BioOne DIGITAL LIBRARY

DataONE Opens Doors to Scientists Across Disciplines

Author: Cohn, Jeffrey P.

DataONE Opens Doors to Scientists across Disciplines

JEFFREY P. COHN

The National Science Foundation (NSF) officially launched the first large-scale data-sharing program in the biological, environmental, and Earth sciences on 23 July when the DataONE project technology became available online to researchers, scientists, educators, and students around the world.

Formally known as the Data Observation Network for Earth, DataONE started in 2009 with a $20-million, five-year NSF grant. It uses cyberinfrastructure to connect researchers sitting at a computer through a server with databases of other scientific institutions. Based at the University of New Mexico, in Albuquerque, DataONE lets users access databases from different disciplines. Many of those databases were previously difficult to find, let alone access. DataONE also lets researchers use, compare, manipulate, store, and back up data in new ways.

DataONE can transform how scientists conduct research, said Bruce Grant, professor of biology and environmental sciences at Widener University in Chester, Pennsylvania, and a member of the viability and assessment working group that helped plan and implement the DataONE project.

For one thing, Grant said, DataONE will encourage more scientists to share their data. He expects fewer sole-authored papers and more multi-authored ones. And that, in turn, may force universities to change how they hire, rate, and promote faculty members. "The problems of today demand interdisciplinary research and action," Grant stated.

Further, DataONE is part of a larger NSF program to develop databases and networks to connect them with scientists and other users, said Alan Blatecki, director of NSF's Office of Cyber Infrastructure. "We recognize data as the basis of science," Blatecki said. "We have built DataONE on [the basis of] how scientists will use it. As the first, [DataONE] will be a model for other data projects."

So far, the DataONE network comprises 10 data centers, or *member nodes*. These include Oak Ridge National Laboratories' Distributed Active Archive Center for Biogeochemical Dynamics, the National Center for Ecological Analysis and Synthesis's Knowledge Network for Biocomplexity, and the Long Term Ecological Research Network at the University of New Mexico. In addition, several government agencies are both member nodes and DataONE sponsors. These include the NSF, NASA, and the US Geological Survey (USGS). Academic members include the Universities of New Mexico, Tennessee, Kansas, and California.

To date, only one data center outside the United States—the South Africa National Parks system—has signed on. But others are expected to join over the next two years, said William Michener, professor and director of the e-science program of the University Libraries at the University of New Mexico and director of the DataONE project. "We expect to have 50 member nodes by 2014. Actually, I hope we have hundreds."

For now, DataONE has been mostly used for demonstration projects, said Robert Chadduck, a program manager in NSF's Office of CyberInfrastructure. Before long, however, researchers will be using DataONE to address some of the grand challenges in biology and the environmental sciences today.

Two such grand challenges have already been studied using DataONE. First, scientists from the Oak Ridge National Laboratories and the Cornell Lab of Ornithology at Cornell University, in Ithaca, New York, accessed 31 different data sources using DataONE to examine changes in the migratory patterns of 250 North American birds on public and private lands. The results were used in the Department of the Interior's annual *Status of the Birds* report.

Second, a joint project by researchers from the US Department of Energy, NASA, and the University of California, Berkeley, developed software that allowed DataONE to serve as a repository for data and computer models from the Intergovernmental Panel on Climate Change. "We are creating the software and data accessibility to meet the needs of a community of users," Michener said.

Speaking of community, DataONE was designed and implemented by a staff of a couple of administrators, some 25–30 software developers, and more than 100 members of the nine working groups who volunteered to offer advice, write papers, and help design and implement DataONE.

Using DataONE "has expanded our knowledge of other scientists and their research," said Michael Frame, a USGS computer engineer and a member of a DataONE working group. "[DataONE] leverages our ability to do research. We could not do our work [nearly as effectively] without DataONE."

"We hope to use DataONE to help change the culture of science so that individual scientists can better recognize the importance of data," Michener said. Toward that end, the DataONE project has spawned learning modules, courses, and outreach programs on data management. "We think we're right on track," he concluded.

For further information on the DataONE project, go to *www.dataone.org* or to *www.nsf.gov*.

Jeffrey P. Cohn (jeffcohn1@verizon.net) is a science writer in Takoma Park, Maryland.