Engagement and Performance Operations Center (EPOC)

(Formerly known as ReSEC)

NSF Award #1826994

Year 1 Annual Report

1 July 2018 through 31 March 2019

PIs: Jennifer Schopf (IU), Jason Zurawski (LBNL), Dave Jent (IU)

Summary

The goal of the EPOC project is to provide researchers and network engineers with a holistic set of tools and services needed to debug performance issues and enable reliable and robust data transfers. It supports five main activities: Roadside Assistance and Consulting, Application Deep Dives, Network Analysis using NetSage, the provision of Managed Services, and Training. In Year 1, highlights included setting up the full Roadside Assistance and Consulting framework and working through multiple cases of each, two completed Deep Dive sessions and planning for seven others, NetSage prototype deployments for the Great Plains Network and iLight, initial work to define Managed Services and partner interest, and several engineering and engagement Training sessions.

1. EPOC Overview

The Engagement and Performance Operations Center (EPOC) is a production platform for operations, applied training, monitoring, and research and education support. EPOC is a collaborative focal point for operational expertise and analysis and is jointly led by Indiana University (IU) and the Energy Sciences Network (ESnet). EPOC provides researchers and network engineers with a holistic set of tools and services needed to debug performance issues and enable reliable and robust data transfers. By considering the full end-to-end data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

EPOC supports five main activities:
- Roadside Assistance and Consulting via a coordinated Operations Center to resolve network performance problems with end-to-end data transfers reactively;
- Application Deep Dives to work more closely with application communities to proactively understand full workflows for diverse research teams in order to evaluate bottlenecks and potential capacity issues;
- Network Analysis enabled by the NetSage monitoring suite to proactively discover and resolve performance issues;

- Provision of Managed Services via support through the IU GlobalNOC and our Regional Network Partners;
- Training to ensure effective use of network tools and science support.

This report details the staffing, collaboration, work in the five activities in Year 1, and planning for Year 2 of the project.

2. Staffing

This project PIs were notified in early July, 2018, that the project had been awarded, but due to administrative overhead, funds were not available within IU for several months, nor was the subaward for LBNL completed until April, 2019. Because of this, for the first several months of the project, personnel were leveraged from other existing projects and not directly funded by EPOC except for the project Director, Jennifer Schopf, and project coordinator, Heather Hubbard, who were back-paid for their time. In October, additional IU staff were funded from this award, including Ed Balas, to assist with the system architecture, Ed Moynihan, to assist with science engagement, Hans Addleman, to oversee the Roadside Assistance and Consulting activities, Doug Southworth, to help with partner coordination and Deep Dives, and Andrew Lee, to help with analysis. Balas left the project February 28, 2019, and the system architect role shifted to Dan Doyle. At that time, this project also began to formally pay for the 0.5 FTE of IU GlobalNOC developer resources under Balas/Doyle that were being used by the project. Dave Jent is a co-PI, but due to his position at IU is unable to formally charge the project. Co-PI Jason Zurawski participated in the project since its start date, but due to the slow set up of the subaward will he back-charge his time in Year 2.

At the end of Year 1, project staff consisted of:
- Jennifer Schopf, IU, PI - overall project director
- Heather Hubbard, IU, project coordinator
- Dan Doyle, IU, system architect - measurement and monitoring
- Hans Addleman, IU, Roadside Assistance Lead and Consulting
- Ed Moynihan, IU, Science engagement support
- Doug Southworth, IU, Partner coordination and deep dives
- Dave Jent, IU, co-PI
- Jason Zurawski, ESNet - co-PI

In Year 2 we expect to ramp up the level of staffing available from Doyle's team, and very likely add on additional help at LBNL as well.

3. Collaboration and Travel

EPOC staff participated in various meetings to support ongoing deployment, collaboration, and training. Note that several of these were funded by other sources but relevant to EPOC. The travel for Year 1 included:

- Schopf and Jent attended the National Research Platform meeting held in Bozeman, Montana, on August 6-7, 2018, http://www.cvent.com/events/national-research-platform-conference-toward-a-national-big-data-superhighway/event-summary-48a69b9807bd46ecb5d4343bcbfa61c5.aspx. She gave two presentations on EPOC - one focusing on the monitoring systems used and the other focusing on science engagement. Several conversations with current and prospective partners also took place.
- Zurawski and Lauren Rotman, LBNL/ESnet, attended and presented at a NOAA Networking Committee Engagement Workshop, in Silver Spring MD, September 11-12, 2018. The primary focus of this event was a workshop on Science Engagement featuring a researcher from NASA.
- Zurawski and Scott Chevalier, IU, attended the SOX PRP workshop in Atlanta GA on Sept 18-21 2018. They both gave presentations on topics including perfSONAR and data transfers.
- Schopf, Zurawski, and Jent attended the Quilt/CC* PI meeting on September 24-27, 2018, at the University of Maryland, College Park, Maryland https://www.thequilt.net/public-event/2018-nsf-esnet-quilt-workshops-meetings/. Schopf gave a talk on EPOC as part of a panel. Schopf and Zurawski also lead a workshop on Science engagement featuring a researcher from the University of Maryland. Both also lead a session of 'speed learning'. In addition to the presentations side meetings were organized with all regional partners (iLight, FRGP, GPN, LEARN, OARnet, and KINBER) to organize activities for the next year.
- Schopf, Doyle, Balas, and Jent attended the Internet2 Technical Exchange, on October 14-19, 2018, in Orlando, Florida https://meetings.internet2.edu/2018-technology-exchange/. Schopf gave a talk on the NetSage flow dashboard and the third party deployments that are part of EPOC. She also met with representatives from NSF, GPN, NIH, and other partners to discuss EPOC project futures.
- Schopf attended the NOAA-EUMetSat Meeting at NOAA in Boulder, CO, October 22-25, 2018. Schopf discussed how to encourage international data sharing of weather data from satellites (note EUMetSat is based in Germany), and how EPOC could support them.
- Schopf attended the American Society of Tropical Medicine and Hygiene (ASTMH) Annual Meeting in New Orleans, LA, October 28-31, 2018 https://www.astmh.org/annual-meeting. Schopf met up with Chris Whalen, NIH, and researchers from Mali to explore additional support work for infectious disease research.
- Moynihan attended the LHCONE meeting at Fermilab in Chicago, IL, October 30-31, 2018, https://indico.cern.ch/event/725706/. Moynihan met with partners to get updates on LHCONE infrastructure and to provide updates on IN@IU LHCONE and science engagement activities.
- Schopf, Zurawski, Rotman, Lee, Addleman, Chevalier, Southworth, and Jent attended the SC18 Conference in Dallas, TX, November 11-18, 2018, https://sc18.supercomputing.org/. Schopf gave a booth talk on the opening night focusing on EPOC and its potential. She also held several one-on-one meetings with project partners. Zurawski was serving in the role of SCinet Chair, and Addleman was

SCinet WAN Team Deputy Chair; both worked during staging, setup, and show as well as additional meetings in 2019.

- Schopf and Zurawski held meetings with leadership and staff at the Texas Advanced Computing Center (TACC) on November 19, 2018. Discussions involved future EPOC funded workshops and a deployment of NetSage for the TACC data ingestion machines.
- Schopf and Zurawski had meetings and presented a symposium talk with the Department of Energy Office of Science on December 11, 2018 in Germantown, MD.
- Schopf and Zurawski had meetings with the National Science Foundation on December 12, 2018, in Alexandria VA.
- Schopf attended the American Geophysical Union (AGU) Fall Meeting on December 13, 2018, in Washington DC https://fallmeeting.agu.org/2018/. She touched base with several geoscience collaborators about their large-scale data transfer issues.
- Schopf and Zurawski co-hosted an EPOC Regional Networking Stakeholders workshop in Tempe, AZ, on January 14, 2019. This event allowed the EPOC leadership to set a roadmap with regional partners (iLight, KINBER, GPN, FRGP, LEARN, and OARnet) for the remainder of the year and plot out strategic programs.
- Moynihan attended the Earth Science Information Partners (ESIP) Winter Meeting in Bethesda, MD, January 15-17, 2019, https://2019esipwintermeeting.sched.com/?iframe=no. He presented a poster on science engagement and EPOC. He also met with researchers and ESIP partners to discuss networking requirements and developments in data transfer needs.
- Schopf attended the TransPacific Research Networks and the PTC '19 conferences in Honolulu, HI, January 18-23, 2018, https://www.ptc.org/ptc19/. She met with University of Hawaii collaborators to discuss EPOC futures.
- Addleman attended the WestNet regional meeting January 28-31, 2019, at Arizona State University, Tempe, Arizona, https://nets.ucar.edu/nets/ongoing-activities/Westnet/prev-mtg/201901.meeting/index.html. He gave an overview presentation and had planning discussions with this regional partner.
- Schopf, Zurawski, and Jent attended the Quilt Winter Member Meeting, in La Jolla, CA, February 4-7,2019, https://www.thequilt.net/public-event/2019-winter-member-meeting/. Schopf and Zurawski held extensive meetings with partners and led a session to help craft the description of "what a *RP session could be", at the request of Quilt leadership, which can be found at: https://docs.google.com/document/d/1sE7gdYxLMMc7QDtfGRJsAG_1lzbz2h6JSpwCZy4FhKE/edit#heading=h.t7iol3vpie4a. We also held a half day meeting with members of the Towards the National Research Platform project to discuss areas of collaboration.
- The EPOC team held an All-Hands meeting in Bloomington, IN, on February 28 and March 1, 2019. All members of EPOC were present to discuss upcoming work and plot out the Year 2 activities.
- Schopf, Zurawski, and Jent attended Internet2 Global Summit in Washington, D.C., on March 3-7, 2019, https://meetings.internet2.edu/2019-global-summit/. Schopf spoke at the Executive Session. Schopf and Zurawski led a panel on EPOC with several partners. Many side meetings with partners took place.

- Zurawski attended the CENIC Member meeting in La Jolla, CA March 18-20, 2019, https://events.cenic.org/march-2019. He gave a talk on engagement and EPOC and held side meetings with project partners.

4. Publications, Project Documents, and Presentations

Throughout the report, we reference these documents and talks by the number listed here.

4.1 Presentations

1. Schopf, Jennifer, "Scaling Science Engagement", Invited Talk, National Research Platform Second Annual Meeting, Bozeman, MT, August 7, 2018.
2. Schopf, Jennifer, "National and International Measurement and Monitoring", Invited Talk, National Research Platform Second Annual Meeting, Bozeman, MT, August 7, 2018.
3. Zurawski, Jason, and Rotman, Lauren, "Science Engagement Workshop", Invited Workshop at NOAA Networking Committee Workshop, Silver Spring, September 11-12, 2018.
4. Zurawski, Jason, and Chevalier, Scott, "perfSONAR and Data Transfer", Invited Workshop at the SOX Member Meeting, Atlanta GA, September 18-21, 2018.
5. Schopf, Jennifer M., "Supporting CI through Roadside Assistance and Proactive Monitoring, "Invited Speaker, Panel Discussion: Models for Cyberinfrastructure Support, Quilt Fall Members Meeting, College Park MD, September 24-27, 2018.
6. Zurawski, Jason, and Schopf, Jennifer, "Science Engagement Workshop", Invited Workshop at CC* PI / Quilt Meeting and Workshop, College Park MD, September 24-27, 2018.
7. Schopf, Jennifer, "Understanding Network Use with the IRNC NetSage Flow Data Dashboards", Invited Talk, Internet2 Technical Exchange, October 18, 2018.
8. Schopf, Jennifer, "The Engagement and Performance Operations Center", Booth Talk, SuperComputing'18, November 13, 2018.
9. Schopf, Jennifer, and Zurawski, Jason, "EPOC Overview", Invited talk at DOE Symposium, Germantown, MD, December 11, 2018.
10. Schopf, Jennifer, and Zurawski, Jason, "EPOC Overview", Invited Workshop at the Internet2 Connectors Meeting, Tempe, AZ, January 14-15, 2019.
11. Moynihan, Edward, "The Engagement and Performance Operations Center", Poster Presentation, Earth Science Information Partners Winter Meeting, Bethesda, MD, January 15-17, 2019.
12. Addleman, Hans, "EPOC Overview", Invited Talk at Westnet Meeting, Tempe, AZ, January 29, 2019
13. Schopf, Jennifer and Zurawski, Jason, "Quilt Session on Engagement and NRP", Invited Session at the Quilt Winter Meeting, San Diego, CA, February 5-7, 2019

14. Schopf, Jennifer "Engagement and Performance Operations Center", Invited Lightning Talk for Executive Session, Internet2 Global Summit, Washington, DC, March 5, 2019.
15. Schopf, Jennifer and Zurawski, Jason, "EPOC Overview", Invited Panel, Internet2 Global Summit, Washington, DC, March 8 2019
16. Zurawski, Jason, "EPOC Overview", Invited Talk, CENIC Members Meeting, La Jolla, CA, Mar 18-20 2019
17. Zurawski, Jason, "Overview: The Engagement and Performance Operations Center", Invited Talk, Quilt Monthly Webinar series, March 29, 2019.

4.2 Project Documents

18. Addleman, Hans, and Schlemmer, Jared, Editors, "Hawaii Pan-STARRS Data Movement Issues Summary, Roadside Assistance Writeup,March 14, 2019. See Appendix A.
19. Addleman, Hans, Editor, "Pakistan-UK LHC Data Movement Issues Summary", March 4, 2019. See Appendix B.
20. Schopf, Jennifer M, "EPOC Network Data Privacy Policy", February 19, 2019. See Appendix C.

4.3 Press Releases

21. "IU, Energy Sciences Network receive $3.5M grant to help researchers accelerate big data sharing", Press Release, June 27, 2018, https://news.iu.edu/stories/2018/06/iu/releases/27-iu-energy-science-network-receive-35-million-to-help-researchers-accelerate-big-data-sharing.html.
22. "IU part of "Transformational" Collaboration", Press Release, June 28, 2019, http://www.insideindianabusiness.com/story/38523247/iu-part-of-potentially-transformational-collaboration.
23. "OARnet partners to help researchers accelerate big data sharing", Press Release, October 5, 2018, https://www.oar.net/press/oarnet_partners_help_researchers_accelerate_big_data_sharing.

5. Project Coordination

5.A Internal Coordination

The EPOC project coordinates internally via four primary mechanisms:
- Synchronous and asynchronous communication via an email mailing list and Slack workspace;
- Project management via shared Trello (digital KanBan board) to track ongoing projects, requests, and record outcomes;

- Weekly project management calls to update the Trello infrastructure and triage new requests; and
- Twice yearly All Hands Meetings for face-to-face discussion on important strategic topics.

Standard synchronous and asynchronous methods are used for communication within the group, including Slack workspaces and email lists. Ongoing active projects, and requests to assist with roadside assistance, are discussed on a daily basis.

A Trello board is used for tracking three critical aspects of EPOC:
1. Internal project status on a number of topics;
2. Roadside Assistance and Consultation requests from the community; and
3. Training and other community-facing events.

All EPOC staff have access to this resource and are expected to make routine updates to project status. Events are archived after completion for reporting purposes.

EPOC project calls are held most weeks to review the Trello board, discuss pending community events, and go over Roadside Assistance and Consultation tickets. Meetings are scheduled weekly for one hour, but can be cancelled to facilitate scheduling conflicts. Typically, the entire team is present or provides electronic updates to critical issues.

EPOC internal partners meet twice a year at Face-to-Face All Hands Meetings to discuss strategic goals, address more complicated topics that cannot be addressed in a phone call, and to plan for the upcoming 12 months. An initial project kick-off meeting was held in July 2018 to prepare ESnet and IU partners for the award and projects that would receive top priority. In February of 2019, the first EPOC All Hands Meeting was held in Bloomington, IN. Topics discussed in detail included setting an internal roadmap for the remainder of 2019, meeting participation, our approach to Managed Services, and strategic goals for Project Year 2. EPOC anticipates holding another all hands meeting in August of 2019.

5.B Work with EPOC Partners

EPOC has three types of partners: Regional Networking Partners, who are deploying the infrastructure EPOC supports and use their members for outreach for EPOC, Infrastructure Partners, who are themselves collaborations that support a variety of cyberinfrastructure (CI) services, and Science Community Partners.

5.B.1 Regional Networking Partners

EPOC is partnered with the following regional network operators:
- iLight, the regional network for Indiana;
- FRGP/WestNet, the networks that cover the western states of Colorado, Wyoming, Arizona, Idaho, Utah, and New Mexico;

- LEARN, the regional network for Texas;
- OARNET, the regional network for Ohio;
- Great Plains Network (GPN), the regional network that serves North Dakota, South Dakota, Nebraska, Iowa, Minnesota, Kansas, Missouri, and Arkansas; and
- KINBER, the regional network for Pennsylvania.

In January of 2019, EPOC conducted a survey of the Regional Networking Partners in preparation for meeting face-to-face with representatives from each of the networks. At that meeting, we discussed strategic goals for EPOC[10], needs for each regional, as well as actions to take for Project Year 1.

Each of the Regional Networking Partners hosts at least one annual meetings with their membership, in addition to other gatherings and events that discuss specific topics. The EPOC team has made plans to attend and present material at their annual meetings in addition to other project activities. The summary of activities in collaboration with each Regional Networking Partner in Year 1 and plans for Project Year 2 are listed in Table 1.

Table 1: Project plans between EPOC team and Regional Networking Partners.

Partner	Roadside and Consulting	Deep Dives	NetSage	Managed Services	Training	Regional Meeting
iLight	Project briefed	Project briefed; Purdue planned 6/19	Flow data deployment-expected 4/19	Discussion will occur in Year 2	Project briefed	Plan to attend 5/19
FRGP/ WestNet	Project briefed; Tribal College (completed); Colorado School of Mines (ongoing)	Project briefed; NOAA 9/18	Discussion to deploy Tstat on NOAA and NCAR data archives	Discussion will occur in Year 2	perfSONAR training event with Tribal colleges Summer 2019	Attended 1/19 [12]; Planning talk 6/19; Palnnign talk Nwave 7/19
LEARN	Project briefed	Trinity University planned 5/19	Tstat on TACC completed	Discussion ongoing, PS, DT, DMZ MS	Discussing a training activity AHM	PlanningDD training and talk 6/19
OARnet	Project briefed	U Cincincati planned 4/19	Discussion will occur in Year 2	Discussion ongoing, PS MS	Project briefed	Planning talk 5/19
GPN	Project briefed	Kansas State University planned 5/19	SNMP deployment 10/19	Discussion with KanREN PS MS	Possible DD training at AHM	Planning DD workshop, and talk 5/19
KINBER	Project briefed	Arcadia University planned 4/19	Discussion will occur in Year 2	Discussion with Franklin Marshall PS MS	Possible DD training at AHM	Planning DD workshop, and talk 4/19

5.B.2 Infrastructure Partners

In the proposal phase, we identified a set of Infrastructure Partners who themselves provide services to end user scientists. Infrastructure Partners are used to leverage different kinds of support offered by each group to expand the set of services available to the broader community.

In order to stay in touch with them, EPOC has instituted an 'Ambassador' organizational structure. Each partner is assigned an EPOC staff member as a primary point of contact, whose responsibilities include:
- Being the initial point of contact for partner on any EPOC matters
- Providing a quarterly update on project status
- Working with partners to coordinate any direct EPOC engagement activities (training, talk requests, roadside assistance)

In the event of activities cross into areas that other EPOC staff are more versed in, those staff are pulled into the interactions as well. However, the EPOC Ambassador will track activities and bring in resources as needed.

The current set of Infrastructure Partner organizations includes:
- The Campus Research Computing Consortium (CaRCC), a consortium of over 30 campuses that facilitate access to cyberinfrastructure. Schopf is the main contact for this group. She is currently tracking many of their mailing lists, and participated in their Cyberinfrastructure Ecosystem workshop in St Louis in April, joint with CNI.
- The NSF Cybersecurity Center of Excellence (CCOE), which supports cybersecurity for NSF funded projects. Addleman is the main contact for CCOE. Schopf has also been in contact with their lead, Von Welch, and plans to speak in a joint panel being planned for the PEARC'19 meeting in July.
- Internet2, which supports solving common technology challenges for their over 200 educational, research and community members, Schopf is the main contact for this organization, and she has held multiple meetings with various staff members to ensure EPOC is understood across the organization, as well as presenting at their member meetings [7,14,15]. Furthermore, EPOC leveraged this partnership in co-locating the January regional meeting with their Tempe stakeholder meeting[11].
- The Quilt, which provides a central organization for networks to share the best practices to support end user science. Zurawski is the primary contact for the Quilt, and has regular meetings with their organization. EPOC staff attended both the Fall and January Quilt Members meetings and presenting at them [5,6,12]. A joint meeting with TNRP was co-located with the February Quilt meeting. Zurawski also gave an invited talk on EPOC as part of their monthly webinar series [17].
- The Science Gateway Community Institute (SGCI), which provides best practice recommendations and support for scientists building and using data portals. Moynihan is the contact for this group and will be attending their meeting in the fall.
- The Extreme Science and Engineering Discovery Environment (XSEDE), which supports a single virtual system and expertise through the Campus Champions. Southworth is the main contact for this group, however in Year 1, Schopf meet with members of the XSEDE leadership team at SC'18, as well as the head Tim Boerner, head of the Data

Transfer Services and various members of the campus champions team. Schopf and Zurawski plan to attend the XSEDE/PEARC meeting in July.

5.B.3 Science Community Partners

In the proposal, we also identified a set of Science Community Partners, each of which consisted of a collaboration of scientists, enabling us to scale our reach to larger community groups. Ambassadors have been assigned for these partners as well. The partners include:

- The Earth Science Information Partners (ESIP), a consortium of over 180 members that provides a forum for the Earth science data and technology community. Moynihan is the primary contact point for this collaboration, and he attended their winter meeting in January, and presented a poster that gave an overview of EPOC[11], and spoke with several groups about their data transfer needs. He plans to attend their Summer Meeting as well.
- The World Climate Research Programme's International Climate Network Working Group (ICNWG), which supports thousands of scientists through using the Earth System Grid Federation's (EGSF) globally distributed climate data repository sites. Zurawski and Eli Dart, LBNL/ESnet, share the contact point for this group. The group has sporadic activity that centers around yearly meetings (typically December) and was a part of the DOE Petascale DTN effort in 2018.
- The IU Grand Challenge Precision Health Initiative, which works with a broad set of precision health applications. Schopf is the primary contact for this team, which experienced a leadership change in 2019 when Bill Barnet left IU to become the Senior Director of Research Computing, at Harvard Medical School. Contact has restarted via the IU Research Technologies team, and we expect to meet with them early in Year 2.
- The University of Hawai'i System Astronomy Community, which supports 15 facilities with hundreds of researchers and experiments every year. Southworth is the primary contact for this group. Addleman oversaw one of our first deep dives with this team[18], examining data flowing between the University of Hawai'i and Johns Hopkins, in Baltimore, Maryland. Schopf also had meetings with the IT lead for this consortium during the year.
- The Midwest Big Data Hub (MBDH), which supports the use of data for a variety of applications and end users across twelve states. Schopf had been the lead contact for this team when Melissa Cragin was the Executive Director. However, Cragin is now at SDSC, and we are re-establishing contact with the new leadership team. Southworth will be the primary contact going forward. We plan to attend their October 2019 All Hands Meeting and present a poster on EPOC.
- The Open Storage Network (OSN), which will support dozens of applications across a broad set of application domains. Southworth is the main contact for this group. This group is still getting organized themselves as they started around the same time that EPOC did. EPOC plans to attend their October AHM in Austin, TX.

5.C External Partners

In addition to the partners that were named in the proposal, the EPOC team is coordinating with a number of additional groups, including the "Toward the National Research Platform" (TNRP) - NSF award #1826967, setting up an external advisory board, and managing various requests from external partners.

5.C.1 Coordination with TNRP

In the original solicitation, NSF expected to make only one award for the Network Performance Engineering and Outreach (NPEO) portion of the CC* solicitation, however two were made. Therefore, we have made sure to coordinate with the other award, "Toward the National Research Platform" (TNRP), NSF award #1826967, led by Larry Smarr and Tom Defanti. EPOC members attended and presented at the National Research Platform meeting in Bozeman [1,2]. Several additional meetings were held by phone and in person, and a coordination document was submitted to NSF - see Appendix D.

In practice, although the goals of the two projects are similar, the day-to-day work by the two groups differs strongly. In general, the TNRP project is more focused on advocacy to move towards container orchestration solutions to enable the use of more advanced network and compute infrastructures, including but not limited to the PRP Nautilus cluster. Their primary science use cases to date (http://pacificresearchplatform.org/nautilus/namespaces/) tend to be associated with the Pacific Research Platform (PRP) or an R1 institution. EPOC is focusing more on working with application scientists in their current environment and debugging existing infrastructure with limited changes to the underlying technologies. Our service is available to anyone working in research or education, not just R1 institutions. We are also promoting the use of advanced infrastructure in a supported way for organizations that don't have the in-house capability to offer these services. Our focus is more strongly on underserved (non-R1) institutions and small to medium sized research collaborations.

We plan to continue meeting with the TNRP staff at regular intervals to ensure the cohesiveness of the two activities.

5.C.2 External Advisory Board

In the proposal, we stated that we would establish an External Advisory Board (EAB) for the project that would meet twice a year to discuss the project metrics and deliverables. This group was outlined in the proposal to consist of five members: two with a specialization in networks or CI, two focused on application end users, and one with a specialization in project management. We have expanded that list to include the following:
- University Networking, Computing, and User Support - Douglas Jennewein
- University Computing -Trey Breckenridge
- Regional Networking - Bill Owens
- End-user Software (Scientific) - Frank Wuerthwein

- End-user Software (Networking) - Jason Leigh
- Outside Representation from Indiana University - Robert Heschel
- Outside Representation from ESNet - Kate Mace

Our first EAB meeting will take place early in Year 2.

5.C.3 Requests for Additional Collaboration

Over the course of the first year of the project, we have been approached by numerous groups requesting support and partnership with EPOC outside of those that we originally collaborated with. Towards this end, we wrote ten letters of collaboration in support of NSF programs including the CC* solicitation (NSF 19-533), the MidScale RI-1 program (NSF 19-542), and AccelNet (NSF 19-501).

At the TechEx meeting in October, we received an urgent request from Dale Smith and Gil Gonzalez to assist the tribal colleges they were working with through support for a perfSONAR deployment. We shipped them 6 NUC5CPYH small perfSONAR nodes and USB keys with installation instructions, and have been told that the full deployment is still being worked on. This effort will continue in Year 2.

Several non-partner groups requested participation in their events by EPOC staff, which whenever possible we tried to accommodate. These included NRP[1,2], SOX[4], EUMetSat, NIH, LHCOne, and CENIC[17].

In Year 1 we were also approached by several additional regional networks with an interest in EPOC, which will be evaluated in Year 2. These include:
- OSHEAN
- LONI
- NYSERNET
- UEN
- SOX

5.D Additional Startup Activities

5.D.1 General Communications and Marketing Planning

EPOC is engaged with the Indiana University Information Technologies Communications Office (ITCO), the communications arm for the Office of the Vice President for Information Technology (OVPIT). We are working with them on ways to better interact with the target user communities. Goals including sharing updates on progress, more broadly, as well as advertising services, events, and future directions. Early work has focused on defining a web and social media presence, as well as preparing marketing collateral for distribution at community events.

5.D.2 Mailing List Dissemination

EPOC will utilize a series of mailing lists to share updates on project status:
- Partners Mailing List: Regular updates sent to EPOC partners
- Community Newsletter List: opt-in mailing list for partners and other entities to receive updates on EPOC services

5.D.3 Web and Social Media Presence

EPOC uses a web site (http://epoc.global) to share several content items:
- Basic information about the project and its partners
- List of upcoming and previous training events, and materials from prior events
- Reports from Deep Dives and Roadside Assistance Cases
- Documentation on CI technology

5.D.4 YouTube Content Delivery

EPOC is working with the CI Engineering Community (https://www.es.net/science-engagement/ci-engineering-brownbag-series/) to catalog the presented webinars from the Brownbag series held from 2017 to present. It is expected that the approximately 70 talks will be uploaded in the spring of 2019 and shared with the R&E community.

5.E Project Coordination Plans for Year 2

In Year 2, we plan to continue the coordination activities we have started in Year 1. We will follow up with our listed partners according to the plans we have developed with them (detailed in Table 1). In addition, we plan to extend our outreach mission to include meetings such as:
- Talk submitted for CNI (April 2019)
- Invited to attend Research Computing and Data Ecosystem Workshop, run by CARCC joint with CNI (April 2019)
- Plan to attend and submitted talk to TNC (May 2019)
- Invited to present at NOAA/NWave summer meeting (July 2019)
- Plan to submit a poster and attend eScience 2019 (September 2019)
- Plan to attend the CC* PI, Fall Quilt Meeting, and NRP3 meeting (September, 2019)
- Plan to attend NSF Cybersecurity Summit for Large facilities and Cyberinfrastructure - (October 2019)
- Plan to attend SuperComputing'19 (November 2019)
- Talk submitted and plan to attend Internet2 Technical Exchange (December 2019)
- Plan to attend and speak at Quilt Winter Members Meeting (February 2020)
- Plan to submit a talk and attend Internet2 Global Summit (March 2020)

We will also be trying to expand our outreach to include work with Science Node (http://sciencenode.org) to publish our success stories, to provide additional support for the use of PerfSONAR for Tribal colleges, and to expand the talks available on the YouTube channel.

6. RoadSide Assistance and Consulting

A key aspect of the EPOC project is the process pipeline for immediate help, referred to as Roadside Assistance. Based on our previous experience with performance issues, the Roadside Assistance approach helps collaborators when data sharing failures occur, since these almost always involve multiple domains and organizations. Over the course of the year, we also received a number of quests for consultation on network engineering aspects, so we rolled that service into this activity. Hans Addleman is the lead for this effort.

6.A Roadside Assistance Overview

In Year 1, we fully developed the Roadside Assistance process, and documented it both internally and externally (available at https://epoc.global/wp-content/uploads/2019/04/Roadside-Assistance-Description.pdf). A Roadside Assistance case can be submitted by email to epoc@iu.edu by anyone associated with either research or educational work who is having difficulties transferring or receiving files with a collaborator at another site. The help request is triaged within 24 hours, basic tracking infrastructure is set up, and a Case Manager and Lead Engineer are assigned.

The infrastructure we are using for this process includes not only a formal ticketing instance within the EPOC project, but a shareable online folder for each case that is made accessible to all engineers working on the case. A shareable folder outside of a ticketing system is required because most performance issues involve more than one organization, and most ticket systems are accessible only to the local users. Within this folder is also a Customer Case Document that is written in language accessible to the submitter and updated frequently to enable the submitter to always be able to access the current status of their issue.

The Case Manager is the primary point of contact for the submitter, and is generally someone who has experience talking with end user scientists or other community members. The Case Manager will update the submitter on a regular basis, through both email and the always-available Customer Case Document, stating what is happening with the issue and what the next steps towards resolution are. The Case Manager also loops in the campus, regional, or national network contacts relevant to the issue.

The Lead Engineer is the assigned staff member who will coordinate the technical aspects of the problem. This often includes involving a set of external engineers, organizations, or network operations centers during the investigation. The Lead Engineer tracks these interactions, any opened ticket numbers, and all associated documentation for the case, making sure the data is copied to the shared folder and is made accessible to the full team debugging the problem.

In general, troubleshooting of these Cases involves understanding the end-to-end path of the data transfer, the use public test and validation services (such as perfSONAR nodes and router proxies) to identify potential issues, and coordination with a wide set of engineering help staff along the data transfer path.

14

Each Case is unique and some of the issues found during an investigation may require more than simple engineering fixes. Training, capital expenditure, network architecture redesign, or policy changes may be required to achieve effective file transfers. If longer term or larger scope issues are identified, the Case Manager will discuss these with the customer and the relevant support staff, and describe possible next steps with help from EPOC. This may include EPOC-led Deep Dives, Managed Service offerings, or larger scale interactions with other organizations. At the end of each Case, a 2-page summary is written up and made available online.

6.B EPOC Consulting Overview

In addition to the more in depth assistance offered by the Roadside Assistance process, EPOC also offers a consultation service to help engineers and management staff make informed decisions about network services and deployments. These include, but are not limited to, questions related to network upgrades, Science DMZ's, data stores, data movement practices, perfSONAR deployments, and other related topics.

Different EPOC staff are engaged as needed for consultations depending on the subject matter expertise required. In addition, if there is an EPOC partner who may be valuable to a consultation they will be engaged as well. For example, we would engage the CCOE regarding security and policy questions or SGCI for questions involving portals.

Consultations can be made via email, video conference, or in person. Consultations may lead into other EPOC offerings such as Deep Dives, Trainings, or Managed Service deployments. Whenever possible, information from a consultation is added to online resource, primarily http://fasterdata.ed.net, for broader dissemination of the information.

6.C Roadside Assistance Cases for Year 1

In Year 1, we had two completed Roadside Assistance cases, one with the Pan-STARRS astronomy project and the other in support for the Large Hadron Collider project. We have one ongoing Roadside Assistance case with Washington State University.

The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) program shares approximately 100 terabytes of data yearly between the Institute for Astronomy (IfA) at the University of Hawaii (UH) and the Space Telescope Science Institute (STSCI) at John Hopkins University in Baltimore, Maryland, to enable researchers to more accurately estimate galaxy redshifts, improving their understanding of the local cosmic expansion and dark energy. They approached the Indiana University (IU) teams for assistance because they were experiencing a maximum transfer rate of only 320 Mbps, despite the fact that they believed the full path between IfA and STSCI was equipped primarily with 100Gbps networks. They hoped to achieve transfer rates in the multi-gigabit range, and looked for help debugging the path. We found multiple 1Gbps network components, buffering problems, MTU mismatches, poorly

placed archives, and other issues. The file transfer was fully re-evaluated in June of 2018 and after our engagement saw a three-fold jump in overall performance seeing a sustained 1Gbps transfer rate, a three-fold improvement. The results of this engagement also led the PIREN project to receive a supplemental National Science Foundation (NSF) award to enhance the capabilities of their overall network, Science DMZ, data transfer hardware, and network testing hardware. They are currently working to procure, design, and put this new architecture in place. The writeup for this case is available in Appendix A.

The National Center for Physics (NCP) at the Quaid-i-Azam University Campus in Islamabad, Pakistan, is a Tier 2 Large Hadron Collider (LHC) Site. NCP contacted EPOC in September, 2018, for assistance with an ongoing performance issue when moving data to and from the GridPP (www.gridpp.ac.uk) Tier 1 site at the Queen Mary University in London. As a Tier 2 LHC site, NCP is responsible for downloading data sets from the Tier 1 site to share with LHC researchers in the region in a timely manner. NCP has a 1Gbps connection to their national R&E network, Pakistan Education and Research Network (PERN), however, they were seeing transfer rates as low as 40Mbps to some Tier 1 LHC sites. These performance problems were intermittent, and have been ongoing for at least a year prior to contacting us. Our engagement with NCP resulted in identifying five main issues, with a performance improvement to ~480Mbps. The writeup for this case is available in Appendix B.

At the end of Year 1, we had one ongoing roadside assistance case with Washington State University in Pullman, Washington. The network engineering team contacted an EPOC staff member because they were having issues achieving consistent high bandwidth file transfers to various remote locations. EPOC is engaged and working with them to baseline their wide area network connections and achieve a consistent transfer rate.

6.C Consultation Cases for Year 1

In Year 1, we had four completed and two ongoing consultations. The primary topics were Science DMZ architectures, data transfer strategies, and perfSONAR setup and deployment.

Completed consultations included:
- Tribal Colleges: Gil Gonzales, former CIO of the University of Mexico, is now consulting for a number of Tribal colleges to upgrade their networks. He contacted EPOC staff to find material related to perfSONAR, ScienceDMZ, and data transfer architectures to share with his partners. EPOC staff engaged via email and video conferencing and discussed these issues in depth. EPOC staff encouraged concentrating on 1G and 10G nodes and network paths. EPOC staff also recommended finding science drivers at the colleges going forward.
- American Museum of Natural History (AMNH): Zurawski attended a meeting between TrustedCI/CCOE, EPOC, Energy Sciences Network (ESnet), and NYSERNet to discuss network design and security concerns that staff at AMNH had around ScienceDMZ architectures.

- University of Florida (UFL): Chris Griffin met with Zurawski and Dart to discuss ScienceDMZ architecture, and firewalls internals, as well as the possibility of improving the research and enterprise networks for UFL. They discussed new network hardware advancements and what performance could be expected with their type of load.
- Compute Canada: Zurawski discussed best practices for securing a ScienceDMZ at 100G line rates. Zurawski provided information on Lawrence Berkeley National Labs (LBL) work with Bro/Zeke as an Intrusion Detection System (IDS).

Ongoing Consultations at the end of Year 1 include:
- Colorado School of Mines: This consultation is still ongoing and includes discussions of perfSONAR, top of rack switches, Science DMZ architecture and security, and firewall hardware. We are also discussing how best to engage with researchers on campus.
- New York University School of Medicine: NYU staff are in the process of engaging with their researchers to identify use cases. They have no ETA but have invited EPOC to review the use cases and provide feedback when they are done.

6.D Metrics

Table 2: Metrics for Roadside Assistance and Consultation activities.

Case Identifier	Type	Start Date	Area of request	Actions taken	Asked by: Eng, Scientist, other	Primarily R(ch), E(du), O(ther)	Size: Large, Med, Small
UH Astro	RA	Nov 17	UHawaii-JHU file transfer issues	Fixed routes; reconfigured equip, firewalls, and ACLS; upgraded full path to 10G: 3X improvement	Eng	R	Large
NCP-LHC	RA	Oct 18	Poor performance Pakistan-London	Relocated archive, fixed routes; found rate limiter; congested backbone still to be addressed with TEIN: 10X improve	Eng	R	Med
Tribal Colleges	Con	Jan 19	DMZ, PS, DTN, NW upgrades, Sci eng	Advised on upgrade possibilities	Other	E	Small
AMNH	Con	Feb 19	DMZ, Security	Consultation on design best practices	Eng	O	Small
UFL	Con	Feb 19	DMZ, Network upgrades	Performance expectations	Eng	R	Large
CC	Con	Mar 19	IDS	Security best practices	Eng	O	Small
CSM	Con	Oct 18	DMZ, PS, HW, Sci eng	Design best practices	Eng	R	Small
NYU Med	Con	Oct 18	Sci eng	Best practices	Eng	R	Large

6.E Roadside Assistance and Consultation Plans for Year 2

The Roadside Assistance team will continue engagements in Year 2. The team is expected to continue to receive requests assistance and consultation at meetings when an EPOC staff member presents. We also continue to engage our Deep Dive customers via ongoing consultations. We are currently developing additional communications materials to share with our partners to make sure that the Roadside Assistance and Consulting services are broadly known across the community, and will be speaking on these topics numerous times in different settings in the upcoming year.

7. Deep Dives

Deep Dives aim to understand the full science pipeline for research teams and suggest alternative approaches for the scientists, local IT support, and national networking partners as relevant to achieve the long term research goals via workflow analysis, storage and computational tuning, and identification of network bottlenecks.

7.A Overview

The Deep Dive approach is based on an almost 10-year practice used by ESnet to understand the growth requirements of DOE facilities (https://fasterdata.es.net/science-dmz/science-and-network-requirements-review/). We have adapted this approach for work with individual science groups, which is documented at: https://epoc.global/wp-content/uploads/2019/04/Application-Deep-Dive-Description-1.pdf. Jason Zurawski is the lead for this area.

The Deed Dive discussion, which takes place through a set of structured data-centric conversations and questionnaires, includes:
- Background information, including a brief overview of the science being performed and where it sits within its discipline.
- Identification of collaborators - both people and institutions that a science group interacts with.
- How the team is currently using both local and remote scientific instruments and facilities.
- The process of science, which includes explicitly explaining 'a day in the life' of the science group. This will tie together the people/collaborators, the use of resources, and the science goals for the team.
- A description of the software infrastructure that is supporting the research goals.
- Details about the network and data architecture between the collaboration sites, as part of the science workflow.
- The use, if any, of cloud services.

18

- A discussion of any outstanding issues and pain points.

At the end of the interaction with the research team, EPOC staff can come away with a good understanding of the research, data movement, who's using what pieces, dependencies, and time frames. This enables us to identify possible bottlenecks or areas that may not scale in the coming years, and to point the research team towards existing resources so that they can reach their goals more effectively. After the Deep Dive, we follow up with the requesting team to check on any changes or questions. We also publish a report containing the full information about the Deep Dive when possible.

7.B Year 1

EPOC completed two Deep Dive activities in 2018, and started planning seven more.

7.B.1 Completed Application Deep Dives

EPOC completed two Application Deep Dives in 2018, both of which were aimed at demonstrating the techniques used in the approach for an audience. These were held at community meetings, and featured volunteer researchers that wanted to be a part of the process.

7.B.1.A NOAA and NASA

The NOAA networking group, operators of N-Wave, requested that EPOC stage a training event during their semi-annual meeting, wherein members of EPOC would instruct an audience of engineering support staff in the methods to properly conduct a Deep Dive. The example used was from Richard Barry, NASA, and involved weather-related research that used data from telescopes and satellites. Approximately 20 NOAA engineers attended the event, which was held in Silver Spring, MD, on Sept 11-12, 2018 [3].

7.B.1.B University of Maryland

The CC* PI meeting program committee requested that EPOC stage a Deep Dive training event during their annual meeting. The example used was from Geoff Ryan, University of Maryland, who was researching black holes and handling astronomy data gathered from NASA resources. Approximately 50 PIs and engineers attended the event, which was held in College Park MD on Sept 24, 2018 [6].

7.B.2 Planned Application Deep Dives

Deep Dive planning typically takes a series of meetings and conversations over several months with the target institutional leadership and research community. After the event, the EPOC team, joint with the participants, produces a report of the events that can be used by the campus and/or regional network to influence future directions of technology support. We have begun planning seven additional Deep Dives to take place in Year 2.

7.B.2.A KINBER and Arcadia University

KINBER requested that EPOC stage a Deep Dive training event during their annual meeting, using an example from Arcadia University related to bioinformatics research. This event is scheduled for April 3, 2019, and will take place in Philadelphia, PA, at KINBERCON, the annual meeting for the KINBER regional network. The final report is expected in the Summer of 2019.

7.B.2.B OARnet and University of Cincinnati

OARnet, with member institution University of Cincinnati (UC), reached out to EPOC in January of 2019 to schedule a Deep Dive at UC to focus on several campus drivers, including:
- High Energy Physics
- Medicine and Bioinformatics
- Mathematics
- Aerospace
- Criminal Justice

The campus wants to take a broad view of data needs for this cross section to be able to inform campus leadership about pending infrastructure needs. This event is scheduled for April 26, 2019, and will take place in Cincinnati, OH, on the campus of the University of Cincinnati. The final report is expected in the Summer of 2019.

7.B.2.C GPN and Kansas State University

The Great Plains Network requested that EPOC stage a Deep Dive training event during their annual meeting, with a researcher from Kansas State University discussing their research related to agronomy. This event is scheduled for May 21, 2019, and will take place in Kansas City, MO, at the GPN Annual Meeting. The final report is expected in the Summer of 2019.

7.B.2.D LEARN and Trinity University

LEARN, with member institution Trinity University, reached out to EPOC in March of 2019 to schedule a Deep Dive that focuses on several campus drivers, including:
- Geology
- Classical Studies and Archaeology
- Computer Science
- Biology
- Physics and Neuroscience

The campus wants to take a broad view of data needs for this cross section in order to better inform campus leadership about pending infrastructure needs. LEARN and Trinity are pursuing an NSF grant, and this information will help inform the campus and network where to focus immediately if funding is awarded. This event is scheduled for May 29, 2019, and will take place in San Antonio, TX, on the campus of Trinity University. The final report is expected in the Summer of 2019.

7.B.2.E Purdue University

Purdue University reached out to EPOC in February of 2019 to schedule a Deep Dive that focuses on two emerging use cases that are users of high-performance computing resources on campus - both from the college of agriculture with a focus on biology. By performing this review in front of a number of research and IT staff, the university hopes to translate the lessons so that the event can be duplicated once EPOC completes the initial work. This event is scheduled for May 31, 2019, and will take place in Lafayette, IN, on the campus of Purdue University. The final report is expected in the Summer of 2019.

7.B.2.F University of Wisconsin

The University of Wisconsin reached out to EPOC in March of 2019 to schedule a Deep Dive that focuses on several campus drivers, including:
- High Energy Physics
- Space Sciences (including support for several NASA and NOAA missions)
- IceCube and Polar Studies
- Bioinformatics
- High Throughput Computing

The campus wants to take a broad view of data needs for this cross section, in order to better inform campus leadership about pending infrastructure needs. This event is scheduled for June 17-19, 2019, and will take place in Madison, WI, on the campus of the University of Wisconsin. The final report is expected in the Summer of 2019.

7.B.2.G PEARC 2019

EPOC submitted a request to do a Deep Dive training at PEARC 2019 in Chicago Illinois on July 29, 2019. If accepted the University of Chicago will participate as the research profile by discussing some research related to bioinformatics and the final report would be expected in the Summer of 2019.

7.C Metrics

Table 3: Metrics for Deep Dive activities.

Date	Audience	Appl name	Offered or requested	Public or private?	Issues identified	Follow up
Sept 11-12, 2018	NOAA Engineers Researchers	NASA Exoplanets	Req	Private	Local connectivity, data transfer, storage, computational power	NASA improved local connectivity; Funding needed to support storage, computation, or data transfer problems.
Sept 24, 2018	CC* Awardees	Simulating Black Hole Accretion Disks	Req	Public	Local connectivity, data transfer	UMD upgraded lab connectivity; Data transfer could not be improved due to far end tool usage

7.D Deep Dive Plans for Year 2

EPOC has three main focus areas for Deep Dives in 2019:
1. Executing planned events
2. Completing and publishing reports
3. Scheduling additional events

7.D.1 Executing Planned Events

The following events are scheduled to take place in Year 2:
- KINBER and Arcadia University, April 2019
- OARnet and University of Cincinnati, April 2019
- GPN and Kansas State University, May 2019
- LEARN and Trinity University, May 2019
- Purdue University, May 2019
- University of Wisconsin, June 2019
- PEARC 2019, July 2019

7.D.2 Report Generation

The following events will result in a final report to share with the sponsoring institutions, and will execute beyond the period of this report:
- KINBER and Arcadia University, April 2019
- OARnet and University of Cincinnati, April 2019
- GPN and Kansas State University, May 2019
- LEARN and Trinity University, May 2019
- Purdue University, May 2019
- University of Wisconsin, June 2019

All reports will be publicly posted to the EPOC web site, along with the training materials.

7.D.3 Additional Events

EPOC has been approached to perform Deep Dives for other regional network partners and campuses, including:
- LEARN and Baylor University (focusing on HPC and medical use cases)
- NERO (the regional network for Oregon) and the University of Oregon and Oregon State University
- NYSERnet (the regional network for New York) and the medical use cases of New York City.

We expect to space these out over the last half of Year 2 and into Year 3 as needed.

8. NetSage Deployments and Analysis

Understanding application performance and network measurement are two sides to a single coin - one doesn't make sense without the other. The EPOC project uses the NetSage tool (http://www.netsage.global) to collect and evaluate common network measurement data. The initial NetSage software was developed and deployed on the NSF-funded international networks. It was meant to work with sparse, international circuits, with circuit owners and operators as the end user. EPOC has expanded the use of this software to more densely defined networks and supports additional analysis and visualization aspects.

8.A Overview

The NetSage testpoint, which gathers the data used in the system, is a collection of software and hardware elements that include passive and active collection techniques. The testpoints communicate with an archiving framework that makes use of an analysis and visualization engine based on Grafana.

The test points collect data from Simple Network Management Protocol (SNMP), an application–layer protocol defined in RFC1157 for collecting and organizing information about managed devices on IP networks, which is used by NetSage dashboards in a simple map display as well as to supply basic bandwidth and uptime information in graphs and as a heatmap.

The second dataset collected by the testpoint makes use of perfSONAR to collect active measurements for throughput, latency, and loss. This data is displayed by NetSage as heatmaps, but more importantly, these measurement points are a fundamental tool for the EPOC Roadside Assistance process to be able to evaluate user-level performance between sites.

The third data set includes passive TCP flow statistics that are de-identified before archiving. We also collect this data on some domain science archives to gather data about specific transfers directly from the source. A variety of NetSage dashboards use flow data to show transfers over a network, to or from an organization, or between two specified endpoints.

The testpoints communicate with an archiving framework that consists of a Time Series Data System archive and the ELK Elastic Stack archive. Data sets can be accessed from these archives directly or through a variety of analytical tools and visualizations that uses the open source tool Grafana, which provides a wide variety of basic analysis and visualization widgets, can interact with multiple backend data sources, and can also provide authenticated access as needed.

Currently, the bulk of the development for the NetSage suite of tools is supported by the NSF IRNC NetSage project, as is the deployment of NetSage on archives domestically. EPOC is only funded for minor adaptations and the domestic deployments associated with the regional

partners, but is supplying significant feedback on the use of the tool as well as expending the NetSage collaborative partnership.

8.B Year 1

We had two test deployments in Year 1 for NetSage for circuits and are inconversations to deploy it on several archival sites. With the Great Plains Network, we deployed the NetSage SNMP dashboard for their associated circuits, as shown in Figures 1, 2, and 3. This is available online at http://gpn.netsage.global. This work was presented at the Internet2 Technical Exchange in October. Discussions are ongoing to extend the deployment to include flow data collection.

With the iLight team, we are in the process of deploying the flow data dashboards to support their collaboration. This will involve getting data from the five Indiana Gigapop-supported routers and setting up flow dashboards both for the full network and to be sorted by organization. We expect to begin collecting data in April, and will be presenting this work at the iLight meeting in May.

The Archive site deployment is funded by the NSF IRNC NetSage project, but will be strongly used by the various EPOC partners. These deployments include:
- TACC: The Tstat software was deployed on four TACC archive head nodes in January, and a dashboard for this information was made public in February: https://portal.netsage.global/grafana/d/mNPduO8mz/flow-data-for-data-archives?orgId=2&var-min_rtt=0&var-dtn_sensors=TACC%20Stampede2%20dtn1%20Tstat&var-dtn_sensors=TACC%20Stampede2%20dtn2%20Tstat
- The University of Hawaii Astronomy group deployed a temporary prototype to gather Tstat information from its DTN to meet an internal deadline, and plan to have a permanent installation in Year 2.
- NOAA is discussing deploying Tstat on one of their DTN sites in Boulder, CO. We expect this to be live before the WestNet meeting in June.
- The National Center for Atmospheric Research group has agreed to deploy Tstat at their site in Wyoming. We expect this work to be completed and available during Year 2.
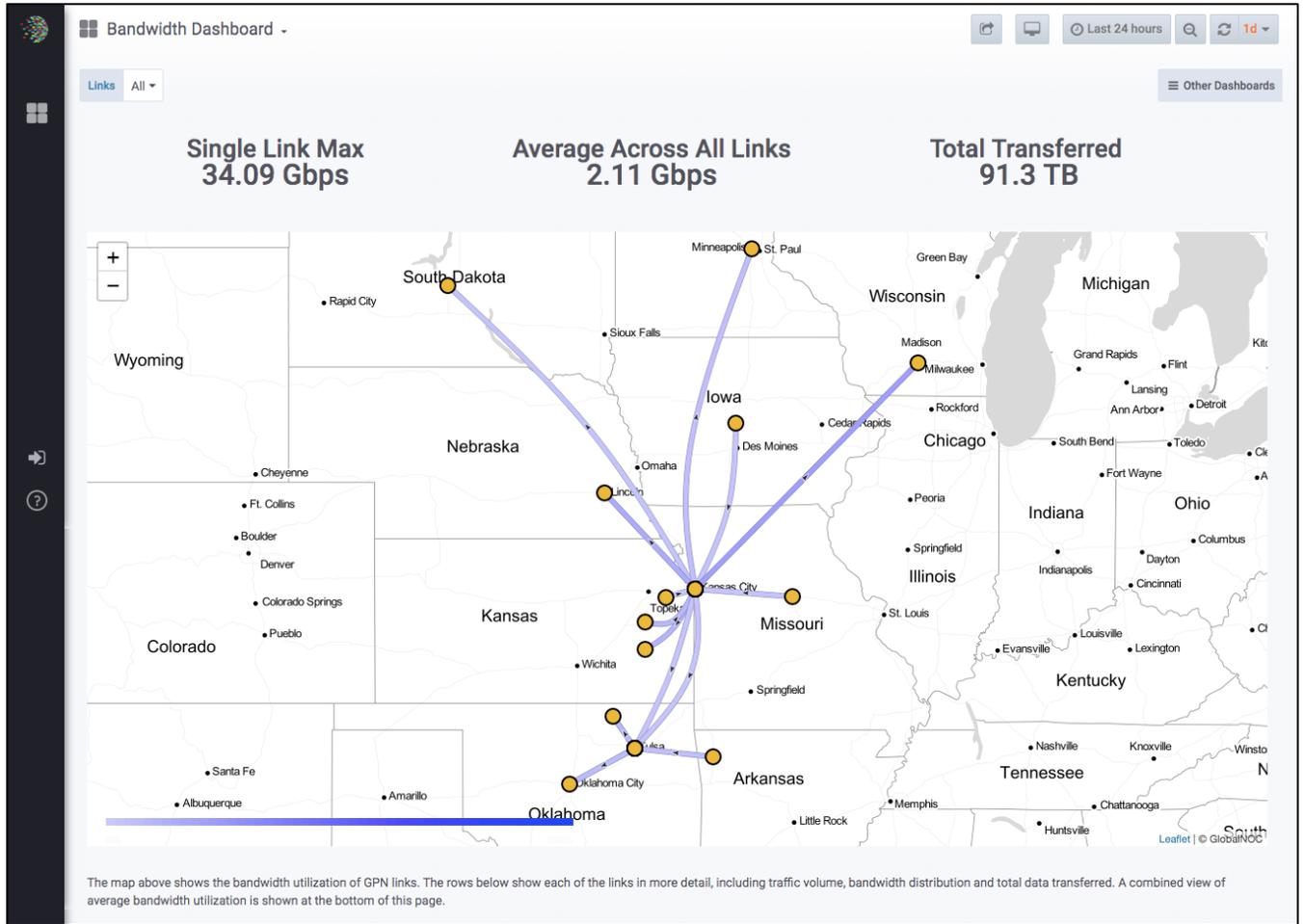
Figure 1: Snapshot of the NetSage SNMP deployment for GPN, available online at http://gpn.netsage.global.
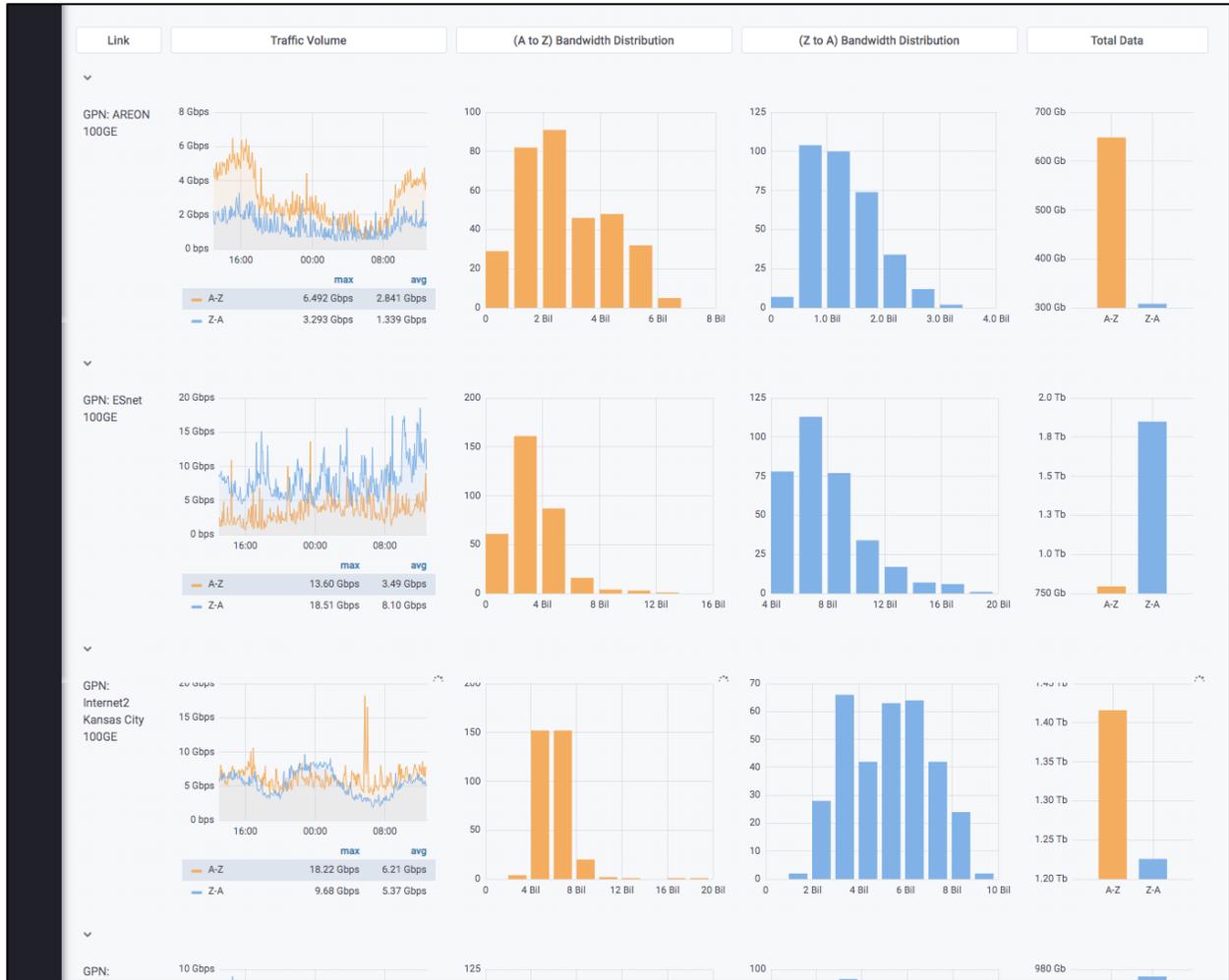
**Figure 2: Screen shot of indivual network data sets for GPN from NetSage deployment.**

26
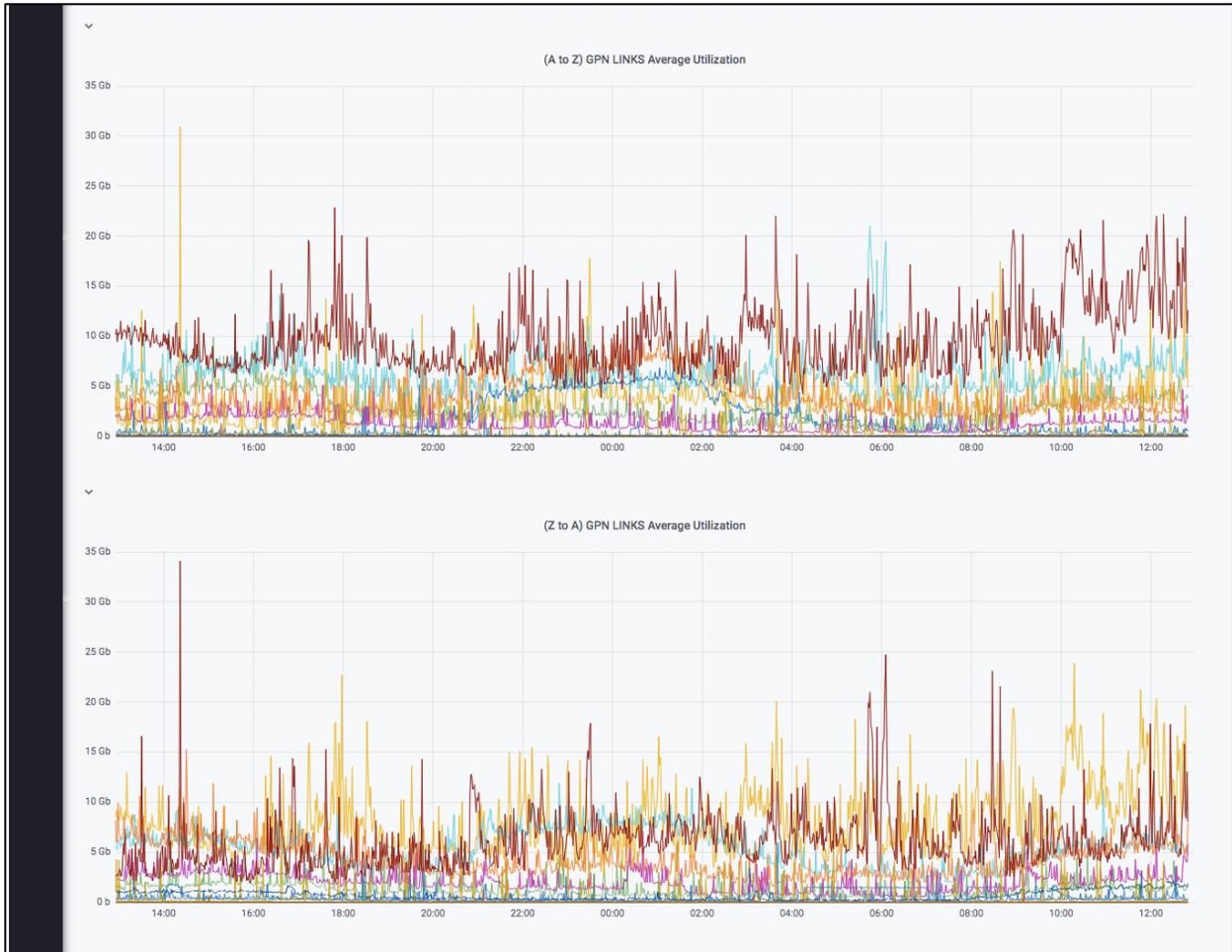
Figure 3: Screen shot of comparitive graphs for GPN NetSage deployment.

8.C Metrics

Table 4: Metrics for the NetSage deployment activities.

Which Regional	What	Website	Date Live	What data	How many devices monitored
GPN	SNMP Dashboard	http://gpn.netsage.global	9/18	SNMP	2 routersX
iLight	Flow Dashboard	To be http://ilight.netsage.global	4/19 (Expected)	Flow	5 routers
LEARN/ TACC	Archive Dashboard		2/19	Tstat	4 head nodes

8.D NetSage Plans for Year 2

In Year 2, we plan to move forward with NetSage deployments for the other regional networks. In collaboration with the IRNC NetSage project, we will also be deploying updated and new dashboards for the data that is already being collected. As we visit the regional networking

27

meetings over the next year, we expect that we will be doing extensive use case collecting, and feeding that data back to the IRNC NetSage project for their development team to consider this work.

9. Managed Services (aka "In a Box")

EPOC is tasked with developing a set of service definitions for common R&E infrastructure components. The goal of these definitions is to provide guidance for our Regional Networking Partners to implement, maintain, operate (potentially for a fee), and upgrade the service as a benefit for downstream connectors. In doing so, the costs associated with design, specification, and installation can be ameliorated for a larger population that would otherwise be reticent to adopt technology due to the burdens of entry which may include not having knowledgeable staff or enough compelling use cases to invest time and money.

9.A Overview

EPOC is targeting four use examples of Managed Services:
1. *perfSONAR*: a widely-deployed test and measurement infrastructure that is used by science networks and facilities around the world to monitor and ensure network performance.
2. *Science DMZ*: Dedicated network infrastructure specifically configured for the security and performance implications required for scientific use cases.
3. *Data Transfer*: PC-based Linux servers built with high-quality components and configured specifically for wide area data transfer.
4. *Network Testset*: specialized hardware used to provision and validate network infrastructure.

There are several known barriers to adoption of these technologies:
1. *Initial Cost of Hardware*: Hardware (servers, routers, switches, etc.) can be an expensive investment given typical drivers that consume resources. Cost is a factor for a medium to small institution that struggles to maintain basic IT infrastructure for mission critical applications.
2. *Expertise Gap*: Knowledgeable staff that are capable of designing and operating infrastructure are often in short supply. The demands on these staff include, but are not limited to installation,operation, ongoing maintenance, data analysis, and troubleshooting both local and external problems.
3. *Lack of Compelling Use Case*: Smaller institutions may not have more than a handful of use cases that involve data-intensive scientific research across a wide area connection.
4. *Inertia*: The typical designs of campus are sufficient for 99% of the users, thus the addition of new technology for a single use case may not be a high priority.

9.B Year One

Activities in 2018 and 2019 centered on the design implications of these potential services, along with discussions with regional partners about the feasibility of deployment, with the exception of the TestSet Service, which is available for use at this time. EPOC has procured a handheld Viavi T-Berd 5800 2 port test set. This test set is capable of 1G, 10G, and 100G testing at line rate on both ports simultaneously. This test set can be requested for use and if approved EPOC will ship the device to them. This testset can be used to validate newly installed circuits, help troubleshoot issues on already installed circuits, or be used to generate test traffic for benchmarking a network span.

In February of 2019, EPOC team members discussed the general model of the services at the All-Hands meeting. Discussion focused on:
- ***Services to pilot***: Data transfer and performance measurement were highlighted as targets
- ***Operational models to pursue***: Discussion on the ownership and operation of hardware, along with the models for configuration and continued operational needs were mentioned
- ***Regional networks to work with***: based on prior conversations, and relative need

Following this planning, the following Regional Networking Partners were engaged in initial discussions.

9.B.1 KINBER: perfSONAR

KINBER is exploring the concept of perfSONAR as a service, with partner institution Franklin and Marshall College. KINBER had an initial meeting with EPOC and staff from Franklin and Marshall College in April of 2018, and will be working closely with EPOC in 2019 to specify and operate the service.

9.B.2 LEARN: perfSONAR, Data Transfer, DMZ

LEARN and EPOC had discussions about operating a combination of services, namely providing Science DMZ, perfSONAR, and Data Transfer, to a selection of small colleges and universities. LEARN is pursuing NSF funding to facilitate the hardware purchase, and has agreed to work with EPOC if funding is awarded. EPOC provided a letter of collaboration to assist in this effort.

9.B.3 OARnet: Data Transfer

OARnet and EPOC had discussions about operating data transfer as a service to several pilot sites that lack the capability to do so currently. OARnet is pursuing NSF funding to facilitate the hardware purchase, and has agreed to work with EPOC if funding is awarded. EPOC provided a letter of collaboration to assist in this effort.

9.B.4 GPN member KanREN: Data Transfer

KanREN is a regional network in the state of Kansas and a connector to the Great Plains Network. KanREN is the only member-based consortium providing Internet and network technologies to research, education, and public service institutions in Kansas.

KanREN has interest in deploying a network of data transfer hardware capabilities around the state, for use by member institutions. Initial conversations discussed the feasibility and scalability aspects of this work. KanREN is pursuing NSF funding to facilitate the hardware purchase, and has agreed to work with EPOC if funding is awarded. EPOC provided a letter of collaboration to assist in this effort.

9.C Managed Services Plans for Year Two

A key deliverable in Year 2 for the Managed Services activity will be the publication of a set of whitepapers that detail the concepts of operating a "service in a box". It is anticipated that EPOC will release documents for each of the four services about how they will be supported by EPOC and their partners.

EPOC will be working with the following regional partners in 2019 to implement, measure, and refine services:
- GPN member KanREN on Data Transfer
- KINBER on perfSONAR
- LEARN on Science DMZ, perfSONAR, and Data Transfer
- OARnet on Data Transfer

Additionally, EPOC will engage with other regional partners on potential service ideas.

10. Training

EPOC will continue the successful training that ESnet and IU lead as part of the Operating Innovative Networks (OIN) program. This includes training for network engineers to be coordinated with existing cyberinfrastructure support teams with a focus not on deployment of services, but on cultivating a deep understanding of how to use today's tools and improve the process of science collaboration.

Broad adoption of CI tools, methods, and resources help to enable new modes of discovery, but also require continuing educational opportunities and dissemination of current practices. The gap in training is felt most at institutions that are smaller in scale, because it is more difficult to spare staff time to attend events or find funding for continuing education. While training programs like OIN emphasized the design and deployment of technology, we have pivoted to train staff on the use of these tools and the improvement of scientific use of networks through them. In addition to training on tools such as perfSONAR, we offer training for network engineers on interacting with their researchers through teaching them how to perform

Application Deep Dives. All EPOC training materials are available online, including lecture materials, exercises, and recorded sessions when possible.

10.A Year One

In the first year, we were involved in both technical training and engagement training. In September, Zurawski and Chevalier attended the SOX PRP workshop in Atlanta, GA [4]. They participated in training sessions on topics including perfSONAR and data transfers. There were approximately 16 people attending, primarily engineers from educational institutions. In the follow-on survey, 87.5% would recommend this training to other people, and comments included
- Really excellent facilitation.
- This was a great workshop, very helpful!
- Organizers where well qualified. Excellent!

We have entered into discussions with the Great Plains Network to assist the Linux Cluster Institute with an Introductory Workshop. Specifically, EPOC partners would host sessions addressing DMZ set up and the use of perfSONAR at this weeklong institute to take place in May 2019.

Zurawski and Rotman did Engagement Training at the NOAA Networking Committee Engagement Workshop, held in September in Silver Spring, MD [3]. The primary focus of this event was a workshop on Science Engagement featuring a researcher from NASA. Twenty people attended from NOAA affiliated campuses. Due to sensitivities surrounding the profiling of NASA science, a final report was not published.

Later in September, Schopf and Zurawski attended the Quilt/CC* PI meeting the University of Maryland, College Park, MD and lead a workshop on Science engagement featuring a blackhole researcher from the University of Maryland [6]. There were approximately 60 people attending, including many CC* PIs.

10.B Training Plans for Year Two

In Year 2, we plan to extend the training offerings as possible. We have been approached by several partners requesting training on science engagement and plan to incorporate these into the regional meetings as appropriate. We are working with our Regional Networking Partners on the identification of locations for training, as well as support they can offer to ensure long-term use and adoption. We will also invite members of the CC* engineering community as well as campus champions to participate in local training as appropriate.

In early 2019, we started discussions with the University of South Carolina, recipients of NSF Award #1829698 ("CyberTraining CIP: Cyberinfrastructure Expertise on High-throughput Networks for Big Science Data Transfers") to offer assistance for a July 2019 workshop on high performance networking technologies. Zurawski is also serving on the advisory committee for this award.

11. Data Privacy and Security

No Personally identifiable information (PII) is shared in the Roadside Assistance or Consultation summaries or reports, which are made public. There may be PII, for example IP addresses, in other documents in a Roadside Assistance Case Folder, but this information is locked down and access is controlled and only shared with specific staff working on a particular issue.

NetSage does not collect PII. We began discussions with several groups about the upcoming role out in Europe of the General Data Protection Regulation (GDPR), but the consensus is that GDPR pertains to data about individuals, and NetSage data only gets down to the level of organizations, so GDPR constraints are not relevant. The IRNC NetSage privacy docs were updated for EPOC and are available online at https://epoc.global/wp-content/uploads/2019/02/EPOC-Data-Privacy-Policy-21919.pdf, and in Appendix C.

Basic security measures are being maintained and there were no security incidents to report for this project.

12. Reporting Against Deliverables

In the original proposal, we listed the deliverables in Table 5 for the project. However, as is well known, "No plan of operations reaches with any certainty beyond the first encounter with the enemy's main force." - nineteenth-century Prussian military commander Helmuth van Moltke. Much of the language we use to describe the project has been updated and clarified, the number of focus areas has been adapted, and what is feasible and in line with our partners goals has had to be taken into account. In Table 6, we give an updated list of deliverables and their current status.

Year 1 focused on project start up and getting the six Regional Networking Partners on board with the project, primarily through the Roadside Assistance and Consulting, Deep Dives, and NetSage deployments. Year 2 will focus more strongly on the Infrastructure Partners and Science Community Partners, as well as moving forward on Managed Services.

Table 5: Original deliverables from ReSEC proposal.

Year	Milestone
1	Adaptation of IRNC NOC PET process with expanded focus
1	Collection of PS, SNMP from all regional networking partners
1,2,3	Adaptation of NetSage analysis for network disturbance detection
1,2,3	2 Deep Dives for applications per regional expected overlife of project
1	Establishment of full business model for Encapsulated Network Service
1	Initial service definition and pilot site deployment for DMZ
1	Adaptation of OIN/PS material for workshop

1	Training workshop 1	
1,2,3	All Attendance at application and technical meetings to extend impact and leverage	
1,2,3	All Evaluation of full project services and progress	
2,3	Expanded outreach to application communities for Roadside assistance	
2	Collection of Flow/Tstat data from all regional partners	
2	Additional analysis as needed for background performance monitoring	
2	Initial service definition and pilot site deployment for DTN, monitoring	
2	Continued support for DMZ offerings	
2	Training workshops 2 and 3	
2	Iniial release of online training resources	
3	Initial service definition and pilot site deployment for network security	
3	Continued support for DTN, monitoring offerings	
3	Training workshops 4 and 5	

Table 6: Re-wored deliverables for EPOC with Year 1 status.

	Deliverables	Status
RA	ROADSIDE ASSISTANCE	
RA.1	Adapt IN@IU, ESnet science engagement, PET and expand focus	Completed Year 1
RA.2	Advertise roadside assistance and consulting	Ongoing
RA 3	Assist with ongoing RAs - Partners	Ongoing
RA 3.1	iLight RA	Ongoing
RA 3.2	FRGP RA	Ongoing
RA 3.2.1	Consultation - Gonzalez Tribal colleges	Completed Year 1
RA 3.2.2	Consultation - Colorado School of Mines	Started Year 1, Ongoing
RA 3.3	LEARN RA	Ongoing
RA 3.4	OARNET RA	Ongoing
RA 3.5	GPN RA	Ongoing
RA 3.6	KINBER RA	Ongoing
RA 3.7	ESIP RA	Ongoing
RA 3.8	ICNWG RA	Ongoing
RA 3.9	IU GC RA	Ongoing
RA 3.10	UHawaii RA	Ongoing
RA 3.10.1	PANStarrs	Completed Y1; 3x improvement
RA 3.11	MWBDH RA	Ongoing
RA 3.12	OSN RA	Ongoing
RA 4	Other RA	Ongoing

RA 4.1	LHC Pakistan	Completed Year 1; 10x improvement
RA 4.2	Washington State Pullman	Started Year 1
RA 4.3	Consultation - American Museum Natural History	Started Year 1, Ongoing
RA 4.4	Consultation- University Florida	Completed Year 1
RA 4.5	Consultation - Compute Canada	Completed Year 1
RA 4.6	Consultation - New York University School of Medicine	Completed Year 1
DD	DEEP DIVE	
DD.1	Adaptation of ESnet facility deep dive process for use with applications	Completed Year 1
DD.2	Over project period, at least 2 deep dives per regional partner	Ongoing
DD.2.1	iLight Deep Dives	Ongoing
DD 2.1.1	Purdue University	Planned May 2019
DD.2.2	FRGP Deep Dives	Ongoing
DD 2.2.1	NOAA and NASA Deep Dive (with Training)	Completed Sept 2018
DD 2.3	LEARN Deep Dives	Ongoing
DD 2.3.1	Trinity University	Planned May 2019
DD 2.4	OARNET Deep Dives	Ongoing
DD 2.4.1	University of Cincinnati	Planned April 2019
DD 2.5	GPN Deep Dives	Ongoing
DD 2.5.1	Kansas State Agronomy (with training)	Planned May 2019
DD 2.6	KINBER Deep Dives	Ongoing
DD.2.6.1	Arcadia Bioinformatics (with training)	Planned April 2019
DD 2.7	ESIP DD	Ongoing
DD 2.8	ICNWG DD	Ongoing
DD 2.9	IU GC RDD	Ongoing
DD 2.10	UHawaii DD	Ongoing
DD 2.11	MWBDH DD	Ongoing
DD 2.12	OSN DD	Ongoing
DD.3	Other Deep Dives	Ongoing
DD.3.1	QUILT/University Maryland (with Training)	Completed Sept 2018
DD.3.2	University of Wisconsin	Planned June 2019
DD 3.3	TBD Y2 (May be one of DD 2.*)	July-August 2019
DD 3.4	TBD Y2 (May be one of DD 2.*)	Sept-Oct 2019
DD 3.5	TBD Y2 (May be one of DD 2.*)	Jan-Feb 2019
NS	NETSAGE	
NS.1	NetSage prototypes for regional partners	Ongoing
NS1.1	NetSage for iLight	Ongoing

34

NS 1.1.1	SNMP for iLight	May not be needed
NS 1.1.2	Flow for iLight	Initial prototype expected May 2019
NS 1.2	NetSage for FRGP	Discussion Year 2
NS 1.2.1	Tstat for NOAA	Expected early Year 2
NS 1.2.2	Tstat for NCAR	Expected Year 2
NS 1.3	NetSage for LEARN	Ongoing
NS 1.3.1	Tstat on TACC archives	Completed Year 1
NS 1.4	NetSage for OARNET	Discussion Year 2
NS 1.5	NetSage for GPN	Ongoing
NS 1.5.1	SNMP for GPN	Released October 2018
NS 1.5.2	Flow for GPN	Planned for Year 2
NS 1.6	NetSage for KINBER	Planned for Year 2
NS 2	NetSage deployments related to other partners	Ongoing
NS 2.1	University Hawaii	Ongoing
NS 2.1.1	Tstat on Astronomy Archive	Prototype deployed
NS 3	Adaptation of NetSage analysis for network disturbance detection	Planned for Year 2
MS	MANAGED SERVICE	
MS 1	Define perfSONAR Managed Service (PS MS)	Started Year 1, Ongoing
MS2	PS MS deployments	Ongoing
MS 2.1	iLight PS MS	TBD
MS 2.2	FRGP PS MS	TBD
MS 2.3	LEARN PS MS	Under discussion
MS 2.4	OARNET PS MS	TBD
MS 2.5	GPN PS MS	TBD
MS 2.6	KINBER PS MS	Started Year 1, Ongoing
MS 2.6.1	KINBER and Franklin Marshal PS MS	Under discussion - planned for Year 2
MS3	Define Data Transfer Managed Service (DT MS)	Planned for Year 2
MS4	Pilot DT MS	TBD
MS 4.1	iLight DT MS	TBD
MS 4.2	FRGP DT MS	TBD
MS 4.3	LEARN DT MS	Under discussion
MS 4.4	OARNET DT MS	Under discussion
MS 4.5	GPN DT MS	Under discussion
MS 4.5.1	GPN and KanRen DT MS	Under discussion
MS 4.6	KINBER DT MS	TBD
MS 5	Define DMZ Managed Service (DMZ MS)	Planned for Year 2

MS 6	Pilot DMZ MS	TBD
MS 6.1	iLight DMZ MS	TBD
MS 6.2	FRGP DMZ MS	TBD
MS 6.3	LEARN DMZ MS	Under discussion
MS 6,4	OARNET DMZ MS	TBD
MS 6.5	GPN DMZ MS	TBD
MS 6.6	KINBER DMZ MS	TBD
MS 7	Testset Loaning	Ongoing
T	TRAINING	
T 1	Set up public site for training materials	Completed Year 1
T 2	Technical training	Ongoing
T 2.1	PS 1- SOX	Completed Y1
T 2.2	PS 2- GPN/LCI - also DMZ	Planned 5/19
T2.3	PS 3- Tribal College WestNet	Planned Year 2
TS2.4	PS 4	TBD Year 3
T 3	Engagement training	Ongoing
T3.1	Deep Dive Training 1-NOAA	Completed Year 1
T 3.2	DD Training 2- QUILT	Completed Year 1
T 3.3	DD Training 3- KINBER	Planned 4/19
T 3.4	DD Training 4-GPN	Planned 5/19
T 3.5	DD Training 5	TBD Year 2
T 3.6	DD Training 6	TBD Year 3
T 3.7	DD Training 7	TBD Year 3
T 3.8	DD Training 8	TBD Year 3
T 4	Other training	TBD as requested
T 4.1	High Performance NW - S. Carolina	Requested for 7/19

13. Reporting against Metrics

Table 7 gives the metrics that were defined in our proposal. Similar to the original deliverables, we no longer believe these to be fully relevant to the work under way. Instead, for each project activity, we have listed metrics as appropriate in each section separately (Tables 2, 3, 4, and Section 10.A). These will be reviewed by the EAB.

Table 7: Metrics form original ReSEC proposal.

Area	Metric
1, 2	Number of tickets for new contacts, both end-to-end engagement and deep dives
1, 2	Count of interactions (emails, phone calls, meetings, etc) associated with the tickets
1, 2	Number of new applications identified for campuses or regional networks
1, 2	Satisfaction surveys for contacts for both end-to-end engagement and deep dives
2	A count of new deployments of NetSage infrastructure
2	A count of anomalies detected by the measurement infrastructure, tickets opened based on that feedback, and activities associated with those tickets
3	Number of Encapsulated Network Services deployed and uptime statistics
4	Number of workshops or training sessions held, along with number and demographics of participants
4	Results of evaluation surveys for events (which will include before/after metrics)
4	Quantity of materials uploaded for training or to http://faster.data.net for general use by community
4	Number of downloads/comments on online training materials

Appendix A

Hawaii Pan-STARRS Data Movement Issues Summary

EPOC Contact Point: Hans Addleman (addlema@iu.edu)

IRNC PET Contact: Jared Schlemmer (jeschlem@globalnoc.iu.edu)

Last edit: March 14, 2019

The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) program shares approximately 100 terabytes of data yearly between the Institute for Astronomy (IfA) at the University of Hawaii (UH) and the Space Telescope Science Institute (STSCI) at John Hopkins University in Baltimore, Maryland, to enable researchers to more accurately estimate galaxy †redshifts, improving their understanding of the local cosmic expansion and dark energy. In November, 2018, they approached the Indiana University (IU) teams for assistance because they were experiencing a maximum transfer rate of only 320 Mbps, despite the fact that they believed the full path between IfA and STSCI was equipped with 10Gbps and 100Gbps networks. They hoped to achieve transfer rates in the multi gigabit range, and looked to IU to help them debug the path and perhaps recommend a parallel file transfer tool.

Over the next 3 months, International Research Network Connections (IRNC) NOC, the IRNC Performance Engagement Team (PET) and International Networks at Indiana University (IN@IU) staff worked closely with network engineers and IT staff from University of Hawaii, John Hopkins University, Indiana University, Internet2, and the Mid-Atlantic Crossroads (MAX) Gigapop (which supports R&E networking in the Maryland, Virginia, and DC area)to actively troubleshoot the issues, identify bottlenecks, and resolve the identified problems.

The team made heavy use of PerfSONAR nodes for ongoing and adhoc testing during the engagement across the full path. These nodes were located at various points of both end networks and the Wide Area Network (WAN) between them, including Internet2, TransPAC, and MAX gigapop nodes. UH had multiple perfSONAR nodes that they moved around their network for more accurate testing as well.

The engagement identified a number of issues, some of which were solved during testing and others cataloged for addressing at a later date. On the University of Hawaii side, these included:
- The Top of the Rack (TOR) switches in the UoH data center Science DMZ were determined to be underpowered for the level of data transmission they were experiencing, so the critical data servers were moved away from this set up.
- Misconfigured access control lists and firewalls in the Science DMZ also contributed to the poor performance. IfA worked to eliminate these bottlenecks by redesigning the equipment layout so that the data transfer nodes were not behind the firewall going forward.
- The default routing between the UH hosts and the JHU hosts were taking a suboptimal and longer route through the Internet2 Network. UH staff moved peering to their PIREN 100G link to Los Angeles and this allowed the traffic to take an optimal path from end to end.

On the JHU side, these included:
- The determination that the JHU portion of the path was actually only a 1G path between JHU and the MAX gigapop. he internal network was upgraded to 10G from the end data receiver host through to the Internet2 connection via the MAX Gigapop.
- PerfSONAR nodes were installed inside and outside the JHU firewall to enable better on ongoing testing and identification of errors.

Both UH and JHU found the following issues with their data transfer methodology and hardware:
- Across the end-to-end path, the Maximum Transmission Unit settings on all of the routers and transfer hosts were upgraded to 9000 byte size frames (Jumbo Frames). This improves network performance by making data transmissions more efficient, because the CPUs on switches and routers can process a larger payload for each frame, but only works if each link in the network path -- including servers and endpoints -- is configured to enable jumbo frames at the same MTU.

- On both ends, the TCP Buffer settings for the end hosts were misconfigured for large scale data transfers, so these also were updated to to the ESnet recommended settings (http://fasterdata.es.net/host-tuning/background/).

In addition, due to the age of the software and system set up for this collaboration, several inefficiencies were identified:
- Because the software was bespoke to PanSTARRS, and written over many years, some aspects of it were ineffective for today's systems. Specifically, the system required manual intervention at various points which could delay the workflow. Full resolution of this will take a significant re-write of the workflow tool.
- Within the bespoke software framework, file transfers were delayed by a per file DNS lookup that would hang due to misconfiguration of the Web Proxy piece of the file transfer mechanism. The configuration of the proxy was updated to resolve this issue.
- The collected data did not exist in a single location, but was instead spread out across over 160 discrete logical storage volumes on 32 hosts, many of which had not been tuned (or could not be tuned) to enable fast data transfers. In addition, some of the hosts had aged to the point of being unreliable and could crash in the middle of data access actions. The project is working towards a new, unified storage system on modern equipment to address these issues.

The file transfer was fully re-evaluated in June of 2018 and after our engagement saw a 3x jump in overall performance seeing a sustained 1Gbps transfer rate, up from the original 320Mbps.

The results of this engagement also led the PIREN project to receive a supplemental National Science Foundation (NSF) award to enhance the capabilities of their overall network, Science DMZ, Data Transfer hardware, and network testing hardware. They are currently working to procure, design, and put this new architecture in place. IN@IU and the PET will remain engaged as needed.

Appendix B

Pakistan-UK LHC Data Movement Issues Summary

EPOC Contact Point: Hans Addleman (addlema@iu.edu)

Last Edit: March 4, 2019

The National Center for Physics (NCP) at the Quaid-i-Azam University Campus in Islamabad, Pakistan, is a Tier 2 Large Hadron Collector (LHC) Site. NCP contacted EPOC in September, 2018, for assistance with an ongoing performance issue when moving data to and from the GridPP (www.gridpp.ac.uk) Tier 1 site at the Queen Mary University in London. As a Tier 2 LHC site, NCP is responsible for downloading data sets from the Tier 1 site to share with LHC researchers in the region in a timely manner. NCP has a 1Gbps connection to their national R&E network, Pakistan Education and Research Network (PERN), however, they were seeing transfer rates as low as 40Mbps to some Tier 1 LHC sites. These performance problems were intermittent, and have been ongoing for at least a year prior to contacting us. Our engagement with NCP started in October 2018 and ended in January 2019. We identified five main issues during that time, which resulted in performance improvement to ~480Mbps. This rate is not consistently achievable due to a congested link in the path detailed in issue five below.

The performance goal requested by NCP was to achieve file transfer rates to LHC sites in line with transfer rates to other locations. NCP had been able to successfully achieve transfers up to approximately 500Mbps to Aarnet (Australia) and up to 280Mbps to ESnet (es.net) Data Transfer Nodes in the United States. It was believed that an acceptable performance goal would be to achieve at least 160Mbps to the GridPP LHC site at Queen Mary University.

EPOC engineers and engagement specialists worked with engineers and systems administrators from five other institutions during the investigation including TEIN*CC (www.tein.asia), PERN (pern.edu.pk), GEANT (www.geant.org), and INFN (home.infn.it).

During the investigation, a perfSONAR (www.perfsonar.net) mesh and dashboard that included servers spanning the NCP, PERN, TEIN, and GEANT networks was created. This mesh allowed for easier ongoing identification of packet loss and network congestion. perfSONAR experts from the GlobalNOC systems team were able to guide the configuration of these hosts and help alleviate systems issues that arose. Adhoc latency and bandwidth testing were also used on an ongoing basis to see the effect of troubleshooting in real time. The perfSONAR MaDDash was left in place for future testing as well.

Five main issues were identified during the engagement:
1. A traffic shaping misconfiguration on the NCP connection to PERN caused research and education traffic to be limited to 50Mbps. PERN identified the error and fixed the issue.

PERN engineers report that R&E (Research and Education) traffic should not have any traffic shaping applied.

2. A top of rack switch was identified as a bottleneck between NCP's file transfer node and their edge router. When the file transfer node was directly attached to the edge router, performance increased from 40Mbps to tests achieving an average of 100Mbps to 500Mbps or better.

3. We also observed in the perfSONAR dashboard small amounts of ongoing packet loss within the campus network supported by NCP. NCP engineers are continuing to track down the cause of this packet loss. NCP moved their data node closer to the edge of their network to alleviate the issue.

4. Packet loss was also identified by the perfSONAR mesh inside the PERN regional network. PERN network engineers worked to resolve the issue, however, they did not report the specific cause of the loss.

5. Additional bottlenecks were observed at the connection between PERN and TEIN networks. This connection currently only supports 1Gbps, and therefore experiences congestion which causes packet loss and that prevents NCP's file transfer speeds from increasing. This was verified by temporarily changing the routing to a commercial path, at which time NCP achieved expected transfer rates. TEIN and PERN are currently discussing an interconnect upgrade to 10Gbps, which would likely alleviate the congestion and allow for greater transfer speeds. It was suggested that in the meanwhile, NCP could work with PERN to have their LHC traffic follow the commercial path while they wait for the interconnection upgrade.

The final result of this engagement was an overall increase in data transfers between NCP and Queen Mary University to ~480Mbps. As identified above, this performance is based on a test during an uncongested time. Additional areas for performance improvements have been identified, but require longer term changes in infrastructure to be achieved. EPOC will stay engaged with all parties and help NCP test when needed.

Appendix C

EPOC Network Data Privacy Policy

February 19, 2019

Contact: Jennifer M. Schopf, jmschopf@iu.edu

I. Introduction

The Engagement and Performance Operations Center (EPOC) uses the NetSage measurement and monitoring tool (http://netsage.global) to gather network data for use in understanding file transfer performance and debugging user data transfer issues. The NetSage tool is an open privacy-aware network measurement, analysis, and visualization service designed to better understand the behaviors and needs of today's research and education (R&E) networks.

Although currently in use by other projects, the NetSage Project was originally funded by the US National Science Foundation (NSF) International Research Network Connection (IRNC) program to better understand the use of the IRNC-funded backbone networks and exchange points. In much the same way as other large-scale NSF facilities track their end users, the NetSage tool was created to understand the use of the IRNC networks. For example, the XSEDE (https://www.xsede.org/) high performance computing platform tracks end users by institution, science domain, and project, as does the Open Science Grid computing consortium (https://www.opensciencegrid.org/). NCAR/UCAR (https://library.ucar.edu/) tracks the use of their data resources in similar ways.

The EPOC project uses the NetSage Tool to understand and visualize large data flows associated with research, education, and science projects for its associated partners, and also uses this data to debug performance issues. The data is de-identified and used primarily to understand the network behaviors of large flows and to better understand the general use and functionality of the monitored networks and exchange points.

EPOC works with their *Regional Partners* to collect data from networks, exchange points, and archives associated with those partners and with their partners' permission. The EPOC data privacy policy for data collected by the NetSage tool strives to balance the privacy interests of users whose data transits the networks that the NetSage tool monitors, the operational needs of the EPOC project and the Regional Partners, and the need to demonstrate the broader benefit of the NSF-funded resources. We are committed to protecting privacy and informing interested parties about our policies and practices.

II. Scope of this Policy

42

This policy identifies:
- The information the NetSage tool collects about data transferred by its infrastructure;
- The ways in which this information may be used and disclosed to third parties; and
- The security measures adopted to prevent unauthorized access to this information.

III. What Information Is Collected?

The NetSage tool captures and collects *active* networking data (for example, latency and throughput from a tool such as perfSONAR) and *passive* network metadata (for example, SNMP and flow data). This data may consist of packet headers in addition to performance data, but will never contain payload data from flows. Data sets are de-identified at the source before being stored. The NetSage tool uses current accepted practices in the R&E community, such as anonymizing the IP addresses in a prefix-preserving manner, to de-identify the data released to researchers to ensure user privacy wherever possible. The data is then highly aggregated and does not contain information about traffic flows specific to individual users. EPOC partners may choose to release additional data to the EPOC team, however that data is not part of this policy.

IV. Disclosure of Data

EPOC is the steward of all the network data it collects. EPOC, at the direction of the EPOC PI, may share network data under the following circumstances:
1. EPOC plans to make summaries of de-identified Regional Partner network traffic data public on the EPOC website (http://epoc.global).
2. Upon request, Regional Partners can have access to the full, de-identified data sets for their site.
3. In rare cases where there is an ongoing performance issue for a specified flow, access to raw data may be needed to debug performance issues. If both endpoints of a flow agree, the EPOC collaboration may, for a limited time, use raw data at a collection point to help identify ongoing performance problems between two sites. Internal procedures exist to ensure this is done securely. The EPOC team takes the following actions when collecting raw data:
   a. Permission is obtained from the authorized representative at Regional Partner.
   b. Affected organizations are notified.
   c. Internal logs are maintained documenting what data is collected, the time period covered, who collected the data, and why the data was collected.
   d. When the issue has been resolved, the Regional Partner will also be informed and the raw data files will be subsequently destroyed using industry best practice data sanitization techniques.

V. How Data Is Collected, Retained, and Protected

All network data is managed under the control of EPOC project members authorized by the EPOC Principal Investigator at Indiana University.

EPOC takes appropriate steps to protect collected network data from unauthorized access or disclosure. Additionally, EPOC employs industry standard security measures to protect against the disclosure, loss, misuse, and alteration of the information under our control.

VI. Notice for Updates and Changes to Policy

This document is derived from the original NetSage project data privacy document (http://www.netsage.global/home/netsage-privacy-policy), which was derived from ESnet's privacy policy (available at https://www.es.net/about/governance/data-privacy-policy/), which itself is derived from the Internet2's policy on privacy of network flow data (available at http://www.internet2.edu/policies/network-flow-data-privacy-policy/). EPOC reserves the right to update this privacy policy at any time to reflect changes in the manner in which it deals with traffic, whether to comply with applicable regulations and self-regulatory standards, or otherwise. Then Privacy Policy posted here will always be current. We encourage you to review this statement regularly.

VII. Who to Contact if You Have Questions

If you have any questions about this privacy policy, please contact Dr. Jennifer M. Schopf, the PI of the EPOC project, at jmschopf@iu.edu.

Acknowledgement: The Engagement and Performance Operations Center (EPOC) is a joint collaboration of Indiana University and the Energy Sciences Network (ESNet), and is funded in part by NSF Award #1826994.

VIII. Glossary
Networking data: Active — Data collected by tools that perform active tests at the user level, such as perfSONAR

Networking data: Passive — Network traffic data, such as netflow or sFlow data, or data collected using passive monitoring tools such as Tstat or from routers directly

Regional Partners — At the time of this writing, EPOC's regional partners included:
- The Indiana State Network (I-Light)
- Ohio State R&E Network (OARnet)
- Keystone Initiative for Network Based Education and Research (KINBER)
- Great Plains Network (GPN)
- The Texas State R&E Network (LEARN)
- Front Range Gigapop (FRGP)

Appendix D

Collaboration Plan - EPOC and TNRP

Contact: Jennifer M. Schopf, jmschopf@iu.edu

Summary
In 2018, the NSF CC* program funded two proposals in response to its request for a Network Performance Engineering and Outreach (NPEO) project submissions (https://www.nsf.gov/pubs/2018/nsf18508/nsf18508.htm). As the solicitation originally specified that there would be only one awardee, neither project plan included a plan for the two funded centers to collaborate. This document gives a brief overview of the two projects, identifies their areas of overlap, and details for collaborative areas going forward.

Project Overviews
Engagement and Performance Operations Center (EPOC) - NSF Award#1826994

The Engagement and Performance Operations Center (EPOC) is a production platform for operations, applied training, monitoring, and research and education support. Over the last decade, the scientific community has experienced an unprecedented shift in the way research is performed and how discoveries are made. Highly sophisticated experimental instruments are creating massive datasets for diverse scientific communities and hold the potential for new insights that will have long-lasting impacts on society. However, scientists cannot make effective use of this data if they are unable to move, store, and analyze it.

The Engagement and Performance Operations Center (http://epoc.global) is a collaborative focal point for operational expertise and analysis and is jointly led by Indiana University (IU) and the Energy Sciences Network (ESnet). EPOC provides researchers with a holistic set of tools and services needed to debug performance issues and enable reliable and robust data transfers. By considering the full end-to-end data movement pipeline, EPOC is uniquely able to support collaborative science, allowing researchers to make the most effective use of shared data, computing, and storage resources to accelerate the discovery process.

EPOC supports five main activities
- Roadside Assistance and Consultation via a coordinated Operations Center to resolve network performance problems with end-to-end data transfers reactively;
- Application Deep Dives to work more closely with application communities to understand full workflows for diverse research teams in order to evaluate bottlenecks and potential capacity issues;
- Network Analysis enabled by the NetSage monitoring suite to proactively discover and resolve performance issues;

- Provision of managed data services via support through the IU GlobalNOC and our Regional Network Partners;
- Coordinated Training to ensure effective use of network tools and science support.

Towards the National Research Platform (TNRP)- NSF Award #1826967

The NSF-funded Pacific Research Platform (PRP) provides an Internet platform with 100-1000 times the bandwidth of today's commodity Internet to major research universities on the West Coast and several locations in the US and Europe/Asia. Now in its fourth year, it has been designed to serve specific application needs of researchers. The disciplines which are engaged in partnering with the PRP range from particle physics to climate to human health, as well as archaeology, digital libraries, and social media analysis. The next stage, now underway, is trying to understand how to scale this prototype cyberinfrastructure to National and Global Research Platforms.

Toward the National Research Platform (TNRP) addresses issues critical to scaling end-to-end data sharing and computing. TNRP instrumenting a large federation of heterogeneous "national-regional-state" networks (NRSNs) to greatly improve end-to-end network performance across the nation, which requires active participation of these distributed regional-level entities to reach out to their campuses, and the help of Internet2 and ESnet. Regionals are trusted conveners of their member institutions, contributing effectively to both the people networking and inter-campus networking that are as necessary to the development of a full National Research Platform as is the stability, deployment, and performance of technology. TNRP's collaborating NRSNs structure leads to engagement of a large set of science applications, identified by the participating NRSNs, Internet2, ESnet and the Open Science Grid. This project offers the Kubernetes-managed PRP Nautilus Cluster as a replicable model. Nautilus includes equipment purchased from many fund sources and is coordinating closely with OSG on technological and science engagement practices.

TNRP is highly instrumented to directly measure performance. Visualizations of disk-to-disk performance with passive and active network monitoring show intra- and inter-NSRN end-to-end performance. Internet2, critical for interconnecting regional networks, provides several instrumented dedicated virtual network instances for the interconnection of TNRP's NRSNs and OSG's use. Cybersecurity is a continuing area of research and deployment; evaluations of advanced containerized orchestration, hardware crypto engines, and novel IPv6 strategies are part of the TNRP plan.

Areas of Overlap

In practice, although the goals of the two projects are similar, the day-to-day work by the two groups differ somewhat. In general, the TNRP project is more focused on advocacy to move towards container orchestration solutions as well as encapsulated and mobile storage and compute infrastructure to enable the use of more advanced network and compute infrastructures, including but not limited to the PRP Nautilus cluster. Their primary science use cases to date (http://pacificresearchplatform.org/nautilus/namespaces/) tend to be associated

46

with the PRP or an R1 institution. They are planning to address scalability issues for user support in part via AI techniques with their RocketChat support room.

EPOC is focusing more on working with application scientists in their current environment and debugging existing infrastructure with limited changes to the underlying technological basis. Their service is available to anyone working in research or education. They are also promoting the use of advanced infrastructure in a supported way for organizations that don't have the in-house capability to offer these services. Their focus is more strongly on underserved (non-R1) institutions and small to medium sized research collaborations. They are focusing on an operations center approach for support.

EPOC	TNRP
Roadside Assistance helpdesk for any research or education project experiencing problems with data transfers; Consultation helpdesk to assist network engineers on design of DMZs, DTNS, etc	Rocket chat to assist end users of Nautilus cluster and to help with deployment of PRP-like regional/national infrastructure; weekly technical engagement of regional and national network engineers and application engineers who work with their users.
Advocacy to end users to understand future cyberinfrastructure needs through the use of Deep Dives	Advocacy to end users to make use of advanced cloud technologies through the use of Kubernetes and Jupyter notebooks and high-speed shared storage technologies.
Science engagement to R&E community, primarily through six science community partners, six regional network partners, and infrastructure partners including Internet2, Quilt, CACR, Science Gateways, CACR, and XSEDE, and CARCC. Engagement with the network engineering community through the cyberinfrastrucure engineering list, campus champions, and strong ties to partner regional networks.	Science engagement to R&E community, primarily through PRP, regional network, Internet2, ESnet, Quilt, and OSG partnerships. Engagement with CIOs through Internet2, CENIC, and regional/national networks.

NetSage and perfSONAR deployment for measurement and monitoring of network infrastructures supported by partners (For example,fhttp://gpn.netsage.global)	perfSONAR and DTN deployments for measurement and monitoring of network infrastructures supported by partners: https://perfsonar.nrp-nautilus.io/maddash-webui/ and numerous Kubernetes-based open source measurement and monitoring tools for networks, computing, and data sharing.
Development and advocacy of advanced services focused on small and medium sized institutions: perfSONAR in a box, DTN in a box, DMZ in a box, Testing Service in a box	Development and advocacy to support science use flows: Kubernetes, Jupyter notebooks, Nautilus cluster and similar federated clusters nationwide
Training for the community in Deep Dives, How IT can talk to scientists, Finding researchers on your network/campus, perfSONAR installations and support - all primarily classroom settings. Additional consultation one-on-one as part of Roadside assistance and consultation track.	One-on-one instruction at Calit2 to support use of Kubernetes, Jupyter notebook, nautilus cluster, FIONA construction and deployment.
Publication of engagement success stories: https://epoc.global/materials/, http://fasterdata.es.net	Publication of engagement success stories: http://pacificresearchplatform.org/nautilus/namespaces/
Participation in alternate Monday PRP/NRP Pilot Engagement Team Meetings (led by Internet2), leading the weekly cyberinfrastrucure engineering brownbag calls, Quilt meetings, and NRP annual meetings.	Participation in alternate Monday PRP/NRP Pilot Engagement Team Meetings (led by Internet2), weekly PRP/NRP Pilot Thursday Technical Engagement Meetings (led by CENIC/PRP/TNRP and Internet2), Quilt meetings, and NRP annual meetings.
Presentations at Internet2 and regional network conferences and workshops on basic and advanced topics. Demonstrations at SC conferences and other venues.	Presentations at Internet2 and regional network conferences and workshops on basic and advanced topics. Demonstrations at SC conferences and other venues.

Points of Collaboration

The two projects held a multi-hour meeting following the Quilt spring meeting in February to discuss how the two groups could collaborate going forward. The following items were identified:

- Continued and ongoing meetings to take place at shared venues (I2, Quilt, NRP, etc)
- Joint participation in the PRP/*RP Engagement calls on Mondays
- Shared notification of training opportunities
- Continued cooperation with partners in common, especially in the area of the *RP Pilot studies