**Project Summary**

**Overview**

In the era of big data and informatics, there is growing awareness of the need for permanent, globally unique identifiers for both physical specimens and digital data, leading to the development of new systems for minting, tracking, resolving, and querying identifiers. Although multiple reports on the characteristics of and best practices for different identifier types have been published recently, there has been little empirical research on the functionality of identifiers or the services that support them, especially pre-publication. Practitioners working with identifiers on a daily basis are undoubtedly aware of the technical and social issues and limitations surrounding identifier systems, but they may not yet have pushed their systems to the limit with very large, multidisciplinary datasets. Our goals for this workshop are to bring together these practitioners to summarize the current state of the field, identify and elucidate the technical issues, and propose solutions. The workshop will incorporate elements of a hackathon in that participants will be expected to develop some initial implementations, but outputs may also include non-technical products like a draft proposal, a survey, or educational materials.

Scientific identifier meetings to date have been primarily been domain focused (e.g., biomedical, biodiversity, earth science, geography), but identifier challenges cross disciplines. This meeting will seek synergy across disciplines by inviting diverse participants from multiple domains, who might otherwise never interact with each other. Representatives from ongoing initiatives such as FORCE11, BD2K, and EarthCube will be asked to participate, as well as people working on specific identifier implementation such as EZID or ISGN. Scientists involved in large-scale data generation or data-reuse will also be asked to participate. Care will be taken to ensure participation by under-represented groups, both at the level of enrollment and through workshop activities designed to be inclusive. The meeting will be hosted by the iPlant Collaborative and will take place in winter 2015/16 at Biosphere 2 at the University of Arizona.

**Intellectual Merit**

Managing data is one of, if not the, major hurdles for scientific advancement at this time. While there has always been some interaction between library scientists and scientists in fields such as biology, physics, or earth science, the growing importance of informatics and data science for answering general scientific questions has brought these disciplines closer than ever. Cross-disciplinary research to address the technical and social barriers of data management -- and the role identifier systems play in those barriers -- is sorely needed. This workshop will bring together researcher at the cutting edges of their fields to identify gaps and bottlenecks in identifier and data management systems and propose solutions. The workshop should be mutually beneficial to all participants and their respective communities, by focusing on real use cases in order to overcome their most pressing challenges.

**Broader Impacts**

Data management, discovery, and re-use are pain points in all aspects of biology and many other scientific disciplines, and all depend on sound solutions for identifying data. The outcomes of this workshop have the potential to impact the way identifiers are applied and used across science and even more broadly. The inclusion of early career scientists will provide training and promote the adoption of new solutions and technologies.