Objectives
Addressing the Earth's environmental problems requires that we change the ways that we do science; harness the enormity of existing scientific data; develop new methods to combine, analyze, and visualize diverse data resources; and create new, long-lasting cyberinfrastructure. DataONE has been designed and built to provide a robust, scalable infrastructure that supports the data archive needs of biological, environmental and earth science researchers. It helps ensure long-term access to data and metadata held in a heterogeneous federation of data repository Member Nodes through a suite of investigator tools that have been developed or adapted to utilize the DataONE services. These services use secure web protocols and enable the DataONE Coordinating Nodes to work with Member Nodes and investigator tools. The result is a resilient and fault tolerant infrastructure with a catalog and index service, controlled access to data on Member Nodes based on inter-institutional identity management, and supporting services to assist with content creation, content replication, and digital object identification.

On-going activities:
DataONE is a federated data network that improves access to, and preserves data about, life on Earth and the environment that sustains it. DataONE supports science by: (1) engaging the relevant science, library, data, and policy communities; (2) facilitating easy, secure, and persistent storage of data; and (3) disseminating integrated and user-friendly tools for data discovery, analysis, visualization, and decision-making. The DataONE architecture is comprised of three components (1) Member Modes (data centers/repositories); (2) Coordinating Nodes that support network wide services (e.g., indexing and replication); and (3) an Investigator Toolkit that provides integrated access to tools that can support researchers in managing all aspects of the data life cycle (e.g., Data Management Planning Tool (DMP Tool), R, Excel, and MATLAB). The DataONE architecture and resources (e.g., data preservation services, best practices, and learning modules) can significantly benefit data-intensive and use-inspired science. DataONE tools and services are designed to promote data and tool interoperability, data preservation, and other features that underpin the movement towards more open data and reproducibility of science.

Results:
Key results include:
• Stable and reliable Coordinating Node and Member Node software stacks ensuring content integrity across the network, reliable coordination of services, and discovery through search interfaces.
• ONEMercury provides a fast and scalable browser-based search interface for the content available in the DataONE network.
• R-Plugin enables read and write access to DataONE from the R analysis environment.
• Morpho is a sophisticated metadata editing application that utilizes the DataONE services for reading and writing data and metadata.
• DataUp helps users contribute Microsoft Excel spreadsheet data along with descriptive metadata for re-use by other investigators.
• Kepler Actor enables read and write of content in Kepler workflows from and to DataONE Member Nodes.
• VisTrails Module enables read and write of data and metadata with DataONE, and provides the ability to store provenance information associated with the origin and processes used to generate new data products in the workflow environment.
• Comprehensive architecture documentation.
• DMPTool, which received recognition as one of the “Top 10 Digital Preservation Developments of 2011” by the Library of Congress.
• Data Management Education Modules for use in instruction and self-learning.

URL: www.dataone.org