

# NetSage Project Network Data Collection Memorandum of Cooperation

This Memorandum of Cooperation is between the NetSage project, and your entity that hosts a monitor(s) for collecting traffic data (Partner Hosting Site).

NetSage is engaging with NSF International Research Network Connection (IRNC) participants to host active and passive monitors as a vital part of NetSage's data collection to better understand the traffic on the NSF-funded IRNC links. The IRNC NOC, which is the NSF-funded network operations center for all IRNC-funded projects, Partner Hosting Site, and NetSage acknowledge that access to this critical data is much needed for the purposes of advancing distributed science. Just as other large scale NSF facilities capture data about their use for science, NetSage will be measuring flows to understand how IRNC investments are facilitating research.

In consideration for your Partner Hosting Site's consent and authorization to share data collections with NetSage for the purposes set forth above, NetSage agrees to the following provisions:

1. While raw data may contain Personally Identifiable Information (PII), data that is archived in the IRNC NetSage/NOC Archive will be de-identified.
2. Passive monitors at Partner Hosting Sites, such as Tstat or sFlow, collect packet headers. A small number of IRNC NetSage and IRNC NOC project personnel trained in protecting user privacy and secure handling of data will have accounts on these monitoring hosts. Any software tools run on a test point will run only strictly necessary services and will be kept up-to-date with security patches and operating system upgrades to minimize security risk.
3. Active monitors at Partner Hosting Sites will be deployed in a manner sensitive to existing traffic and other active testing as to not cause hardship or overwhelm available bandwidth. Any software tools run on a test point will run only strictly necessary services and will be kept up-to-date with security patches and operating system upgrades to minimize security risk.
4. No packet payloads will be recorded without specific permission from the Partner Hosting Site. Because packet headers have dynamic lengths, a few bytes of payload may be initially recorded during an attempt to capture the full length of packet headers, but this information will be filtered and discarded. Raw data on a host site will reside there no more than 24 hours (or less if the Hosting Site Owner requires).
5. De-identification of data will be performed by truncating all IP addresses to remove at least the last 8 bits for IPv4 or 64 bits for IPv6. More bits can be removed if requested by the Partner Hosting Site Owner. This takes place

- before the data is shipped to the IRNC NetSage/NOC Archive.
6. Flows smaller than a set threshold (currently 500M) will also be discarded before the data is shipped to the IRNC NetSage/NOC Archive.
  7. Upon request, Partner Hosting Site Owners can have access to the full, de-identified data sets for their site.

NetSage makes summaries of de-identified Partner Hosting Site network traffic data public on the NetSage portal (<http://portal.netsage.global>).

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 IRNC NOC may, for a limited time, use raw data at the collection point to help identify ongoing performance problems between two sites. Should this occur, a Hosted Site Owner will also be informed, and the raw data files will be subsequently destroyed when the issue has been resolved.

The NetSage project recognizes the need for progress on greater understanding of NSF-funded links to support large-scale science. The access you provide to real data supports this pursuit, and assists in the rendition of your services as critical information infrastructure.

If you have any questions about this document, please contact International Networks at [INTL-l@iu.edu](mailto:INTL-l@iu.edu).

## **Glossary**

Networking data: Active — Network data collected using tools 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

Partner Hosting Site — IRNC-funded projects including the NOC, exchange points, and backbone providers. This includes AmLight, TransPAC, PIREN, StarLight, AmPATH, the PacificWave Exchange, and the IRNC NOC

Partner Hosting Site Owners — the PIs of the IRNC-funded Partner Hosting Sites