



CASE STUDY:

Design, Deploy, and Support a National Scientific Research Network and High-Performance Computing Environment



Challenge

A federal government agency was having problems supporting a large number of research laboratories across the U.S., each with significant, growing computational requirements and aging infrastructure. The agency knew that in order to fulfill its mission into the future, it needed to address these problems. The solution: deploy an advanced national scientific research IT network with high-performance computing capability and dedicated scientific IT support.

Approach

During this multi-year engagement, BioTeam provided consulting, engineering, and implementation support. In collaboration with the organization's IT leadership, scientists, and third-party vendors, BioTeam designed, built, maintained, and further expanded a high-speed national scientific research network, a high-performance computing infrastructure, and dedicated scientific IT support services for researchers.

PHASE 1: Needs Assessment

The first phase consisted of assessing the needs of multiple agency stakeholders (scientists, leaders, IT staff) across the U.S. to understand their requirements given barriers faced as production of research data increased in volume and complexity.

- Conducted extensive interviews to understand the limitations of the current network and computing infrastructure from the IT and the scientists' perspectives.
- Identified and illustrated the specifics of limitations. Findings indicated that:
 - Existing networks were not designed to provide the performance and capacity that the next frontier in research required to manage, store, and interpret Big Data.
 - Security policies that may not be necessary for the scientists' research were inhibiting internal and external collaboration efforts.
 - A large number of agency researchers across the U.S. would benefit from access to high performance computing (HPC).
- Provided a vision for how to overcome barriers in the form of recommendations. BioTeam recommended the creation of a new scientific network and computational center that should:
 - Operate as a high-speed and highly scalable scientific network that will span all the organization's locations.
 - Operate with higher efficiency security policies than the current networks in order to enable collaboration and maintain clean network paths for big data.



- Exist as a physically separate network from other networks to ensure continued compliance of other networks with existing security policies and ensure that those policies did not need to be imposed on the scientific network to maintain its high-performance networking environment.
 - Function as a network for scientific use only and carry non-classified, non-sensitive, and non PII-related data due to the implementation of alternate security policies (i.e., non-firewalled access to external resources)
 - It will provide high-speed access to external computational and data resources (i.e., NCBI, XSEDE).
 - Create a state-of-the-art computing cluster that could be accessed by any agency laboratory.
 - Provide agency researchers with dedicated scientific IT support and training.
- Created a design and timeline that contained detailed plans for the implementation for each of three assessment topics (Research Computing Support, Hybrid HPC Cloud, Science Network), including: complete designs, bills of material, costs, and procurement data for hardware implementations and modeling, service, and staffing strategies.
 - Led the development of an actionable and detailed methodology for the implementation of each of the major recommendations outlined in the Needs Assessment.
 - The final deliverable of this phase was a single cohesive document that outlined the detailed implementation plan for each of the three major recommendation areas from the assessment.

PHASE 2: Science Network Design and Deployment

BioTeam led the design and initial deployment of the agency's science network. BioTeam's networking and HPC engineers, bioinformatics and scientific IT consultants led the design, deployment, integration, and testing of the science network infrastructure, equipment, software, and services recommended in Phase 1.

At the core of the science network was a national high-speed network connecting six major data-collecting locations throughout the organization at speeds between 10 and 100 Gbps. The network facilitated data transfer between sites, to the HPC core facility, and external collaborators.

The scientific network also included the following:

- A state-of-the-art high-performance computing (HPC) core facility to support the scientific IT needs of agency scientists. The HPC system included a wide range of pre-installed software tools and applications in genomics, image analysis, or data processing that agency scientists could use for their research.
- A petabyte-scale on premises data storage infrastructure directly connected to the HPC system that agency scientists could use to store and analysis the data.
- A virtual IT and research support core staffed by BioTeam consultants and also supported by a third-party vendor provided agency scientists with bioinformatics, data management, and data analysis support and was also responsible for installing and maintaining scientific applications on the HPC system. The research support core also provided researchers with scientific IT training and collaborated closely with agency labs on onboarding them to the science network.
- Developed and maintained a Science Gateway based on the Galaxy platform that allowed agency scientists with a web-based and easy-to-use mechanism to conduct scientific analyses on the HPC core.
- Provided agency leadership with ongoing scientific IT and strategic advice to maximize the impact of the science network for agency researchers.

PHASE 3: Science Network Support and Expansion

BioTeam was responsible for operating and expanding the science network. Throughout this phase, we worked closely with the agency on transitioning long-term operations and support of the science network to agency staff and third-party vendors. BioTeam provided consulting, engineering, and implementation support for continued science network maintenance, operation, buildout, and enhancements including:

- Support and continual enhancement of the high-speed network backbone, network border security, HPC core, and scientific data storage infrastructure.
- Deployed data transfer nodes to support efficient data transfer between the HPC core and agency labs connected to the science network.
- Continued development of the Galaxy-based Science Gateway working toward seamless integration into the HPC core.

- Developed a dedicated AWS-based Cloud platform integrated into the science network and worked closely with several agency labs to develop scientific services in AWS. BioTeam also provided dedicated Cloud training and support services for agency scientists.
- Provided scientists with bioinformatics, data management, and data analysis support and continued support for installation, integration, optimization, and support for scientific applications on the HPC core.
- Training and support of scientists, and continued coordination with leadership, governance bodies, and selected support personnel for operations of the virtual research support center.
- Provided agency leadership with ongoing scientific IT and strategic advice to maximize the impact of the science network for agency researchers.

Outcomes

BioTeam was instrumental in helping the agency develop, deploy, and operate a national science network. Previously, agency scientists could not efficiently share and analyze their scientific data sets, greatly hampering their ability to fulfill the agency's mission.

The deployed high-speed network connects agency locations across the US and has allowed scientists to efficiently transfer and share their ever-growing data sets with their peers and move them to centralized storage and computing resources.

In addition, the science network offers agency scientists access to centralized data storage and a range of HPC and science gateway resources for data analysis and modeling.

The science network's research support team maintains the infrastructure and scientific applications and offers scientists support, advice, and training for their scientific IT needs. In active use at the agency, the science network continues to be expanded.

Find more case studies at bioteam.net/tag/case-studies/