

### ***What is NKN?***

The Northwest Knowledge Network (NKN) is a unit within the University of Idaho (UI) Office of Research and Economic Development (ORED) that provides research data management and computing support for UI researchers and their regional, national, and international collaborators. NKN operates as a partnership between ORED, the UI Information Technology Services Unit, and the UI Library. In January 2019, NKN merged with the UI Institute for Bioinformatics and Evolutionary Studies (IBEST) now operates as a third core facility within that institute.

The NKN technical team is comprised of software developers, systems administrators, web developers, and metadata experts that are actively enabling data-intensive science through the development and deployment of innovative technologies, services, and infrastructure.

### ***NKN Core Equipment and Facilities***

NKN maintains two identical cyberinfrastructure deployments at distinct geographic points in Idaho: one at the UI Library in Moscow, Idaho and another at the Idaho National Laboratory (INL) in Idaho Falls, Idaho. These two NKN facilities are connected at 10 gigabit-per-second (Gbps) speeds through the Idaho Regional Optical Network (IRON). Each server cluster consists of a small number of enterprise-grade servers with multiple 10 Gbps network interfaces. NKN physical servers are directly interconnected over a local dedicated 10 Gbps managed switch. All NKN servers utilize a dedicated, multi-path, high-performance iSCSI Storage Area Network (SAN) comprised of NetApp FAS 2554 storage arrays. In total, NKN manages more than a *petabyte* (PB) of storage between UI and INL that can be easily expanded to several PB without investments in additional network cyberinfrastructure.

Both server clusters are configured in a flexible, highly virtualized environment using the commercial VMWare vSphere/ESXi hypervisor that allows for dynamic resource provisioning and allocation, monitoring, and remote administration. Each NKN server/storage cluster currently resides behind an enterprise-grade 10Gbps WatchGuard XTM 1535 firewall managed directly by NKN staff. Research data is both synchronously and asynchronously replicated between clusters for the purposes of redundancy, disaster recovery, and load balancing. NKN virtual machines can also take advantage of a separate high-performance Science DMZ network operating at 10Gbps.

NKN currently manages over 60 virtual machines (VMs) for a variety of purposes, installed with either Red Hat Enterprise Linux 6/7 or Windows Server 2008/2012 operating systems. The core NKN services provided by these virtual servers include: multiple ESRI ArcGIS geospatial servers, file sharing servers, dedicated data transfer nodes (DTNs), the NKN data sharing portal (<http://www.northwestknowledge.net>), private cloud storage via OwnCloud, metadata harvesting, web servers, database servers, THREDDS/OPeNDAP, and various development and testing servers.

NKN is actively engaged with national and international scientific data management initiatives such as NSF DataONE, CUAHSI, and EarthCube. NKN operates a tier-4 member node in the NSF DataONE project, connecting our data management repository to similar repositories at a global scale. Through our involvement with DataONE, data published at NKN can be automatically

replicated to other DataONE member nodes, increasing the exposure, resiliency, accessibility, and discoverability of these data. NKN contributes many datasets and over 5TB of storage capacity to the growing DataONE network.

In addition, NKN provisions and hosts dedicated virtual machine servers to research projects in order to extend delegated, controlled access to our compute and storage infrastructure. In many cases, this allows research projects to manage their virtual server environments to meet their project needs without having to purchase, deploy, and manage their own physical hardware.

All NKN server, network, and storage infrastructure is administered by dedicated NKN systems administrators that work in collaboration with the UI Information Technology Services (ITS) units, the Idaho National Lab, and IRON technical personnel. New NSF-funded network infrastructure upgrades ensure that NKN servers and data can be accessed at 10 Gbps network speeds to serve data and provide virtual server access to any location with similar network capabilities. NKN deployed two “Science DMZ” networks at both the University of Idaho and Idaho National Laboratory. These Science DMZ networks allow for unfettered, low-latency, high-throughput access to our growing data collection and associated web services.

### ***NKN Facilities for Interns and Trainees***

Interns and trainees working with NKN will be given workspace and personal computing equipment during the duration of their tenure with NKN. These workspaces will be co-located along with all NKN staff and students in the University of Idaho Library and each will include a desk, chair, network access, laptop, monitor, keyboard, and other peripherals as needed. Interns and trainees will be given appropriate access to all NKN data management cyberinfrastructure (as described above) and will be invited to participate in all bi-weekly NKN Staff Meetings and Technical Talks. The shared office space includes two networked printers, conference table, video projector and audio equipment, office supplies, and more.

### ***NKN Data Management***

NKN manages a freely available public website that allows researchers to upload data, document the data using a standard metadata format, and subsequently publish these data into the public domain. NKN provides a powerful public search interface for discovering datasets within its catalog. Published data at NKN can be associated with an official and unique Digital Object Identifier (DOI), which enables that data to be permanently citable.

All data contributed to NKN is accompanied with descriptive, standards-based documentation (metadata). Data and associated metadata are backed up frequently and replicated to NKN storage infrastructure hosted at the DoE Idaho National Laboratory (INL) in Idaho Falls. Finally, NKN operates as a member node in the NSF DataONE project, allowing NKN to replicate and disseminate data across a global network of distributed data repositories. This comprehensive fault-tolerant storage architecture ensures the long-term persistence of all data hosted at NKN and increases the exposure and discoverability of these data.

For some data types (images, raster data, hydrologic point source, and more) NKN can provide value-added web services to expose these data via standards-compliant web services. NKN fully supports a wide array of Open Geospatial Consortium (OGC) web services including WMS, WCS,

WFS, CSW, and more. NKN also operates a THREDDS OPeNDAP server for providing interoperable access to NetCDF-formatted data.

***Staff Resources to Conduct Stated Work Objectives***

Dr. Luke Sheneman is the NKN Director and serves as principal technical architect for NKN and directs day-to-day activities including all NKN personnel and projects. Dr. Sheneman has a Ph.D. in Bioinformatics and Computational Biology from the University of Idaho.

NKN maintains a diverse and talented team of programmers, data managers, systems administrators and web developers in close collaboration with technical personnel from the units of IBEST, UI Library, Information Technology Services, Idaho Regional Optical Network, and Idaho National Laboratory. A full description of NKN, including staff profiles, current projects, and service center rates is available on the NKN website at <http://www.northwestknowledge.net>.