESnet’s MaDDash visualization. Other data transfer methods need to be monitored, measured, and visualized.

**Security arose as a thread which is only partially addressed by the trust section above.** There are complex (technical and policy) security issues that remain barriers to a NRP.

### 5 Acknowledgements
This workshop was supported in part by NSF Cooperative Agreement, ACI-1541349. Additional support was received from Montana State University, the Corporation for Education Network Initiatives in California (CENIC), the Center for Information Technology Research in the Interest of Society (CITRIS) and the Banatao Institute, the California Institute for Telecommunications and Information Technology (Calit2), Calit2’s Qualcomm Institute, and Pacific Interface, Inc.
6 Appendices

6.1 Workshop Presentation Materials and Videos

Archival information for this workshop is available at the PRP website http://pacificresearchplatform.org. All presentation materials are posted at http://prp.ucsd.edu/presentations/nrp; a prefix and speaker’s name precedes the file names, with the prefixes S (for Session number), SS (for Special Session), D (for Deep Dive Session number), and K (for Keynote).

The event was live-streamed. Videos from the live streaming and slide presentation materials are linked to the agenda shown in the ‘Presentations’ tab (http://prp.ucsd.edu/presentations/nrp).

The workshop used an event website for registration and participant information (http://www.cvent.com/d/v5qz2z?lang=en&sms=7&cn=Lpj3pt1FgUmfMcP8bDq_QA). This website will remain available for a period of time, but will not be maintained long-term.

6.2 Selected Attributed Opinions of the Participants

The First NRP Workshop participants offered many suggestions for future efforts. In order to be maximally inclusive and impactful, the opinions are assembled in this section, in the order offered during the Workshop. These detailed notes led to both the Findings and the Recommendations of this report.

Qualters stated, “this is the time for big, bold ideas.”

Smarr emphasized that “one should not underestimate” the extensive people-to-people interaction required to establish this collaboration on a regional level.

Newman envisions a major opportunity to exploit the synergy between global operations data and workflow management systems, deeply programmable agile software-defined networks (SDN), and machine learning/modeling/game theory. In his address, Newman stated that the National Research Platform can have profound benefits to

- Expand the reach and shorten the path to the next rounds of scientific discovery
- Reimagine data intensive networks with real-time analytics in science and education and integrate them more deeply in daily life
- Spark a new generation of students and developers and drive the emergence of AI

Frank Würthwein:
Be open to all – resource providers and user communities at all scales, and accommodate institutions with differing business models for their services.

- One tool does not fit all – services must be flexible; seamless integration is key.
• Open source is mandatory – need multiple sources and need to survive transience of software providers (or their licenses).

• Stay engaged with software providers and IT shops.

Inder Monga:

• Protect your large science (a.k.a. elephant) flows which require almost lossless networks.

• Unclog your data taps – i.e., the Science DMZ architecture is designed to accomplish, simplify, and effectively on-ramp science data to a capable WAN.

• Prepare your data cannons – i.e., DTNs are designed for high-performance data transfers (and should be limited to that function).

• Keep flossing the network – i.e., use tools such as perfSONAR to monitor performance and identify issues.

• Build a bridge between science and infrastructure: science engagement is a human-intensive endeavor and includes partnerships, education, consulting, documentation, and knowledge bases.

• Share, train, listen, learn, and share – e.g., training provided by the Operating Innovative Networks model. Again, in-person training is human-intensive; virtual training has been initiated in some places and can scale to broader audiences with comparable effort.

• Well-tuned end-to-end science infrastructure is critical for next-generation science.

The construct of design patterns can be an effective tool for scaling up. A systems approach is required to identify these patterns, crossing boundaries of compute/storage/networking, as well as observation and persistence. Finally, it is critical to engage end users – positive impact on their science is the goal.

Rob Gardner: Latency in updates across sites makes it difficult to deploy and innovate platform services; distributed expertise in specialized software stacks is required, and resource providers need to become experts in a range of software services. Work with partners to inform technology choices, align with emerging Science DMZ standards, address security issues, promote usability and integration with campus resources, and facilitate science gateway developers and distributed service deployment teams.

Ian Foster recommended moving to cloud services where possible.
Rob Gardner recommended focusing on reducing operational costs (e.g., cloud services).
Cheatham reiterated that scaling is hard, and people/organizations need to work together. For example, ACI-REF did better facilitation on campuses by borrowing expertise from elsewhere when appropriate.

Foster recommended finding common elements and design patterns then building further on those. Capture things you see people doing all the time, then communicate the patterns and the tools, so others know about them.

Würthwein noted that all the speakers see the need for integration, for example, the need for a national platform in the first place. Next steps will be to identify the need for experts, develop a model for the general community to access those experts (so you don’t have to replicate expertise), and minimize operational effort. There should be a mix of proactive and responsive measures.” The campus network community is in a good position to be proactive, since they can measure who uses the network and how much. In a reactive mode, it can help to have meetings on a regular schedule where people can hear what’s going on and perhaps see opportunities.

Deep Dive #1: From an organizational perspective, regionals should start working from existing science collaborations and let those drive the effort to extend beyond the west coast. The more these efforts can then leverage existing regional network collaborations and other national projects, such as CASC, ACI-REF, CaRC, etc., the better. From a technical perspective, having a list of the minimal standard infrastructure and systems needed to get started would prove very helpful. Training is essential, and the effort should take advantage of OIN workshops, the ESnet Science Engagement Team, and other outlets to help with the skills shortage. It was suggested that focusing on building MadDash dashboards for specific collaborations and focusing on “getting to green” is a great way to motivate a team. Building trust among the technical team, between the researchers and support folks, and generally creating a team with a clear set of common objectives was highlighted as essential to success. Strategies such as developing a cadence of regular technical calls to discuss status and help each other work through issues and utilizing social media tools to foster collaboration are key.

People who maintain DTNs on campus should consult with security folks; many don’t understand high performance computing (HPC) or the performance implications of legacy firewalls.

Patrick Schmitz: One needs to scale networks to move data among researchers, instruments, and data processing facilities, and scale up computing and storage resources. Science engagement and consulting resources must also keep pace to make sure researchers have what they need and can make productive use of physical resources.

Alex Feltus (Clemson University), a domain scientist in agricultural genomics, works with Clemson software engineer Claries Castillo; their talks provided their perspectives on the collaboration. The initial point was that the research network must have comparable reliability and robustness to the campus enterprise network in order to be adopted by scientists. It cannot be an experiment. It is better to embed active end users moving and processing large amounts of
data within agile cyberinfrastructure developer teams. • Claris Castillo – it’s important to let scientists step back from scaling challenges and for them to be able to delegate them to infrastructure experts. Feltus said that NCBI has been depositing sequence data since the 1980s. He can dump raw data easily, but he has had problems managing the analysis data over the years, and he could really use help from a university library/librarian.

Shawn McKee: Part of the challenge is to figure out what data can be discarded. For data that must be preserved, it is a challenge to ensure it is documented and archived and that analysis tools remain accessible (e.g., even on a ten-year timeframe, let alone “long-term”).

Ray Idaszak said NSF-funded cyberinfrastructures are important because one can’t initially anticipate the longer-term value of data. Communities can seek funding for the data that emerge as very important.

Larry Smarr (UCSD) asked McKee about collaborations across DIBBS awards on the Ceph work. McKee replied that this was a good idea.

According to Dart, the role of science engagement is for “technology professionals working with scientists to help make technology a productive scientific tool.” Instead of technologists providing a “bag of Swiss Army knives” to the scientist to understand and integrate, he recommends a collaborative relationship between the technologists and scientists with each contributing expertise. Dart pointed out that it is important to engage key facilities, such as supercomputer centers and data repositories, that are hubs for major data traffic. Since engagement has a broad mission scope and domain knowledge, it is crucial for participants to work together. Many groups are doing this in various forums (e.g., Indiana University, ACI-REF, OSG, XSEDE, etc.) and can share best practices and knowledge and develop trusting collaborations across these efforts.

Marla Meehl strongly endorsed the ESnet science engagement reports and the OIN training.

Camille Criuttenden: there is a need to develop better methods to scale up the human infrastructure for cross-institution efforts. As solutions to some of these challenges, Crittenden advocated continuing to work with campus research IT, ESnet, and related groups, attending conferences and sharing best practices across various efforts, documenting and sharing case studies for successful projects, and growing a workforce in engagement, particularly starting with students and other tactics for workforce development.

John Towns (UIUC/NCSA) asked if there has been formal documentation of use cases and research needs in order to develop requirements. Camille Crittenden replied that the PRP is probably at a good point now to start doing that. Towns noted that XSEDE has documented various science use cases that could be reviewed.

Tom DeFanti commented on the science engagement staffing challenges. He said that it is easier for universities to hire/retain people in the domain sciences than it is in IT. Maybe recruitment for science engagement should emphasize domain scientists rather than IT experts.
Alex Szalay: it is more effective to empower the users and provide a workbench for them to build and share their own databases, analysis tools, and archival capabilities. We need to make the technologies more sharable.

Harvey Newman (Caltech) recalled that at the beginning of LHC planning within DOE, the first step was to convince people that the cyberinfrastructure budget would not be zero. He sees numbers around 20% for many projects. He also suggested that, based on his experience, it would be valuable to form scenarios of network needs, and derive examples of science-driven requirements.

During a breakout it was suggested that storage repositories could be NRP-facilitated.

Eli Dart commented on Jamison’s talk and said that he is glad the marketplace is responding to the need for elephant flows in both the research and enterprise community. This technology will help sites that are required to have a firewall. He thinks it is important to come up with best practices and is anxious to see how this equipment performs in SCinet 2017 in November.

Tim Lance (NYSERNet) said that this may be a time to look at these network security issues again, and he wants to talk with researchers at his institutions.

Michal Krsek (CESNET) commented that, in his experience working with audio/video data transport, non-permanent data streams are fragile with respect to delay/jitter. He said it is okay to bypass a firewall, but he needs signaling, e.g., there are six firewalls to bypass to get data from the Czech Republic to San Diego. He encouraged hardware developers to think about multi-domain environments and signaling.

Jerry Sheehan: it is important for smaller institutions to be early adopters and add invaluable use cases from a more resource-constrained environment. “If not,” warned Sheehan, “the NRP will be built for the 1% by default (not by intent).” the national CI community, such as the PRP, FIONA designs, Globus, perfSONAR, and technical mailing lists.

Tracy Futhey stated the NRP concept can clearly increase the rewards to researchers via collaborations and access to data, but the “risk people have veto power.” And arguments must be convincing for them not to use that veto power. “The NRP is unlikely to succeed in the high reward medical arena (except anecdotally) unless it reduces the risk as well.” Medical center administrators are very risk-averse and for the NRP to have impact in medical research, it must substantively address their risk concerns.

Von Welch: Start with the premise that the goal of cybersecurity is to support a mission by managing risks to IT, and research and science are part of the mission in the same way that enterprise functions are part of the mission. In this context, an enterprise network and a science network, both with appropriate cybersecurity for their uses, makes sense. For scaling the NRP to a large number of institutions, the NRP needs to get more CIOs and CISOs behind this different perspective. To do so, it would help for the CIOs and CISOs from early adopter institutions to communicate early success stories to their peers.

Jason Arviso: Tribal colleges and universities STEM programs must address K-12 system shortcomings, provide remedial classes, engage students in their community (projects that affect
them), build student competence, stimulate job creation, and be responsive to local/regional STEM workforce needs.

Mount Allen advocated equitable treatment across organizations and reminded the networking community not to give the equivalent of “second-hand books” to the communities that have historically gotten them.

Von Welch replied it is a tough challenge to get the right security expectations from sites: If the bar is not high enough, they don’t get enough value from the federation, but if the bar is too high, then it is too difficult to join the federation. He acknowledged there may be tiers to that problem and observed it is “easier to pull the Band-Aid once, up front,” and ensure that the difficulties are clear in the beginning and don’t stretch out over time. Finally, he said it helps to build on other available widespread tools, such as InCommon.

The Open Storage Network (OSN) proposal is for NSF to fund (~$30M–$40M) a national, distributed storage system with standard 1–2PB storage racks deployed across ~100 universities with high-speed network connections. Tim Lance (NYSERNet) also thought the OSN is a great idea and highlighted the two comments that policy management is hard, and wide buy-in and trust are required. He thinks the big data hubs are too new and inadequately funded to serve as catalysts for this, but regional collaborations, as illustrated by the PRP, already have a community of trust. Glenn Ricart (US IGNITE) weighed in with his support for the proposal, commenting that big data is evolving to be at least as important as big computing in scientific research, and in the context of this workshop, high-performance networking is a key enabler of the distributed storage system. Alex Szalay: Many universities now have diverse backup solutions, and the OSN could serve as a cost-efficient backup service, with geographically-distributed sites. Chris Hoffman (UC Berkeley) thought that in selling this concept to universities, it would be important to highlight not only petabyte-scale datasets, but also terabyte- or even gigabyte-scale datasets. Also, the OSN could work with libraries and other units on campuses that are not involved in extreme-scale science.

Alex Szalay replied that Andrew Moore at Carnegie Mellon University is interested in petabyte-scale datasets for machine learning, not datasets that can be easily downloaded. His experience at Johns Hopkins demonstrates that many researchers on campus need at least terabyte-scale datasets.

Larry Conrad (UC Berkeley) asked what the right funding source is for the OSN. Alex Szalay proposes the NSF seed the OSN with standard storage building blocks and software stacks, with perhaps a planned refresh in five years. Universities could augment the NSF-provided resources with their own funding.

Tim Lance (NYSERNet) is not convinced that NSF should seed the proposal. Separately, he commented that if campuses are to distribute their data to other sites, there needs to be a believable way to get data quickly back to the original campus.

Barr von Oehson sees positive developments with science engagement efforts, the CC* CI-Engineer awards, and the Science DMZ and DTN concepts (including FIONAs) but would like more standardization and best practices to emerge.

Inder Monga (ESnet) opened his comments with “culture eats strategy for lunch.” Institutional cultures, particularly with regard to security and risk management, must be addressed. A uniform understanding of security requirements and approaches is needed. A cultural transformation is
also required with respect to science engagement; more generalists are needed who can work across multiple technical systems and work with scientists to solve their problems.

Many institutions lose excellent staff to positions with better security (especially relative to soft money positions) and better pay. Many advocated getting staff into more stable positions (e.g., permanent funds), with good career prospects for pay and advancement.

Celeste Anderson (USC) commented that consultants had helped convince the administration that the networking infrastructure is critical to all functions of the campus enterprise – it shouldn’t be viewed just as a cost center – and will need regular technology refresh going forward. Steve Fulkerson noted that state/regional networks may be able to play a role in this dialog across their member campuses.

Jim Bottum expressed the need to engage and retain staff in this community, including professionalizing the career paths of campus cyberinfrastructure experts and the importance of determining what is required to scale up an NRP community nationally and beyond.

Jeff Weekley noted it is important that NRP builders think “from a user perspective, with an emphasis on user service, quality of service, and quality of experience.”

Larry Conrad (UC Berkeley) observed that the NRP “is more of a mesh architecture than a top-down architecture,” and it is necessary for individual universities in that mesh to want to make this happen. It makes sense for the regional networks to step in to support that type of architecture when there’s a modest number of participants who know and trust one another. It’s important for the regionals (and the Quilt, an organization of regionals) to be active participants in the NRP.

Harvey Newman (Caltech) said he “thinks we’ve lost the thread in this conversation.” The NRP should be addressing some challenges that have not been discussed. For example, many people have not been empowered to solve problems in a way that others have been. Newman said the NRP needs to be about new abstractions, new workflows, configuring end systems well, the nature of networks, the intelligence of networks, and how networks are used and how they react.

Cees DeLaat (UvA) noted that proposals in the European Union now need research data management plans. He advocated that researchers should be required also to state how they would use national or international cyberinfrastructure, in order to ensure awareness of available capabilities.

Wendy Huntoon (KINBER) commented that while many people have talked about funding, not everyone will be able to get grants, especially at smaller campuses. KINBER encourages its members to integrate ideas into their campuses and plan for additional capabilities, whether they get grants or not. In addition, she stressed the significant role that “people networking” plays. While she knows many people in the national community, not everyone has the benefit of that people network. By building trust relationships regionally and then nationally, we can move forward more quickly.

Larry Smarr: each campus that wants to really participate needs to have the CIO’s commitment of support.
Inder Monga (ESnet) noted that the first thing the PRP project did was to ask domain scientist users “are we doing the right thing for your science and is the return-on-investment there?” The PRP has brought together science and network experts in a conversation about needs and technology. Rather than just deploying a box and forgetting about it, the PRP needs scientists to use the capability. Monga also observed that as instruments get larger, there are fewer of them – one LHC, one LSST, a couple exascale machines – and it will be crucial for people to get access to data from those instruments to do research. Freeing data so it can be used is a key value-added issue for the PRP and its evolutions. The more facilities/data sources that connect, the greater the value to the scientific community.

Mark Berman (GENI Project Office) stated that the primary goals of this workshop and immediate next steps naturally focus on community development and planning for deploying basic NRP capabilities across a broader, national footprint. However, the pace of change is rapid in research infrastructure, as demonstrated by several promising technologies presented during the workshop’s “Scaling Science Communities” and “What Scaling Means to the Researcher” sessions. Berman stated that a long-lived NRP must include lifecycle planning for new capabilities, including the opportunity for staged and controlled rollout of candidate capabilities. By incorporating a testbed environment within the NRP and/or working with other research testbeds, first adopters within the NRP community can help to better define, validate, integrate, and harden emerging capabilities.

Camille Crittenden (UC Berkeley) commented that, as a champion of science engagement, the PRP needs to do a better job in science communications. She would like to look at good models for how to explain the science that is being done, collect use cases from scientists, develop good materials that could be posted on a website, and converge on coherent messages. She recognized that ESnet does a good job of this and invited others to suggest ideas.

John Towns (UIUC/NCSA) suggested that to “get to what ‘it’ is, ‘it’ should be driven by researcher needs.” While that is probably implicit, it had not been made explicit in the discussions so far. He stated the PRP needs to start building the use cases that can drive this. XSEDE has been building scientist-driven use cases for some time, and some of these existing use cases could probably be used for the NRP.

Jim Pepin (Clemson University) suggested using a Slack channel to find support. The NRP may need something at a higher bandwidth for scientists.

Srteve Huter: The model of Science DMZ/DTN/Engagement can be done in a systematic way across not only in the U.S. but also internationally and would improve the distributed cyberinfrastructure and accelerate scientific research. He cautioned that it is important to respect your peers – both in the network sense of peering and in the human sense – to cultivate a community of peers that benefits all parties.

Wendy Huntoon cited some of the conclusions from the Deep Dive 1 discussion on the role of regional networks in the NRP:

- Work on science collaborations
- Get FIONA boxes and start
• Organize communities – e.g., monthly calls, gain traction

• perfSONAR dashboard is effective – peer pressure works, especially when visible

Gil Gonzales concluded that “you need continuous engagement, rethinking the environment so that it’s aligned with what institutions need.”

Jen Leasure suggested that the PRP capture an operational list of “things to do to get started,” then convene a “coalition of the willing” of additional institutions and regionals that can identify science drivers, leverage institutions with knowledge/experience, and build on existing regional partnerships. Leasure noted that the NRP should “accept that coalition efforts may not look the same,” and there will be a process of building trust among that coalition. Existing engagement efforts (ESnet, Internet2, ACI-REF, CaRC, XSEDE, CASC, NSRC, etc.) need to be engaged, as well as training (e.g., ESnet, OIN workshops) and the availability of standardized affordable hardware (e.g., FIONA, FIONette). Leasure concluded that some level of funding will be required, whether institutional or federally provided.

Larry Smarr: Anyone can create an NRP session to get more input from the community and to keep the momentum going forward. Smarr continued, there are opportunities for pre-existing structures to begin their own efforts toward an NRP or GRP. For example, if a federation of regionals is an effective model for the NRP, regionals could take the initiative. Parallelism is the key to scaling rather than central control.

Smarr also noted Harvey Newman’s comments that the NRP should not pursue just incremental changes over the next 3-5 years but rather transformative changes. For example, ESnet has been engaged in a strategic planning exercise about its future, and there are likely to be major changes.

Smarr made some observations about several current and prospective NSF programs, saying “if we think that the country needs an NRP over the next three to five years, perhaps the NSF should issue a call for proposals to address the tough technical issues that will need to be addressed in an NRP; in fact, this could be a tremendous opportunity for DOE and NSF to work together.”

Smarr thanked NSF for the cyberengineer awards and said he hoped NSF will continue that program. Even if NSF doesn’t continue funding these positions, campuses should consider hiring cyberengineers anyway, and he encouraged people to go to NSF with ideas (e.g., emerging from the regionals).

Jim Bottum said that he was thinking about what got them motivated at Clemson to push on their cyberinfrastructure and recommended that “if you don’t have one already, find an Alex Feltus” on your campus, a scientist who needs advanced cyberinfrastructure and is willing to work with you. Feltus became a “burr under the saddle” at Clemson and now has evolved to a burr under the saddle of his collaborators’ campuses to bring them up to speed.
6.3 Workshop Agenda (Original)

The agenda below is the planned workshop agenda, with the original titles for presentations and sequence of speakers. The workshop notes above (Section 3) reflect actual titles and sequence of speakers.

Monday, August 7, 2017

8:30 AM-8:50 AM Welcome: Irene Qualters

8:50 AM-9:25 AM Leading Up to the NRP
Speaker: Larry Smarr

9:25 AM-10:00 AM Keynote: Scaling Challenges for Research Cyberinfrastructure
Speaker: Harvey Newman

10:00 AM-10:20 AM Break

10:20 AM-12:00 PM Session One: Scaling Science Communities--Lessons Learned and Future Plans.
Chair: Ruth Marinshaw
- Tom Cheatham, ACI-REF
- Frank Würthwein, Fixing The "Last Mile Problem" of The LHC In California
- Inder Monga, ESnet's Scaling Strategies to Accelerate to the NRP
- Ian Foster, Scaling the PRP Software Infrastructure to the NRP
- Rob Gardner, Services at The Edge (SLATE)
- Q&A

12:00 PM-2:00 PM Lunch and Optional Deep Dive Sessions

12:00 PM-1:00 PM Deep Dive One: Roles of Regionals, QUILT, LEARN, FLR, MREN, PNWGP & Others
Moderators: Ron Hutchins, John Moore

12:00 PM-1:00 PM Deep Dive Two: Scaling and Security: Not At The App Level, Please!
Speakers: Philip Papadopoulos, Von Welch

1:00 PM-2:00 PM Deep Dive Three: Opportunities in Research Cyberinfrastructure at NSF
Speaker: Amy Walton

1:00 PM-2:00 PM Deep Dive Four: How Does PRPv1 Work? --An Explanation of
DMZs, DTNs... Chair: Celeste Anderson
Speakers: Eli Dart, John Graham, John Hess, Thomas Hutton

2:00 PM-3:40 PM Session Two: What Scaling Means to the Researcher
Chair: Patrick Schmitz
- Chris Hoffman, Cyber-archaeology, Immersive Visualization, and the UC Catalyst Project
- Shawn McKee, OSiRIS: Scalable, Computable, Collaborative Storage
- Ray Idaszak, Scaling Up Water Science Research
- Alex Feltus and Claris Castillo, Linking the PRP to two other NSF projects: SciDAS (CC*) and Tripal Gateway (DIBBS): Engaging Multiple Domains.
- Q&A

3:40 PM-4:00 PM Break

4:00 PM-5:40 PM Session Three: Engaging Communities of Researchers
Chair: Thomas DeFanti
- Eli Dart, ESnet Science Engagement
- Marla Meehl, NCAR Climate Data Distribution Engagement Strategies
- Camille Crittenden, Scaling and Evaluating our Engagement
- Tom DeFanti, Machine Learning Community Infrastructure Engagement
- Q & A

6:00 PM-9:00 PM Dinner: Museum of the Rockies
600 West Kagy Blvd Bozeman, Montana 59717
Host: Jerry Sheehan

Tuesday, August 8

8:20 AM-8:55 AM Keynote: The Need for Big Data, Networking, Storage, and Machine Learning
Speaker: Alexander Szalay

9:00 AM-9:45 AM Special Session: High-Speed Equipment
Chair: John Hess
- JJ Jamison, Science DMZs behind firewalls: Heresy?
- Azher Mughal, Arista Deployments at USC
- Tim Martin, Cisco Research Computing & Networking Strategy for Campus IT

9:45 AM-10:00 AM Break
10:00 AM-12:00 PM  Session Four: NRP Impact on Big and Small Campuses  
Including MSI's  
Chair: Gil Gonzales  
• Jerry Sheehan, Supporting the NRP with Minimal Staffing  
• Tracy Futhey, NRP Med School Challenges  
• Von Welch, Security for DMZs  
• John Hess, What Measurements Can We Report to Show Engagement?  
• Jason Arviso, The NSF Tribal Colleges and Universities Program  
• Q&A

12:00 PM-2:00 PM  Lunch and Optional Deep Dive Sessions

12:00 PM-1:00 PM  Deep Dive Five: Creation of Community Data Resources  
Speakers: Jerry Sheehan, Alexander Szalay

12:00 PM-1:00 PM  Deep Dive Six: What's Worked and What Hasn't For High-Performance Networking  
Moderator: Richard Moore  
Speakers: Eli Dart, Steven Fulkerson, Inder Monga, James Barr von Oehsen

1:00 PM-2:00 PM  Deep Dive Seven: Strategies for Moving Forward; How To Build A Network of Regional DMZs  
Speakers: Jim Bottum, Larry Smarr

1:00 PM-2:00 PM  Deep Dive Eight: SC17 Demos  
Speakers: John Graham, Azher Mughal

2:00 PM-3:15 PM  Session Five: Towards A Global Research Platform (GRP)  
Chair: Joe Mambretti  
• Jeonghoon Moon, Towards a Global Research Platform (GRP)  
• Heidi Morgan, LSST Scaling Issues and Network Needs  
• Phil Papadopoulos, PRAGMA’s DMZ Strategy  
• Jennifer Schopf, International & National Network Roles/Requirements  
• Q&A

3:15 PM-3:40 PM  Break

3:40 PM-5:00 PM  Session Six: Democratizing Collaboration  
Chair: Maxine Brown
• Wendy Huntoon, Regional Research Platforms (RRPs)
• Gil Gonzales, Equity and Access
• Jen Leasure, National & Regional Network Roles / Requirements
• Steve Huter, Research Collaboration Opportunities in a Global Research Platform
• Cees de Laat, Global Collaborative Research Groups (CRGs)
• Q&A

5:00 PM-6:00 PM  Closing Session: What's Next
    Speakers: Jim Bottum, Larry Smarr
### 6.4 Workshop Registrants

A list of registrants, with affiliations and titles, is included below.

Note that speaker bios may be available at the original workshop website [http://www.cvent.com/events/the-first-national-research-platform-workshop-toward-a-national-big-data-superhighway/speakers-cf48695e2c074b488f3d7a33a0b01b87.aspx?lang=en&sms=7&cn=Lpj3pt1FgUmfMcP8bDq_QA](http://www.cvent.com/events/the-first-national-research-platform-workshop-toward-a-national-big-data-superhighway/speakers-cf48695e2c074b488f3d7a33a0b01b87.aspx?lang=en&sms=7&cn=Lpj3pt1FgUmfMcP8bDq_QA).

<table>
<thead>
<tr>
<th>Last Name</th>
<th>First Name</th>
<th>Affiliation</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allen</td>
<td>Mount</td>
<td>San Francisco Jazz Organization [SFJAZZ]</td>
<td>Director of Operations</td>
</tr>
<tr>
<td>Amalfi*</td>
<td>Franco</td>
<td>Oracle</td>
<td>Director, Strategic Government Programs</td>
</tr>
<tr>
<td>Anderson</td>
<td>Celeste</td>
<td>Pacific Wave</td>
<td>Director, Customer Relations</td>
</tr>
<tr>
<td>Arviso</td>
<td>Jason</td>
<td>Navajo Technical University</td>
<td>Director of Information Technology</td>
</tr>
<tr>
<td>Barber</td>
<td>David</td>
<td>Oregon State University</td>
<td>Senior Program Manager</td>
</tr>
<tr>
<td>Barker</td>
<td>Mike</td>
<td>University of North Carolina at Chapel Hill</td>
<td>Chief Technology Officer</td>
</tr>
<tr>
<td>Baru</td>
<td>Chaitan</td>
<td>National Science Foundation</td>
<td>Senior Advisor for Data Science</td>
</tr>
<tr>
<td>Benedetto</td>
<td>Michael</td>
<td>American Museum of Natural History</td>
<td>Director of Information Technology and Deputy CIO</td>
</tr>
<tr>
<td>Berman</td>
<td>Mark</td>
<td>GENI Project Office</td>
<td>GENI Project Director</td>
</tr>
<tr>
<td>Berman</td>
<td>Michael</td>
<td>California State University, Channel Islands</td>
<td>Vice President for Technology and Innovation</td>
</tr>
<tr>
<td>Bezerra</td>
<td>Jeronimo</td>
<td>AmLight Project, Florida International University</td>
<td>Chief Network Engineer</td>
</tr>
<tr>
<td>Bottum</td>
<td>Jim</td>
<td>Internet2</td>
<td>Internet2 Presidential Fellow</td>
</tr>
<tr>
<td>Boyd</td>
<td>Eric</td>
<td>University of Michigan</td>
<td>Director of Research Networks</td>
</tr>
<tr>
<td>Breen</td>
<td>Joe</td>
<td>University of Utah Center for High Performance Computing</td>
<td>IT Architect</td>
</tr>
<tr>
<td>Brown</td>
<td>Maxine</td>
<td>University of Illinois at Chicago</td>
<td>Director, Electronic Visualization Laboratory</td>
</tr>
<tr>
<td>Bruce</td>
<td>David</td>
<td>University of Arkansas</td>
<td>Deputy CIO</td>
</tr>
<tr>
<td>Castillo</td>
<td>Claris</td>
<td>RENCI - UNC Chapel Hill</td>
<td>Senior Scientist</td>
</tr>
<tr>
<td>Chandramoulswaran*</td>
<td>Ishwar</td>
<td>NIAID, NIH</td>
<td>Project Officer, Bioinformatics</td>
</tr>
<tr>
<td>Chase</td>
<td>Wallace</td>
<td>Clemson University</td>
<td>Executive Director of Networking &amp; Telecommunications</td>
</tr>
<tr>
<td>Cheatham</td>
<td>Thomas</td>
<td>University of Utah</td>
<td>Professor/Director</td>
</tr>
<tr>
<td>Cheetham</td>
<td>Jan</td>
<td>University of Wisconsin-Madison</td>
<td>Research Cyberinfrastructure Liaison</td>
</tr>
<tr>
<td>Christian</td>
<td>Patrick</td>
<td>University of Wisconsin-Madison</td>
<td>Assistant Director of Networks</td>
</tr>
<tr>
<td>Cleveland</td>
<td>Sean</td>
<td>University of Hawaii</td>
<td>CI Research Scientist</td>
</tr>
<tr>
<td>Name</td>
<td>First Name</td>
<td>Last Name</td>
<td>Affiliation</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Collins</td>
<td>Kurt</td>
<td></td>
<td>California State University, San Bernadino</td>
</tr>
<tr>
<td>Conrad</td>
<td>Larry</td>
<td></td>
<td>University of California, Berkeley</td>
</tr>
<tr>
<td>Corbato</td>
<td>Steve</td>
<td></td>
<td>Oregon Health &amp; Science University</td>
</tr>
<tr>
<td>Cothren</td>
<td>Jack</td>
<td></td>
<td>University of Arkansas</td>
</tr>
<tr>
<td>Crittenden</td>
<td>Camille</td>
<td></td>
<td>UC Berkeley, CITRIS and the Banatao Institute</td>
</tr>
<tr>
<td>Cruz*</td>
<td>Jared</td>
<td></td>
<td>Docomo Pacific</td>
</tr>
<tr>
<td>D'Angelo</td>
<td>Cas</td>
<td></td>
<td>Southern Crossroads</td>
</tr>
<tr>
<td>Dart</td>
<td>Eli</td>
<td></td>
<td>ESnet</td>
</tr>
<tr>
<td>de Laat</td>
<td>Cees</td>
<td></td>
<td>University of Amsterdam</td>
</tr>
<tr>
<td>Deaton</td>
<td>James</td>
<td></td>
<td>Great Plains Network</td>
</tr>
<tr>
<td>DeFanti</td>
<td>Thomas</td>
<td></td>
<td>University of California, San Diego</td>
</tr>
<tr>
<td>Edelman*</td>
<td>Adam</td>
<td></td>
<td>Montana State University</td>
</tr>
<tr>
<td>Eklund</td>
<td>Daniel</td>
<td></td>
<td>University of Michigan</td>
</tr>
<tr>
<td>Eldayrie*</td>
<td>Elias</td>
<td></td>
<td>University of Florida</td>
</tr>
<tr>
<td>Exleben*</td>
<td>Jennifer</td>
<td></td>
<td>USGS</td>
</tr>
<tr>
<td>Feltus</td>
<td>Alex</td>
<td></td>
<td>Clemson University</td>
</tr>
<tr>
<td>Finne</td>
<td>Alan</td>
<td></td>
<td>University of Arkansas for Medical Sciences</td>
</tr>
<tr>
<td>Foster</td>
<td>Ian</td>
<td></td>
<td>University of Chicago &amp; Argonne National Lab</td>
</tr>
<tr>
<td>Fox</td>
<td>Louis</td>
<td></td>
<td>CENIC</td>
</tr>
<tr>
<td>Fulkerson</td>
<td>Steven</td>
<td></td>
<td>Arkansas Research &amp; Education Optical Network</td>
</tr>
<tr>
<td>Futhey</td>
<td>Tracy</td>
<td></td>
<td>Duke University</td>
</tr>
<tr>
<td>Gardner*</td>
<td>Mark</td>
<td></td>
<td>Virginia Tech</td>
</tr>
<tr>
<td>Gardner</td>
<td>Rob</td>
<td></td>
<td>University of Chicago</td>
</tr>
<tr>
<td>Genung</td>
<td>Scott</td>
<td></td>
<td>University of Chicago</td>
</tr>
<tr>
<td>Gonzales</td>
<td>Gil</td>
<td></td>
<td>Gonzales Consulting</td>
</tr>
<tr>
<td>Gorman</td>
<td>Trisha</td>
<td></td>
<td>Pacific Interface</td>
</tr>
<tr>
<td>Graham</td>
<td>John</td>
<td></td>
<td>University of California, San Diego</td>
</tr>
<tr>
<td>Harden</td>
<td>Ryan</td>
<td></td>
<td>University of Chicago</td>
</tr>
<tr>
<td>Herr</td>
<td>Laurin</td>
<td></td>
<td>Pacific Interface Inc.</td>
</tr>
<tr>
<td>Hess</td>
<td>Gregory</td>
<td></td>
<td>Montana State University</td>
</tr>
<tr>
<td>Hess</td>
<td>John</td>
<td></td>
<td>CENIC/Pacific Wave/PRP</td>
</tr>
<tr>
<td>Hilmer</td>
<td>Jonathan</td>
<td></td>
<td>Montana State University</td>
</tr>
<tr>
<td>Name</td>
<td>First Name</td>
<td>Affiliation</td>
<td>Position</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Hofer</td>
<td>Erik</td>
<td>University of Michigan, School of Information</td>
<td>Clinical Assistant Professor &amp; CIO</td>
</tr>
<tr>
<td>Hoffman</td>
<td>Chris</td>
<td>University of California, Berkeley</td>
<td>Program Director, Research IT</td>
</tr>
<tr>
<td>House</td>
<td>Harry</td>
<td>USGS</td>
<td>IT Specialist</td>
</tr>
<tr>
<td>Hunsinger</td>
<td>Ana</td>
<td>Internet2</td>
<td>Vice President, Community Engagement</td>
</tr>
<tr>
<td>Hunter</td>
<td>David</td>
<td>Indiana University</td>
<td>Network Design Engineer</td>
</tr>
<tr>
<td>Huntton</td>
<td>Wendy</td>
<td>KINBER</td>
<td>President &amp; CEO</td>
</tr>
<tr>
<td>Hutchins</td>
<td>Ron</td>
<td>University of Virginia</td>
<td>Vice President, IT</td>
</tr>
<tr>
<td>Huter</td>
<td>Steven</td>
<td>Network Startup Resource Center (NSRC)</td>
<td>Director</td>
</tr>
<tr>
<td>Hutton</td>
<td>Thomas</td>
<td>San Diego Supercomputer Center</td>
<td>HPC Network Architect</td>
</tr>
<tr>
<td>Hwang</td>
<td>Tae</td>
<td>Cisco Systems</td>
<td>Architect</td>
</tr>
<tr>
<td>Idaszak</td>
<td>Ray</td>
<td>RENCI; University of North Carolina at Chapel Hill</td>
<td>Director of DevOps</td>
</tr>
<tr>
<td>Jamison</td>
<td>JJ</td>
<td>Juniper Networks</td>
<td>Chief Technology Officer, Strategic Verticals - Americas</td>
</tr>
<tr>
<td>Janssen</td>
<td>Jerry</td>
<td>NOAA</td>
<td>Executive Director</td>
</tr>
<tr>
<td>Jennewein</td>
<td>Doug</td>
<td>University of South Dakota</td>
<td>Director of Research Computing</td>
</tr>
<tr>
<td>Jent</td>
<td>David</td>
<td>Indiana University</td>
<td>Associate Vice President, Networks</td>
</tr>
<tr>
<td>Johnson</td>
<td>Mark</td>
<td>MCNC</td>
<td>Chief Technology Strategist</td>
</tr>
<tr>
<td>Johnson</td>
<td>Ron</td>
<td>University of Washington &amp; PNWGP</td>
<td>Chief Executive Officer &amp; Professor</td>
</tr>
<tr>
<td>Kettimuthu*</td>
<td>Raj</td>
<td>Argonne National Laboratory</td>
<td>Computer Scientist</td>
</tr>
<tr>
<td>Kimm</td>
<td>Jeffrey</td>
<td>Montana State University</td>
<td>Network Analyst</td>
</tr>
<tr>
<td>Klingenstein</td>
<td>Ken</td>
<td>Internet2</td>
<td>Director</td>
</tr>
<tr>
<td>Kneifel</td>
<td>Charle</td>
<td>Duke University</td>
<td>Senior Technical Director</td>
</tr>
<tr>
<td>Kovalchick</td>
<td>Ann</td>
<td>University of California Merced</td>
<td>Associate Vice Chancellor &amp; CIO</td>
</tr>
<tr>
<td>Kowalski</td>
<td>Karl</td>
<td>University of Alaska</td>
<td>Chief Information Technology Officer</td>
</tr>
<tr>
<td>Krsek</td>
<td>Michal</td>
<td>CESNET</td>
<td>Senior Researcher</td>
</tr>
<tr>
<td>Lance</td>
<td>Timothy</td>
<td>NYSERNET</td>
<td>Chief Research Officer</td>
</tr>
<tr>
<td>Lazor</td>
<td>Joseph</td>
<td>Florida LambdaRail</td>
<td>Chief Executive Officer</td>
</tr>
<tr>
<td>Leasure</td>
<td>Jen</td>
<td>The Quilt</td>
<td>President</td>
</tr>
<tr>
<td>Lee</td>
<td>Joyce</td>
<td>The National Coordination Office</td>
<td>CTR</td>
</tr>
<tr>
<td>Livny</td>
<td>Miron</td>
<td>University of Wisconsin-Madison</td>
<td>Professor of Computer Science</td>
</tr>
<tr>
<td>Llovet</td>
<td>Pol</td>
<td>Montana State University</td>
<td>Associate Director of Research Cyberinfrastructure</td>
</tr>
<tr>
<td>Macdonell</td>
<td>James</td>
<td>California State University, San Bernadino</td>
<td>Information Security Analyst</td>
</tr>
<tr>
<td>Name</td>
<td>First</td>
<td>Institution</td>
<td>Position</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------</td>
<td>------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Mambretti</td>
<td>Joe</td>
<td>Northwestern University</td>
<td>Director International Center for Advanced Internet Research</td>
</tr>
<tr>
<td>Manzoul*</td>
<td>Mahmoud</td>
<td>Jackson State University</td>
<td>Professor</td>
</tr>
<tr>
<td>Marble</td>
<td>David</td>
<td>OSHEAN</td>
<td>Chief Executive Officer</td>
</tr>
<tr>
<td>Marentette</td>
<td>Joseph</td>
<td>Washington University, St. Louis</td>
<td>Network Engineer 3</td>
</tr>
<tr>
<td>Marinshaw</td>
<td>Ruth</td>
<td>Stanford University</td>
<td>Chief Technology Officer - Research Computing</td>
</tr>
<tr>
<td>Martin</td>
<td>Tim</td>
<td>Cisco Systems</td>
<td>Product Specialist</td>
</tr>
<tr>
<td>McClure*</td>
<td>Mimi</td>
<td>National Science Foundation</td>
<td>Program Director</td>
</tr>
<tr>
<td>McKee</td>
<td>Shawn</td>
<td>University of Michigan</td>
<td>Research Scientist</td>
</tr>
<tr>
<td>McMullen</td>
<td>Rick</td>
<td>Texas A&amp;M University</td>
<td>Associate Director High Performance Research Computing</td>
</tr>
<tr>
<td>Meehl</td>
<td>Marla</td>
<td>UCAR</td>
<td>Manager of NCAR/UCAR/FRGP/BiSON Networking</td>
</tr>
<tr>
<td>Merrifield</td>
<td>David</td>
<td>Arkansas Research and Education Optical Network</td>
<td>Chief Technology Officer</td>
</tr>
<tr>
<td>Miller</td>
<td>Ken</td>
<td>Penn State</td>
<td>CI Engineer</td>
</tr>
<tr>
<td>Minton</td>
<td>Jessie</td>
<td>University of Oregon</td>
<td>Vice Provost &amp; Chief Information Officer</td>
</tr>
<tr>
<td>Monga</td>
<td>Inder</td>
<td>ESnet</td>
<td>Director</td>
</tr>
<tr>
<td>Moon</td>
<td>Jeonghoon</td>
<td>KISTI</td>
<td>Researcher</td>
</tr>
<tr>
<td>Moore</td>
<td>Richard</td>
<td>University of California, San Diego</td>
<td>(Emeritus)</td>
</tr>
<tr>
<td>Moore</td>
<td>John</td>
<td>Internet2</td>
<td>AVP - Network Architecture &amp; Planning</td>
</tr>
<tr>
<td>Morgan</td>
<td>Heidi</td>
<td>University of Southern California - Information Science Institute</td>
<td>Senior Computer Scientist</td>
</tr>
<tr>
<td>Mughal</td>
<td>Azher</td>
<td>University of Southern California</td>
<td>Senior Internet Engineer</td>
</tr>
<tr>
<td>Murali</td>
<td>Viji</td>
<td>University of California, Davis</td>
<td>Vice President &amp; CIO</td>
</tr>
<tr>
<td>Newman</td>
<td>Harvey</td>
<td>Caltech</td>
<td>Professor of Physics</td>
</tr>
<tr>
<td>Nordmark</td>
<td>Robert</td>
<td>OneNet</td>
<td>Director</td>
</tr>
<tr>
<td>Oswald*</td>
<td>David</td>
<td>FDA/CFSAN</td>
<td>Program Manager</td>
</tr>
<tr>
<td>Owens*</td>
<td>Courtney</td>
<td>USGS</td>
<td>Cloud Hosting Solutions Migration Specialist</td>
</tr>
<tr>
<td>Papadopoulos</td>
<td>Philip</td>
<td>UCSD/Calit2/SDSC</td>
<td>Associate Research Scientist</td>
</tr>
<tr>
<td>Pappin*</td>
<td>Nick</td>
<td>Washington State University - CAHNRS</td>
<td>Systems Engineer</td>
</tr>
<tr>
<td>Pepin</td>
<td>James</td>
<td>Clemson University</td>
<td>Chief Technology Officer</td>
</tr>
<tr>
<td>Pfeffer</td>
<td>Howard</td>
<td>Internet2</td>
<td>President &amp; CEO</td>
</tr>
<tr>
<td>Polterrock</td>
<td>Josh</td>
<td>University of California, San Diego</td>
<td>Manager Scientific Projects</td>
</tr>
</tbody>
</table>

79
<table>
<thead>
<tr>
<th>Name</th>
<th>Last Name</th>
<th>Organization and Affiliation</th>
<th>Position/Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poppe</td>
<td>Yves</td>
<td>NSCC (National Supercomputing Centre Singapore)</td>
<td>Transcontinental Supercomputing Networking</td>
</tr>
<tr>
<td>Prior</td>
<td>Fred</td>
<td>University of Arkansas for Medical Sciences</td>
<td>Chair – Department of Biomedical Informatics</td>
</tr>
<tr>
<td>Qualters</td>
<td>Irene</td>
<td>Office of Advanced CyberInfrastructure</td>
<td>Director</td>
</tr>
<tr>
<td>Quire</td>
<td>Kevin</td>
<td>Utah Education and Telehealth Network</td>
<td>Senior Network Engineer</td>
</tr>
<tr>
<td>Reese</td>
<td>David</td>
<td>CENIC</td>
<td>Vice President, Regional &amp; International Activities</td>
</tr>
<tr>
<td>Ricart</td>
<td>Glenn</td>
<td>US Ignite</td>
<td>Founder &amp; CTO</td>
</tr>
<tr>
<td>Riley</td>
<td>Matt</td>
<td>University of Montana</td>
<td>Chief Information Officer</td>
</tr>
<tr>
<td>Riter</td>
<td>Stephen</td>
<td>UTEP</td>
<td>Vice President &amp; CIO</td>
</tr>
<tr>
<td>Robb*</td>
<td>Chris</td>
<td>Indiana University</td>
<td>Senior Network Manager</td>
</tr>
<tr>
<td>Rosolen</td>
<td>Brett</td>
<td>AARNet Pty Ltd</td>
<td>Data Program Manager, eResearch</td>
</tr>
<tr>
<td>Royal</td>
<td>Von</td>
<td>Oklahoma State Regents for Higher Education</td>
<td>Chief Information Officer</td>
</tr>
<tr>
<td>Russell</td>
<td>John</td>
<td>AAAS @ National Science Foundation</td>
<td>AAAS S&amp;T Policy Fellow</td>
</tr>
<tr>
<td>Schmitz</td>
<td>Patrick</td>
<td>University of California, Berkeley</td>
<td>Associate Director, Research IT</td>
</tr>
<tr>
<td>Schopf</td>
<td>Jennifer</td>
<td>Indiana University</td>
<td>Director, International Networks</td>
</tr>
<tr>
<td>Schopis</td>
<td>Paul</td>
<td>Ohio Academic Resource Network</td>
<td>Interim Executive Director</td>
</tr>
<tr>
<td>Schroeder*</td>
<td>Tracy</td>
<td>Boston University</td>
<td>Vice President, IS&amp;T</td>
</tr>
<tr>
<td>Sedore*</td>
<td>Christopher</td>
<td>NYSERNet.org, Inc.</td>
<td>President</td>
</tr>
<tr>
<td>Shah</td>
<td>Pankaj</td>
<td>LEARN</td>
<td></td>
</tr>
<tr>
<td>Sheehan</td>
<td>Jerry</td>
<td>Montana State University</td>
<td>Vice President &amp; CIO</td>
</tr>
<tr>
<td>Smarr</td>
<td>Larry</td>
<td>Calit2 at University of California, San Diego</td>
<td>Director, Calit2</td>
</tr>
<tr>
<td>Smith*</td>
<td>Steve</td>
<td>University of Nevada, Reno</td>
<td>Chief Information Officer</td>
</tr>
<tr>
<td>Sosnkowski</td>
<td>Mike</td>
<td>Virginia Commonwealth University</td>
<td>Deputy Director Network Services</td>
</tr>
<tr>
<td>Stanton</td>
<td>Michael</td>
<td>RNP (Brazilian National Research and Education Network)</td>
<td>Director</td>
</tr>
<tr>
<td>Szalay</td>
<td>Alexander</td>
<td>The Johns Hopkins University</td>
<td>Professor</td>
</tr>
<tr>
<td>Towns</td>
<td>John</td>
<td>NCSA</td>
<td>XSEDE director and NCSA Deputy Director</td>
</tr>
<tr>
<td>Vania</td>
<td>Gi</td>
<td>University of Texas at Dallas</td>
<td>Director Enterprise Architecture</td>
</tr>
<tr>
<td>Vaughn</td>
<td>Ryan</td>
<td>Florida LambdaRail</td>
<td>Senior Network Engineer</td>
</tr>
<tr>
<td>Verrant</td>
<td>Jeff</td>
<td>Ciena Government Solutions</td>
<td>Vice President &amp; General Manager</td>
</tr>
<tr>
<td>von Oeelsen</td>
<td>Barr</td>
<td>Rutgers University</td>
<td>Associate Vice President</td>
</tr>
<tr>
<td>Walton</td>
<td>Amy</td>
<td>National Science Foundation</td>
<td>Program Director</td>
</tr>
<tr>
<td>Name</td>
<td>Title</td>
<td>Institution/Position</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Wang*</td>
<td>Kuang-Ching</td>
<td>Clemson University, Associate Professor</td>
<td></td>
</tr>
<tr>
<td>Weekley</td>
<td>Jeffrey</td>
<td>UC Merced, Director of CI &amp; RC</td>
<td></td>
</tr>
<tr>
<td>Weilhamer</td>
<td>Caroline</td>
<td>I-Light Network / Indiana Gigapop, Manager of Operations</td>
<td></td>
</tr>
<tr>
<td>Welch</td>
<td>Von</td>
<td>Indiana University, Director, Center for Applied Cybersecurity Research (CACR)</td>
<td></td>
</tr>
<tr>
<td>White*</td>
<td>Heath</td>
<td>Montana State University, Network Analyst</td>
<td></td>
</tr>
<tr>
<td>Woo</td>
<td>Melissa</td>
<td>Stony Brook University, Vice President for IT &amp; Chief Information Officer</td>
<td></td>
</tr>
<tr>
<td>Würthwein</td>
<td>Frank</td>
<td>SDSC/UCSD, Professor</td>
<td></td>
</tr>
<tr>
<td>Yashar</td>
<td>Mark</td>
<td>CITRIS (Center for Information Technology Research in the Interest of Society), UC Berkeley, Specialist</td>
<td></td>
</tr>
<tr>
<td>Zysman*</td>
<td>Joel</td>
<td>University of Miami, Director of Advanced Computing</td>
<td></td>
</tr>
</tbody>
</table>

* Registered for workshop but unable to attend.