# Data and Network Integration

Infrastructure Integration Roadmap Task

**Task Type(s)**: Technology, Support  
**Start by phase**: Integration  
**Complete by phase**: Operations, Ongoing  
**RP role(s)**: Data storage, networking, and sysadmin contact(s)

## Summary

Meeting data transfer requirements requires an understanding of storage system, application requirements, and site network connectivity. While available data transfer options are determined by site preferences, ACCESS Networking and Data Transfer Services (NDTS) network engineers and data transfer specialists will continue to facilitate those options by: 1) offering a Globus subscription through at least Project Year 1 (ending 31-Aug-2023); 2) identifying, evaluating, and encouraging the use of promising data transfer applications; 3) providing the CONECTnet Internet2 overlay network to improve connectivity between RPs, and 4) consulting with sites upon request to help them tackle data transfer application and infrastructure challenges.

## Prerequisite tasks

None

## Support Information

For assistance with this task see the *Support Information* section in the *Integration Roadmap Description*.

## Detailed Instructions

### Understand your requirements

To help in planning for adequate data movement and networking capability, the following aspects of your resource, cyberinfrastructure, and applications should be considered.

#### Storage and file transfer applications

1. What part(s) of your resource needs to be accessible for data transfer into and/or out of your site?
   1. The RP should identify what resources it wishes to make available over CONECTnet. Ideally these are just “ACCESS resources”, probably including a perfSONAR network performance measurement server. However, it is not inappropriate to include the entire Science DMZ.
2. For bulk data transfer (datasets in ~100GB range and larger on ACCESS CONECTnet):
   1. Describe your storage system characteristics, types, and partitions
      1. POSIX filesystems are currently the norm in ACCESS, but some new resources have other storage—tape archives, object stores, cloud storage—so please think about which should/should not be available for ACCESS data transfer. Are any of these parallel/high-performance filesystems, such as HPSS, Lustre, or GPFS?
   2. Do you have a throughput goal for data transfers?
   3. What transfer throughput do you anticipate your ACCESS users will need in order to accomplish their research objectives?
   4. Does your site have a [Science DMZ](https://fasterdata.es.net/science-dmz/) for supporting bulk data transfer?
3. Other than bulk data transfer, does your resource provide applications that have particular network throughput requirement characteristics, e.g., real-time streaming data, interactive, streaming video, etc?
4. Do any of these applications require special network features (e.g., QoS, reservations)?
5. Does your site currently, or are you willing to, host a network performance monitor (perfSONAR) to help ACCESS monitor and diagnose end-to-end network behaviors?

### Data and networking survey and consultation

The ACCESS ecosystem offers a Layer3 VPN (L3VPN) provisioned on Internet2 (CONECTnet) to provide connectivity between RPs. Participation in CONECTnet provides valuable performance metrics collection/reporting as well as availability statistics and error reporting. NDTS will be working with Internet2 to identify and expand the available metrics in an effort to increase our visibility and understanding of the types and volumes of traffic traversing CONECTnet.

1. Please contact the NDTS team t3-ndts@access-ci.org to inform us that you intend to connect to CONECTnet
2. We will send you a link to a network connectivity survey. Your responses are valuable to help us understand your site’s connectivity and data transfer priorities.
3. When you complete the survey we will schedule a consultation to review the information and follow up on any questions.

### Network Connectivity - WAN connections

#### To integrate with CONECTnet follow these steps:

1. The RP should identify a router (RP router) “suitably close” to their ACCESS resources. The definition of “suitably close” is very site specific. Members of the NDTS team are available to help with this determination.
2. The RP, possibly in conjunction with their campus and/or Internet2 connector (i.e., their regional network provider), needs to configure a VLAN from the RP router to the interface on the router that peers with Internet2. Note the VLAN tag value. Jumbo frames (9000-octet IP MTU) **should[[1]](#footnote-0)** be supported by the devices in the VLAN.
3. The RP (or its campus or Internet2 connector) needs to set an ACL in OESS to allow access by the CONECT workgroup.
4. The RP needs to pick IP addresses for the point-to-point connection between the RP router and the Internet2 router. Either public or private (ULA/RFC 1918) addresses will work, as will any convenient prefix length (for example, PSC and NCSA both use /127 and /31). Clearly, IPv4 needs to be configured; IPv6 is optional but recommended.
5. The RP now needs to pass along to the CONECT NDTS group the IP address block(s) for the point-to-point connection (indicating which addresses are for the RP end and which are for the Internet2 end), along with the VLAN tag and an autonomous system number (ASN; either public or private) for BGP peering. Optionally, a password for BGP peering can be specified, as can the desire to use BFD (bidirectional forwarding detection).
6. The CONECT NDTS group will provision the connection on the CONECTnet L3VPN.
7. The RP can now bring up BGP peering. Note that no prefix filtering is done on the network.
8. If the RP wishes to configure connections to multiple Internet2 routers, the above steps can be repeated for each connection. The RP is responsible for any traffic engineering (eg, local preference or MED).
9. The RP…

### Network Performance Measurement

(will add text pointers to existing perfSonar documentation)

### Data Transfer

The supported file transfer applications for ACCESS are currently scp, sftp, rsync, and Globus. scp, sftp, and rsync are commonly used file transfer applications. Other than verifying that the servers are running (system administrators) and accessible (network engineers for site firewalls; system administrators for IPtables, firewalld, etc.), these applications likely will not require special configuration and management for use by ACCESS participants.

Globus requires specific system, application, and (potentially) hardware configuration. Please see the ACCESS CONECT document [*Deploy Globus Endpoint*](https://docs.google.com/document/d/19xv0ahgH8m4pFsu5LabYdOVSaNjmB6Ja1Q7I7cc_dM8/edit?usp=sharing) for detailed guidance.

## Document Management

**Status**: Draft

**Official date**: <mm/dd/yyyy>

**Coordinators**: Kathy Benninger, ACCESS Operations

**Last revised date**: 3/11/2023

**Retired date**:

1. If jumbo frames are not configured, Path MTU Discovery must be enabled across the end-to-end path. [↑](#footnote-ref-0)