UC SANTA BARBARA

THE Current

July 24, 2012

Sonia Fernandez

NCEAS' DataONE Streamlines Search and Analysis of Massive Amounts of Ecological Data

In response to the growing need for a way to easily access and analyze massive amounts of heterogeneous data in the fields of earth and environmental sciences, UC Santa Barbara's National Center for Ecological Analysis and Synthesis (NCEAS), a core partner in a joint effort to streamline such research, presents DataONE, the Data Observation Network for Earth. DataONE is capable of providing researchers access to globally distributed, networked data from a single point of discovery.

"People have been gathering and synthesizing ecological data for decades," said NCEAS Director of Informatics Research and Development and DataONE co-investigator Matthew Jones. Much of the problem –– an issue that NCEAS has been working to address since its inception in 1995 ––

is the time and effort spent on

locating, gathering, checking, and transforming data of interest for synthesis.

It's an effort that can take researchers nearly a year to complete, as they examine and analyze various forms of information, from remotely sensed data, to hundreds of published papers, to historic observational field data. Simultaneously, these researchers search remote repositories, check for duplicates, and integrate the

information, as they try to find answers to complex problems that affect both science and society.

"Right now researchers have a hard time even finding the right data to answer complex environmental questions, and when they do, the work necessary to integrate really different types of data can be overwhelming," said NCEAS Deputy Director and DataONE co-investigator Stephanie Hampton. "DataONE provides the type of platform we need, to propel environmental science into the digital age."

DataONE, through the knowledge and infrastructure provided by library, computer, and environmental science experts, currently integrates information held by South Africa National Parks; the Knowledge Network for Biocomplexity; the Ecological Society of America; Dryad; Oak Ridge National Laboratories Distributed Active Archive Center, the United States Geological Survey, the Long Term Ecological Research Network; the Partnership for Interdisciplinary Studies of Coastal Oceans; and the California Digital Library. In the coming months, more organizations are joining as members to make their data accessible.

"In addition to broad data accessibility, DataONE also provides an interoperability framework that allows these diverse repositories to work together, share tools, and preserve data," said Jones. DataONE is an open network and encourages institutions and projects –– with data to share –– to become members of the federation.

Scientists and other users, meanwhile, will experience massive gains in efficiency, ease of access, and reductions in redundancy, as the data submitted to one repository will be easily available from multiple participating repositories. Users will also have the security of data persistence, thanks to better data curation and institutional diversity, which ensure that data do not disappear when organizations shift priorities or lose funding.

The data will also be available to a wide variety of audiences, Jones added. K-16 educators, those who could use the information as the basis for policy and management decisions, funders, and stakeholders will also have access to data from DataONE.

NCEAS is one of three national coordinating nodes, housing large data storage and computing resources in the UCSB data center at the California Nanosystems Institute. The two other coordinating nodes are located at University of Tennessee and University of New Mexico. With the sponsorship of the Davidson Library, NCEAS plans to move its data center to the North Hall Data Center on the UCSB campus.

DataONE is an outgrowth of a series of repository efforts, starting with the creation of the Knowledge Network for Biocomplexity (KNB) in 1998, which is the repository housing output from NCEAS' synthesis efforts. The KNB repository is open to submissions from ecologists and environmental scientists throughout the world, and represents a streamlined way for investigators to preserve and share their data with colleagues. As a participating node in DataONE, any data added to the KNB is automatically accessible.

DataONE is supported by a $20 million award, made as part of the National Science Foundation's (NSF) DataNet program.

Related Links

National Center for Ecological Analysis and Synthesis

Data Observation Network for Earth

The KNB repository

About UC Santa Barbara

The University of California, Santa Barbara is a leading research institution that also provides a comprehensive liberal arts learning experience. Our academic community of faculty, students, and staff is characterized by a culture of interdisciplinary collaboration that is responsive to the needs of our multicultural and global society. All of this takes place within a living and learning environment like no other, as we draw inspiration from the beauty and resources of our extraordinary location at the edge of the Pacific Ocean.