PROJECT SUMMARY

Many individual investigators and small research programs do not have the resources or technical expertise to link with data repositories serving coordinated communities such as the Long Term Ecological Research (LTER) Network. In this case, the LTER Network developed the Provenance Aware Synthesis Tracking Architecture (PASTA) to serve as the core infrastructure for its Network Information System (NIS). This proposal will enhance PASTA (as "PASTA Plus") to serve three communities that do not have a central data management approach: 1) the Long Term Research in Environmental Biology (LTREB) program, 2) the Macrosystems Biology (MSB) program, and 3) members of the Organization of Biological Field Stations (OBFS). Members of these communities will be assisted in archiving ecological data in PASTA+ through a proactive program of training and sustained support. This effort will foster comparison and synthesis by expanding access to key data sets produced by these communities, preserving these data sets in a persistent long-term repository, and providing community members access to a broad range of tools for information management and scientific analysis. Data that are published in PASTA+ will be synchronized through LTER's DataONE Member Node making it accessible through the DataONE search interfaces and eventually replicated to other peer member nodes.

Intellectual Merit
When data are archived in a persistent, accessible repository, both data producers and consumers benefit. Data in a managed repository are available for reuse and data producers receive credit for the use of their data. Data consumers are able to find data more easily and are able to conduct scientific syntheses requiring complex, varied types of data more economically. All of the communities involved in this proposal (LTREB, MSB, and OBFS) produce data that document important ecological processes such as long-term changes in ecosystem structure or biodiversity. Moreover, these communities are still searching for a structured approach to archive and make their data accessible through a common, persistent repository. As a data repository designed by a community of ecologists, PASTA+ will provide an appropriate framework to link LTREB, MSB, and OBFS data with analytical tools and persistent storage. PASTA+ will extend identity management for inclusion of these new communities, streamline data documentation through improved interface design and conversion of science metadata standards into the Ecological Metadata Language, and expand rules for quality checking uploaded data packages to the standards of these communities.

Broader Impacts of the Proposed Research
By adding data from three new research communities to PASTA+, we will greatly expand the geographic scope of the repository and thus, its value to society. The improved analytical power resulting from expanding the pool of accessible data will increase society's capability to predict and mitigate future changes in ecosystem services, climate, and biodiversity. Through eight data immersion workshops, we will provide support to senior and early career scientists and students to document data and make it accessible through the PASTA+ data repository. Feedback from participants in these workshops will inform improvements and harden the open source PASTA+ software to increase its usability by the scientific community. Our project will establish enduring collaborative relationships between the LTER Network and our three target communities and thus, enhance infrastructure to address ecological questions over broad spatial and temporal scales. We will disseminate information about best practices in data stewardship to a broad audience of early career scientists through an online course based on the data immersion workshops. Workshops at national scientific gatherings will train graduate students and post-doctoral researchers to use national repositories of ecological data (e.g., LTER, DataONE) in their future research. We will organize a workshop on careers in informatics for graduate and undergraduate students at the annual conference of the Society for Advancement of Hispanics/Chicanos and Native Americans in Science (SACNAS). We will conduct formative and summative assessments of data immersion and training working groups by employing a mixture of qualitative and quantitative approaches. The formative component will generate feedback about what works and what does not, thereby enabling us to modify plans or redesign activities to increase the likelihood of meeting working group objectives. The summative component will use qualitative and quantitative measures to produce an objective analysis of project outcomes.