Workshop Award #1540112 Report
Building the Pacific Research Platform:
A Workshop Towards Deploying a
Science-Driven Regional ‘Big Data Freeway’

[Image of Pacific Research Platform diagram]
October 14-16, 2015

California Institute for Telecommunications and Information Technology (Calit2)
University of California, San Diego

This workshop is supported by NSF workshop award #1540112, as well as NSF grants ACI-1246396 and ACI-1541349. Additional support is greatly appreciated from the University of California Office of the President (UCOP), the Corporation for Education Network Initiatives in California (CENIC), Center for Information Technology Research in the Interest of Society (CITRIS), and the California Institute for Telecommunications and Information Technology (Calit2).

Table of Contents

1 Executive Summary ............................................................................................................... 3
2 Workshop Context and Objectives ......................................................................................... 4

3 Workshop Presentations and Panel Discussions ................................................................. 6
3.1 Workshop Overview ............................................................................................................. 6
3.2 PRP Introduction Session ................................................................................................ 8
3.3 Science Engagement Session .......................................................................................... 10
  3.3.1 What is Science Engagement? (Eli Dart) ................................................................. 12
  3.3.2 Particle physics (moderator Frank Würthwein) ......................................................... 12
  3.3.3 Astronomy and astrophysics (moderator Mike Norman) ....................................... 14
  3.3.4 Biomedical Sciences (moderator Larry Smarr) ....................................................... 15
  3.3.5 Earth Sciences (moderator Camille Crittenden) ...................................................... 16
  3.3.6 Scalable Visualization, Virtual Reality, and Ultra-resolution Video (moderator Tom DeFanti) ......................................................................................................................... 17
  3.3.7 General Discussion (moderators Greg Bell and Camille Crittenden) ................. 20
3.4 Technical Approach Session .......................................................................................... 21
  3.4.1 The PRPv1 Architecture Model (moderator John Hess) ....................................... 21
  3.4.2 Securing the PRPv1 (moderator Eli Dart) ............................................................... 22
  3.4.3 PRPv2 Architecture and Security (moderator John Haskins) ............................. 24
  3.4.4 International Perspective (moderator Joe Mambretti) ........................................... 25

4 PRP Requirements .............................................................................................................. 26
5 Findings and Recommendations ......................................................................................... 29
6 Acknowledgements ............................................................................................................ 32
7 References .......................................................................................................................... 33
8 Appendices .......................................................................................................................... 34
  8.1 Workshop Agenda, Presentation Materials and Videos ............................................. 34
  8.2 Workshop Registrants ................................................................................................. 38
1 Executive Summary

The Pacific Research Platform (PRP) is an ambitious project driven by the high-speed networking needs of collaborative, big-data science. Many research disciplines are increasingly multi-investigator and multi-institutional and need ever more rapid access to their ultra-large heterogeneous and widely distributed datasets. In response to this challenge, the Department of Energy’s ESnet developed the Science DMZ model, a network system optimized for high-performance scientific applications rather than for general-purpose or enterprise computing. The PRP will extend the campus Science DMZ model, which is widely funded on individual campuses by NSF’s CC-NIE and CC-NII programs, to a regional model for data-intensive networking. It will enable researchers to quickly and easily move data between collaborator labs, supercomputer centers, and data repositories, creating a big-data freeway that allows the data to traverse multiple, heterogeneous networks without performance degradation. The PRP’s data sharing architecture, with end-to-end 10–100Gb/s connections, will enable region-wide virtual co-location of data with computing.

The main focus of the PRP project is to build a researcher-defined and data-focused network whose requirements are driven by direct engagements with sophisticated, cyberinfrastructure-knowledgeable Science Teams chosen from the fields of particle physics, astronomy, biomedical sciences, earth sciences, and scalable data visualization. It is a partnership of more than 20 institutions, including four National Science Foundation, Department of Energy, and NASA supercomputer centers. The purpose of this workshop was to bring together representatives from all PRP partner institutions, participating Science Team members, technical staff, and network architects and implementers to discuss deployment of the PRP and address interoperability of Science DMZs at a regional level.

The PRP Workshop was held October 14–16, 2015 at the Qualcomm Institute on the UC San Diego campus. The workshop began on Wednesday morning (Oct. 14) with an introduction to the Pacific Research Platform from the viewpoint of partner administrators, campus IT managers, and the National Science Foundation, which has provided funding for the project. Details of that session are provided in Section 3.2 of this report. On Wednesday afternoon, workshop attendees were invited to attend a UC San Diego event celebrating the San Diego Supercomputer Center’s 30th anniversary and the launch of their newest high-performance computing system, Comet. Thursday (Oct. 15) was the first full-day of the workshop and was devoted to researchers and technical experts in science-driver applications describing their needs for high-speed data transfer and their successes and frustrations. Sessions included groups from key initial areas targeted for the PRP, including high-energy physics, astronomy, biomedical sciences, earth sciences and visualization. Those sessions are described in detail in Section 3.3. On the final day of the workshop, discussions focused on requirements from the domain scientists and the networking architecture, policies, tools and security necessary to deploy the 20-institution Pacific Research Platform. The workshop technical sessions are described in Section 3.4 and a summary of PRP technical requirements is provided in Section 4. All sessions were followed by Q & A panels where participants were encouraged to ask questions and share their thoughts on the topics under discussion.

This first PRP workshop was successful in gathering key stakeholders in the PRP project, including administrators, scientists, network engineers and researchers, to ascertain the science-driven requirements of the project and discuss ideas and priorities for the project going forward. Section 5 of this report discusses the findings and recommendations from the workshop. Overall, it was agreed that the PRP project is a social engineering project as well as a technical
networking/IT project and that engagement with the Science Teams that are ready to begin using
the PRP is crucial to the project’s success. It was also noted that scientists want to do science, not
networking or IT. Therefore, an effective partnership will have cyberinfrastructure experts
working with scientists at their interface and understanding the desired scientific outcomes, rather
than viewing the technology as an end to itself. With these facts in mind, it was recommended
that the PRP work with science teams to build a system that meets their immediate needs rather
than take a “build it and they will come” approach. It was also recommended that the PRP
identify common functionality that can be leveraged between science applications to make the
PRP more efficient and effective. Additionally, it was suggested that the PRP should prioritize
high-performance access to supercomputer centers and also create a forum for ongoing discussion
of issues such as software and security.

2 Workshop Context and Objectives

Many research disciplines are increasingly multi-investigator and multi-institutional, and need
rapid access to large, heterogeneous and widely distributed datasets (‘Big Data’). High-
performance networking is a key enabling element for this research, which requires large, fast
transfers of data on a regional, national and international scale. Even with advances in network
backbone speeds, it has always been a technical challenge to achieve high-speed networking
performance between end-points where researchers typically work (labs, instruments, computers).
In response to this challenge, the Department of Energy’s ESnet developed the Science DMZ
model [1], which is a major advance in developing designs and policies that reduce network
roadblocks for scientific research data within a campus and out to regional/national network
backbones. In addition, through a series of solicitations over the last 4 years, the National Science
Foundation (NSF) has made tremendous networking investments in more than 100 U.S. campuses
to increase bandwidth capabilities and deploy the Science DMZ model [2].

The next logical step is to build on these investments and establish a science-driven, high-
capacity ‘Big Data Freeway System.’ The analogy to the freeway system is intentional. While the
US investment in the interstate freeway system did not add any new destinations that could not
already be reached by car, that investment with its vastly increased speed and lack of
impediments (e.g. stoplights) transformed automotive travel, commerce, personal mobility, and
national defense, and what once were transformative capabilities are now taken for granted. The
Pacific Research Platform (PRP) partnership, recently funded by the NSF [3], will extend the
campus Science DMZ model to an interoperable regional model for data-intensive networking,
enabling researchers to quickly move data without impediments to/from their labs and their
collaborators’ sites, supercomputer centers or data repositories. The PRP project will build a
researcher-defined network, with requirements driven by direct technical engagements with
sophisticated cyberinfrastructure-knowledgeable Science Teams. The PRP will facilitate big data
transfers as they traverse multiple, heterogeneous campus Science DMZs and networks at much
higher performance.
The PRP is a partnership of more than 20 institutions (see Figure 1 for the initial members), centered on the U.S. West Coast, but extending westward to Hawaii and Korea/Japan and eastward to Chicago and Amsterdam. Virtually all the west coast partners are members of CENIC, a regional optical network with a 100 Gbps backbone (shown in Figure 1), which is central to the PRP implementation. The PRP includes four NSF/DOE/NASA supercomputer centers, as well as three initial international partners.

Fully realizing high-performance, data-focused network interoperability and enabling broader access to data has a democratizing effect in advancing educational and research opportunities. Researchers and students at any PRP campus can gain access to massive research datasets and participate in sharing such datasets among institutions in distributed collaborations without being limited to local resources. Current PRP partners include a broad range of top research institutions, six of which are Hispanic-Serving Institutions (UCM, UCR, UCSB, UCSC, UIC, and SDSU), one is a Native Hawaiian Serving Institution (UHM) and two universities (U. Hawaii and Montana State U.) are in EPSCoR states. The PRP partnership includes the California Institute for Telecommunications and Information Technology (Calit2) and the San Diego Supercomputer Center (SDSC), which for over a decade have held summer workshops with the Minority-Serving Institution Cyberinfrastructure Empowerment Coalition. The PRP is building working regional-scale models to prove the feasibility of future national-scale, network-aware, data-focused approaches and capabilities. The lessons learned from the PRP deployments will improve the future technological, educational, and social learning curves for the nation’s institutions and data-centric research collaborations, ultimately giving benefits to organizations far beyond this partnership.

The PRP team has been building the foundation for this initiative for more than a year, with internal collaboration meetings, multiple proposals, and a demonstration of the initial (‘v0’) PRP prototype at the annual CENIC conference in March 2015 [4]. This workshop, while in the planning stages for many months, comes at an opportune time with NSF’s recent five-year award.
to the PRP team under the CC*DNI DIBBS (Data Networking and Innovation, Data Infrastructure Building Blocks) solicitation (program solicitation number 15-534) [3]. Participation in the workshop was by invitation only, focused on the PRP members and collaborators, and in some respects this workshop served as a kick-off for the NSF-funded PRP project. Participants represented federal agencies, all ten campuses of the University of California, the major California private research universities, Cal State campuses, three national supercomputer centers, national labs, several leading universities outside California, three international institutions, regional and national networking organizations (CENIC, ESnet, and Internet2), as well as industry. While the workshop participants were primarily PRP partners, the presentations and outcomes of this workshop have relevance well beyond the PRP project and institutions.

The workshop was structured into three primary sections across the 2.5 days of the event: Introduction, Science Engagement, and Technical Approach. A crucial element of this workshop was to engage representative domain scientists that will use the PRP to define the requirements to enable their science and to interact with technical implementers at the outset of the program to ensure success. The workshop objectives included:

- Engage the PRP team 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 five scientific domains identified as initial applications for the PRP:
  - Particle Physics  
  - Astronomy and Astrophysics  
  - Biomedical Sciences  
  - Earth Sciences  
  - Scalable Visualization, Virtual Reality, Ultra-Resolution Video  
- Identify common science-driven technical requirements for the PRP  
- Exchange technical ideas for the PRP’s technical implementation, including
  - Network architecture  
  - Data transfer nodes  
  - Software and tools  
  - Security  
  - Near-term and longer-term PRP capabilities

The next section of this report describes the presentations and panel discussions for the three major sections of the workshop. (All presentation materials and videos of the presentations are available online for reference; see Appendix.) Since establishing science-driven requirements for the PRP was a key objective of this workshop, a separate section is devoted to the discussion of those domain-driven requirements. We then summarize the conclusions and recommendations emerging from the workshop.

## 3 Workshop Presentations and Panel Discussions

### 3.1 Workshop Overview

The Pacific Research Platform is a science-driven regional partnership with a focus on networking capabilities to enable scientific collaborations across the partnership. While the PRP initiative is receiving $5M of funding over five years from an NSF grant, that funding is heavily
leveraged with voluntary efforts by administrators that recognize the importance of achieving the PRP’s capabilities and vision for their own institutions and researchers. Thus, there are many stakeholders within the partnership, from the scientists who will set the requirements for and use the PRP, to the campus and regional/national network technical experts who will design and deploy the PRP, to campus and organizational leaders who must support the socio-technical, policy, and funding issues necessary to create a collaborative virtual organization. This workshop was intended to bring together all of these stakeholders to establish requirements and set the course forward for the project.

As this workshop focused on the specific PRP program, attendance was by invitation only and included the original PRP member institutions as well as some external collaborators that had expressed an interest in the PRP. Approximately 137 people registered for the workshop with ~130 actual attendees – far exceeding the initial projections of ~50 attendees. A total of 40 organizations were represented, including virtually all of the initial PRP partners shown in Figure 1 and several organizations that already represent outreach and extensibility for the project. All ten University of California (UC) campuses were represented, as well as the UC Office of the President and UC Agricultural and Natural Resources. Eleven additional US universities were represented, including the three major private universities in California (Caltech, USC and Stanford), two campuses from the California State University system (Fullerton and San Diego), and six non-California universities (Clemson, University of Illinois at Chicago, Montana State, Northwestern, Oregon State, and University of Washington). There was strong representation from the regional and national networking organizations CENIC, ESnet, and Internet2, as well as three national labs: Lawrence Berkeley National Laboratory (LBNL), National Center for Atmospheric Research (NCAR), and NASA Jet Propulsion Laboratory (NASA/JPL). Four international organizations (from Amsterdam, Canada, Korea, and Japan) were represented, as well as five industries (Kelvin 1687, KLM, NTT, Pacific Interface, and SAIC), and at least one government agency (NSF). A complete list of registrants is provided in the Appendix to this report.

The agenda for each of the 2.5 days of the workshop followed the three primary sections of the workshop, Introduction, Science Engagement, and Technical Approach. The first half-day session was an Introduction, setting the vision for the PRP and engaging senior stakeholders crucial to the program’s success. The second day focused on engaging the domain scientists who will use the PRP for their collaborations. Sessions included groups from key initial areas targeted for the PRP, including high-energy physics, astronomy, biomedical sciences, earth sciences, and visualization. This extended exchange between scientists/users and the implementation team was a crucial element for achieving the goals of this workshop and for designing a PRP that will respond to researchers’ requirements. The third day focused on requirements from the domain scientists and the networking architecture, policies, tools, and security necessary to deploy the PRP.

High-level summaries of the presentations and panel discussions for the three days are provided in the subsections below. Complete presentation material from this workshop is available at the PRP website http://prp.ucsd.edu/. Presentation slides are posted on the website under the “Presentations” tab, and the reader is encouraged to review these directly in conjunction with this report. Videos of the speakers’ presentations are available directly on YouTube at https://www.youtube.com/playlist?list=PLbbCsk7MUlGcrHp3t_ia0At9c2T3Unn_R.
3.2 PRP Introduction Session

The first day’s Introduction session provided an overview of the Pacific Research Platform and the context for this initiative by some of its senior administrative stakeholders.

UCSD’s Vice Chancellor for Research Sandra Brown welcomed the workshop participants and discussed the importance of this initiative from a UC perspective. Within the ten-campus University of California system, there has been a history of collaborative efforts at the intersection between the Vice Chancellors for Research (VCRs) and the Chief Information Officers (CIOs), including a Summit meeting at UCLA this past spring that brought together researchers and cyberinfrastructure experts from across all campuses (report pending). The PRP was highlighted at that meeting, is a key topic amongst the UC Information Technology Leadership Council (ITLC), and has the strong support of both UC VCRs and CIOs. She stressed the importance of the PRP being science-driven, addressing the socio-political and administrative issues, especially as a virtual collaborative organization and being a vehicle for training and workforce development.

Larry Smarr, who conceived the PRP program and is the Principal Investigator of this workshop and the recent PRP NSF award, gave an overview of the PRP’s background, motivation, and progress to date. He noted the analogy of the “Big Data Freeway System” to the interstate freeway system, which did not add any destinations that could not already be reached by car, but by providing high-speed, unimpeded safe travel, completely transformed automotive travel and commerce. He highlighted key enablers for the PRP project, including NSF’s extensive investments in campus research networking under the CC-* programs [2], CENIC’s connectivity and capabilities, the Science DMZ concept promulgated by ESnet, and pathfinder programs for the team such as UCSD’s PRISM. PRP version 0 (V0) is already a reality, with a number of PRP partners working together to conduct a demonstration of high-performance cross-campus networking at the CENIC meeting in March 2015 [4]. In some respects, this workshop was a kick-off for the NSF-funded PRP project, which is expected to deliver a Version 1 (V1) PRP in two years that will support many of the initially-identified science teams and a Version 2 (V2) PRP, which will transition to full-IPv6 and extend the PRP to additional science use cases and partners, by the five-year end of the award.

Larry Smarr then served as moderator for a panel of senior stakeholders to discuss the PRP from their perspectives, including Greg Bell, (ESnet Director), Louis Fox (CENIC CEO), Tom Andriola (Vice President and Chief Information Officer for the UC System), and Amy Walton (NSF/CISE/ACI and PRP Program Officer). The panelists did not make presentations but their verbal comments are summarized below.

Greg Bell described ESnet’s vision for a “world of scientific discovery unconstrained by geography,” a vision that focuses on the goal rather than the means (i.e., networking) to achieve that goal. ESnet is enthusiastic about the PRP as a grassroots movement across campuses, supercomputer centers, labs, and other facilities and believes that the PRP will be a powerful “existence proof” of the Science DMZ concept of frictionless networking to enable scientific collaborations. He also acknowledged the substantial, successful collaborations in networking in recent years between the Department of Energy (DOE) and NSF, which have enabled major advancements in US research cyberinfrastructure.
Louis Fox described CENIC’s broad role in supporting all educational institutions in California, from K-12 to universities as well as public libraries. CENIC’s board, composed of member representatives, decided a couple of years ago to invest in an upgrade to 100 Gbps capability and to bring that capability to campuses with a focus on enabling research. The “reason for those investments has come along in the PRP,” and CENIC is committing substantial technical staff time and other resources to the PRP. In addition to its infrastructure within California, CENIC partners with other organizations, such as the Northwest GigaPOP, PacWave, Asian-Pacific networks, StarLight (Chicago), and a loose confederation of other western regionals. Many of these organizations and their member universities are PRP collaborators.

Tom Andriola works closely with CIOs across the UC campuses and the Office of the President. While relatively new to the community, he recognizes that “great organizations take innovation and scale it,” that UC is “doubling down on the network,” and the PRP is an example of innovation that can have a transformational impact on research across the UC system and the entire PRP partnership.

Amy Walton is working on a number of data activities within NSF. From NSF’s perspective, the PRP project is an important experiment, moving beyond investments in individual institutions to a regional level. The focus is “not about speed of individual pieces, it’s what’s the time to science, how can people do their job better?” To realize these capabilities, both the individual science areas/teams and technical cyberinfrastructure must be strong. It is hoped that the PRP experiment will benefit the scientific community far beyond California and the PRP partners. Also, NSF is discussing the general issue of project sustainability beyond the period of their grants, and the PRP should plan for its post-award sustainability. Larry Smarr noted that the PRP has many elements of sustainability built in as it is a highly leveraged NSF award with independent commitments from its members, including all UC campuses, strong partnership with CENIC’s well-established infrastructure, and robust collaboration with ESnet.
3.3 Science Engagement Session

The second day of the workshop focused on science engagement, with a discussion of the engagement process and a series of panels representing diverse scientific projects from five exemplary scientific domains that can potentially use the PRP for their collaborations.

Camille Crittenden, Pacific Research Platform co-PI and Deputy Director of the Center for Information Technology Research in the Interest of Society (CITRIS), a multi-campus institute headquartered at UC Berkeley, defined the objectives for this day’s discussions as follows:

- Meet each other in person: the other kind of ‘networking’
- Gain a common understanding of ‘Science Engagement’
- Explore data storage, computing and networking requirements across five representative domains

She challenged the audience to keep a long-term perspective in mind, looking ahead to what can be done five years from now that cannot be done now. The PRP will accelerate scientific discovery, document cultural/architectural heritage, inspire new forms of artistic expression, and open the door for unanticipated applications. This exchange between scientists/users and the implementation team is a crucial element for designing a PRP that will respond to researchers’ requirements and for establishing mutual understandings between users and cyberinfrastructure experts.

Prior to the workshop, general guidance was provided to the science panel participants in the form of a template to suggest common issues and questions that would help inform technical requirements for the PRP. This template is included below.

*The panels will discuss the needs of the intercampus research communities in the following terms:*

- Size of data & frequency of transfer, ad hoc, daily, or periodic?
- With whom is data exchanged (local, global, storage back ends?)
- External PRP connectivity to supercomputer centers, clouds?
- Tools used to access and move data?
- What speeds to date have been achieved?
- What’s screwed up?
- *In an ideal world, we’d get…*

*Title (name of science project)*

- **Project or collaboration structure, from a data perspective**
  - Describe the way in which the project or collaboration uses data, how the data sets are analyzed in terms of data acquisition, storage, analysis, publication, etc.
- **Geographic information**
  - Collaborating sites with data movement needs
  - Include site names or departments, if that information would be useful, and contacts/collaborators at the sites the goal is to establish onsite “coordinates” for the data transfers that will be traversing the PRP. If the resources used at the different locations are known, please describe them (e.g., SDSC computation (Comet), disk storage at SDSC, long-term storage on tape (HPSS) at NCSA).
National and commercial resources
• If the project makes use of national resources (e.g., national HPC facilities), commercial clouds, or other resources, please describe the data flow and usage. The goal is to help establish the set of major external resources that PRP projects use.

Data management
• Data scale, frequency of transfers, and similar information
  o Describe the amount of data that must be transferred between sites, the composition of the data sets transferred, and the frequency of data transfers.
  o Please also describe the performance requirements (both what you need and what you can get today). For example, are the transfers regular and periodic, or are they ad hoc? Are they done hourly, daily, weekly, or on some other time scale?
  o The goal is to help people outside the collaboration understand what you are trying to do, how often you need to do it, what performance you need, and how that differs from the performance you can get today.
• In terms of the data sets, how large are the files, how many of them are there, and what does the directory structure on the filesystem look like? Does the data come from a portal or other data repository, or is the data simply in a directory on a filesystem at both ends of the transfer?
  o The goal is to help systems and networking people understand the structure of your data sets so that the systems you use can be properly configured to support your project.

Data transfer tools
• Describe the data transfer tools you use, and how you use them. Examples might include Globus, SCP, rsync, LHC specific tools (PanDA, etc), http/wget, and so forth.
• Data transfer performance - Describe the performance you get from the data transfer tools you use (or include this information in the tools section above). In addition, indicate what performance is needed, whether the tools provide the features you need (e.g., automatic failure recovery, integrity verification, encryption during transfer, notifications/callbacks for use in a larger workflow, etc).
• Unmet needs and outstanding issues - Impediments and problems
• Please describe any network or data-related issues that adversely affect the productivity of the collaboration.

Wish list
• In an ideal world, what would be different?
• What capabilities would you have that you do not have today?
• If data movement were instantaneous, how would that affect/change your collaboration?

Evidenced by the presentation materials, some presenters in the domain sessions explicitly addressed elements of these guidelines in their presentations.
Note that this workshop highlighted five scientific domains that potentially provide specific near-term applications for the PRP. However the PRP will by no means be limited to these scientific domains or the specific projects discussed in this workshop. These five areas were selected to help capture technical requirements for the PRP and to identify some projects as pathfinders for the initial PRP development, which would then be extensible to other domains/projects.

### 3.3.1 What is Science Engagement? (Eli Dart)

Eli Dart from ESnet’s Science Engagement group described the process that ESnet has developed to facilitate the interactions between the scientists that drive the need for ESnet’s high-performance networking and those who develop and operate that network [5]. While there is exponential growth in data, “humans are not exponential,” and new tools and capabilities are required for scientists to keep pace. Scientists “want to do science, not IT/networking,” and historically it has been challenging for scientists to easily make effective use of wide-area networks. The cyberinfrastructure necessary to support science is a broad system of compute/storage/networking systems, software and workflows. While scientists can and often do invest their time/staff to become cyberinfrastructure experts as well as domain scientists, ideally domain scientists and cyberinfrastructure experts can bring their respective areas of expertise and interests to work collaboratively and enable scientists to focus on their science objectives, not IT.

The vision of ESnet, as earlier defined by Greg Bell, is that “scientific progress is completely unconstrained by the physical location of instruments, people, computational resources, or data.” ESnet has developed several approaches to science engagement: Partnerships, Training and Community Resources. For example, Partnerships include a series of ‘CrossConnects’ workshops, each of which focuses on a specific science domain and network-related requirements, and a set of in-depth collaborations between cyberinfrastructure/networking experts and competitively-selected projects [6]. To a large extent, these partnership approaches serve as models for the PRP, including activities at this workshop and the Science Teams approach. It is important for technical experts to focus on the desired outcomes, such as enabling new science and reducing the time-to-science, and keep the perspective that the cyberinfrastructure technology is a platform rather than an end to itself. This can change the nature of the conversation between scientists and network experts, to the benefit of both groups.

### 3.3.2 Particle physics (moderator Frank Würthwein)

The high-energy physics community, particularly those that analyze data from the CMS and ATLAS experiments at CERN’s Large Hadron Collider (LHC), provides an excellent example of broadly-distributed researchers collaborating on large-scale datasets. Just within the current PRP partnership, nine universities have large-scale LHC research efforts, including local compute/storage resources, and five facilities host large-scale LHC compute/data resources.

This panel included high-energy physicists from five UC campuses and Caltech. Frank Würthwein from the UCSD Physics Department and SDSC (and Executive Director of the Open Science Grid) opened this session with an overview of the LHC data/experiments. The PRP can make LHC’s public data readily accessible from a researcher’s lab and will focus on the “last mile” problem: from private data in a researcher’s lab to publication. Würthwein has already initiated a collaboration within some PRP partner universities, SDSC, and the Open Science Grid to develop tools to transparently access data and compute on data across local or regional facilities. Servers that serve the role of Data Transfer Nodes (DTNs) in the ScienceDMZ parlance have already been distributed to a number of PRP sites and can be remotely administered. Tools
such as XRootD will allow federated seamless access to data sources, and Open Science Grid (OSG) tools (e.g., HTCondor) allow federated access to compute resources. There are three different caches on an HTCondor DTN for LHC experiments: an application/library cache, a data cache (data from elsewhere, for local compute), a data publication or “origin server” (data produced locally and served outward).

Jason Nielsen of UC Santa Cruz (UCSC) works on the ATLAS experiment, with the primary analysis done using the Athena Framework. A diagram of the data scales at various levels of analysis informs where the PRP is most applicable to its members – working on Tier 3 clusters with datasets on the order of tens of terabytes and analysis timescales of a day (e.g., think during the day, analyze overnight). The XRootD remote I/O described by Würthwein would be a major win for Tier 3 centers. He also noted that future LHC datasets in the 2025 timeframe and beyond will be ~100X the size of current datasets, with even higher computing loads.

Owen Long of UC Riverside (UCR) works on the CMS experiment and has modest scale cluster/storage at UCR. Analyses using current compute/networking resources could be vastly improved with PRP network capabilities, combined with tools such as XRootD for federated access to required data and federated compute resources, leading to higher productivity and a faster pace for innovation.

Chris West of UC Santa Barbara (UCSB) is also a CMS researcher with modest-scale local compute/storage/networking resources. Poor network performance incurs significant costs, including human-intensive workflows to deal with failures, doing more data analysis “just in case” so you don’t have to go back and do something else later, and there is a secondary data reduction step in case data is too large and there are failures. His wish list, perhaps exemplary for smaller-scale sites such as UCSB, includes minimal maintenance, improved data transfers from distant nodes, ability to use remote storage and compute resources (e.g., at UCSD) transparently, LHCOne connectivity, and access to other experiments outside LHC.

Maria Spiropolu of Caltech broadened the discussion to include data-intensive challenges associated with instruments such as LSST or DESI in astronomy, or the explosion in genomics data. She also highlighted that scientists do their work in an ecosystem that includes compute, storage, networking, and software, and a project such as the PRP must address the ecosystem, not just the network, in order to satisfy researchers’ requirements.

Frank Würthwein concluded the presentations in this session with a response to the template questions related to high-energy physics (see Figure 3).

A general question during the panel Q&A was whether this planned capability for high-energy physics could be extended to other domains/projects. Würthwein replied that the key concepts of “submit locally, compute globally” and having data available as a cache as opposed to put/get type of access can certainly apply to other projects, even if the specific software artifacts are adapted to different projects and evolve over time. Spiropolu identified several applications that could benefit from these concepts, such as fMRI, astronomical sky survey analyses, and the BRAIN initiative. Eli Dart noted that the science engagement team should understand what those structures look like, build a functional toolset that is usable, and enable common adoption of tools across science collaborations.
3.3.3 Astronomy and astrophysics (moderator Mike Norman)

Mike Norman, Director of SDSC and member of UCSD Physics Department, opened this session with a summary of key issues at the interface between astronomy research and the PRP. These included data volume and rates over time, data access and archiving over time, the processing workflow (e.g., what computers and where), the size of data caches near processing, the data products distribution, and use of the network to enable collaboration. He also noted that the Comet supercomputer at SDSC will be fully integrated as a resource within OSG and will have XRootD installed for data access.

Brian Keating (UCSD Physics Department) discussed the POLARBEAR experiment to measure the cosmic microwave background B-modes at degree and sub-degree scales. The current instrument is located in the high-altitude Atacama Desert in Chile. Data transfer rates are relatively modest, but because of network limitations, data is currently transferred from Chile by shipping hard drives monthly. There are plans to improve the detector arrays, as well as expand to the three-telescope Simons Array; these upgrades will increase data acquisition rates up to ~1TB/month. Having direct network access for data transfers would improve time-to-science. (There are discussions about utilizing fiber from the nearby ALMA telescope, but this is not currently available.)

Peter Nugent (LBNL and UC Berkeley) discussed data, computing, and workflow requirements arising from the Palomar Transient Factory (PTF) and the follow-on Zwicky Transient Factory (ZTF) scheduled for mid-2016. The PTF conducts large-scale sky surveys nightly and compares new images to reference images to detect transient events, which can then be followed-up for observations by a host of other astronomical observatories. While supernovae are perhaps the most broadly known transient events for their cosmological implications, an extensive set of projects track additional transient events, from solar system events to variable stars to distant galaxies. Time is of the essence for data transfer, analysis and reporting as events trigger follow-up observations at other observatories. About 100 GB of raw data is currently transferred from
Palomar Mountain to SDSC (via HPWREN microwave relay), then to the National Energy Research Scientific Computing Center (NERSC) via ESnet. Analysis at NERSC results in ~500 GB/night of data, which is then scanned for transient events. Those events are currently published in less than an hour from the time of image acquisition at Palomar for notification and follow-up by other observatories (see Figure 4). Three drivers will increase the data/processing requirements. First, the ZTF upgrade in 2016 will increase data rates and processing load by at least an order of magnitude. Second, there are increasing requests to re-process PTF/ZTF data after an event is detected, above and beyond the normal single-pass workflow. This computing load cannot currently be accommodated by NERSC, and it would be preferable to push data out to requestors and let them re-process data on distributed local resources. Third, the [www.legacysurvey.org](http://www.legacysurvey.org) project would like to make historical survey data accessible for professional and public use, significantly increasing the potential scientific impact of the data, as well as the data transfer requirements.

Sharon Brunett of Caltech presented a talk about data/compute requirements for the Laser Interferometer Gravitational-Wave Observatory (LIGO). There are three models of computing on the data, all of which can be decomposed into “pleasantly parallel” tasks. Low-latency computing is required for detector characterization and astrophysical events, such as compact binary coalescence or gamma ray bursts. These are executed on dedicated LIGO resources. Longer-latency processing can be accommodated on national resources (e.g. XSEDE) or, for calculations with low data/processing ratios, by Einstein@Home public community computing. Data rates are generally reasonable, with ~1 PB/yr (after compression) of raw data and ~2 PB/yr generated by various search groups. She noted that stable, reliable networks/computing would significantly increase the productivity of researchers, as significant time is now spent monitoring results, re-sending data, and re-submitting compute jobs.

### 3.3.4 Biomedical Sciences (moderator Larry Smarr)

Rob Knight (UC San Diego Pediatrics and Computer Science Departments) presented a talk about microbiome research and illustrated its big data requirements. The NIH Human Microbiome Project (HMP) is a $173M initiative to capture DNA sequences of the vast number of human microbes and research their impact on human health. With the dramatic decreases in the cost of sequencing, data acquisition is being democratized, and additional microbiome sequence data is being captured from a wide variety of sources. Knight is building and leveraging substantial compute/storage/networking resources across UCSD to support microbiome research and is establishing extensive inter-disciplinary collaborations across the campus. Easy, high-speed access is crucial to enabling those collaborations, not only within UCSD but also with collaborators worldwide.

Sergio Baranzini (UC San Francisco Neurology Department) described the challenges and current cyberinfrastructure used by his group in the study of multiple sclerosis pathogenesis and, more
generally, in biological sciences. Their datasets are already on the order of 10–100 TB. The lab and local compute resources already have excellent connectivity, and connectivity within UCSF end-points is generally good. However, firewalls and other impediments significantly reduce bandwidth outside UCSF, e.g., to Amazon EC2 resources, and it is hoped that PRP can address these external networking performance limitations.

JJ Garcia-Luna-Aceves of UCSC was scheduled to present a talk about the Cancer Genome Hub (CGHub). However, he was unable to attend, and his charts were presented by Larry Smarr (see Figure 5). CGHub has been an incredibly successful data repository in heavy demand by researchers across the country (and many in the PRP partnership). More than 30 PB of data have cumulatively been downloaded from CGHub and sustained transfer rates have steadily grown and recently reached 17 Gbps. But CGHub faces significant challenges going forward, with increasing data/transmission requirements driven by personalized medicine, and limitations of the internet protocol stack that was not designed for large data transfers over paths with large bandwidth-delay products, including TCP throughput, DDoS vulnerabilities, caching versus privacy (e.g., https), and static directory services (DNS versus content directories). The CGHub group and other researchers at UCSC are evaluating Content Centric Network (CCN) architectures to address these limitations and recommend the PRP evaluate the use of CCN architectures for future designs.

During the panel Q&A, participants discussed privacy restrictions frequently associated with biomedical data and the requirements this could place on the PRP. In general, de-identification techniques address the privacy issue, although the effectiveness of these techniques is a subject of discussion in the community. Larry Smarr noted efforts across the UC system and other campuses to establish secure procedures for sharing private data, and the PRP is not funded by the NSF to solve this challenging problem.

### 3.3.5 Earth Sciences (moderator Camille Crittenden)

The session opened with a talk by Frank McKenna of the Pacific Earthquake Engineering Research (PEER) Center at UCB. PEER is a multi-institutional research and education center focused on performance-based earthquake engineering and includes many universities in the PRP partnership, as well as government and industry partners. Engineering studies are based on experiments, simulations, and hybrid simulations that incorporate experimental data. PEER uses a number of national, local, and commercial (Amazon) compute/storage facilities, much of it managed by the Pegasus workflow software. Experimental data volumes are typically up to a few TB, and simulation data volumes are 1–10 TB; data transfers are currently performed with a
variety of tools including Globus, GridFTP, Condor I/O, iRods, scp, symlink, hkp, local file systems, and, at times, shipping disks by FedEx. A driver for increased data volumes in coming years is the fact that earthquake engineering is moving from single-building analyses to "resilience," which is the ability of a community to recover from a disaster. A PEER wish list related to PRP support included (a) enable remote files/applications to look local on the desktop, (b) the ability to do more simulations and to learn from other simulations via knowledge-based artificial intelligence, and (c) the ability to create and move more data, share simulation data more easily, and avoid using FedEx to ship data.

Eric Nienhouse of the National Center for Atmospheric Research (NCAR) talked about the Coupled Model Intercomparison Project (CMIP) widely used for analysis of global climate models. The current CMIP5 repository includes ~2 PB total data, kept in situ at ~30 contributing modeling centers, with typical research projects requiring transfers of 10-100 TB; data are accessed via the Earth System Grid Federation (ESGF) software stack. The next-generation CMIP6 repository is expected to contain ~16 times this volume, or ~30 PB of data. The CMIP Analysis Platform is a shared resource hosted at NCAR with dedicated storage, analysis clusters and workflows for CMIP research. The end objective is to allow scientists to do science (and not data preparation and management) by offloading IT from each individual science project to dedicated data curators and infrastructure. Any researcher with access to NCAR’s Yellowstone system will be able to use the analysis platform and analyze these datasets. In addition to CMIP6-level data volumes, the platform will require data transfers from 30 different data nodes; current transfer rates from those nodes range from <1 Mbps up to 1 Gbps. NCAR’s wish list included (a) an intelligent network with data reduction capabilities, (b) an “invisible” security model, (c) rich data use metrics for application and workflow development, and (d) data compression that retains scientific content.

Dan Cayan of UCSD/Scripps Institute of Oceanography is a climate modeling researcher, with an emphasis on California and the western US. Following the general template, the data transfer volumes for his research are ~from 100s of GB to 100s of TB, amongst several scientific collaborators and national facilities (e.g., LLNL, NERSC and Oak Ridge National Laboratory), using tools such as secure copy rsync and httpget/wget. In an ideal world, there would be (a) better catalogue describing available data, (b) better access tools for range of users (small to large), (c) remote processing and analysis tools, and (d) more accessible expert network knowledge. He noted that many problems are “not technical issues per se, but are more limited by the quality of the info available and the community’s ability to understand it.”

Christopher Paolini of San Diego State University (SDSU) works on simulations of geological substrate CO2 sequestration. SDSU participated in the PRP V0 demonstration and has already been collaborating with the OSG to enable distributed computing for the simulations. The simulations generate 1–10 TB of data that must be transferred from the OSG compute site using Globus to SDSU’s DTN for post-processing and visualization. Transfers of partial datasets range from 1 GB to 5 TB daily, while the final datasets are ~1–10 TB on a biweekly basis. Current application data transfers have only achieved 70 Mbps, showing substantial “room for improvement.” Many OSG compute sites and national centers are within the PRP, and improved performance will directly improve these data transfers.

### 3.3.6 Scalable Visualization, Virtual Reality, and Ultra-resolution Video (moderator Tom DeFanti)

Maxine Brown of the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago opened the session and described the high-speed (100 Gbps) between StarLight and
EVL’s visualization facilities and their development of the Scalable Amplified Group Environment (SAGE2) software to access, display, and share high-resolution digital media on scalable resolution display environments (sage2.sagecommons.org). Five use cases were defined for applications that would use the PRP, including (a) collaborators using SAGE2 to transfer movies, pictures, or video streams, (b) using OmegaLib virtual reality middleware, which is being extended as a platform-independent framework for scientific visualization to stream real-time interactive environments among large-scale displays, such as CAVE2, WAVE, etc., (c) large data transfers between facilities, (d) using SAGE2 and UltraGrid for low-latency, high-resolution video transmissions, and (e) sharing data captured with the planned Sensor Environment Imaging camera (SENSEI) [7], which can generate more than a terabyte of data per minute. General challenges faced include (a) storage devices not often accessible on the network edge (hidden behind firewalls, multiple gateways, or login systems), (b) storage not on the right network, (c) network misconfigurations; e.g. no jumbo-frame network end-to-end, (d) UDP enables low latency but can be problematic for networks, and (e) firewalls not configured for UDP. In an ideal world, the EVL group wants intelligent, scientist-friendly tools to access Big Data files from home, office, lab (the goals of SAGE2 and OmegaLib), and fast protocols and transfer tools.

Tom Levy of UCSD’s Archaeology Department and Calit2/Qualcomm Institute has conducted all-digital archaeological digs in Jordan since 1999. These digital archaeological digs have four primary functions: Acquisition, Dissemination, Curation and Analysis, of which the PRP is particularly valuable for dissemination and curation. Typical datasets from an archaeological dig are on the order of a terabyte, currently managed via tools such as Autodesk Memento. A substantial amount of the data is GIS data. In practice, they continue to use hard drives for data transfers with external collaborators and often sneaker-net within UCSD. The PRP includes many campuses with strong archaeology programs, and it would be valuable to move cultural heritage data around these sites, particularly if visualization facilities like CAVEs are available at those sites.

Jurgen Schulze (UCSD and Calit2/Qualcomm Institute) provided a summary of the bandwidth requirements for photorealistic virtual environments. The equivalent bandwidth of the human visual system is \(\sim 15 \text{ Tbps} \) (over \(4\pi\)), an interesting limit for visualization systems. Current and planned sensors for capturing video data include WAVE (100 Gbps), GoPro Odyssey (50 Gbps), CAVECam (400 Gbps), SENSEI (1,000 Gbps), and the Oculus Rift 2016 (6 Gbps, or \(>60 \text{ Gbps} \) for sphere). The conclusion is that live streaming for current sensors requires \(\sim 100 \text{ Gbps} \) end-to-end network bandwidth, and this will increase for near-term sensors (but the theoretical limit required to match human visual acuity is \(\sim 15 \text{ Tbps}\)).

Falko Kuester (UCSD Structural Engineering Department and Calit2/Qualcomm Institute) illustrated various exemplary applications of visual analytics designed to “enable analytical reasoning through interactive, collaborative visualization.” He noted that it would be great if “the network was easy, and it’s just about the data.” He provided quantitative requirements for three potential PRP applications (see Table 1): the “100 Island Challenge” led by SIO to study coral reefs, underwater real-time microscopy also led by SIO, and “Drones for Good” to provide remote imaging in emergency situations. These exemplary applications illustrate the breadth of use cases, for example from real-time streaming video to large, asynchronous data transfers, or with end-points within UCSD to external field sites, supercomputer centers and commercial clouds.
### Table 1: Response to 'template' questions for three science use cases (Falko Kuester)

<table>
<thead>
<tr>
<th>Application</th>
<th>100 Island Challenge</th>
<th>Underwater Microscopy</th>
<th>Drones for Good</th>
</tr>
</thead>
</table>
| **Size of Data and Frequency of Transfer** | • Data Intake: 2-5 Tb per year  
• Data Transfer Frequency: Daily  
• Data Transfer Size: 0.1-0.5 Tb per transfer  
• Projected Needs: 10x increase in data assets and bandwidth | • Data Assets: 20,000,000 images  
• Data Intake: 100,000 images per day  
• Data Transfer Rate: 3.2 Gb/s  
• Data Transfer Frequency: Continuous | • Data Assets: > 1,000,000 images + derivative data (10x)  
• Data Intake: 500 – 5,000 images per flight @ 30 MB per image  
• Data Transfer Rate: Site specific for fieldwork; 10Gb/s post-processing  
• Data Transfer Frequency: Daily |
| **Data Exchange Sites** | • Existing: UCSD labs (SIO, JSoE, Birch Aquarium)  
• Projected: Field sites – Living Laboratories | • Existing: UCSD labs via SIO Pier fiber optic network; multi-site collaboration over commodity networks.  
• Projected: Global | • Existing: UCSD labs  
• Projected: Global |
| **External Connectivity needed to Supercomputer Centers** | • Existing: Batch processing at different supercomputing sites  
• Future: Gordon and Comet at SDSC | • Existing: N/A  
• Future: Comet | • Existing: in-house (WAVE, VROOM)  
• Future: Gordon / Comet |
| **Tools to Access and Move Data** | • Existing: “Hard-Drive Centric” & “FedEx”  
• Projected: Cloud Services | • Existing: "Cloud Centric" (Dropbox, GoogleDrive)  
• Projected: Cloud Centric | • Existing: FIONAs (Flash I/O Network Appliances) via local subnet  
• Projected: FIONAs via the PRP |
| **Data Transfer Speeds Achieved to Date** | • Local: USB 3.0 transfer rates (<10Gb/sec)  
• Remote: 1Gb/sec | • Local: 3.2 Gb/sec (instrument side)  
• Remote: 1Gb/sec | • Local: ~40Gb/sec  
• Remote: ~10Gb/sec |
| **Where are the Problems** | • Storage  
• Bandwidth  
• Processing cycles  
• Workflow(s) | • Too early to tell  
• Workflow(s)  
• Parallelism of algorithms (or lack thereof) | |
| **In an Ideal World...** | • Cyber-infrastructure (CI) supporting seamless workflows  
• CI providing access to the academic community  
• CI enabling/empowering education and outreach | • Interdisciplinary Collaboration  
• CI enabling/empowering, fully immersive instrument access to the public | • CI supporting interdisciplinary collaboration  
• CI supporting seamless workflows  
• CI providing access to the academic community  
• CI enabling/empowering education and outreach  
• Shared-use infrastructure |

Walt Scacchi (UC Irvine) researches applications of game-based virtual worlds for scientific research and education. In contrast to many scientific applications that have large data sets and infrequent asynchronous transfers between specific end-points, these applications generally have modest datasets (e.g., 50 MB–2 GB) but frequent data transactions with hundreds of distributed users. Workarounds are needed to limit user-data interaction streams due to effective latency limitations, and new data distribution mechanisms (e.g., ‘install’) are required as opposed to traditional put/get, copy, or rsync mechanisms.

Cees de Laat (University of Amsterdam) has been deeply engaged in fundamental networking research and worldwide high-performance networking. He emphasized that scientists/users do not want to know the underlying details of networking and data transfers; they just want it to work with minimal effort on their part – a “magic data carpet” that provides transport, storage, security, integrity, and curation. In cinematic applications, data volumes range from < 1 GB for film dailies to ~10 TB for feature films, and up to 150 TB for raw film acquisition prior to editing. Currently, the bandwidth for many small file transmissions over long distances is limited by the latency and error-checking; new techniques are required to better fill pipes with multiple small files. He also described the Service Provider Group (SPG) concept, an organizational structure providing a defined service only available if its members collaborate, which has relevance to the PRP (networking examples include Internet2 or eduroam). Members (e.g., campuses) are autonomous but together provide a service none could provide on its own. To the customer, the organization appears as a single provider, with common interfaces and policies wherever possible.
3.3.7 General Discussion (moderators Greg Bell and Camille Crittenden)

Both the morning and afternoon sessions concluded with opportunities for general discussion about science engagement.

Greg Bell noted the similarities of these discussions to ESnet’s CrossConnects workshops, with the emphasis on science even though an intended outcome is understanding networking requirements to support the science. An important skill for the IT/networking community to develop is to extract IT requirements at the science interface. He advised that the PRP should look for low-hanging fruit where something can be done quickly with significant return for scientists, projects where the PRP could have a transformative impact on science, and collaborations with eager partners who will not only work closely with the IT engineers but also market the success amongst their colleagues to encourage adoption.

There was some discussion about how best to access data across a network – e.g., traditional put/get transfers, remotely mounting file systems, and smart caching. Frank Würthwein characterized three general use patterns as uploading, downloading and remote access – each of which would have distinct requirements. Eli Dart noted another use case which is high-performance access to central data repositories (e.g., astronomical surveys, climate data, CGhub) – particularly to increase use of archived data. Inder Monga suggested these could be considered good primitives to be applied to various projects and their workflows. Cees de Laat mentioned he has analyzed workflows in this context, particularly looking at computations/byte for various types of data.

Alternative “primitives” or dimensions also characterize science-driven use cases. For example, the Palomar Transient Factory uses pipeline processing with an extensive compute/storage/networking workflow with time constraints. Most of the applications discussed involve asynchronous data transfer, but some applications require real-time streaming of video/audio data. Additional concerns include the distinct requirements of publicly accessible data (e.g., open repositories) vs. data with restricted user access, and issues of regulatory privacy and security requirements for protected data such as clinical health records and patient-identifiable information.

Peter Nugent commented on the benefits of encapsulating analyses pipelines into “containers” that could be run anywhere, from laptops to national supercomputer centers, particularly if data could be seamlessly accessible to the container, wherever it runs, as opposed to needing to specify where data (and compute) are located.

Participants discussed how to approach problems as they are encountered while using the PRP for scientific applications. First, participants endorsed the general approach of using a few specific science applications to test and debug the PRP, rather than design/build a master solution for all institutions and applications. For example, for distributing the microbiome data collected by Rob Knight across other PRP collaborators, researchers should just pick a few sites and work through the problems. With that approach, a problem registry for the project could capture problems as they are encountered (slow data transfers, security issues, incompatible software) and their eventual solutions, and develop a usable knowledge base that can then be leveraged as similar issues are identified at other institutions/applications both within and outside of the PRP partnership.
Larry Smarr made a general request for the science applications to diagram their workflows, with explicit end-points, common primitives (e.g., upload/download/cache), and compute/storage facilities; such diagrams could directly inform the requirements for the PRP design.

3.4 Technical Approach Session

The third day of the workshop focused on the technical approaches that have been and will be implemented to meet the science-driven use cases and their requirements highlighted the previous day, i.e. a switch from the scientists’ to the IT/network experts’ perspectives.

As discussed by Larry Smarr on the first day of the workshop, the PRP is envisioned to have three phases. Version 0 (V0) has already been demonstrated at the CENIC annual meeting in March 2015, and the networking team established in that effort – with participation by CENIC, ESnet and 14 of the PRP research institutions – has continued to work together to establish connectivity, deploy data transfer nodes (DTNs), and work towards interoperability across campus Science DMZs prior to the NSF PRP award. The recently-awarded NSF grant calls for a Version 1 (V1) PRP in two years that will support many of the initially-identified science teams, and a Version 2 (V2) PRP by the five-year end of the award, which transitions to IPv6 and extends the PRP to additional science use cases and partners.

3.4.1 The PRPv1 Architecture Model (moderator John Hess)

John Hess is the lead network engineer for CENIC’s support of the PRP and has been the project manager for the collaborative V0 effort to date. He led off this session with a presentation primarily about the V0 efforts, which has been a grass-roots, collaborative effort to benchmark high-performance interconnections across partner Science DMZs traversing the CENIC network, using (primarily) FIONA boxes as data transfer nodes (see Section 3.4.2), GridFTP, and FDT as the file transfer software, and a mesh of perfSONAR toolkits deployed across the sites. Some site pairs have already established excellent connectivity, while other pairs are still being tested and debugged (see Figure 6).

Phil Papadopoulos (UCSD/SDSC and Calit2/Qualcomm Institute) described the Flash I/O Network Appliances (FIONAs) that are being deployed at many PRP sites to serve as Data Transfer Nodes (DTNs) in the Science DMZ model. (There are various types of DTNs and they are not limited to FIONA boxes, even within the PRP.) FIONA boxes are built of commodity parts, and a basic box (~$7K) can drive/accept data to/from a network at ~40 Gbps.

The basic box can be configured to include additional flash or disk storage.

- DTNs loaded with Globus Connect Server suite to obtain GridFTP tools.
- cron-scheduled transfers using globus-url-copy.
- ESNet-contributed script parses GridFTP transfer log and loads results in an esmond measurement archive.
- FDT – developed by Caltech in collaboration with Polytehnica Bucharest

![PRPv0: Transfer Results from March 2015](image)

**Figure 5:** Summary of PRP V0 Data Transfer Results (John Hess)
storage, or to add GPUs. Debugging the initial systems and getting expected performance was a challenge. However they are now being readily built/distributed with a suite of installed data transfer tools (e.g. FDT, GridFTP, Globus, UDT-based, XRootD), data access tools (e.g. NFS, Samba, SCP), and (optionally) managed by the Rocks toolkit.

Tom Hutton (UCSD/SDSC) described UCSD’s Science DMZ and external research connectivity (up to 100 Gbps) that has been developed under two CC-* awards PRISM and CHERuB. He laid out the network architecture for PRP V1 – a single Layer 2 VLAN with limited BGP peering – and discussed three options still under consideration for V2, and their pros/cons.

Eli Dart (ESnet) posited that a routed PRP V1, with campus Science DMZs connected to a 100 Gbps backbone, can make the networking part relatively simple, allowing the project to focus on the user interfaces and science outcomes. According to Dart, the PRP WAN needs to “stay big, fast, and clean.” He recommended that the PRP identify early adopters that can achieve some quick science successes, which can then be leveraged to recruit additional scientists/applications to adopt the network capabilities. It is important that IT/networking people and science groups on each campus communicate effectively to understand scientific requirements, respond to those requirements, and mutually understand expectations and capabilities. Finally, he highlighted the value of incorporating DTNs in compute cluster designs, mounting the parallel file system directly, and connecting directly to the Science DMZ.

Eric McCroskey (UCB) represented the perspective of the CENIC High-Performance Research (HPR) network Technical Advisory Committee (TAC). PRP V0 was a peering mesh over Pacific Wave. The CalREN HPR network is now being upgraded to 100 Gbps, and it is expected to meet PRP V1 requirements. The backbone needs to be “big, fast, and clean” moving traffic and “getting out of the way.” Most of the networking effort for the PRP will be “at the edges” at the interfaces to campuses and the science groups. PRP future requirements, as well as other requirements such as data streaming, are being incorporated into the next-generation CalREN network design. (Tom Hutton noted that meetings such as this workshop are valuable to the network engineers to identify those requirements.)

Larry Smarr commented that the work by the TAC and the HPR investments by CENIC are excellent examples of the substantial leverage of external efforts that benefit the PRP. There was also discussion about the decisions to adopt Layer 2 or 3 for PRP. With the right routers, they are no longer an impediment to high-performance, and a Layer 2 system is easier to debug and maintain than Layer 3. In addition, a Layer 2 solution is better-suited to a small environment, and a more scalable solution is required to meet future PRP needs. Harvey Newman of Caltech noted that there has not yet been discussion of dynamic circuits or software-defined networking and cautioned that the pursuit of the highest performance should not “fuzz out the vision,” which is to ensure that a number of science groups can reliably utilize reasonable bandwidth (e.g. ~10 Gbps).

### 3.4.2 Securing the PRPv1 (moderator Eli Dart)

Bruce Vincent (Stanford’s Senior Technology Strategist & Chief IT Architect) opened the session with some verbal comments about the PRP in the broader context of a campus IT environment. Incorporating common identity and security tools such as InCommon makes it easier for campus users and, thus, lowers the barriers to adoption. He has had discussions with the Globus group, which plans to develop automatic account generation using InCommon credentials. There also is a group identity service contemplated but not yet funded for the Globus group.
There was further group discussion of Globus’ role within the PRP, particularly with its central role as a frequently-used data transfer tool. Pol Llovet (Montana State) endorsed Globus as a common data transfer tool across all PRP partners. Eli Dart endorsed the responsiveness of the Globus group in their collaborations with ESnet. Larry Smarr noted that Ian Foster had contacted him and offered to help with the PRP collaboration.

Michael Duff (Stanford Chief Information Security Officer) opened his talk stating that the primary reason he was attending this workshop was to provide a strong endorsement of the PRP from the perspective of a campus Chief Information Security Officer. He believes that it is important to reduce network impediments for research data flows and that this can be done without compromising security. He discussed Stanford’s general security approach (see Figure 7), recommended that PRP equipment (including DTNs) be kept as simple as possible in terms of installed software, and suggested that a few key security practices (e.g. software updates, two-factor authentication, encryption in transit, and anomaly detection) will have a high return-on-investment.

Leon Gommans (Air France KLM) discussed the Service Provider Group concept as a possible way to help organize Service Federations that are driven by Cooperating Research Groups (CRGs, or Science Teams in other parlance) within the context of Science DMZs. From a networking perspective, the PRP is a federation of autonomous campus-based Science DMZs (and the network backbone); from a scientist/user perspective, the PRP provides secure access amongst the collaborating researchers and meets specific requirements for that group’s workflows, security, etc. There needs to be common policies and agreements amongst the Service Providers (campuses and their Science DMZs) that balance local autonomy and control with enabling interoperability and scalability for both the Service Providers and the Science Teams. The Service Provider Group is an organization structure providing a defined service only available if its members collaborate, and its framework is a mechanism to organize inter-organizational trust. It was recommended that this model be considered for the PRP, and there are organizational research questions that this implementation could help address for similar federations.

Rick Wagner (UCSD/SDSC) described a “trusted file system” project currently being developed with funding by UCSD’s Integrated Digital Infrastructure program to better integrate compute and storage resources at SDSC with instruments and labs across campus. The project has three components: SDSC’s Data Oasis (a 7 PB high-performance scalable parallel file system),
UCSD’s campus backbone and PRISM network as a Science DMZ, and a team staffing approach with system administrators from both SDSC and the participating labs. Data Oasis is built on the “magical ZFS” that can easily partition common metadata and storage pools across effectively distinct file systems. In addition, Indiana University has recently developed a Lustre “nodemap” capability that maps user IDs across administrative domains; this opens the door to workflows using geographically distributed resources and easy data sharing. This capability is being deployed with SDSC’s Comet and Data Oasis, where labs across campus will have their own effective storage on SDSC’s Data Oasis, seamlessly accessible (mounted) from their lab systems or from other SDSC resources (e.g., Comet). In principle, this model can be extended across the PRP for data sharing and distributed access to high-performance, large-scale data repositories.

3.4.3 PRPv2 Architecture and Security (moderator John Haskins)

Phil Papadopoulos (UCSD/SDSC and Calit2/Qualcomm Institute) discussed some high-level issues with PRP architecture and data sharing. As the PRP is knitting together autonomous campus Science DMZs, scalability and labor demands are important. Solutions need to be extensible to new DMZs/Science Teams, and labor-intensive solutions should be avoided unless generally applicable across DMZs/Science Teams. It is likely there will be multiple data sharing modes for various Science Teams: Share only within the group, share with anyone in PRP, share with anyone on Internet2, or share to the world. Peering VLANs are generally not a scalable architectural solution. The PRP team must establish mechanisms for managing PRP access (e.g., BGP, SciPass, leverage IPV6 or SDN features). Finally, he would recommend that PRP V2 transition to an IPV6-only system, rather than try to accommodate both IPV4 and IPV6 (see further discussion below).

Darrell Newcomb (CENIC) then presented two talks, first about Pacific Wave activities and second about SDN in the context of PRP V2. Pacific Wave is a collaboration between CENIC and the Pacific Northwest GigaPOP and is a distributed exchange with interconnections between participating networks. Twenty-two major research and education networks connect into the Pacific Wave fabric. NSF recently made an award to Pacific Wave under the International Research Network Connections (IRNC) program for (a) continued enhancement, upgrade and evolution of Pacific Wave to support more 100G connections, (b) additional 100G capacity between exchanges points along West Coast, (c) SDN/SDX deployment on parallel infrastructure to enable experimentation while maintaining production use of the Pacific Wave exchange, and (d) collaboration with other IRNC awardees on SDX development, measurement, and monitoring. In the second talk, Newcomb recommends that the approaches available with SDN be considered in designing V2 of the PRP, particularly in the context of enhancing end-to-end performance for specific science applications. He cited relevant work by Joe Mambretti at StarLight and Vandervecken SDN instance with Australia’s Academic and Research Network (AARNet).

John Graham (UCSD and Calit2/Qualcomm Institute) presented a talk about Trusted Platform Modules (TPMs) and SDN in the context of the PRP. The team currently uses Rocks for software builds for FIONAs and other PRP equipment, both for convenience in distributions but also as an initial step in deploying “trusted systems.” TPMs are inexpensive hardware devices that can be installed on system motherboards that provide a set of tools for trusted systems, such as high-quality cryptography, hashing engines, random number generation, key generation, and power detection. OpenStack is incorporating TPM capabilities into their environment. Another important development is that many systems such as XSEDE or Globus are providing CILogin authentication services that allow users to use their InCommon credentials.
Andrea Zonca (UCSD/SDSC) talked about developments to enable broad access to Jupyter, backended by SDSC’s Comet supercomputer. Jupyter is an application that can combine many different types of data (e.g. text, equations, software code, plots, and interactive visualizations) into a single document. This is a powerful new collaboration tool in that an entire analysis process can be easily shared with collaborators or for publication. While it normally functions as a single-user on a laptop, it can be connected to a server, enabling multiple users on the same server. In order to host more Jupyter users on more powerful resources, there is a project to create a backend for Jupyter that can run on Comet, specifically taking advantage of Comet’s 24-core/128 GB nodes. Users can authenticate to the system via CILogon and can spawn Jupyter notebooks as a SLURM job on Comet. Once the job is running, the user sees the same Jupyter notebook, but with the power of a Comet node.

Valerie Polichar (UCSD/ACT) led a discussion focused on the transition from IPV4 to IPV6 in the PRP project. IPV6 is a key feature of the PRP V2 design, but there is debate as to whether PRP V2 will be exclusively IPV6 or will still accommodate IPV4 legacy applications as well as IPV6. A significant contingent of participants argued that, while many users may want the relief valve to defer the transition to IPV6, a hard stop to IPV4 is necessary sooner than later, and the PRP V1-V2 transition can serve as that hard stop. Others countered that scientists are the PRP customers, and if they need more time to transition, it should be feasible to continue taking advantage of the PRP under IPV4. A semi-compromise position is that PRP V1 (IPV4) will overlap with PRP V2 (IPV6) for some finite period of time, so people wanting to defer the transition could continue to use the PRP V1 architecture for some to-be-determined overlap period. Eli Dart commented that ESnet has been running in production on IPV6 for some time, and the transitions for users have ranged from relatively straightforward to quite challenging (e.g. complex web sites), and there is still work for Globus to do with respect to IPV6. Harvey Newman (Caltech) emphasized that the PRP team must look for opportunities to deploy strategically important services as we find out what works for scientists, and Joe Mambretti (Northwestern) noted the macro-revolution from provider-centric managed services to distributed platforms, on top of which users can create services.

### 3.4.4 International Perspective (moderator Joe Mambretti)

The workshop included six international participants from Canada, Japan, Korea, and The Netherlands, and this session included talks by four of these participants.

Joe Mambretti (Northwestern/StarLight) opened the session with a description of the StarLight facility and a number of projects they are currently involved in, with a focus on SDN/SDX. The StarLight SDX will “provide the services, architecture, and technologies designed to provide scientists, engineers, and educators with highly advanced, diverse, reliable, persistent, and secure networking services.” He would like to take the PRP model of interoperable Science DMZs with high-performance data exchange, and “go global,” for example across the Global Lambda Integrated Facility (GLIF).

Akihiro Nakao (University of Tokyo) has been a collaborator on many previous networking efforts within the US and identified four potential areas of collaboration on the PRP project: utilizing within PRP V2 (or a global version of PRP) relatively inexpensive FLARE “software-defined data plane boxes,” network functions (security/firewall and redundancy elimination by intelligent networks), potential contributions to the FIONA DTNs from the FLARE architecture, or utilizing IBM’s True North neural network processor for machine learning or pattern recognition. His group has developed FLARE boxes that provide programmable (and sliceable) data planes on top of commodity hardware. Challenges are programmability, performance, and
cost. Initial boxes used many-core (36–72) core chips for early implementations, and recent experiments with more traditional, less-expensive Intel CPUs have been successful in demonstrating high-performance. FLARE boxes have been deployed in a network testbed across Japan, and, if successful, there could be applications for the PRP V2.

Minsun Lee (Korea Institute of Science and Technology Information) presented a talk about the Korea Research Environment Open NETwork (KREONET), a national R&D network for science and technology information exchange and supercomputing-related collaboration. Currently KREONET member institutions are over 200 organizations and 12 regional network centers. KREONET has several direct links to foreign countries such as the US (StarLight, 1.2Gbps) and China (Chinese Academy of Science/CNIC, CSTNet, 155Mbps). Functions include high-performance data transfers, high-reliability networks for collaborative research, and user-oriented dynamic and flexible networks for research. Scientific applications include many of those discussed in this workshop, including high-energy physics, astrophysics, biomedical data, earth sciences data, and ultra-resolution video.

Cees de Laat (University of Amsterdam) made some general verbal comments for the PRP going forward. The PRP needs to think about scalability from the outset, as domain scientists within the PRP will inevitably need to collaborate with institutions outside the PRP, and more and more institutions will want to participate. He recommended developing a testbed that can be used for innovation and experimentation without interfering with production scientific usage. There are some useful ideas from GENI that can apply to the PRP. Finally, he noted that, while much of the discussion in this workshop has “focused on the edges,” the team should consider the implications that future PRP requirements will have for the core network architectures.

Leon Gommans (Air France KLM) outlined an analysis framework for researching innovations that could be applied to future networking concepts, Service Provider Groups concepts, and autonomous response cyber-security. He provided examples using the Open Infrastructure Architecture Method (OIAm).

4 PRP Requirements

A primary objective of this workshop was to identify requirements for the PRP, particularly from the applications presented by domain scientists on the second day of the workshop. While the PRP approach is generally to work bottom-up with specific science teams to accomplish their objectives and then extend that work to other applications/institutions, rather than design and build a top-down system that meets a set of requirements defined in advance, it is important to identify common needs across the various scientific applications and to establish a clear vision for PRP functionality and performance that will support scientific collaborations and enable transformative science.

Following the five domain-based panels on the second day of the workshop, Frank Würthwein opened the third day with his summary of science-driven requirements for the PRP and moderated a discussion with the workshop participants to further define those requirements. This section starts with a summary of that presentation/discussion, and then captures additional requirements that arose during the technical implementation discussions later in the third day of the workshop.
Würthwein started with a general (albeit non-comprehensive) wish list of capabilities from across the science panels (Figure 8). Fast access to data, ease of use, and reliability were common themes. Most of these capabilities are only indirectly related to network capabilities per se.

Würthwein cautioned that the requirements and wish lists from the science applications cover a broad range of topics across the IT ecosystem (networking, storage, compute, software, etc.), and the reality is that NSF’s PRP award provides only enough funding for ~3 FTEs over the next five years, i.e. the PRP award alone can not be expected to solve all research IT problems for scientists and universities. However, the PRP can and will leverage many other investments, for example by CENIC, ESnet, universities within the PRP, NSF’s funding of CC-*, and other awards to PRP partners, large-scale national facilities, and development efforts by PRP organizations and science teams. (These independent investments and projects also are important elements for sustainability of the PRP beyond the term of the NSF award.) The PRP will “pick low-hanging fruit” to advance specific science teams’ use of the PRP, take advantage of the lessons learned to extend to other science applications and institutions, and advertise the successes to increase adoption by other users.

For clarity, the remainder of Würthwein’s science-driven requirements that relate more directly to the PRP are listed verbatim in the bulleted list below:

- Connectivity within the PRP
  - Most science drivers are OK now with 10 Gbps (achieved) as long as it is consistent and reliable across all ScienceDMZs within the PRP.
  - Some want to push the envelope all the way to 100 Gbps and beyond, especially the visualization folks.
- Connectivity beyond the PRP
  - Many scientists want to connect at 10 Gbps (achieved) to XSEDE resources and US National labs
  - LHC/HEP locations want to route across LHCOne.
  - Connectivity to international Global Climate Model archives at PRP quality
  - Connectivity to Atacama Large Millimeter Array (ALMA) at PRP quality
  - Connectivity to Amazon Web Services, etc. at PRP quality
  - Feeding data to large compute resources is a widely shared requirement.
- Size of Data
  - While the full range of sub-TB to multi-PB was mentioned, most current needs for data transfers seem to be in the O(10)–O(100) TB range.
  - Needs will scale by x10 or more within five-year lifetime of PRP.
  - Needs likely to scale faster than TB/SS growth.
• Starting out with single FIONA is ok, but scale out into distributed cluster of DTNs will happen on campuses.

• “Requirements” for university CIOs and alike
  • Science DMZ must reach the instruments on campus.
    • This is not just a centralized data center IT issue!
  • There will be a strong push on your campus to buy more DTN hardware over time, and locate them in places you did not expect.

• Security and Privacy
  • Generally, security and privacy concerns are secondary to networking, storage, and compute issues in most scientific use cases, except some types of biomedical data.
  • Probably not surprising, security is typically more a concern for resource providers than resource consumers.
    • However, PRP needs to satisfy not just resource consumers!
  • Can we build trusted systems from the ground up?

• Science-driven use cases
  • Bring my data to a large compute resource for processing, and bring the output back to me when done.
  • Make my data available for others to download and process with their local compute resources.
  • Bring my friend’s data to me for processing with my local resources.
  • It’s probably true that nobody cares to manage these transfers, but would rather use caches that make transfer management superfluous.
  • Support incorporating all of the above into orchestrated pipelines and workflows
  • Support this all to happen in (quasi-)realtime, within human attention spans, and in ‘batched’ mode ‘overnight’.

• How do we support science beyond the initial drivers?
  • Other Sciences at participating institutions?
  • Same sciences at other institutions?
  • Other Sciences at other institutions?

There was lively discussion during and after this presentation. Harvey Newman (Caltech) remarked that many people don’t understand the realities of networks, including the complexities of operations or the finite bandwidth often shared across many users. He also notes that LHCONe is often users using a few 10 Gbps links even though the LHCONe dashboard shows 200 Gbps bandwidth. Jim Pepin (Clemson) expressed the concern that there’s lots of general traffic on Internet’s AL2S and it’s filling up; adding high-volume research data to AL2S could create bottlenecks and also increase costs for university CIOs. He also cautioned that some campuses will encounter resistance to removing “network impediments” unless research data utilizes a completely independent network. Eli Dart (ESnet) commented that while some users already have things working such that they can immediately scale up with added bandwidth, other users need to get things reliably working in their environment before added performance makes any difference to them.

Frank Würthwein (UCSD) concurred that getting things working reliably is often the priority for science groups, more than achieving the highest possible bandwidth; and few users can currently use the highest speeds like 100 Gbps – most of those pipes will be happily shared by a number of 10 Gbps users. Cees de Laat (UvA) counters that when providers increase the pipe bandwidth, users adapt to the new ceiling and exploit it. There were several comments about connectivity to commercial cloud providers such as AWS and Azure; while there are some high-speed links via PacWave, the outbound data transfer fees are often prohibitive. Larry Smarr (UCSD) has had
recent conversations with NCSA about joining the PRP partnership as another national supercomputer facility (which has very high bandwidth to Starlight). Richard Moore (UCSD) notes that the summary indicates national supercomputer centers only as data destinations, but they are also data sources for the large volumes of simulation data. With respect to security for protected data, a participant noted that encryption in flight across the network is an important requirement.

There were several comments about DTNs. Cees de Laat reminded the audience that this equipment will need to be updated quickly as required performance increases with bandwidth; Tom DeFanti (UCSD) countered that at least the DTN hardware has modest costs, particularly compared to the labor for updating the networks. Eli Dart added that “while technology changes quickly, people and protocols change slowly”; it is important to get the architecture right, so users don’t have to adapt their workflows significantly as hardware changes. Harvey Newman would add to the science use case sub-bullets that projects such as the Large Synoptic Survey Telescope (LSST) would like quick data transfers (~1 minute), even if not in real time. Eli Dart commented that “quick” data transfers, for the purpose of verifying that an experiment is working and good data are being collected, are an important use case that is quite different from overnight bulk data transfer.

Larry Smarr concluded this discussion with a reminder that “we have seen this movie before.” When the NSF created the NSFNet, NSF made some upfront investments, then campuses had to contribute funding to participate and sustain the capability; campuses invested a total of 10X the initial NSF investment. When vBNS came along, a similar model was used with initial NSF investments, but sustainability from campuses (and in that case, the creation of Internet2). NSF’s current CC-* networking investments are a tremendous investment and the PRP will help show the way for what can be created from those investments - but campuses will need to invest.

The following list represents additional requirements for the PRP technology and the project identified during the Technical Approach sessions on the third day of the workshop:

- Keep the PRP wide-area network “big, fast and clean.”
- Build an architecture that is readily scalable (i.e., low labor) to additional institutions and additional science applications.
- Implement IPV6 for PRP V2 – but decide soon how to handle IPV4 legacy applications.
- Incorporate common identity management tools such as InCommon to make it easier for campus users and lower barriers to adoption.
- Ensure that Globus is deployed as a common data transfer tool across all PRP partners; other tools are likely to also be in this toolkit.
- Develop and adopt effective common security practices across the PRP equipment, including use of TPMs.
- Incorporate DTNs in compute cluster designs, mounting the parallel file system directly and connected directly to the Science DMZ.

5 Findings and Recommendations

The 2.5-day workshop was successful in gathering key stakeholders in the PRP project – administrators, scientists, network engineers and researchers – to capture science-driven requirements and discuss ideas and priorities for the project going forward. This section
summarizes the findings and the recommendations from the workshop. Note that the requirements laid out in the previous section are also key findings from the workshop and are incorporated here by reference.

Findings

- Many examples of large-scale science collaborations require high-performance end-to-end networking, and with that capability, there are opportunities for transformative advances in the science. (And many more are sure to arise beyond those highlighted in this workshop.)
- The Science Engagement process is crucial to project success. 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.
  - This workshop was successful in bringing together a diverse set of Science Teams and network/IT experts from across the PRP partnership to develop a mutual understanding of requirements and implementations of the PRP.
- The PRP project 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/management: 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 all PRP partners and CENIC, etc.
  - Stakeholders may occasionally have conflicting interests or priorities, for example network security issues and funding.
  - High-level administrators within the PRP partnership (campus CIOs and VCRs, CENIC and ESnet management) have expressed clear support of PRP objectives, and their support is a key requirement for the project’s success. This commitment indicates a willingness to resolve potential conflicting interests and priorities.
- 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.
- Having full engagement of the regional network (CENIC) is a tremendous advantage to the project. CENIC provides the high-speed backbone and the interfaces to the various campus networks, and has demonstrated strong support for this project.
- While the PRP is intentionally a regional partnership to build realistic-scale collaborations and leverage the CENIC infrastructure, opportunities/pressures to extend the successes of interoperable Science DMZs beyond the current partners and beyond CENIC are likely to emerge.
  - Participating scientists will have collaborators or data repository users beyond the PRP partnership, will want access to national/international supercomputer centers and data repositories, and other institutions will want to adopt this capability across their own Science DMZs.
- The PRP is a modest-scale grant and its leverage of other major investments by NSF, CENIC, campuses, ESnet and other organizations is crucial to its success, as well as to the sustainability of the PRP beyond the award.
  - A potential detriment to this leverage is that only a small fraction of efforts necessary for success are funded and managed by PRP staff. This reinforces the need for high-level management support of this project across the partnership.
- There is general consensus that a Layer 3 architecture is best-suited to the PRP.
- There is consensus that the collaboration should incorporate IPV6 capability into the PRP V2, but there was not yet consensus on whether V2 should be exclusively IPV6 or...
accommodate both IPV4 and IPV6 for some period of time. This transition will be disruptive for some applications.

- SDN/SDX are important technologies to incorporate in the design of PRP V2.
- The concept of primitives to characterize various dimensions of data transfers is useful to identify common tools/processes across different scientific use cases.
- It would be valuable for campuses to deploy local cyberinfrastructure to allow additional scientists (those not supported under the PRP grant) to effectively use the PRP.
- Successfully support of science outcomes on the PRP will require the deployment and operation of higher level services on top of the network infrastructure. Those are likely to include filesystems, disk caches, and job and workflow management infrastructure, in addition to gridftp.
- One of the important outcomes of the PRP is likely a better understanding of how to differentiate application/science team needs based on the higher level services most crucial for their success.

Recommendations

- Rather than “build it and hope they will come,” use the Science Engagement process to capture requirements from science teams that will use the system, deploy the system responsive to those requirements, and work with the science projects all along the way to build and use the system.
- Pursue the proposed strategy of starting with a few ‘low-hanging fruit’ science projects that have specific collaborations and technical requirements, science staff willing and able to work with PRP engineers, and clear scientific advances to be made with higher-performance data transfers.
  - Many unanticipated challenges will arise with any of the science projects. Do something real and quickly on a few initial projects, and then plan to iterate as challenges are encountered.
  - It is expected that the initial projects will serve as pathfinders for solving both social and technical issues in deploying the PRP, and the lessons learned will be extensible to other projects/domains, resulting in quicker, easier integration of subsequent science teams and sites.
  - Add stories on early PRP-enabled science drivers to the PRP web site and send email announcements of the stories to all PRP participants on a regular basis.
  - Be alert for new application drivers that can begin using the PRP before some of the ones described in the grant.
- Keep the CENIC network backbone “big, fast and clean” and focus efforts on “the edges” – the campus Science DMZs and the network interfaces to the science users and their data.
- Prioritize high-performance access to/from the national supercomputer facilities within the PRP (e.g., NERSC, NCAR, SDSC) as their facilities are used or can be used by many different science teams. This includes high-performance access to/from large parallel file systems and data repositories at those facilities.
- Develop a PRP problem repository (e.g., instances of low data transfer performance) with follow-up on solutions that can be developed into a knowledge base and used for future PRP institutions/applications, as well as regional partnerships beyond the PRP.
- Evaluate the Service Provider Group concept as a construct for the PRP to establish a well-defined federation of autonomous entities (e.g., campuses), particularly for scalability when the PRP is extended to include a broader set of institutional partners.
6 Acknowledgements

This workshop is supported by NSF workshop award #1540112, as well as NSF grants ACI-1246396 and ACI-1541349. Additional support is appreciated from the University of California Office of the President (UCOP), the Corporation for Education Network Initiatives in California (CENIC), and Calit2.
7 References

[2] Recent NSF campus cyberinfrastructure networking solicitations (CC-*)
8 Appendices

8.1 Workshop Agenda, Presentation Materials and Videos

Complete information about this workshop is available at the PRP website http://prp.ucsd.edu.

Presentation slides are posted on the website under the ‘Presentations’ tab (http://prp.ucsd.edu/presentations). Videos of the speakers’ presentations are available directly on youtube at https://www.youtube.com/playlist?list=PLbbCsk7MULGcrHp3t_ia0At9c2T3Unn_R.

The agenda below is the planned workshop agenda, updated to reflect substitute speakers and the actual order of panelist presentations. (Actual times are not updated, so this can be used only as a guideline when correlating times with the video presentations.)

Wednesday, October 14, 2015 — Pacific Research Platform Introduction

10:00am VCR/CIO Collaborations in UC
Sandra Brown, UC San Diego Vice Chancellor for Research

10:15am Overview of the Pacific Research Platform
Larry Smarr, PRP Principal Investigator, Calit2

10:45am Panel Discussion
• Greg Bell, ESnet Director
• Louis Fox, CENIC CEO
• Tom Andriola, UC Office of the President VP and CIO
• Amy Walton, PRP NSF Program Officer

Noon Lunch

(Wednesday afternoon: Workshop attendees were invited to participate in the SDSC 30th Anniversary/Comet event hosted by SDSC.)

Thursday, October 15, 2015 — Science Engagement

9:45am Registration

10:00am Opening remarks
Camille Crittenden, CITRIS
10:10am  **What is Science Engagement? A Summary & Introduction to Goals**  
Eli Dart, ESnet

10:30am  **Particle Physics**, Frank Würthwein, UC San Diego, moderator  
- Jason Nielsen, UC Santa Cruz  
- Owen Long, UC Riverside  
- Chris West, UC Santa Barbara  
- Maria Spiropolu, Caltech

11:15am  **Astronomy and Astrophysics**, Michael Norman, San Diego Supercomputer Center, moderator  
- Brian Keating, UC San Diego, POLARBEAR  
- Peter Nugent, Telescope Surveys Team: LBNL  
- Jocelyn Read, CSU-Fullerton, Gravitational Wave Astronomy Team (presented by Sharon Brunett, Caltech)

11:50pm  **Discussion**, Greg Bell, Moderator

12:30pm  Lunch

1:30pm  **Biomedical Sciences**, Larry Smarr, Calit2, UC San Diego, moderator  
- Rob Knight, UC San Diego, Microbiome and Integrative ‘Omics Team  
- Sergio Baranzini, UC San Francisco  
- JJ Garcia Luna, UC Santa Cruz, Cancer Genomics Hub/Browser Team (presented by Larry Smarr, UC San Diego/Calit2)

2:20pm  **Earth Sciences**, Camille Crittenden, moderator  
- Frank McKenna, UC Berkeley, Pacific Earthquake Engineering Research Center (PEER), Data Analysis and Simulation for Earthquakes and Natural Disasters  
- Eric Nienhouse, NCAR/UCAR, Climate Modeling  
- Dan Cayan, UC San Diego, Scripps Institute of Oceanography, California/Nevada Regional Climate Data Analysis  
- Christopher Paolini, San Diego State University, CO2 Subsurface Modeling

3:15pm  Break

3:30pm  **Scalable Visualization, Virtual Reality, and Ultra-Resolution Video**, Tom DeFanti, UC San Diego, moderator  
- Maxine Brown, University of Illinois, Chicago, coordinating content with Jason Leigh, University of Hawaii, Manoa  
- Tom Levy, UC San Diego
Friday, October 16, 2015 — PRP Technical Approach and Project Plan

7:45-8:30am  Breakfast, Registration

8:30-9:15am  End-to-End Technical Requirements  Frank Würthwein, UC San Diego
              Summarize end-to-end technical requirements from the key science use cases presented on Thursday.

9:15-10:15  The PRPv1 Architecture Model Panel w/ Discussion  John Hess (moderator):
              The PRPv1 architecture as it stands now, mid-October 2015 moving to a routed layer 3 implementation in time for CENIC 2016 in March; what does CalREN need to do? (10 min each + Q&A.)
              ● Phil Papadopoulos (FIONAs)
              ● Tom Hutton (UCSD’s DMZ)
              ● Eli Dart (From the ESnet POV)
              ● Erik McCroskey (from the HPR TAC POV)

10:15-10:45  Break

10:45-12:16  Securing the PRPv1  Eli Dart (moderator) (15 min)
              General: How closed/open should the PRPv1 be? What transfer software is mandatory/optional? Read global, write local?
              InCommon/CILogon Federated access:
              ● Bruce Vincent / Stanford Incommon (15 min)
              ● Michael Duff/security (15 min)
              Service Provider Group Framework:
              ● Leon Gommans / KLM, Cees De Laat / UvA (15 min)
              DTN architecture and relation to storage back ends  Rick Wagner SDSC (15min):
- Lustre, CEPH, Cloud (SWIFT), GPFS, FIONAs with 10s of TBs

Discussion (15 min)

Plan for next PRPv1 tech workshop—
- When, what, and where--Tom DeFanti

12:16-1:20  
Lunch

1:20-3:00  
PRPv2 Architecture and Security (John Haskins, moderator)
- Read/Write abilities; general concepts Phil Papadopoulos (10 min)
- SDN over PacificWave Darrell Newcomb (10 min)
- SDN OpenFlow Darrell Newcomb (10 min)
- Trusted Platform Modules (TPM) for DTNs John Graham (10 min)
- CILogon authentication and TPM certificate encryption for Jupyterhub Andrea Zonca and John Graham (10 min)
- Pv6 security concepts on PRPv2 Valerie Polichar & guests (30 min)
- Discussion: (remainder)

3:10-3:30  
Break

3:30-4:30  
International Perspective (Joe Mambretti, moderator, 12 min)
Analysis and feedback regarding the architecture and plans put forth at the workshop, from both an applications and technical perspective. Panelists (12 min each):
- Japan - Akihiro Nakao, U Tokyo
- Korea - Min Sun Lee, KISTI
- Netherlands - Cees de Laat, UvA
- Netherlands – Leon Gommans, KLM

4:30-5:30  
Outlining the Workshop Report
### 8.2 Workshop Registrants

<table>
<thead>
<tr>
<th>First Name</th>
<th>Last Name</th>
<th>Affiliation</th>
<th>Job Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Celeste</td>
<td>Anderson</td>
<td>U Southern California</td>
<td>Director External Networking Group (unable to attend)</td>
</tr>
<tr>
<td>Tom</td>
<td>Andriola</td>
<td>UC Office of the President</td>
<td>CIO</td>
</tr>
<tr>
<td>Naveen</td>
<td>Ashish*</td>
<td>U Southern California</td>
<td>Associate Professor</td>
</tr>
<tr>
<td>Paul</td>
<td>Averill</td>
<td>NASA JPL</td>
<td>Network Architect</td>
</tr>
<tr>
<td>Corey</td>
<td>Baker</td>
<td>UC San Diego</td>
<td>Postdoc</td>
</tr>
<tr>
<td>Sergio</td>
<td>Baranzini</td>
<td>UC San Francisco</td>
<td>Professor</td>
</tr>
<tr>
<td>Schyler</td>
<td>Batey</td>
<td>U Washington</td>
<td>Network Engineer</td>
</tr>
<tr>
<td>Greg</td>
<td>Bell</td>
<td>LBNL / ESnet</td>
<td>Director</td>
</tr>
<tr>
<td>Joe</td>
<td>Bengfort*</td>
<td>UC San Francisco</td>
<td>Chief Information Officer</td>
</tr>
<tr>
<td>Dawn</td>
<td>Boyd</td>
<td>Caltech</td>
<td>Director, Voice &amp; Data Networks</td>
</tr>
<tr>
<td>Maxine</td>
<td>Brown</td>
<td>U of Illinois at Chicago</td>
<td>Director, Electronic Visualization Lab</td>
</tr>
<tr>
<td>Sandra</td>
<td>Brown</td>
<td>UC San Diego</td>
<td>Vice Chancellor for Research</td>
</tr>
<tr>
<td>Sharon</td>
<td>Brunett</td>
<td>Caltech / LIGO</td>
<td>Computing Analyst</td>
</tr>
<tr>
<td>Richard</td>
<td>Caasi</td>
<td>UC San Diego</td>
<td>Programmer Analyst</td>
</tr>
<tr>
<td>Dan</td>
<td>Cayan</td>
<td>UC San Diego / Scripps Inst of Oceanography &amp; USGS</td>
<td>Researcher</td>
</tr>
<tr>
<td>Jim</td>
<td>Chen</td>
<td>Northwestern Univ / International Center for Advanced Internet Research</td>
<td>Associate Director</td>
</tr>
<tr>
<td>Sung-En</td>
<td>Chiu</td>
<td>UC San Diego</td>
<td>GSR</td>
</tr>
<tr>
<td>Vivian</td>
<td>Chu</td>
<td>UC San Diego</td>
<td>Developer</td>
</tr>
<tr>
<td>William</td>
<td>Clebsch</td>
<td>Stanford University</td>
<td>Associate Vice President</td>
</tr>
<tr>
<td>Larry</td>
<td>Conrad*</td>
<td>UC Berkeley</td>
<td>Associate Vice Chancellor &amp; CIO</td>
</tr>
<tr>
<td>Camille</td>
<td>Crittenden</td>
<td>UC Berkeley / CITRIS</td>
<td>Deputy Director</td>
</tr>
<tr>
<td>Matteo</td>
<td>Danieleto</td>
<td>UC San Diego</td>
<td>PostDoc</td>
</tr>
<tr>
<td>Eli</td>
<td>Dart</td>
<td>LBNL / ESnet</td>
<td>Network Engineer</td>
</tr>
<tr>
<td>Cees</td>
<td>de Laat</td>
<td>University of Amsterdam</td>
<td>Researcher</td>
</tr>
<tr>
<td>Thomas</td>
<td>DeFanti</td>
<td>UC San Diego</td>
<td>Research Scientist</td>
</tr>
<tr>
<td>Raheleh</td>
<td>Dilmaghani*</td>
<td>SPAWAR Systems Center Pacific/UC San Diego</td>
<td>Researcher</td>
</tr>
<tr>
<td>Thanh</td>
<td>Doan</td>
<td>U Southern California</td>
<td>Senior Network Engineer</td>
</tr>
<tr>
<td>Sudip</td>
<td>Dosanjh</td>
<td>LBNL / NERSC</td>
<td>Director</td>
</tr>
<tr>
<td>Michael</td>
<td>Duff</td>
<td>Stanford University</td>
<td>Chief Information Security Officer</td>
</tr>
<tr>
<td>Michael</td>
<td>Dula</td>
<td>UC Office of the President</td>
<td>IT Planning Executive</td>
</tr>
<tr>
<td>Clinton</td>
<td>Edwards</td>
<td>UC San Diego / Scripps Oceanography</td>
<td>Staff Research Associate</td>
</tr>
<tr>
<td>Alvaro</td>
<td>Espinoza</td>
<td>UC Merced</td>
<td>I.T. Infrastructure Services Manager</td>
</tr>
<tr>
<td>Louis</td>
<td>Fox</td>
<td>CENIC</td>
<td>President &amp; CEO</td>
</tr>
<tr>
<td>Scott</td>
<td>Friedman</td>
<td>UC Los Angeles</td>
<td>Chief Technologist for Research Computing</td>
</tr>
<tr>
<td>JJ</td>
<td>Garcia-Luna-Aceves</td>
<td>UC Santa Cruz</td>
<td>Baskin Professor of Computer Engineering (unable to attend)</td>
</tr>
<tr>
<td>Name</td>
<td>Affiliation</td>
<td>Position</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Leon Gommans</td>
<td>Air France KLM</td>
<td>Science Officer</td>
<td></td>
</tr>
<tr>
<td>Nancy Grady*</td>
<td>SAIC</td>
<td>Principal Data Scientist (unable to attend)</td>
<td></td>
</tr>
<tr>
<td>John Graham</td>
<td>UC San Diego</td>
<td>Senior Development Engineer</td>
<td></td>
</tr>
<tr>
<td>David Greenbaum</td>
<td>UC Berkeley</td>
<td>Director, Research Information Technologies</td>
<td></td>
</tr>
<tr>
<td>Russ Harvey</td>
<td>UC Riverside</td>
<td>Computing Infrastructure</td>
<td></td>
</tr>
<tr>
<td>John Haskins</td>
<td>UC Santa Cruz</td>
<td>Network</td>
<td></td>
</tr>
<tr>
<td>Thomas Heetderks</td>
<td>Montana State University</td>
<td>Software Engineer</td>
<td></td>
</tr>
<tr>
<td>Laurin Herr</td>
<td>Pacific Interface Inc.</td>
<td>President</td>
<td></td>
</tr>
<tr>
<td>John Hess</td>
<td>CENIC</td>
<td>Network Engineer</td>
<td></td>
</tr>
<tr>
<td>John Hicks</td>
<td>Internet2</td>
<td>Research network engineer</td>
<td></td>
</tr>
<tr>
<td>Greg Hidley</td>
<td>UC San Diego / CALIT2</td>
<td>Project Manager</td>
<td></td>
</tr>
<tr>
<td>Chris Hoffman</td>
<td>UC Berkeley / Research IT</td>
<td>Manager, Informatics Services</td>
<td></td>
</tr>
<tr>
<td>Thomas Hutton</td>
<td>UC San Diego / SDSC</td>
<td>Network Architect</td>
<td></td>
</tr>
<tr>
<td>Ron Johnson</td>
<td>U Washington &amp; Pacific NorthWest Gigapop</td>
<td>Professor UW &amp; CEO PNWGP</td>
<td></td>
</tr>
<tr>
<td>Gary Jung</td>
<td>LBNL</td>
<td>HPC Services Manager</td>
<td></td>
</tr>
<tr>
<td>Jonathan Kaufman</td>
<td>UC San Diego</td>
<td>Postdoctoral Researcher</td>
<td></td>
</tr>
<tr>
<td>Brian Keating</td>
<td>UC San Diego</td>
<td>Professor</td>
<td></td>
</tr>
<tr>
<td>Rob Knight</td>
<td>UC San Diego / Pediatrics and Computer Science &amp; Engineering</td>
<td>Professor</td>
<td></td>
</tr>
<tr>
<td>Falko Kuester</td>
<td>UC San Diego / Qualcomm Inst.CISA3.GRAVITY</td>
<td>Professor</td>
<td></td>
</tr>
<tr>
<td>Bill Labate</td>
<td>UC Los Angeles / Research Technology Group</td>
<td>Director</td>
<td></td>
</tr>
<tr>
<td>Minsun Lee</td>
<td>KISTI</td>
<td>Sr. Researcher</td>
<td></td>
</tr>
<tr>
<td>Thomas Levy</td>
<td>UC San Diego / Center for Cyber-Archaeology and Sustainability</td>
<td>Director</td>
<td></td>
</tr>
<tr>
<td>Clark Lin</td>
<td>UC San Diego</td>
<td>Student</td>
<td></td>
</tr>
<tr>
<td>Pol Llovet</td>
<td>Montana State University</td>
<td>Assoc. Director of Cyberinfrastructure</td>
<td></td>
</tr>
<tr>
<td>Owen Long</td>
<td>UC Riverside</td>
<td>Professor</td>
<td></td>
</tr>
<tr>
<td>Jim Madden</td>
<td>UC San Diego</td>
<td>Network Architecture</td>
<td></td>
</tr>
<tr>
<td>Joe Mambretti</td>
<td>Northwestern University</td>
<td>Director International Center for Advanced Internet Research</td>
<td></td>
</tr>
<tr>
<td>Surya Manchikanti</td>
<td>UC San Diego</td>
<td>Student</td>
<td></td>
</tr>
<tr>
<td>Harry Mangalam</td>
<td>UC Irvine</td>
<td>Senior Geek</td>
<td></td>
</tr>
<tr>
<td>Eddie Mardon</td>
<td>UC San Diego</td>
<td>Director Enterprise Network</td>
<td></td>
</tr>
<tr>
<td>Ruth Marinshaw</td>
<td>Stanford University</td>
<td>CTO - Research Computing</td>
<td></td>
</tr>
<tr>
<td>Kevin Mayeshiro</td>
<td>UC Davis</td>
<td>Senior Network Engineer</td>
<td></td>
</tr>
<tr>
<td>Erik McCroskey</td>
<td>UC Berkeley</td>
<td>Principal Network Engineer</td>
<td></td>
</tr>
<tr>
<td>Frank McKenna</td>
<td>UC Berkeley / PEER</td>
<td>Research Engineer</td>
<td></td>
</tr>
<tr>
<td>Inder Monga</td>
<td>LBNL / ESnet</td>
<td>CTO</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Affiliation</td>
<td>Position</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Richard Moore</td>
<td>UC San Diego</td>
<td>Emeritus</td>
<td></td>
</tr>
<tr>
<td>Azher Mughal</td>
<td>Caltech</td>
<td>Network Engineer</td>
<td></td>
</tr>
<tr>
<td>Paul Murray</td>
<td>Stanford University</td>
<td>Network Engineer</td>
<td></td>
</tr>
<tr>
<td>Akihiro Nakao</td>
<td>The University of Tokyo</td>
<td>Professor</td>
<td></td>
</tr>
<tr>
<td>Darrell Newcomb</td>
<td>CENIC</td>
<td>Internet Services</td>
<td></td>
</tr>
<tr>
<td>Harvey Newman</td>
<td>Caltech</td>
<td>Professor</td>
<td></td>
</tr>
<tr>
<td>Jason Nielsen</td>
<td>UC Santa Cruz</td>
<td>Associate Professor of Physics</td>
<td></td>
</tr>
<tr>
<td>Eric Nienhouse</td>
<td>National Center for Atmospheric Research</td>
<td>Group Head Software Applications and Gateway Engineering</td>
<td></td>
</tr>
<tr>
<td>Michael Norman</td>
<td>UC San Diego / San Diego Supercomputer Center</td>
<td>Director</td>
<td></td>
</tr>
<tr>
<td>Peter Nugent</td>
<td>LBNL, UC Berkeley</td>
<td>Division Deputy</td>
<td></td>
</tr>
<tr>
<td>Anikesh Pal</td>
<td>UC San Diego</td>
<td>Student</td>
<td></td>
</tr>
<tr>
<td>Christopher Paolini</td>
<td>San Diego State University</td>
<td>Staff Scientist</td>
<td></td>
</tr>
<tr>
<td>Philip Papadopoulos</td>
<td>UC San Diego / SDSC, Calit2</td>
<td>Program Director</td>
<td></td>
</tr>
<tr>
<td>Florent Parent</td>
<td>Compute Canada, Université Laval</td>
<td>Networking team lead</td>
<td></td>
</tr>
<tr>
<td>George Peek</td>
<td>UC Santa Cruz</td>
<td>Network Engineer</td>
<td></td>
</tr>
<tr>
<td>Jerry Peng</td>
<td>UC San Diego / ECE</td>
<td>GSA</td>
<td></td>
</tr>
<tr>
<td>James Pepin</td>
<td>Clemson University</td>
<td>Chief Technology Officer</td>
<td></td>
</tr>
<tr>
<td>Wayne Pfeiffer</td>
<td>UC San Diego / San Diego Supercomputer Center</td>
<td>Distinguished Scientist</td>
<td></td>
</tr>
<tr>
<td>Valrie Polichar</td>
<td>UC San Diego</td>
<td>Program Manager</td>
<td></td>
</tr>
<tr>
<td>Dave Rapp</td>
<td>UC San Diego</td>
<td>Network Architect</td>
<td></td>
</tr>
<tr>
<td>Jocelyn Read</td>
<td>Cal State Univ Fullerton</td>
<td>Assistant Professor</td>
<td></td>
</tr>
<tr>
<td>David Reese</td>
<td>CENIC/Pacific Wave</td>
<td>Vice President</td>
<td></td>
</tr>
<tr>
<td>Phil Reese</td>
<td>Stanford University</td>
<td>Research Computing Strategist</td>
<td></td>
</tr>
<tr>
<td>Tad Reynales</td>
<td>Internet2</td>
<td>Community Engagement Manager</td>
<td></td>
</tr>
<tr>
<td>Fuzzy Rogers</td>
<td>UC Santa Barbara</td>
<td>HPC Sys Admin</td>
<td></td>
</tr>
<tr>
<td>Dana Roode</td>
<td>UC Irvine</td>
<td>Chief Information Officer</td>
<td></td>
</tr>
<tr>
<td>Walt Scacchi</td>
<td>UC Irvine / Institute for Virtual Environments and Computer Games</td>
<td>Director of Research</td>
<td></td>
</tr>
<tr>
<td>Allen Schiano</td>
<td>UC Irvine / Office of Information Technology</td>
<td>Director, Instructional and Research Infrastructure</td>
<td></td>
</tr>
<tr>
<td>Kevin Schmidt</td>
<td>UC Santa Barbara</td>
<td>Network Manager</td>
<td></td>
</tr>
<tr>
<td>Patrick Schmitz</td>
<td>UC Berkeley</td>
<td>Associate Director, Research IT</td>
<td></td>
</tr>
<tr>
<td>Jurgen Schulze</td>
<td>UC San Diego</td>
<td>Research Scientist</td>
<td></td>
</tr>
<tr>
<td>Woojin Seok</td>
<td>KISTI</td>
<td>Director</td>
<td></td>
</tr>
<tr>
<td>Horst Simon</td>
<td>LBNL</td>
<td>Deputy Laboratory Director</td>
<td></td>
</tr>
<tr>
<td>Shailendra Singh</td>
<td>UC San Diego</td>
<td>Graduate Student</td>
<td></td>
</tr>
<tr>
<td>Julie Slusser</td>
<td>UC Riverside</td>
<td>Network Analyst</td>
<td></td>
</tr>
<tr>
<td>Larry Smarr</td>
<td>UC San Diego / Calit2</td>
<td>Director</td>
<td></td>
</tr>
<tr>
<td>Chris Sneatthen</td>
<td>UC Santa Barbara / Enterprise Technology Services</td>
<td>Assoc. Director, CCVIS</td>
<td></td>
</tr>
<tr>
<td>Manish Sompura</td>
<td>UC San Francisco</td>
<td>Manager - Network Engineering</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Affiliation</td>
<td>Title</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Josh Sonstroem</td>
<td>UC Santa Cruz</td>
<td>Unix Sysadmin</td>
<td></td>
</tr>
<tr>
<td>Andrew Spector</td>
<td>SAIC</td>
<td>VP for Solution Architecture</td>
<td></td>
</tr>
<tr>
<td>Maria Spiropulu</td>
<td>Caltech</td>
<td>Professor of Physics</td>
<td></td>
</tr>
<tr>
<td>Stephen Squires</td>
<td>Kelvin 1687 Inc</td>
<td>President</td>
<td></td>
</tr>
<tr>
<td>Adam Stone</td>
<td>LBNL</td>
<td>Deputy Chief Information Officer</td>
<td></td>
</tr>
<tr>
<td>Rune Stromsness</td>
<td>LBNL</td>
<td>LBLnet Manager</td>
<td></td>
</tr>
<tr>
<td>Bill Strossman</td>
<td>UC Riverside</td>
<td>Research Technologist</td>
<td></td>
</tr>
<tr>
<td>Jean Su</td>
<td>UC San Diego</td>
<td>Web developer</td>
<td></td>
</tr>
<tr>
<td>Atsushi Takahara</td>
<td>NTT Electronics Corporation</td>
<td>Senior Vice President, Sales &amp; Marketing Group</td>
<td></td>
</tr>
<tr>
<td>Michael Van Norman</td>
<td>UC Los Angeles</td>
<td>Senior Director, Networking &amp; Communications</td>
<td></td>
</tr>
<tr>
<td>Johan van Reijendam</td>
<td>Stanford University</td>
<td>Network Engineer</td>
<td></td>
</tr>
<tr>
<td>Bruce Vincent</td>
<td>Stanford University</td>
<td>Senior Technology Strategist &amp; Chief IT Architect</td>
<td></td>
</tr>
<tr>
<td>Rick Wagner</td>
<td>UC San Diego / San Diego Supercomputer Center</td>
<td>HPC Systems Manager</td>
<td></td>
</tr>
<tr>
<td>Amy Walton</td>
<td>National Science Foundation</td>
<td>Program Director</td>
<td></td>
</tr>
<tr>
<td>Chang-Heng Wang</td>
<td>UC San Diego</td>
<td>Student</td>
<td></td>
</tr>
<tr>
<td>Conway Wang</td>
<td>UC San Diego / ECE</td>
<td>PHD student</td>
<td></td>
</tr>
<tr>
<td>Yadong Wang</td>
<td>U Southern California</td>
<td>Senior Internet Engineer</td>
<td></td>
</tr>
<tr>
<td>Jim Warner</td>
<td>UC Santa Cruz</td>
<td>Network</td>
<td></td>
</tr>
<tr>
<td>Lee Ann Weber</td>
<td>CENIC</td>
<td>Director Communications &amp; Engagement</td>
<td></td>
</tr>
<tr>
<td>Chris West</td>
<td>UC Santa Barbara</td>
<td>Postdoc</td>
<td></td>
</tr>
<tr>
<td>Stephen Wolbers</td>
<td>Oregon State University</td>
<td>Research Computing Manager</td>
<td></td>
</tr>
<tr>
<td>Frank Würthwein</td>
<td>UC San Diego / SDSC</td>
<td>Professor</td>
<td></td>
</tr>
<tr>
<td>Gabe Youtsey*</td>
<td>UC Agriculture and Natural Resources</td>
<td>Chief Information Officer</td>
<td></td>
</tr>
<tr>
<td>Jessica Yu</td>
<td>UC Irvine</td>
<td>Network Architect</td>
<td></td>
</tr>
<tr>
<td>Li Yu</td>
<td>UC San Diego</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Zhiyong Zhang</td>
<td>Stanford University</td>
<td>Computational Scientist</td>
<td></td>
</tr>
<tr>
<td>Andrea Zonca</td>
<td>UC San Diego / San Diego Supercomputer Center</td>
<td>HPC Applications</td>
<td></td>
</tr>
<tr>
<td>Jason Zurawski</td>
<td>LBNL / ESnet</td>
<td>Science Engagement Engineer</td>
<td></td>
</tr>
</tbody>
</table>

* Registered for workshop but unable to attend.