DataONE: A Virtual Data Center for Biology, Ecology, and the Environmental Sciences

**Authors:** PATRICIA CRUISE, William Michener, John Cobb, Suzie Allardi, Paul Allen, Peter Buneman, Randy Butler, Robert Cook, Ewa Deelma, David DelRoue, Cliff Duke, Mike Frame, Carole Goble, Stephanie Hampton, Donald Hobens, Peter Honeyman, Jeffery Horsburgh, Vivian Hutchinson, Matthew Jones, Steve Kelling, Jeremy Kranowitz, John Kunze, Bertram Ludescher, Lorraine Normore, Ricardo Pereira, Line Pouchard, Robert Sandusky, Ryan Scheller, Mark Serrilla, Kathleen Smith, Carol Tenopir, Dave Viegla, Von Welch, Jake Weltzin, and Bruce Wilson

**University of New Mexico: National Center for Ecological Analysis and Synthesis • University of California - Santa Barbara: Atlas of Living Australia • University of Michigan: Cyberinfrastructure that enables its use) about Earth’s biota and environments covering a range of spatial and temporal scales, and scales of biological organization. DataONE provides an ‘investigator’s toolbox’ that includes metadata management, scientific visualization, and other tools that empower scientists and organizations to more easily and effectively manage, analyze, and synthesize data. Innovative training and outreach to scientists and students include best-practice videos, podcasts, online certificate programs, downloadable best practice guides and exemplars of data management plans. Through working group meetings, computer and information scientists are engaged in developing and promulgating ontologies that will facilitate data integration and simplify creation of complex scientific workflows. The DataOME portal simplifies the process of acquiring and using appropriate scientific workflow software like Kepler and Taverna, as well as publishing and sharing new workflows via mechanisms such as myExperiment that allows workflows to be re-used and possibly adapted for other uses.

**The Vision:** "DataONE will be commonly used by researchers, educators, and the public to better understand and conserve life on earth and the environment that sustains it."

By creating an infrastructure of technology and standards, people, and institutions to support the full life cycle of biological, ecological, and environmental data and tools that enable universal access, DataONE will accelerate use of earth observational data in research, education and decision-making. In so doing, DataNetONE will transform our understanding of ecological processes and conserve life on earth and the environment that sustains it.

**Cyberinfrastructure**

**Data Sources**
- Research networks and environmental observatories
- Biological specimens
- Individual scientists
- Citizen scientists’ data
- Natural resources and conservation data
- Observational data
- Global and continental land cover/land change and biochemical data

**Data Types**
- Biological (genomes to biomes)
- Environment
- Atmospheric
- Ecological
- Hydrological
- Oceanographic

**Coordinating Nodes**
- Retain complete metadata catalog
- Subset of all data
- Perform basic indexing
- Provide network wide services
- Ensure data availability (preservation)
- Provide replication services

**Member Nodes**
- Diverse institutions
- Serve local community
- Provide resources for managing their data

**DataONE Observation Network for Earth**, MSCI 3180, University of New Mexico, Albuquerque, NM 87131-0001

**Education and Outreach**

**Career-Learning Opportunities**
- Best practice guides
- Exemplary data management plans
- Podcasts, webcasts
- Workshops and seminars
- Downloadable curricula

**Use Case Scenario:**

**Forecasting the spread of an invasive species across North America**

**The European Collared Dove (Streptopelia decaocto) was introduced in the Bahamas in 1966. Since then it has spread across North America. There are concerns that the European Collared Dove could compete with native dove species. Long-term analysis workflows are used to identify factors that best predict species occurrence and drive hypotheses generation and project planning.**

**Community Engagement**

**Building Global Communities of Practice: Creating Long-Lived CI Enterprises**

- Broad active community engagement
- Involvement of library and science educators engaging new generations of students in best practices
- Leverage existing outreach and education programs
- Transparent, participatory governance
- Adoption/creation of innovative and sustainable business and organizational models

**The Working Group Model**

- Structure
- 10 – 20 participants
- Deep analysis
- Intensive collaboration
- Neutral territory
- 2 weeks per year
- Remote collaboration
- Social networking

**Success**
- Community-driven
- High productivity
- High impact

**DataONE Collaborators**

- Libraries & digital libraries
- Academic institutions
- Research networks
- Synthesis & supercomputer centers/networks
- Governmental organizations
- International organizations
- Data and metadata archives
- Professional societies
- NGOs
- Commercial sector