Until very recently, almost all efforts to digitize insect (and other) collections focused entirely on capturing label data (retrospectively) and organizing it in a database (various custom collection database solutions have been developed and are currently in use). This involved (and still does to a large extent) manually keying label data into the appropriate fields of the database (e.g., locality, date, collector, identification) and physically attaching an additional label to each specimen containing a unique identifier (UID, a number and/or barcode) to facilitate tracking of individual specimens. For specimens older than a decade or so (i.e., those collected prior to the availability of hand-held GPS devices), retrospective georeferencing (obtaining map coordinates for the collecting locality) is a further step in the databasing process to facilitate mapping of geographic distribution data. This basic approach is problematic because:

Manual data entry is prohibitively time consuming and, thus, expensive
Manual data label entry is prone to data entry errors
Handling specimens to attach UID/barcode labels (a task often performed by hourly technicians with limited experience) involves substantial risk of specimen breakage

Nevertheless, this remains the dominant paradigm. Hence, the need for new approaches as outlined in our proposal.

Improved methods should seek to increase the speed and accuracy, and reduce the cost, of data capture and, at the same time, minimize unnecessary handling of fragile specimens (each specimen is unique and irreplaceable). One solution theoretically possible is, instead of using UID labels, to use unique characteristics of the individual specimens themselves to recognize and facilitate tracking of those specimens. Although different specimens of the same species are usually extremely similar in appearance they vary at the individual level, having subtle differences in size, coloration, and patterns of surface texture. Theoretically, if individual specimens can be imaged at sufficient resolution, they should be individually recognizable, just as individual humans are recognizable by fingerprints and, as with fingerprints, the available data should be sufficiently redundant to facilitate identification even if only partial data are available. The problem then becomes one of scale. How difficult will it be to find an individual insect specimen in a database that will ultimately contain several million species and, potentially, billions of individual specimens? This is something to think about for the future, but it is not clear whether this approach is feasible given existing technologies.