

## Abstract: DataNet Federation Consortium

Major science and engineering initiatives are dependent upon massive data collections that comprise observational data, experimental data, simulation data, and engineering data. To support science and engineering collaborations, a policy driven national data management infrastructure will be implemented. The implementation prototype will address both the life cycle of science and engineering data and the sustainability of data collections and repositories over time, across changes in technology and changes in usage. The motivation for building the national infrastructure comes from the data management requirements of the NSF Ocean Observatories Initiative (real-time data streams, simulation output, video), the NSF Consortium of Universities for Advancement of Hydrologic Science (point data), engineering projects in education and CAD/CAM/CAE archives, the iPlant collaborative (genome databases), the Odum social science institute (statistics), and the NSF Science of Learning Centers (EEG / MRI sensor data, video).

The approach is based on a bottom-up federation of existing data management systems through use of the integrated Rule-Oriented Data System (iRODS). Each of the referenced national initiatives has implemented a core data management system based upon the iRODS data grid technology. Through federation, the independent systems can be assembled into a national data infrastructure that integrates collections across project-specific technology (such as real-time sensor data acquisition systems), institutional repositories, regional data grids, federal repositories, and international data grids. The resulting infrastructure will enable collaborative research among researchers in academic institutions and federal agencies, and across national boundaries.

Evolution of the policies (computer actionable rules) and procedures (computer executable workflows) that govern each stage of the data life cycle will be supported. Specific policies and procedures will be implemented for each domain to support their community standards for managing data in their local data grid. The project will develop the interoperability mechanisms required to share data between the domains, develop sets of policies and procedures to govern the data life cycle stages, and develop policies and procedures that enable re-use of collections. The national data management infrastructure will demonstrate enforcement of data management policies that comply with NSF Data management and preservation requirements.