

Making Big Data Accessible and Reliable for Data-Intensive Science

The greatest challenge for 21st-century science is responding to the new era of data-intensive science, a new paradigm beyond experimental and theoretical research and simulations of nature, requiring new tools, techniques, and ways of working.

- Douglas Kell, University of Manchester

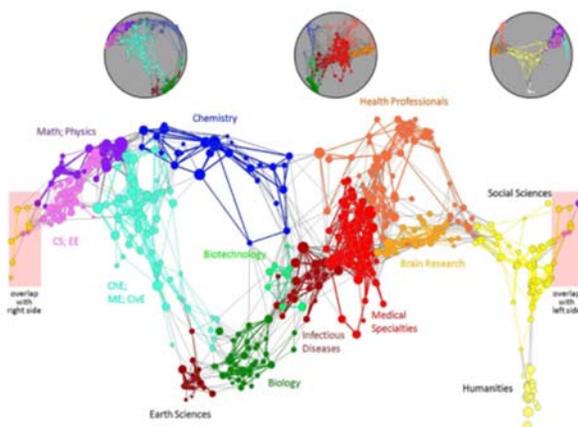
A data-rich world and data-intensive science environment have opportunities and challenges that public institutions need to address. Researchers want to combine big data from different sources and of different types and scales to increase the power of science for understanding complex issues. Researchers, decision-makers and policymakers want to better understand possible outcomes and consequences of complex scenarios by visualizing vast data collections. To accomplish these new levels of inquiry and application, researchers, policymakers and the public need to know the value and quality of data sets to ensure credible science and informed decision making. To access and comprehend big data, users must be able to leverage significant existing cyber-infrastructure, software tools and technical expertise.

Public institutions of research and education need to provide infrastructure to support users in reliably accessing and using large data collections. Anticipating the potential for big data to impact businesses, agencies and science, the University of Idaho stepped out as a leader among universities nationwide, launching in 2010 the *Northwest Knowledge Network* (NKN, www.northwestknoweldge.net). NKN is a suite of secure data storage and software services, data-sharing partnerships and fundamental data-intensive science research that facilitates use of big data techniques by researchers and decision-makers in Idaho, the region and beyond. In short, NKN helps people apply large-scale data and tools to address complex societal challenges.



The Power of Partnerships and Networks

Through networks, NKN serves researchers at the University of Idaho and Idaho National Laboratory, and also provides services and access to big data for users at regional, national and global scales. NKN uses innovative software and data architectures, policies and partnerships to help researchers collaborate within networks of scientists, distributed data storage and high-performance computing systems. This adds capacity to Idaho's research enterprise and supports the needs of business, agencies and government in Idaho and beyond. NKN serves public, non-profit and private enterprises, occupying a niche that brings



Scientists now work in networks that link expertise and data across disciplines to better address complex problems.

concepts from the private sector's dynamic, cloud-based data-management services to researchers and stakeholders.

Idaho National Laboratory provides hardware and software backup for NKN. The Idaho Regional Optical Network (IRON) partners with NKN to provide high-throughput networking across Idaho and large regional and national fiber network backbones. NSF Idaho EPSCoR collaborates with NKN on statewide cyber-infrastructure strategic planning that bridges Idaho's research universities and state government.

Investing in the future

Increasing demand for data services is driven by new federal requirements for data curation, and by scientist and practitioner desires to fully utilize and apply big data. To accommodate this demand and sustain NKN's leadership in the data-services realm, we need to make strategic investments that yield direct benefits in the competitiveness and performance of businesses, agencies and educational institutions in Idaho and the broader region. The next best investments in Idaho's data-management capability should be in:

- Internet Connectivity - Invest in the growth and development of the high-speed fiber networks that connect Idaho universities, schools, agencies and businesses and connect Idaho to the world. NKN partners with Idaho IRON on continued improvements to statewide internet connectivity.
- Data and Metadata Quality - Invest in hardware, software and software development to build secure storage platforms that emphasize quality metadata. This will ensure that original research products, funded by public dollars, are protected and available for reuse by everyone. NKN collaborates with the USDA, USDI and other agencies to develop common policies and systems for data management.
- Data Security - Invest in technical staff to build and maintain robust security systems for the long-term protection of highly valuable data products.
- Data Interoperability - Invest in cluster hires of software engineers, domain scientists and data scientists who develop innovative methods for combining large data collections across disciplines and spatial and temporal scales. NKN, Idaho EPSCoR and INL have begun building a core group of data scientists with a focus on data interoperability.
- Visualization and virtualization - Invest in developing and deploying technologies that demonstrate the potential impacts and outcomes of management and policy scenarios. NKN is cooperating with grant-funded projects across the state to link data-management systems with demonstration technologies for use by decision-makers, policymakers and stakeholders.



For more information, please contact:

John K. "Jack" McIver, Vice President for Research and Economic Development
vpresearch@uidaho.edu | 208.885.6689 | www.uidaho.edu/research/federal-relations