**4.B.3. Improving Natural History Museum Data Collections:** Informatics Tools for Biodiversity Data Conversion

Maintaining and accessing biodiversity data is an extremely important issue for many institutions. Such data is usually stored in different formats, making access difficult. Additionally, the multiple databases where the data are stored have non-uniform entry tags that make data provenance problematic.

Creating a centralized database that contains all data in a unified format addresses these problems. This project was a collaboration between the San Diego Natural History Museum (SDNHM), Balboa Park Online Collaborative (BPOC), the National Institute of Information and Communications Technology (NiCT), and the UCSD PRIME program.

SDNHM has multiple collections of taxonomy, each stored in its own database. Therefore creating a central database containing all the information that can be updated periodically in an efficient manner is critical for the future use of SDNHM collections.

This project used Specify 6.5 (University of Kansas) as a centralized database that aggregates all the biodiversity data from various departments at SDNHM. Specify was chosen because of its ability to store more information than Darwin Core Archive (DCA), its ability to export to DCA, and the availability of premade tools and a support team to help in the process. The aggregated data can then be used to create a central database, allowing easy access to all biodiversity data from SDNHM, without having the data in different formats. Furthermore, this allows the museum to create a customized Web portal to display its data to the public, making it user friendly to all members of the public.

Python-based tools were used to convert the different data formats to a Microsoft Excel format. The data were then inspected and reformatted so that any unneeded information was removed and the data conform to Specify requirements. The data were then successfully imported into a single Specify database. A simple web portal was also created allowing access to this database.

In summary, this project created a centralized database for biodiversity information. The tools developed will continue to be used as a model for data aggregation and unification for other museums. SDNHM plans to expand the functionality of its Web portal and use the centralized database to provide metadata for future interactive exhibits.

**Participants:** PRIME 2013: Kittinan Ponkaew; UCSD: Jason H. Haga; BPOC: Vivian Kung Haga; NICT: Shinji Shimojo; SDNHM: Michael Wall, Brad Hollingsworth



