

# Collaborative Research: BiSciCol Tracker: Towards a tagging and tracking infrastructure for biodiversity science collections.

## Project Summary

“Scientific collections created and used in basic research are an integral part of the nation’s scientific infrastructure. They hold specimens of plants, animals, microbes, fossils, minerals and other artifacts that together comprise a national legacy of biological diversity”. (NSF Scientific Collections Survey, 2009). Individual specimens in these collections serve as the anchor for an expanding array of information that grows and changes with time about the specimen and the group that the specimen represents. Unfortunately, specimens and subsamples are scattered geographically across institutions. Taxonomic, genomic, geospatial, and other information about the specimens are also scattered across independent computer systems and on paper and are very difficult to access or synthesize. Current data sharing systems such as DigIR are one-way channels and do not allow for quick and easy two-way linking of information or updates as new knowledge is gained.

**Intellectual Merit:** We will take the appropriate next steps to address a community-wide challenge facing the biological collections community – linking and tracking scientific collection objects (specimens, sequences, images, etc.) and their digital metadata across multiple institutional collections with heterogeneous information management systems. In current distributed data systems (e.g., GBIF, MANIS, HerpNET, ORNIS), information is passed one-way from data providers to users. No mechanism exists to tag or annotate collection objects and link information to other collection objects or data records and back to the original collections. The **BiSciCol** team would **1) develop a tracking and annotation system** based on globally unique identifiers (GUIDs) and ontological relationships; **2) *deploy this system and others in a Virtual Information Appliance (VIA)*** as a Virtual Machine (VM); and **3) *document and implement a set of use cases and practices***, based on characteristic physical and digital workflows in the community.

The need to provide access to validated biodiversity information has been documented in a number of workshops, reports, etc., but as yet there is no single implementation that would support collections and research information management using the proposed approach. **BiSciCol** is designed on the simple premise that changes to data objects are trackable with GUIDs, and that semantic relationships are assignable and discoverable among physical and data objects, for example when a specimen is imaged or sampled for DNA extraction. This project directly addresses a key objective of the IBRC, enabling discovery, accessibility, and networking of collections, in order to advance semantic interoperability for collection information systems.

**Broader Impacts:** The **BiSciCol** collaborative represents a broadly trained team of biologists, collections curators, and information and technology specialists with a common vision of tracking and annotating information about the billions of specimens in natural history collections worldwide. Deliverables are designed to benefit the entire biological collections community by taking initial steps to implement core information infrastructure based on established challenges in the community. Collections data are critical to land management decisions, maintenance of biodiversity, and analysis of the causes and consequence of climate change. **BiSciCol** will improve data quality and quantity for these communities of non-scientists and scientists.

**BiSciCol** will actively engage use communities through training workshops, summer student internships, community BioBlitz enhancements, and a partnership with NESCENT on a Google Summer of Code challenge.