

NetSage

Award #1540933

Year 5 Quarter 4 and Annual

1 February 2019 through 31 January 2020

PIs: Jennifer Schopf (IU), Andrew Lake (LBNL), Jason Leigh (UHM)

Summary

The goal of the IRNC NetSage project is to collect data from the IRNC-funded backbone and exchange points to better understand the use of the resources. In addition, this collected data is also made available for use by the NOC for day-to-day operations and to support end-to-end performance troubleshooting. Highlights of Year 5 included three releases for Dashboards, significant work in displaying flow data, additional use of Sankey graphs, and the collection of usage analytics showing over 1,700 unique users in 89 countries between July 2019 and January 2020.

1. NetSage Overview

NetSage is building and deploying advanced measurement services that will benefit science and engineering communities, focusing on:

- Better understanding of current traffic patterns across IRNC links;
- Better understanding of the main sources and sinks of large flows to know where to focus outreach and training; and
- Better understanding of where packet loss is occurring, whether or not the loss is caused by congestion or other issues, and the impact of this on end-to-end performance.

The NetSage software consists of a set of open source tools that follow a basic monitoring tool architecture, as shown in Figure 1. NetSage TestPoints are a collection of software and hardware components that gather data from SNMP, perfSONAR, and flow devices. The data from the TestPoints is sent to the Data Ingest Pipeline, where additional tags are added, including information from the MaxMind GeoIP database and the NetSage Science Registry. Data collection is discussed in Sections 5.1-5.5. The data is then stored in the NetSage Archive, a storage framework consisting of several different databases, including a Time Series Data System (TSDS) archive and an Elasticsearch archive, discussed in Section 5.6. A variety of Dashboards, built on top of the open source Grafana analysis and visualization engine, access the data from the NetSage Archive to visualize the results of queries. The Dashboards deployed for the IRNC resources are available online at <http://portal.netsage.global> and are discussed in Section 6.

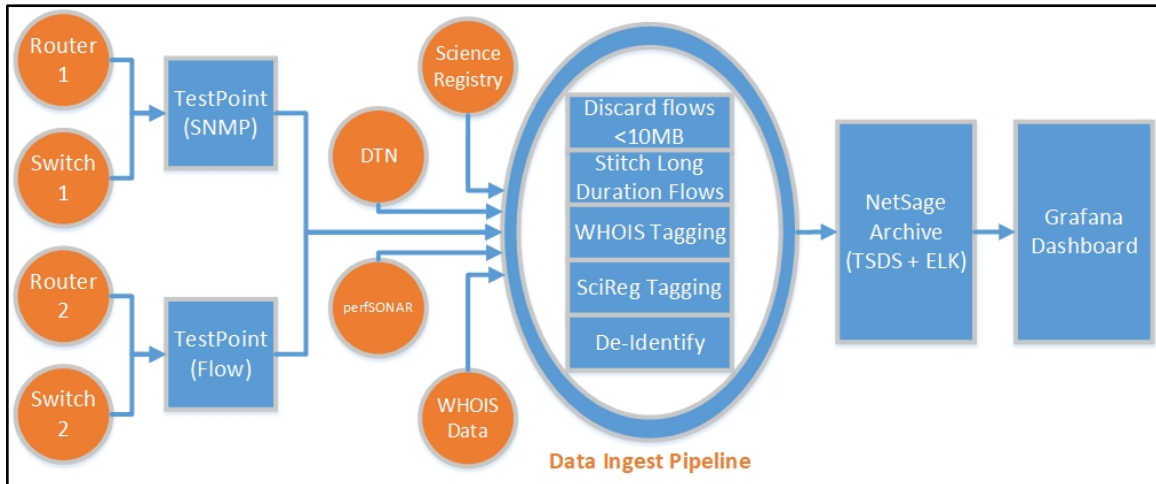


Figure 1: The current software architecture for NetSage.

Year 5 focused on extending the data collected from instrumented archives and the science registry, and planned to address the following questions:

- 4.c - What are the highest retransmit rates between organizations/subnets over a timeframe?
- 6.g - Who have been the top talkers each year (longitudinal study)
- 7.f - Is retransmit data a proxy for packet loss?
- 9.b - What level of retransmits is an archive experiencing during data flows?
- 9.c - What are the patterns for retransmits on an archive?
- 9.d - For a pair of endpoints, what is the retransmit behavior for the individual flows?
- 11.a - What is the bandwidth of a GridFTP file transfer between two backbone end points?
- 12.a - How are active tests between two sites performing? (a replacement for perfSONAR MaDDash)

Year 6 will focus on supporting prior user-requests for additional functionality, hardening the current offerings, and making NetSage more robust for uptake by other projects. It is expected that the IRNC backbone and exchange points will end their current projects on or around December 2020, at which point the NetSage project will hand off the data and code resources to third-parties who want to continue using NetSage on their own.

2. Staffing

At the end of Year 4, funded staff included:

- Jennifer Schopf, IU, PI - overall project director
- Ed Balas, IU, system architect - collection and reporting
- Scott Chevalier, IU, IRNC perfSONAR mesh support
- Dan Doyle, IU, developer - collection and reporting
- Lisa Ensmann, IU, developer - Science Registry
- Heather Hubbard, IU - Staff support
- Sangho Kim, IU, system engineer - collection and reporting

- Andrew Lee, IU, network data analysis development
- Ed Moynihan, IU, Science Registry Data support
- Sreemuka Taduru, IU, Grafana map development
- Andy Lake, LBNL, co-PI
- Sean Peisert, LBNL, Security advisor
- Jason Leigh, UH Mānoa, co-PI - visualization oversight
- Mahesh Khanal, UH Mānoa, graduate research assistant - developer
- Katrina Turner, UH Mānoa, graduate research assistant - developer
- Alan Whinery, UH System – perfSONAR, PIREN coordination

In March, 2019, the co-PI shift from Peisert to Lake was finally completed and made official. Lake had been functioning in this role for the last 6 months already. At IU, Ed Balas, Andrew Lee, and Sreemukha Taduru left the project. Dan Doyle stepped into the system architect role. Doug Southworth joined for analysis work. The IU development team also shifted FTEs over the course of the year, from 1.8 FTE to 1.0 FTE, at the end of Quarter 1, and then raised to 1.5 FTE in Quarter 4. These changes were due to available staff in the Network Data Collection and Analysis team at IU. On the LBNL team, Sartaj Baveja and Samir Faci joined the project as software developers. At University of Hawai'i, Tyson Seto-Mook graduated in May 2019 and Katrina Turner joined in April as a Masters student.

At the end of Year 5, funded staff included:

- Jennifer Schopf, IU, PI - overall project director
- Scott Chevalier, IU, IRNC perfSONAR mesh support
- Dan Doyle, IU, developer - collection and reporting
- Lisa Ensmann, IU, developer - Science Registry and data ingest pipeline
- Heather Hubbard, IU - Staff support
- Sangho Kim, IU, system engineer - collection and reporting
- Ed Moynihan, IU, Science Registry Data support
- Doug Southworth, analysis
- Andy Lake, LBNL, co-PI
- Sartaj Baveja, LBNL, developer
- Samir Faci, LBNL, developer
- Sean Peisert, LBNL, Security advisor
- Jason Leigh, UH Mānoa, co-PI - visualization oversight
- Mahesh Khanal, UH Mānoa, graduate research assistant - developer
- Katrina Turner, UH Mānoa, graduate research assistant - developer
- Alan Whinery, UH System – perfSONAR, PIREN coordination

3. Collaborations, Travel, and Training

NetSage 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 NetSage. Travel in Quarters 1-3 were detailed in the previous reports, and included:

- Schopf attended the Quilt Winter member meeting in La Jolla, CA, February, February 4-7, 2019, <https://www.thequilt.net/public-event/2019-winter-member-meeting/>.
- Schopf, Whinery, and Lee attended APAN 47 in Daejeon, Korea, on February 18-22, 2019, <https://2019.apricot.net/apan47/>.
- Dan Doyle attended GrafanaCon meeting in Los Angeles, CA, February 25-26, <https://www.grafanacon.org/>.
- Schopf, attended the Internet2 Global Summit meeting in Washington DC, March 5-8, 2019, <https://www2.internet2.edu/e/66332/l-66332-2018-11-01-blhxn4/blhzn2/366390041>.
- Schopf attended the Spring 2019 CNI meeting in St, Lois, Missouri, April 8-11, 2019, <https://www.cni.org/event/cni-spring-2019-membership-meeting>.
- Schopf attended the iLight Members Meeting, in Indianapolis, IN, on May 7-8, 2019, <https://ilight.net/members-meeting/>.
- Schopf attended the Great Plains Network (GPN) Annual Meeting, in Kansas City, MO, on May 20-25, 2019, <https://www.greatplains.net/gpn-annual-meeting-2019/>.
- Lake, Baveja, Chevalier, and Southworth attended the perfSONAR Face-to-Face developer meeting in Bloomington, IN from May 20-May 22, 2019.
- Chevalier attended the 1st European perfSONAR User Workshop, in London, UK, June 5-6, 2019, <https://wiki.geant.org/display/perfSONAR/1st+European+perfSONAR+User+Workshop>.
- Schopf attended TNC 2019, in Tallinn, Estonia, on June 16-20, <https://tnc19.geant.org/>.
- Schopf attended the FRGP/WestNet Summer Meeting, in Salt Lake City, UT, June 24-30, 2019.
- Schopf attended the first annual NOAA NWave Meeting in Boulder, Colorado, on July 8-11, 2019.
- Schopf attended the Quilt, CC* PI, and NRP meeting in Minneapolis, MN, on September 23-26, 2019, <https://www.thequilt.net/public-event/2019-nsf-nrp-and-the-quilt-workshops-and-meetings/>.
- Tierney visited the IU development team for NetSage on October 14-16, 2019, to assist with future planning.

In Quarter 4, travel included:

- Leigh, Turner, Schopf, Chevalier, and Southworth attended SC'19 in Denver, CO, November 17-22, 2019, <https://sc19.supercomputing.org/>. Southworth worked with SCinet to support perfSONAR and to implement a GridFTP file transfer demonstration. Schopf met with additional NetSage partners to discuss the latest release and next steps.
- Doyle, Lake, Baveja, and Tierney attended TechEx in New Orleans, December 2020, <https://meetings.internet2.edu/2019-technology-exchange/>. They met with stakeholders including other IRNC project staff, perfSONAR developers, and regional network engineers. They also had dedicated time

for technical design discussion about upcoming project tasks. Doyle also presented a NetSage update at the Advanced Networking track of the conference.

A list of presentations and publications in Year 5 includes:

- “Research at the Laboratory for Advanced Visualization & Applications”, Jason Leigh, invited Presentation, LandSAGE Annual Meeting, KIST, February 21, 2019.
- “The Engagement and Performance Operations Center: Overview and Opportunities”, Panel - coordinated by Jennifer Schopf, Internet2 Global Summit, Washington, DC, March 8, 2019.
- “The Engagement and Performance Operations Center (EPOC)”, Jennifer M. Schopf, Invited presentation, Coalition for Networked Information (CNI) Spring Meeting, April 9, 2019.
- Schopf, Jennifer, “Helping Network Operators Identify Researchers Using Their Resources”, Invited workshop, 2019 I-Light and Indiana GigaPOP Members Meeting, Indianapolis, IN, May 7, 2019.
- Schopf, Jennifer, “What are GPN Folks Doing Internationally and Who’s Using my Networks?”, Invited Talk, GPN Annual Meeting, Kansas City, MO, May 23, 2019
- Lake, Andy, “An Introduction to NetSage”, remotely presented for the European perfSONAR Workshop, London, UK, June 5, 2019.
- Schopf, Jennifer, “The Engagement Performance and Operations Center (EPOC), and using NetSage Internationally”, Invited Talk, TNC 2019 Conference, Tallinn, Estonia, June 18, 2019.
- Leigh, Jason, “Research at the Laboratory for Advanced Visualization & Applications”, LandSAGE meeting, National University of Laos and University of Engineering and Technology in Vietnam, June 20, 2019.
- Schopf, Jennifer, “The Engagement Performance and Operations Center (EPOC), and Using NetSage”, Invited Talk, FRGP/WestNet 2019, Salt Lake City, UT, June 25, 2019.
- Lake, Andrew, “New NetSage Update is Live”, NetSage Bog Post, July 3, 2019, <https://www.netsage.global/post/new-netsage-update-is-live>
- Schopf, Jennifer, “The Engagement and Performance Operations Center and NetSage”, NOAA NWave Meeting, Boulder, CO, July 9, 2019.
- Tyson Seto-Mook, Alberto Gonzalez, Mahesh Khanal, Jason Leigh, Andrew Lake, Brian Tierney, Lisa Ensmann, Daniel Doyle, Ed Balas, and Jennifer M. Schopf, “Deploying NetSage - a Network Monitoring Framework on Large Scale Network Backbones And Exchange Points”, submitted to SC2019.
- Schopf, Jennifer, “NetSage - A tool to understand data transfers”, Invited Presentation, CI Engineering Brown Bag Series, August 23, 2019.
- Lake, Andrew, “NetSage update improves the way you see flows between countries”, NetSage Blog Post, September 4, 2019, <https://www.netsage.global/post/netsage-update-improves-the-way-you-see-flows-between-countries>

- Lake, Andy, “An Introduction to NetSage”, Invited presentation, TransLight/PacWave (TLPW) perfClub User Group, October 21, 2019.
- Lake, Andrew, “Navigate Data More Easily with Latest NetSage Update”, NetSage Blog Post, November 11, 2019, <https://www.netsage.global/post/navigate-data-more-easily-with-latest-netsage-update>
- Turner, Katrina, “NetSage at SC19”, NetSage Blog Post, November 26, 2019, <https://www.netsage.global/post/netsage-at-sc19>
- Doyle, Dan, “Understanding End-to-End Science with NetSage: Archives, Science Registry, and even more Flow Data”, Invited presentation, Internet2 Technology Exchange, December 12, 2019.
- “IU’s GlobalNOC offers network monitoring service, improving research collaboration”, Indiana University Press Release, Jan 31, 2020, <https://itnews.iu.edu/articles/2020/GlobalNOC-offers-network-monitoring-service-improving-research-collaboration%20-.php>

4. Project Coordination

4.A Internal Coordination

Internal project coordination continued with weekly meetings of the majority of the team. We also hold weekly technical calls to be able to dive-down into more detailed topics with those NetSage members who are interested. These two calls are complementary to the twice-yearly face-to-face meetings that concentrate on more strategic planning.

We held our summer All Hands Meeting at the University of Illinois Chicago on June 10-12. This included a full day Hack-a-thon to unify the existing Dashboards and discuss options for updating the development and release process. On the second day, we reviewed all work in progress and set plans for the following 6 months, including the possibility of working with the IU Global NOC to support NetSage as a managed service for third party deployments.

We held our winter All Hands Meeting at the University of Hawai’i Mānoa on January 27-29, 2020. This included, again, a full day Hack-a-thon to complete various tasks for the release of 1.3.0, followed by a day and a half of planning for the NetSage support for the IRNC projects to end in December 2020.

4.B Coordination with IRNC Partners

Work with the IRNC-funded backbones continued, and we now have SNMP and perfSONAR data from all of the original circuits as well as their latter-year additions. Sampled flow data is being collected from NEAAR, TransPAC, Pacific Wave, PIREN (Hawai’i links), and AmPath. StarLight and the PIREN Australia links will not share flow data.

At the Internet2 Global Summit in April, 2020, NetSage participated in an IRNC PI meeting that included time for feedback on the latest Dashboards. In general, the

feedback on the flow data Dashboards was extremely positive, with many IRNC PIs interested in how these could be used more extensively with their own end users. Several additions were requested and included in the following releases.

NetSage leadership is planning to hold meetings with each IRNC project PI to establish how they would like to hand off their data and support of deployments as part of the conclusion of the IRNC funding.

We continue ongoing discussions and coordination with the IRNC NOC. In August, we also began discussions with the IU GlobalNOC to evaluate if they would be interested in supporting NetSage as a managed service for deployments that were not part of NSF-supported partnerships. This was agreed to and announced at the Internet2 Technical Exchange in December 2019, and in January again, with additional details sent to current IRNC PIs who were applying for the 2020 IRNC solicitation, NSF 20-535.

4.C Coordination with External Partners

Additional third party deployments of network data also took place this quarter. In addition to the IRNC Dashboards, deployments (some supported by other funding) include:

- SNMP data for the core Advanced North Atlantic consortium (ANA), available online at <https://ana.netsage.global>. SINET was added to this Dashboard.
- SNMP data for the core Great Plains Network (GPN), available online at <https://gpn.netsage.global>.
- Flow data for the Indiana GigaPOP and I-Light networks, available online at <https://ilight.netsage.global>.
- Flow data for KINBER/PennRen, available online at <https://pennren.netsage.global>.
- Flow data for Front Range GigaPop, available online at <https://frgp.netsage.global/>

At the APAN meeting in February, it was confirmed that the AARNet team/PIREN would not be sharing flow data with NetSage for the Hawai'i-Australia links. This is not unexpected as conversations had been ongoing for several years and there appeared to be issues with the Australian legal data sharing requirements. Furthermore, AARNet made a request to mask the data showing specific Australian universities and instead have them all listed together as AARNet. This was implemented shortly after the request.

Year 5 also saw the significant updates for the NetSage Project landing page, <https://www.netsage.global>. This site is the central launch point for all other NetSage Dashboards and general information. A NetSage blog was added in July 2020 to better communicate information related to releases and events.

5. Data Collection

NetSage staff are involved in the development and deployment of various pieces of software to support collecting active and passive measurements. This section details that work.

5.1 Collecting Simple Network Management Protocol (SNMP)

The Simple Network Management Protocol (SNMP) is an application-layer protocol defined in RFC1157 for collecting and organizing information about managed devices on IP networks. SNMP is used by routers and switches to monitor networks for conditions that warrant administrative attention. This data is commonly collected and openly archived by most R&E networks.

During Year 5, we updated the data collected for Starlight to address a number of changes that had occurred at their exchange point. No other SNMP deployments were added in Year 5 for IRNC partners.

Looking forward to Year 6, we will continue to update SNMP collection to meet the IRNC project expansion plans. For example, we are working with the AmLight network to put in place monitoring adaptations as they update their topology in the next 6 months.

5.2 Collecting Flow Data from Routers (sFlow, NetFlow)

Network Flow data collected using NetFlow or sFlow data consists of IP traffic information to better understand where network traffic is coming from and going to and how much traffic is being generated. Flow data collection relies on the Data Ingest Pipeline, shown in Figure 1, to filter data, add tags, and de-identify the flows.

During Quarters 1-3, several releases of the Flow Ingest Pipeline took place to reduce the complexity and increase in performance for data ingestion. This included converting the flow-stitching stage of the Pipeline to Logstash. After finding a bug in the Logstash aggregation module, and waiting on the ELK team to release a new version, the first production instance of the Pipeline that incorporated the changes was released in Quarter 4. Year 6 will include updates for tagging edge cases, such as multicast traffic, ignoring RFC1918 addresses, and updating the data schema to make the tagging more efficient.

In order to simplify the deployment of the Data Ingest Pipeline, we have begun work on creating a containerized Docker image of the entire Pipeline. Initial creation and testing of this container started at the very end of Year 5, and we expect that we will test and begin early deployments in Year 6. EPOC partner, the Great Plains Network, is slated to be our first deployment test of this new technology, but is it expected that this approach will simplify the longer-term maintenance for NetSage.

During Year 5, we were able to test lowering the threshold of what defines a large flow for NetSage. This threshold had been set to 500MB previously but was dropped to 10MB for traditional sFlow and NetFlow and 1MB for Tstat data collected from science archives. The change of threshold resulted in a significant increase in the number of flows collected, and as a result, directly improved the overall effectiveness of the data analysis and presentation. Figure 2 shows the difference in the volume of flows collected as a result of this change.

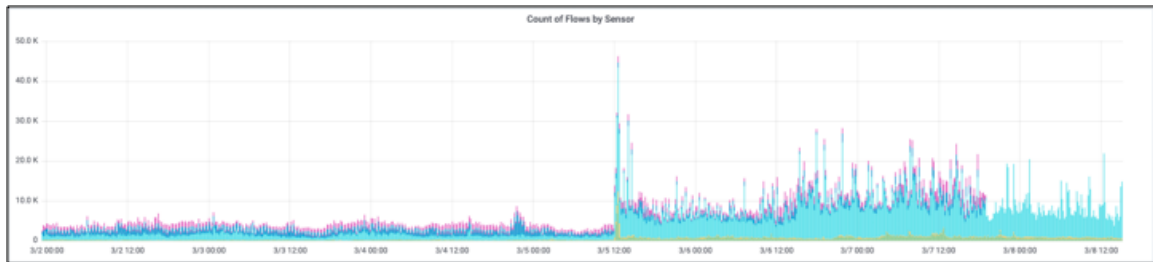


Figure 2: Graph showing number of flows collected for IRNC resources from March 2, 2019, to March 8, 2019, highlighting the threshold change introduced on March 5.

During Quarter 4, we spent time working with CENIC to fix and add more robust monitoring for the Pacific WAVE flow data we receive from them. This work was done as part of the turn-up of NetSage for them, available online at <http://pacwave.netsage.global>. The original flow collection was set up some time ago and had suffered from bitrot as their network evolved.

5.3 Collecting Tstat Data from Archives

We collect archive-based TCP flow statistics using Tstat, a tool that was developed as part of the EU Measurement Plane (mplane) FP7 project by Munafó and Mellia at Politecnico di Torino. Tstat examines all packet headers, similar to flow data collection on a router only unsampled, and reports the number of bits and packets transferred, the duration of the flow, the flow type, protocol and port used, and TCP retransmits.

In Year 5, work was done on the backend of the collector to store the percentage retransmits reported by Tstat in addition to the count of retransmitted packets. This required a change on the backend due to limitations in the active version of the Grafana frontend to perform mathematical operations on multiple fields, and was needed by the Dashboards for this data.

The deployment for the Texas Advanced Computing Center (TACC) that was started in Year 4 is now available online at: 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. Based on the successful third party Tstat deployment with TACC, we refined the documentation using their feedback.

We moved forward with the UH Astronomy group as well, first with a temporary deployment to meet an internal deadline, then with a permanent installation that was purchased using the remaining UCD funding. This data is available online at https://portal.netsage.global/grafana/d/mNPduO8mz/flow-data-for-data-archives?orgId=2&var-dtn_sensors=University%20of%20Hawaii%20Tstat&var-min_rtt=0

We worked with NOAA to gather data from an archive, which went live in March 2019. However, the selected initial deployment was on a backup archive so that the software could be verified from a security standpoint, and the backup archive ceased to be accessed in October, so the deployment was turned off in November. We have had follow-up discussions for a new deployment but have not yet been able to move forward.

The Tstat box for the National Center for Atmospheric Research (NCAR) was built and shipped to their site. We worked with the NCAR engineers to sort out issues related to their aggregator and are now successfully ingesting flows. This data is available online at https://portal.netsage.global/grafana/d/mNPduO8mz/flow-data-for-data-archives?orgId=2&var-dtn_sensors=UNET%20UCAR%20Tstat&var-min_rtt=0.

We have started work on creating a containerized image for Tstat collections, similar to the work with the Ingest Pipeline to enable easier deployment, upgrade, and support of third party deployments particularly in cases where the end site may not have technically capable people readily available to help with installation and troubleshooting. This will be part of the Year 6 deliverables.

5.4 perfSONAR

perfSONAR (<http://www.perfsonar.net/>) is a network measurement toolkit designed to provide federated coverage of paths, and help to establish end-to-end usage expectations. The NetSage project uses perfSONAR for its active measurements of throughput, latency, and loss, and archives them in the NetSage archive using TSDS. The IRNC projects participate in the IRNC perfSONAR mesh, available at <http://data.ctc.transpac.org/maddash-webui/index.cgi?dashboard=IRNC%20Mesh>.

Year 5 saw the release of four new versions of perfSONAR, including perfSONAR 4.2 in August 2019, which added the added ability to perform file transfer tests. A beta-version of this functionality was demonstrated by NetSage staff at the SC'19 conference, which included visualizing disk-to-disk transfers between DTNs on the SC'19 show floor to ESnet hosts around the US. However, this functionality only tests the file transfer performance of the perfSONAR testing node, not an actual data source being used in practice. At this time, we do not plan to include this functionality for the IRNC resources.

There were a number of changes to the IRNC perfSONAR infrastructure in Year 5. During Quarter 1, perfSONAR test results from AmLight were successfully re-added back into the latency displays. Due to internal-project work, AmLight is no longer participating in the IRNC perfSONAR mesh, however NetSage is able to still collect their data for the NetSage archive and display it in the other Dashboards. In addition, the perfSONAR nodes at the Guam Open Research and Education eXchange (GOREX) were replaced, restoring the ability to test for throughput and latency to this endpoint.

In Year 6, we plan to investigate ways to more efficiently ingest perfSONAR results. The pScheduler component of perfSONAR now supports pushing results directly to NetSage-like Data Ingest Pipelines as opposed to periodically polling from the perfSONAR archive. This change will reduce the need to support a custom collection script and will also allow for easier addition of new data types in the future by eliminating the need to update an ad hoc script. The perfSONAR project in general is encouraging users to pursue models where data is pushed directly to open source software stacks like ELK, and NetSage is a leader in this space.

5.5 Science Registry

The Science Registry is a system we have developed to document known network endpoints, organizations, and science projects that are users of network resources. The system supports collaborative and crowd sourced data entry and is a key component for finding higher fidelity information about endpoints than what existing MaxMind database can provide. In particular, flows with endpoints in the Science Registry can be tagged with a science discipline.

During Year 5, we continued work on adding additional Top Talkers into the Science Registry, with over 70 new entries. In addition, we worked with the University of Hawai'i Institute of Astronomy to add 37 resources to the Registry, each consisting of multiple IP blocks. As part of this work, we developed a simplified excel spreadsheet as an easier way for outside groups to contribute data.

The data schema for the Science Registry was simplified. The science discipline list now maps to the NSF recognized research areas, with a few additions, for example the use of "Multi-Science Facility" for resources that support multiple disciplines such as supercomputer centers. In Year 6, we will adapt the schema to allow listing sub-organizations to more accurately model cases where a higher degree of accuracy is needed, such as a specific department at a university, without losing the association back to the higher level organization.

5.6 Time Series Data System (TSDS)

The Time Series Data System (TSDS) (<http://globalnoc.iu.edu/software/measurement/tsds.html>) is a software suite that provides well-structured and high performance storage and retrieval of time series

data, including interface throughput rates, flow data, CPU utilization, and number of peers on a router. Along with the raw data, the TSDS suite is capable of tracking and reporting based on metadata, for example viewing interface throughput from the viewpoint of a VLAN or BGP peer session of a particular ASN.

In Year 5, there were seven releases of the TSDS Grafana driver code that included bug fixes and optimizations, as well as compatibility updates for Grafana version 6, which was released in May. In addition to the driver code, the core TSDS packages saw two releases, which improved performance and resolved some edge case bugs, particularly in the case of histogramming.

Year 6 will include updates to be compatible with the next version of Grafana, the 7.0 release expected in May 2020. We anticipate that this will also enable us to use several anticipated features to streamline queries, resulting in faster and more responsive Dashboards.

6. Visualization And Analysis Dashboards

6.1 Development Process and Environment

A primary focus of Year 5 was to make the development and release process more formal and able to produce more reliable products. A detailed process was defined to more efficiently track changes and work on Dashboards in parallel, with additional effort spent on quality assurance (QA). The release process now includes explicit code reviews, a well-defined style sheet to ensure consistency across Dashboards, and iterative walkthroughs of the Dashboards prior to the release.

The Grafana infrastructure used by the NetSage Dashboards was upgraded to Grafana 6, the most recent major release of the Grafana project. This was a large update from Grafana 5 and required multiple code changes to drivers maintained by the NetSage project as well as verification and minor adjustments to the Dashboards. Updates in Grafana also enabled the NetSage team to develop base templates for the Dashboards to more easily and accurately support the integration of individual differences between deployments, such as what the default organization is or the default list of sensors for a deployment. This was rolled out in Quarter 4, and we expect additional functionality will be added in Year 6. In Year 6, we will also update to using Grafana 7 in order to take advantage of the planned functional additions as well as bug fixes. We also plan to pursue contributing back to Grafana some of our additional development work, including the various plugins that are under development.

The NetSage development environment was updated to address the growing complexity of supporting numerous Dashboards and deployments, and to reduce the need for manual maintenance and merges. A portable development environment using Vagrant VM kickstarting was created to enable developers to work on Dashboards from their own laptops. It also incorporates git hub as the version control system.

6.2 Dashboard Releases

There were three major releases of the NetSage Dashboards in Year 5.

NetSage 1.0 was released in February 2019. This was the first release using the updated software development and release process. It contained additions for the Flow Data Dashboards, including one focused on Tstat data gathered at science archives, with the goal of answering the question “What sources/destinations are receiving data from an archive?” In addition to the statistics common to other flows such as Top Talkers by rate and volume, this Dashboard provides additional data including retransmits and round trip time (RTT), which is available from Tstat but not from NetFlow or sFlow. This data can provide deeper insights into whether or not a particular path is experiencing poor performance.

NetSage 1.1 was released in July, and included two new Dashboards: the Science Discipline Patterns Dashboard, with Heatmaps to show the volume of data being moved by different science disciplines over time, and the Individual Flows Dashboard, which displays statistics on individual flows between two selected organizations. The 1.1 release started our process toward unifying the look and feel of the Dashboards by adding descriptive text at the top of each Dashboard, and other unifying clarifications.

NetSage 1.2 was released in November, and was the first release to experience the full QA pipeline now in place, which greatly increased the consistent look and feel across the Dashboards. This release also included navigation between Dashboards for the first time via links to relevant information, which made using the Dashboards for analysis workflows much simpler.

In Year 6, we plan four releases:

- 1.3, in February, will focus on the introduction of a new Flow Analysis dashboard designed to correlate SNMP and flow data as well as providing greater selection criteria, a new visualization to help with understanding top talker pairs, and a host of clarifications and consistency adjustments.
- 1.4, in May, will focus on updates required to get to the latest version of Grafana 6 in preparation for the jump to Grafana 7, as well as major updates to the flow pipeline to optimize the way data is stored and presented.
- 1.5, in September, will primarily include updates for compatibility with Grafana 7, which is planned for release in May 2020.
- 1.6, likely the final IRNC-focused release scheduled for December, which will focus on stability and documentation.

6.2 Released NetSage Dashboards

NetSage Dashboards are designed to answer specific types of questions about the state of a resource. Each Dashboard, divided here into six related Sets, is made up of a suite of complementary components, sometimes called widgets, which are detailed in Section 6.3-6.7. For most of these Dashboards, Year 5 saw only minor stylistic

changes to simplify or clarify the data being shown after user feedback, to update the data to include additional data being collected, or other small changes to make them more consistent with each other.

The current Sets of NetSage 1.2 Dashboards deployed for the IRNC projects include:

- **Set 1: Bandwidth Dashboard:** How heavily used are the circuits and exchange points?
 - <http://portal.netsage.global>
 - Map, showing the path of the data, with information about the current use of the links from SNMP, as well as indications if the relevant pieces of the link are experiencing loss.
 - A line chart for each IRNC circuit showing capacity and usage information from SNMP throughput data.
 - A line chart showing the use of all of the circuits by average and maximum use, in both directions.
- **Set 2: Flow Data Dashboards (by circuit, archive, organization, or country):** What are the top ten senders/receivers of flows (ranked by volume or rate) by source and destination?
 - Top Sources/Destinations for all Flow data:
https://portal.netsage.global/grafana/d/-l3_u8nWk/individual-flows
 - Top Flows by Archive: *This Dashboard was first deployed in Year 5*
<https://portal.netsage.global/grafana/d/mNPduO8mz/flow-data-for-data-archives>
 - Top Flows by Organization:
<https://portal.netsage.global/grafana/d/QfzDJKhik/flow-data-per-organization>
 - Top Flows by Country:
https://portal.netsage.global/grafana/d/fgrOzz_mk/flow-data-per-country
 - Table and Bar chart showing Top Talker data by volume and rate.
 - Top Pairs of Talkers, if appropriate.
 - Summary statistics about ports, protocols, and sensor use.
 - A Sankey graph of Top Talkers, for some use cases, detailed in Section 6.4.
- **Set 3: Individual Flows Dashboard:** For a particular organization or country, what is the per-flow data?
 - Individual Flows by Organization: *This Dashboard was first deployed in Year 5*
https://portal.netsage.global/grafana/d/-l3_u8nWk/individual-flows?orgId=2
 - Individual Flows by Country: *This Dashboard was first deployed in Year 5*
<https://portal.netsage.global/grafana/d/80IVUboZk/individual-flows-per-country>
 - Table listing data specific to individual flows, including subnet for source and destination.

- Heatmap showing data transfers over time, by volume and rate, as detailed in Section 6.3.
- Summary statistics about ports, protocols, and sensor use.
- **Set 4: Flow Data by Science Discipline Dashboard:** What are the flows for a defined Science Discipline (using the NetSage Science Registry)?
 - *This Dashboard was first deployed in Year 5:*
<https://portal.netsage.global/grafana/d/WNn1qyaiz/flows-by-science-discipline?orgId=2>
 - Summary statistics about how many flows were matched in the Science Registry.
 - Sankey graph of Top Talkers, detailed in Section 6.4.
 - Summary statistics by discipline.
- **Set 5: Pattern Dashboards (for Bandwidth, Loss, Latency, Science Discipline):** What are the recurring patterns of behaviors?
 - Bandwidth data from SNMP:
<https://portal.netsage.global/grafana/d/000000004/bandwidth-patterns>
 - Loss data from perfSONAR:
<https://portal.netsage.global/grafana/d/000000006/loss-patterns>
 - Latency data from perfSONAR:
<https://portal.netsage.global/grafana/d/000000005/latency-patterns>
 - Science Discipline data from Flow data collection and Science Registry data: *This Dashboard was first deployed in Year 5*
<https://portal.netsage.global/grafana/d/ufIS9W7Zk/science-discipline-patterns>
 - Heatmaps by relevant item (circuit, exchange point, discipline) showing data use over time, as detailed in Section 6.3.
- **Set 6: Flow Statistics Dashboard:** What are the current flow data statistics?
 - *This Dashboard was first deployed in Year 5:*
<https://portal.netsage.global/grafana/d/CJC1FFhmz/other-flow-stats?orgId=2>
 - Summary statistics for number of flows, source and destination information, and Science Registry use.
 - Summary statistics about ports, protocols, and sensor use

In addition to these Dashboards, we deployed separate Dashboards to visualize the flow data for the Pacific Wave Exchange Point, specifically Sets 2, 3, and 4. While flow data had been collected for Pacific Wave for some time now, we had not been showing it on the IRNC Flow Data Dashboards since it includes a large amount of domestic data that overwhelms the international data. Our original plan had been to take advantage of a planned Grafana update that would allow us to easily tag flows as domestic or international. When this update was not included in the Grafana v6 update, we deployed separate Dashboards specifically for Pacific Wave data instead

of waiting any longer. These Dashboards are available, separate from the standard IRNC Dashboards, at <https://pacwave.netsage.global/>.

In Year 6, the IRNC Dashboards are expected to experience minor updates to keep them consistent with the state of the projects and the other NetSage Dashboards in development. Some Dashboards will add updated displays, as discussed in Section 6.3-6.8. However, the bulk of Year 6 will be spent in increasing the stability of the Dashboards and documenting them for easier maintenance in the future. New Dashboards under development are listed in Section 6.9.

6.3 Heatmaps

We use Heatmap visualizations to show changes in values over time and to easily identify patterns of behavior. Heatmaps are used in Sets 3, 4, and 5.

Heatmaps for Science Disciplines were added in Quarter 2 to display information about Flows that have been tagged with information from the Science Registry on a per-link basis (Set 4). Heatmaps for the Starlight, AmLight, and Pacific Wave Exchange points were added in Quarter 3 to the Bandwidth Patterns Dashboard (Set 5), which had previously only shown information for circuits.

Concerted development during Year 5 was spent to try to produce a visualization that displays a Heatmap showing the transfer volumes over periods of time for multiple countries in a single chart, where the vertical axis represents each country, and the horizontal axis is time. This would enable users to be able to answer the question “What are the recurring patterns of data flow from a source country to a list of destination countries?” and to easily compare transfer transfers and patterns across multiple countries over long periods of time. This visualization was requested by several IRNC projects in Year 4. However, we had limited success due to the limitations with the standard Grafana Heatmap plugin. To address this, a new plugin will be developed during Year 6.

6.4 Sankey Graphs

Sankey graphs show relationships between items using a ribbon graphic, where the width shows the quantity proportionately. We use Sankey graphs to show data flows over the IRNC resources in Dashboard Sets 2 and 4. The widget to support Sankey graphs was fully developed by NetSage team members and will be contributed back to Grafana in Year 6.

In Quarters 1-3, improvements to the Sankey graph plugin included:

- Ability to view data when hovering over a data point.
- More intuitive node header names for all Sankey charts.
- Ability to distinguish one Sankey ribbon from another by having a ribbon highlight when hovered over.
- Adaptations for Sankey graphs to work with additional data types, including the protocol, port number, or science discipline.

In Quarter 4, in addition to significant code cleanup that increased the responsiveness of the plugin, functionality for an optional monochromatic color scheme was added.

In Year 6, we will complete the Sankey plugin development effort by adding user interface features such as mouse hovering to reveal data underneath Sankey paths, as well as bug fixing, optimization, and documentation.

6.5 Flow Data Tables using sFlow, NetFlow, Tstat

In Year 5, data from Tstat was added to the Flow Dashboards (Set 2), including retransmits and round trip time (RTT) in addition to the statistics common to other flows providers. Retransmit and RTT data is available from Tstat but not from NetFlow or sFlow, and can provide insights into whether or not a particular resource is experiencing poor performance.

6.6 Bump Chart

A bump chart, as shown in Figure 3, can answer queries such as “How has the use of a resource changed over time?”. This type of chart provides a longitudinal view that can be used, for example, to show which science disciplines remain as the top users of the network over time.

Work on a Bump Chart plugin started in Year 5. In Year 6, we will complete the Bump Chart plugin and deploy it on a new Dashboard to enable users to see the consistency at which organizations are sending or receiving the highest volume of data over long periods of time, such as multiple years. This may also involve work on the data side to roll up, or summarize, longer-term data for scalability.

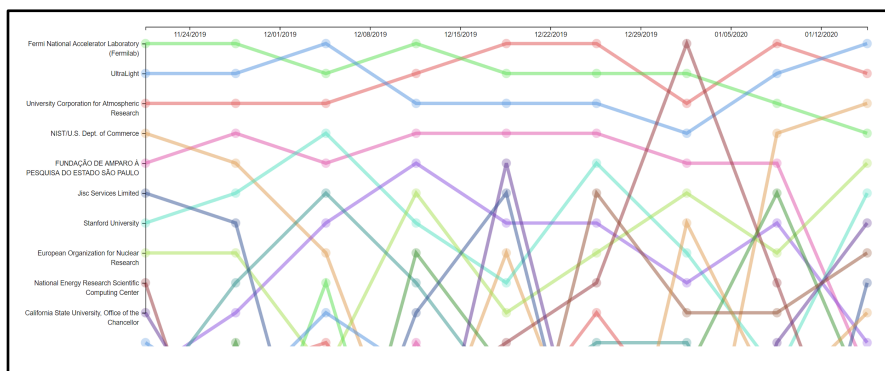


Figure 3: Prototype Bump Chart showing how Top Talkers change their status over time.

6.7 Maps for Arbitrary Locations

A new map plugin to answer the query of “Where science disciplines are sending their traffic to and at what volume” was prototyped. The map uses data tagged by the Science Registry and displays it, as shown in Figure 4. This map was produced

with a different mapping tool than the one used to produce the main bandwidth map so that it could draw arcs between arbitrary map locations automatically. The map tool used in the main bandwidth tool requires developer intervention to specify the control points to draw the arcs properly, in the same way the well known GLIF (www.glif.is) map was produced. This approach cannot be applied to Science Registry data as there are many possible source and destination arcs and it would be impossible to manually specify every combination of them. In Quarter 4, the map layout was streamlined and the performance (visual frame rate) was improved. Both a light and dark color scheme were developed.

In Year 6, this map will be included as part of the Flows by Science Discipline Dashboard (Set 4) in the planned 1.4 release. The map plugin will be contributed back to Grafana.

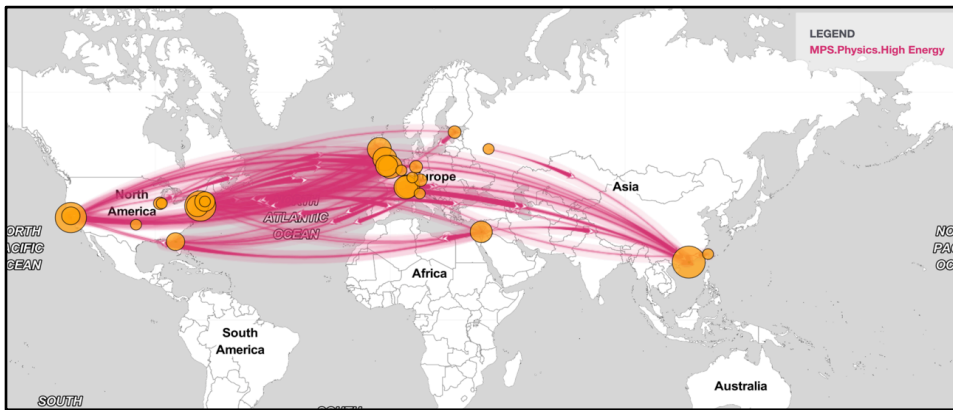


Figure 4: Example of the map with arbitrary endpoints being used to show science disciplines, in this case, filtered high energy physics flows.

6.8 Slope Graph

A Slope Graph is a visual way of showing relationships between two lists. It can contribute to answering questions such as “What are the top ten senders/receivers of flows by source and destination?” In this graph, a darker color saturation for a line indicates a larger volume of data transferred.

During Quarter 4, the graph, as shown in Figure 5, was being prototypes, with the intention of publishing it in Year 6 Quarter 1 as part of the Flow Data Dashboards (Set 2).

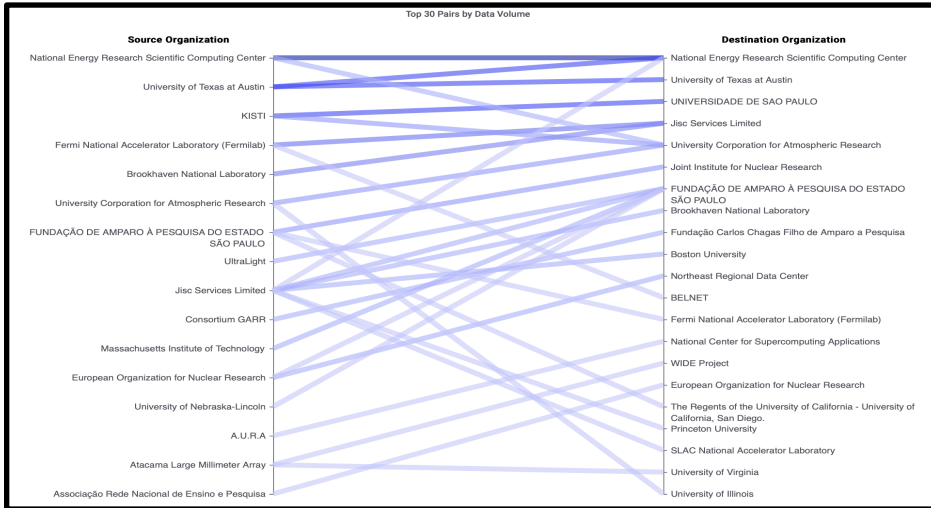


Figure 5: Slope Graph Showing Top 30 Pairs by Volume.

6.9 New Dashboard - Analytics

In Year 5, we began development of a more advanced flow analytics Dashboard to answer the query of “What is this anomaly?”. This Dashboard will display both SNMP and flow data that can be viewed by continent, country, organization, or ASN. The workflow for this Dashboard, shown in part in Figure 6, is expected to involve a user selecting a timeframe of interest on an SNMP line chart, for example to zoom in on a performance spike. This will also adjust the timeframe for the corresponding Flow data graphs, including Top Pairs, Top Sources/Destinations, and Individual Flows information, to show additional detail to help determine what data transfers may be related to the performance change. Further query refinement is possible using supplemental filters to remove a specific value or view only that specific value.

Major development of this Dashboard took place during the last half of Year 5, and it will be released in February as part of the NetSage 1.3.0 release. Further refinement of this Dashboard is planned, specifically in the areas of user interface simplification and improved navigation once it is in use beyond the development team.

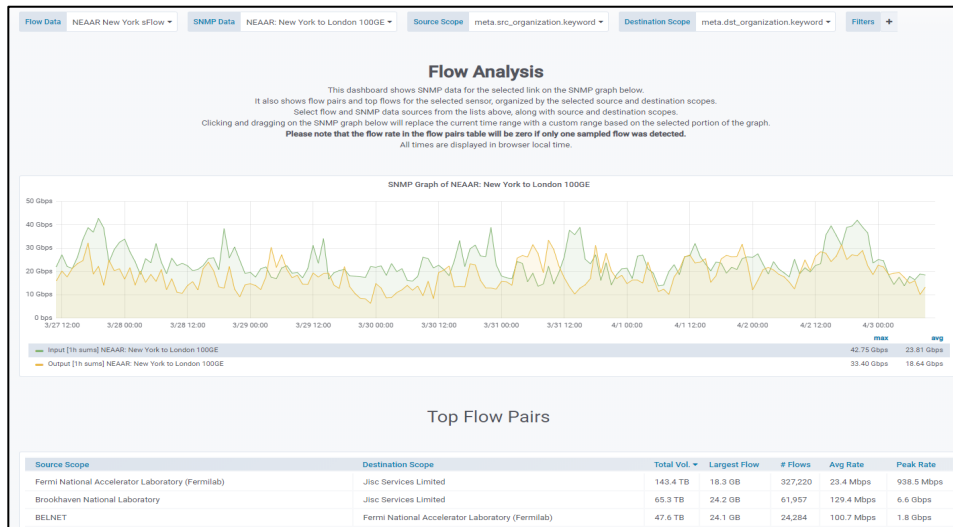


Figure 6: Prototype display for the Advanced Flow Analysis Dashboard.

7. Examining The Effects Of TCP Retransmissions In perfSONAR Test Results

In order to address the question: "how do TCP retransmissions correlate to degraded performance?" a large sample of perfSONAR throughput test results were collected and analyzed, and compared with a series of data transfer tests. It was observed that retransmissions can be linked to a cause of degraded performance, or as a side effect of near-limit performance, or unrelated to performance. Across a spectrum of throughput performance levels, retransmissions appear in distinctly different contexts, and we posit that it may be able to examine different patterns of retransmissions on a transfer to gain diagnostic insight to be used in troubleshooting.

Several networking trends, such as the incipient change from traditional congestion control algorithms (CUBIC, HTCP), to innovative algorithms like BBR (<https://cloud.google.com/blog/products/gcp/tcp-bbr-congestion-control-comes-to-gcp-your-internet-just-got-faster>), and the upgrading of a majority of test and transfer nodes from 10 G to 40G and 100G interfaces will probably affect the study of TCP retransmissions at large.

8. Use of NetSage

In Quarter 2, the NetSage project began collecting data about people accessing and using the Dashboards using Google Analytics. These statistics show that the project has had success in reaching its global community, not only in the number of views but also in how it is being used. Between July 2019 and January 2020, over 1,700 unique users in 89 countries visited the NetSage Dashboards, as shown in Figure 7.

The United States led the way accounting for 50.1% of the Dashboard traffic, followed by Brazil (5.3%), Singapore (4%), France (2.7%), and Germany (2.7%). What is even more promising is that the average user spends over 9 minutes on the site and on average visits 4.59 pages in a single session. This seems to indicate that users are spending time with the site and navigating the available data in several ways for each session, which is the intended use of the NetSage Dashboards. The most visited Dashboard is not surprisingly the initial Bandwidth Dashboard, which is the default starting page for IRNC NetSage. The next most popular Dashboards are those relating to flow data, specifically by organization, by science discipline, the individual flows, and then per country summaries. The success of the Dashboards that make the most use of the data collected by the Ingest Pipeline and Science Registry information, which indicates that the project effort in these areas to make them more efficient is well spent.

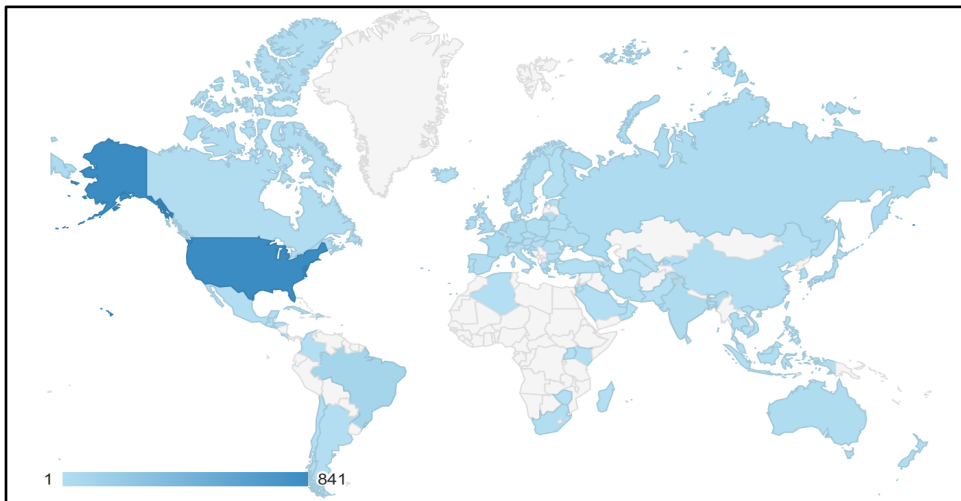


Figure 7: Map of the countries that accessed NetSage Dashboards between July 2019 and January 2020. Darker blue indicates higher usage within that country for individual users.

9. Data Privacy and Security

Basic security measures are being maintained and there were no security incidents to report. As a reminder, NetSage does not collect PII and therefore are compliant with the European General Data Protection Regulation (GDPR).

Beyond the masking of the AARNet data described in Section 4.C, no data privacy or security changes were made during the Year.

10. Updated WBS for Year 5 and Year 6

Item	Y5 and Y6 WBS	Notes
Data Collection	1	
PerfSonar Related Tasks	1.4	Ongoing

Define and deploy PS test mesh for backbones	1.4.2	Completed
Add AmLight resources directly to NetSage archive	1.4.2.12	Completed Y5Q1
Ongoing support for IRNC PS mesh	1.4.3	Ongoing
Update PS Ingest	1.4.4(NEW)	Expected Y6
SNMP related tasks	1.5	Ongoing
SNMP data from Backbones	1.5.2	Ongoing
Tstat/Flow deployment	1.7	Ongoing
Need dashboard for exchange point flow data	1.7.14.6	Data added to existing Dashboards Y5Q3
Lower flow data collection threshold from 500M	1.7.19	Completed Y5Q1
Instrumentation of Data Archives	1.8	Ongoing
Generate an RPM and/or better documentation on how to install tools on archives that will forwards tstat data to IU	1.8.2	Needs final documentation
Deploy Tstat on Hawaiian astronomy archives	1.8.3	Completed Y5Q2
Deploy Tstat on CENIC/PRP archives	1.8.4	OBE
Use top talkers list to identify likely DTNs	1.8.5	OBE
Instrument NCAR Archive	1.8.6	Completed Y5Q3
NASA DTN instrumentation	1.8.7	OBE
Other possible DTNs from IRNC partners	1.8.8	Completed
NOAA Data archives	1.8.8.2	Completed Y5Q1
Additional software framework Upkeep	1.12	Ongoing
TSDS maintenance	1.12.1	Ongoing
Add a keep alive notification for Tstat sensors (modify package)	1.12.5	Ongoing
Data transfer information (ie fiona) as additional data source	1.15	Demo Y5Q4
Find guinea pigs for data transfer inclusion	1.15.3	OBE
Hand off data sets to IRNC owners	1.16 (NEW)	Planned Y6
Dockerize Ingest Pipeline	1.17 (NEW)	Planned Y6
Dockerize Tstat	1.18 (NEW)	Planned Y6
Additional tagging	1.19 (NEW)	Planned Y6
Investigate Ingest Scalability	1.20 (NEW)	Planned Y6
Adapt ingest for additional scalability	1.20.1(NEW)	Planned Y6
Data Rollups	1.21 (NEW)	Planned Y6
Analysis	2	
Data cleaning	2.2	Ongoing

Recreate AS to Science Project Database (Science Registry)	2.4	Ongoing
input data	2.4.3	Ongoing
Get TransPAC to add data to science registry	2.4.3.1	Ongoing
Get NEAAR to add data to science registry	2.4.3.2	Ongoing
Get Ampath to add data to science registry	2.4.3.3	Will query in Y6
Get PIREN to add data to science registry	2.4.3.4	Astronomy data added Y5Q2
Get CENIC to add data to science registry	2.4.3.5	Will query in Y6
Extensions to basic science registry framework	2.4.4	Ongoing
More science disciplines and ability to edit list	2.4.4.6	Completed Y5Q4
More roles and ability to edit list	2.4.4.7	Completed Y5Q4
Notes field for SR	2.4.4.8	OBE
URL field for SR	2.4.4.9	OBE
Admin section functionality for SR	2.4.4.10	Ongoing
Stitching with Elk Stack	2.4.4.12	Expected completion Y6Q1
Rework SR Data model	2.4.4.13 (NEW)	Planned Y6Q2
Transition plan for MaxMind	2.4.4.14 (NEW)	Planned Y6Q1
Dashboards for Tstat data from archive showing retransmit	2.7.1	Release Y5Q2
Updates to Tstat archive dashboard	2.7.1.2	Ongoing
Heatmap for Tstat archive data	2.7.2	Year 5
Americas Greatest Networks - most reliable	2.13.5	OBE
Updates to top X based on flow	2.13.9	Ongoing
Filter out Australian university names from flow data	2.13.9.1	Completed Y5Q4
Analysis of buffer size issues	2.16	Completed Y5Q4
PIREN analysis for astronomy data	2.20.	Ongoing
Dashboard Tasks	3	
Dashboard management tasks	3.11	Ongoing
Develop template for standard dashboard	3.11.4	Expected Y6Q1
Sensor health dashboard	3.11.5	Available Y5Q1
IRNC Statistics dashboard	3.11.6	Completed Y5Q4
Discussions for running as managed service	3.11.7	Completed Y5Q4
Updated for Grafana 7.0 (May 2020)	3.11.8 (NEW)	Planned Y6Q2
Map updates	3.12	Ongoing

Map for science registry data	3.12.2	Started Y5, Compl Y6
Bugs and Fixes	3.15	Ongoing
Dashboard for PS File Transfer data	3.25	OBE
Basic graph dashboard of file transfer data	3.25.1	OBE
Heatmap of PS file transfer data	3.25.2	OBE
Viz for Max sending vs retransmits	3.16	OBE
Sankey Next Steps	3.20.8 (NEW)	Expected submission to Grafana Y6Q2
Analysis Dash	3.26	Expected Y6Q1
Slope Graph	3.27	Expected Y6Q1
Dash for Volume >X, rate <y	3.28	Expected Y6
Navigator Plugin	3.29 (NEW)	Ongoing
Add additional dash to Navigator	3.29.1(New)	Ongoing
Submit Navigator to Grafana after cleanup	3.29.2 (NEW)	Expected Y6
Dash - Top Talkers by Year	3.30 (NEW)	Expected Y6
Dash - Science Registry Projects	3.31 (NEW)	Expected Y6
Dash - Country vs Time	3.32 (NEW)	Expected Y6
Review over the questions that guide Viz	3.21	Completed
Which are still valid?	3.21.1	Completed
Gather additional questions	3.21.3	Completed
Design additional dashboards in order to answer the questions.	3.21.4	Completed
Third party deployments	3.23	Ongoing
ANA Deployment	3.23.1	Ongoing
Add new link to ANA SNMP dashboard	3.23.1.3	Completed Y5Q4
ANA Flow deployment	3.23.1.2	OBE
Help with EPOC deployments	3.23.2	Ongoing
iLight- Flow	3.23.2.2	Completed Y5Q1
LEARN	3.23.2.3	Expected Y6Q1
KINBER	3.23.2.4	Deployed Y5Q4
Asia Pacific Ring deployment	3.24.	OBE
Project Coordination	4	
Project management and coordination	4.1	Ongoing
Weekly project meetings	4.1.1	Ongoing

Refresh NetSage website home page	4.1.2	Ongoing
REU funding for testers	4.1.3	OBE
Coordinate with NOC	4.2	Ongoing
Year 6 reporting	4.17	Ongoing
Y6Q1 report	4.17.1	Planned Year 6
Y6Q2 report	4.17.2	Planned Year 6
Y6Q3 report	4.17.3	Planned Year 6
Y6Q4 report	4.17.4	Planned Year 6
Y6 Annual report (with Y6Q5)	4.17.5	Planned Year 6
Year 6 travel plans	4.18	Ongoing
AHM in June	4.18.1	Completed
SC'20 in November	4.18.2	Completed
Google Analytics	4.19	Ongoing

10. Financials

Table 2 shows the expenditures for Year across the full team.

Item	Univ	Feb 2019	Mar 2019	Apr 2019	May 2019	Jun 2019	Jul 2019	Aug 2019	Sep 2019	Oct 2019	Nov 2019	Dec 2019	Jan 2019	TOTAL
STAFF COSTS (INCLUDING BENEFITS, F&A)														
Schopf, Jennifer	IU	4,918	4,918	4,918	4,921	4,921	5,039	5,039	5,039	5,039	5,039	5,039	5,039	59,873
Moynihan, Ed	IU	2,323	2,323	2,323	2,325	2,325	2,380	2,380	2,380	2,380	2,380	2,380	2,380	28,275
Chevallier, Scott	IU	1,945	1,945	1,945	1,946	1,946	1,946	1,946	1,946	1,946	1,946	1,946	1,946	23,346
Doyle, Dan	IU		1,922	1,922	1,923	1,923	1,940	1,940	1,940	1,940	1,940	1,940	1,940	21,274
IU Dev Team	IU	23,401	11,151	11,212	7,422	7,396	7,486	10,730	10,732	10,815	23,850	23,612	21,487	169,294
Hubbard, Heather	IU	1,287	1,351	1,415	19,487	1,287	1,308	1,381	1,337	1,317	1,977	1,318	1,318	34,783
Lee, Andrew	IU	1,976	1,976	1,348	0	0	0	0	0	0	0	0	0	5,300
Balas, Edward	IU	1,287		0	0	0	0	0	0	0	0	0	0	1,287
Lake, Andrew	LBNL	7,429	7,075	14,602	11,123	11,339	10,801	7,449	6,565	5,735	7,701	8,885	14,057	112,760
Baveja, Sartaj	LBNL	0	0	11,401	10,273	10,246	9,619	5,066	6,757	8,478	8,643	8,801	8,339	87,622
Faci, Samir	LBNL	0	0	0	0	0	0	0	0	0	0	18,596	26,067	44,683
Whinery, Alan	UH					22,163	22,163							44,326
Leigh, Jason	UH					8,557	8,557							17,113
Kanal, Mahesh	UH	3,950	3,950	3,950	4,144	4,144	4,144	4,144	4,144	4,144	4,144	4,144	4,144	49,147
Turner, Katrina	UH					4,144	4,144	4,144	4,144	4,144	4,144	4,144	4,144	33,152
Seto-Mook, Tyson	UH	4,144	4,144	4,144										12,432
TOTAL STAFFING		52,660	40,756	59,179	63,564	80,390	79,527	44,219	44,983	45,938	61,765	80,806	90,881	744,667
Travel and Other Costs (including overhead)														
Travel - Doyle - AHM July 2018	IU					387								387
Travel - Schopf - AHM Jan 2019	IU	6,267												6,267
Equipment to support Tstat in UH	UCD		530											530
Dell memory upgrade	IU				4,945									4,945
OSI optic for HI Tstat	IU				2,607									2,607
Dell - power supply for Tstat ncar	IU				259	259								517
FedEx - Tstat to HI	IU				61									61
fedEx - optic to HI	IU				139									139
FedEx Tstat to NCAR	IU					91								91
Travel-Schopf-Nwave Jul'19	IU						2,899							2,899
Travel - Doyle AHM Jun'19	IU					1,827								1,827
Travel - Schopf AHM Jun'19	IU						2,632							2,632
Travel - Southworth AHM Jun'19	IU						2,149							2,149
Travel - Ensman- AHM Jun'19	IU						2,070							2,070
Travel - Baveja - PS AHM May'19	LBNL					1,639								1,639
Travel - Lake - AHM Jun'19	LBNL					1,973								1,973
Travel - Baveja - AHM - Jun'19	LBNL						3,070							3,070
Travel - Turner - AHM - Jun'19	UH					3,907								3,907
Travel - Kanal - AHM - Jun'19	UH					3,203								3,203
Travel - Leigh - AHM - Jun'19	UH					2,621								2,621
Travel-Whinery-AHM-Jun'19	UH					2,815								2,815
Elastic SW	IU										65,934			65,934
Travel - Baveja - AHM Jan'20	LBNL												2,880	2,880
Travel - Lake - AHM Jan'20	LBNL												2,131	2,131
Travel - Faci - AHM Jan'20	LBNL												3,326	3,326
Travel - Turner - SC 2019	UH										4,614			4,614
Travel - Leigh - SC2019	UH										0			0
Travel-Tierney-TechEx Dec'19	IU												449	449
TOTAL TRAVEL		6,267	530	0	8,010	18,722	12,820	0	0	0	70,548	0	8,785	125,681
PARTICIPANT SUPPORT														
Travel-Tierney-SC-Nov18	IU		1,514											1,514
IRNC PI Mtg- 12-Mar19	IU				3,154									3,154
Hospitality - NOAA Boulder							65							65
Hospitality - TechEx													731	731
TOTAL PARTICIPANT SUPPORT		0	1,514	0	3,154	0	65	0	0	0	0	0	731	5,464
EQUIP OVER \$5K														
Tstat box for HI	UCD			68,978										68,978
TOTAL EQUIP		0	0	68,978	0	0	0	0	0	0	0	0	0	68,978
TOTAL EXPENDITURES		58,927	42,800	128,157	74,727	99,112	92,411	44,219	44,983	45,938	132,313	80,806	100,397	944,790