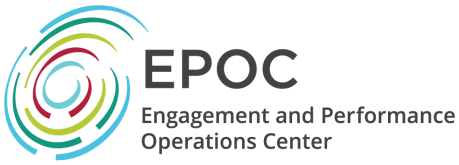
EPOC Network Data Privacy Policy

Contact Point: Jennifer M. Schopf (jmschopf@iu.edu)
Audience: General
Last updated: August 17, 2022



ABOUT EPOC

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 was established in 2018 as 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.

SCOPE OF THIS POLICY

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.

EPOC COLLECTION OF NETWORK DATA USING NETSAGE

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.

This material is based upon work supported by the National Science Foundation under Grant No. 1826994

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.

The NetSage tool captures and collects active networking data, including latency and throughput from perfSONAR testpoints, and passive network metadata, including SNMP and flow data from routers or switches. 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 as part of the NetSage Ingest Pipeline (see Figure 1) before being stored in the NetSage Archive. Following the current accepted practices in the R&E community, we remove the low order bits from the IP address for any flow-related data. For IPv4 this includes the lowest 8 bits. For IPv6 this includes the lowest 64. In doing so, we remove any personally identifiable information (PII) from the information that will be stored. This approach also meets the guidelines for the European Union's General Data Protection Regulation (GDPR).

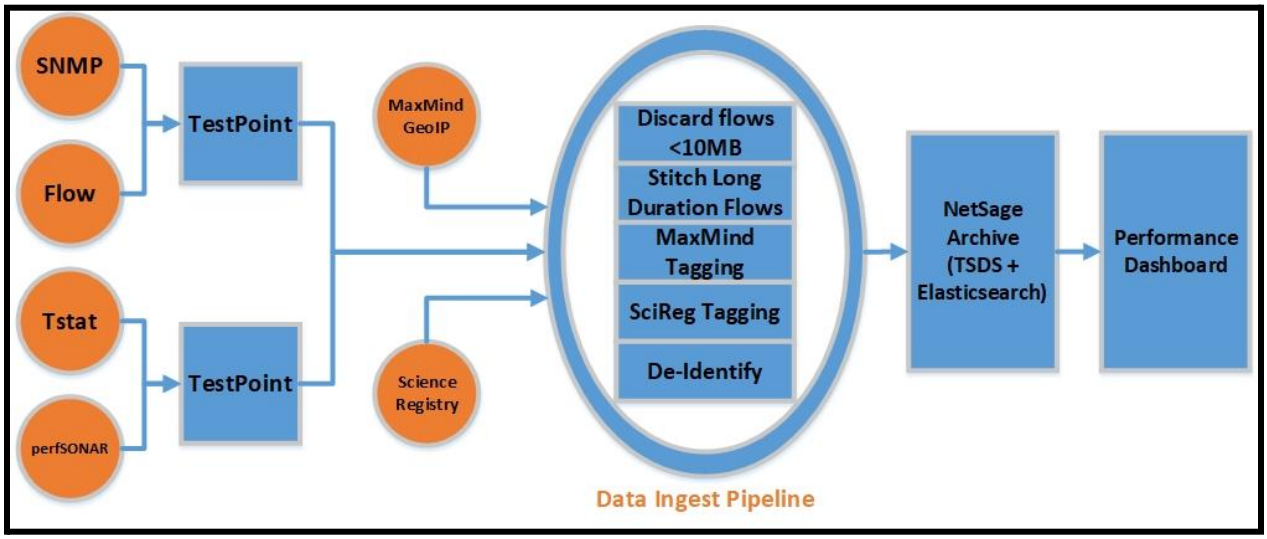


Figure 1: The NetSage logical architecture.

This material is based upon work supported by the National Science Foundation under Grant No. 1826994
info@epoc.global

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 makes summaries of de-identified Regional Partner network traffic data public on various NetSage Portals and Dashboards. These include:
    a. Front Range GigaPop (FRGP) - https://frgp.netsage.global
    b. Great Plains Network (GPN) - http://gpn.netsage.global
    c. iLight/Indiana GigaPop - http://ilight.netsage.global
    d. Lonestar Education and Research Network (LEARN) - https://learn.netsage.global
    e. Pacific NorthWest GigaPop/PacificWave - https://pacwave.netsage.global
    f. Southern Crossroads (SoX) - https://sox.netsage.global
    g. Texas Advanced Computing Center (TACC) - https://tacc.netsage.global
    h. Plus archive data from NCAR, NERSC, TACC and U. Hawaii – Archive data

2. Upon request, Regional Partners can have access to the full, de-identified data sets for their site.

If any other third party wishes to obtain access to the data directly from the NetSage archive, the EPOC team refers this request to the Regional Partner.

HOW DATA IS 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.

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. The Privacy Policy posted here will always be current. We encourage you to review this statement regularly.

This material is based upon work supported by the National Science Foundation under Grant No. 1826994
info@epoc.global

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:

- **Front Range GigaPop (FRGP),** the regional collaboration of networks that cover the western states, including Colorado, Wyoming, Arizona, Idaho, Utah, and New Mexico.
- **The Great Plains Network (GPN),** the regional network that serves North Dakota, South Dakota, Nebraska, Iowa, Minnesota, Kansas, Missouri, Oklahoma, and Arkansas.
- **iLight,** the regional network for Indiana.
- **The Keystone Initiative for Network Based Education and Research (KINBER),** the regional network for Pennsylvania.
- **The Lonestar Education and Research Network (LEARN),** the regional network for Texas.
- **National Oceanic and Atmospheric Administration (NOAA) N-Wave,** the R&E network for NOAA. Joined Year 4 Quarter 1.
- **The Ohio Academic Resources Network (OARnet),** the regional network for Ohio.
- **Pacific Islands Research and Education Network (PIREN) and the University of Hawai'i System Network,** which provides R&E network capacity to interconnect Pacific Islands with each other and the global R&E network fabric, including links to Australia and Guam, in addition to connectivity for the University of Hawai'i system and Mauna Kea and Haleakala astronomy observatories.
- **Pacific Northwest GigaPop (PNWGP),** which provides access to next generation internet services and technologies throughout the Pacific Rim, but in the US primarily in California, Oregon, and Washington State.
- **Southern Crossroads (SoX),** the regional network for much of the southeastern part of the US, including parts of Alabama, Georgia, South Carolina, and Tennessee.
- **Sun Corridor Network (SCN),** the regional network for the state of Arizona.
- **Texas Advanced Computing Center (TACC),** an advanced computing research center

This material is based upon work supported by the National Science Foundation under Grant No. 1826994
info@epoc.global