1 Project Summary

Continuing advances in computation and communication are transforming the scientific process. Biodiversity research, like the fields of molecular, structural and proteomic biology, is re-inventing itself with new technology applications, and is evolving into an increasingly predictive and integrative science focused on important research and policy issues. This international ITR collaboration takes on key challenges for biodiversity researchers by modeling, designing and implementing the data discovery, integration and visualization components for a semantic web in environmental science. The initiative stands on the foundation of substantial and productive NSF investment in ecological and biodiversity informatics and it brings together four highly collaborative, forward-looking institutions in a partnership committed to inventing and supporting a global computing infrastructure for environmental biology. The project involves a multidisciplinary team of computer scientists, ecologists and technologists collaborating in complementary IT R&D and working group units.

This five-year initiative will investigate methods for semantic enrichment of scientific data, information services and analysis resources. Research will develop and test the use of formal ontologies to facilitate scientific analysis (“Semantic Mediation”) by discovery and automated integration of highly relevant but heterogeneous data. We will create a system that permits advanced inferences based on parameter ontologies and constraint languages. A second focus will investigate the scientific applicability of extensions of service languages to a Grid-based framework (“EcoGrid”) by developing integrated views over sets of currently incompatible data resources using metadata wrappers and emerging Web-service technologies (SOAP, WSDL, UDDI). Finally, our researchers will develop and investigate how formally described analytical pipelines (“Analysis and Modeling”) can capture scientific workflow as structured objects in digital libraries for re-use and extension. These analysis objects will be deployed to execution environments for processing of the formalized code on multiple platforms using various software environments. The goal of these research thrusts is to develop technology that enables new classes of environmental research.

These IT research foci will be integrated within a common web-enabled technology framework we call the “Science Environment for Ecological Knowledge” (SEEK). Perhaps no field stands to gain more from these advances than ecology, where researchers are grappling with questions that range over extremes of spatial and temporal scales, and where investigations encompass all of the physical and life sciences, as well as recently emerging areas such as chaotics and complexity theory. Ecology is the meeting ground for studies in biocomplexity, where understanding of critical issues such as the processes governing ecosystems, natural resource sustainability and earth’s biodiversity emerges from the union of behavioral, biological, chemical and physical perspectives.

SEEK will lead to fundamental improvements in how researchers can 1) gain global access to data and information, 2) rapidly locate and utilize distributed computational services, and 3) exercise powerful new methods for capturing, reproducing, and extending the analysis process itself. SEEK will also specifically provide ecologists and other researchers access to a large-scale network of information resources and computational services, via powerful data discovery and analysis tools that operate from desktop computers. These capabilities will significantly build research capacity to more effectively address global research, management and policy issues in environmental biology that increasingly require much more efficient, automated access to distributed and heterogeneous data.