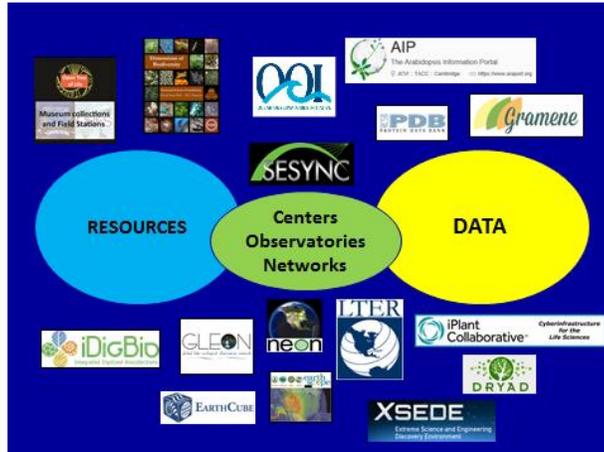# BioData: A Network for Understanding Life

What is BioData? BioData is intended to be a federated network of data and biological resources and that will allow researchers to seamlessly access tools to develop and address grand challenge questions in the biological sciences through exploration and exploitation of linkages among biological patterns and processes across multiple scales of complexity and over ecological and evolutionary time.

Why is BioData needed? 21st Century biological research is evolving into a data-driven science that is transdisciplinary. While there have been tremendous technological advances in some areas of biology, for example in genomics and synthetic biology, the integration and analysis of the resulting data and testing of outcomes are limited by the ability of researchers to seamlessly access data and associated research resources. At the same time, there is a need to ensure the preservation of important biological data now and for future generations, to foster the development the new tools necessary to use and analyze biological data, to educate and training biologists to be both proficient with and prepared to use the most advanced cyberinfrastructure available, and to do so in a sustainable, extensible, and scalable manner.

What will BioData enable? BioData will network biological research facilities and related collaborative activities with other relevant legacy and new extant data sources and biological resources to enable the synthesis of fundamental research data and new knowledge from a wide spectrum of activities including large-scale molecular, phylogenetic, environmental, and behavioral databases, as well as digitized collections of biological specimens.

How will BioData be established? The design and development of BioData will be based on network integration best practices and community input. It is intended to be modular and develop around activities where there are demonstrated community needs. Four current BIO activities will be used to spearhead the broader networking goals of BioData:

- Community workshops sponsored through the **MacroSystems Biology** Program to develop networking strategies for large-scale data generated by NEON (when

operational), Long Term Ecological Research (LTER) sites, Field Stations and Marine Labs, iDigBio (for digitized collections) and iPlant.

- The new **Genealogy of Life** (GoLife) program to resolve the phylogenetic history of life and to integrate this genealogical architecture with underlying organismal data. The resulting knowledge infrastructure will enable synthetic research on biological dynamics throughout the history of life on Earth, within current ecosystems, and for predictive modeling of the future of life on Earth.

- The **Arabidopsis Information Portal** (AIP) model for development of a modular, federated resource to ensure that responsibility for generation and maintenance of data remains in the hands of the individual data providers and spreads the burden of supporting such resources across a potentially wider range of funding agencies and countries.

- BIO-sponsored community workshops to design and plan for the development of a **BRAIN/neuroscience** data sharing and analysis facility modeled on the iDigBio approach.

What are the anticipated outcomes of BioData? The National Academy of Sciences 2009 report entitled "*A New Biology for the 21st Century*" asked, "What are the implications for the life sciences research culture of a newly-integrated approach to biology?" and offered the answer:

> The essence of the New Biology . . . is integration— re-integration of the many sub-disciplines of biology, and the integration into biology of physicists, chemists, computer scientists, engineers, and mathematicians to create a research community with the capacity to tackle a broad range of scientific and societal problems.

BioData will be an essential component of BIO's support of this new biology, serving as a framework for connecting existing and future data and research resources that will ensure long-term sustainability and community access to research outcomes.