Field Station Collection
Digitization Workflow

Misha Leong
Momentum of CalBug

Essig Museum of Entomology
California Academy of Sciences
California State Collection of Arthropods
Bohart Museum, UC Davis
Entomology Research Museum, UC Riverside
San Diego Natural History Museum
LA County Museum
Santa Barbara Museum of Natural History
“Quantifying ecological effects of land use and climate change using historical collections”

UC ANR Grant, PIs Kip Will, Rosemary Gillespie, George Roderick, and Maggie Kelly
UC ANR Research and Extension Centers (n=9)

UC NRS (n=39)

UC Center for Forestry Sites (n=4)

Essig Museum of Entomology

Jepson Herbarium

Museum of Vertebrate Zoology

Natural Reserve System
UNIVERSITY OF CALIFORNIA
Where to start?!?

https://www.idigbio.org/content/workflow-modules-and-task-lists

Avoid “reinventing the wheel”, but everything needs to be personalized to the needs of your collection

• Tradeoff of volume vs. completeness
• Institutional/collection policies, needs, goals
• Constraints of personnel, expertise, funds, physical layout
• Type of collection
Workflow

1. Assess Current Collection
2. Data Decisions
3. Digitizing
   a. Imaging/Handling
   b. Data capture/Databasing
   c. Data Manipulation
4. Collection Maintenance
Step 1

ASSESS CURRENT COLLECTION
Damage?
Access?
Level of identification?
Who is using?
Organization of collection?
Size?
Step 2

DATA DECISIONS
Not everything needs to be digitized.
Things to Consider

- Future maintenance of collection?
- Who will have access? And how difficult?
- IT support?
- Legacy data?
- Size of collection?
- What needs to be digitized?
- What will data be used for?
- Might be different for all collection types
Spectrum of Options

• Donating collection to larger research collection
• Accessioning to local research collection (and permanently “on loan”)
• Independent, but digital data housed within a local research collection (a specific collection within the museum)
• Independent, but linked into a broader network (ARCTOS, North American network for small herbaria, etc)
• Completely independent
Step 3

DIGITIZING: imaging, databasing, & data manipulation
Digitization workflow

Handling & Imaging

(Optional)
Sort by locality, date, sex, etc.

Remove labels, add unique identifier

Take digital image, name and save file

Replace labels, return to collection

Data Capture

Manually enter data into MySQL database

Online crowd-sourcing of manual data entry

Optical Character Recognition (OCR) & Automated data parsing

Data Manipulation

Error checking

Geographic referencing

Aggregate data in online cache

Temporospaital analyses
The Jepson Herbarium has an efficient in-house plant specimen digitizing set-up.

Therefore, for plant collections, we relied on this partner, and packed up all the specimens and brought them to campus for digitizing. All specimens went through the Jepson specific workflow, and were returned when completed on a following trip.

Need: connection and funding to pay for staff time.
Arthropods and vertebrates were imaged on site with our portable imaging set-up.

Main tools: dino-lites, laptops, extra light sources, imaging boxes, photo stand, digital camera, tools to work with specimens
We brought our digitization set-up to all 3 field stations.
Imaging

1. Select taxa for databasing

2. Sort specimens by location & date

3. Arrange labels to view all text, add catalog # label

4. Take, name, and save digital image of labels
University of California
Sagehen Creek Field Station

#014

University of California
Museum of Vertebrate Zoology

Scapanus latimanus
Broadfoot mole
University of California
Sagehen Creek Field Station

# 014

87 1731 A. S. Leopold
Sagehen Creek Field Sta., 6400', 3 mi.
NW 1/4 Sect. 16, T20N, R37W, Nevada Co., Calif.
164-37-20-0 62m June 27, 1971
Databasing/Data Capture

5a. Manually enter data into MySQL database with some error checking

5b. Online crowd-sourcing of manual data entry

5c. Optical Character Recognition & data parsing

-Transcription
-Crowd Sourcing
-OCR
Efficient and user-friendly set-up through Essig Museum for databasing insect specimens with uploaded images.
Collecting Event Data
- eventID (DC)
- country (DC)
- stateProvince (DC)
- county (DC)
- locality (DC)
- minimumElevationMeters (DC)
- maximumElevationMeters (DC)
- decimalLatitude (DC)
- decimalLongitude (DC)
- coordinateUncertaintyMeters (DC)
- geodeticDatum (DC)
- verbatimCoordinateSystem (DC)
- georeferenceSources (DC)
- georeferencedBy (DC)
- georeferencedDate
- georeferenceRemarks (DC)
- collectionBeginDate (*)
- collectionEndDate (*)
- recordedBy (DC) = collectors
- samplingProtocol (DC)
- associatedTaxa (DC)
- sex (DC)
- individualCount (DC)

Specimen Data
- catalogNumber (DC)
- institutionCode (DC)
- kingdom (DC)
- phylum (DC)
- class (DC)
- order (DC)
- family (DC)
- genus (DC)
- specificEpithet (DC)
- subspecies
taxonIDCertainty
- scientificNameAuthorship (DC)
- identifiedBy (DC)
- dateIdentified (DC)
- eventID (DC)

Bold = required
Normal = recommended

(DC) = Darwin Core field
(*) = Darwin Core recommends one field that accommodates several date options. We prefer “begin” and “end” dates.
Besides ID, date, and location, there is also:

1. Type of part (skin)
2. Sex
3. Total length
4. Tail length
5. Hind foot with claw
6. Ear from notch
7. Weight
8. Reproductive data
9. Unit of measurement, who determined, and when

Vertebrate specimens have many more columns to database than arthropods.
Some wet collections were donated to UC Berkeley’s MVZ
Step 4

COLLECTION MAINTENANCE
Collections of all types were enhanced through:

- Cleaning
- Reorganization
- Acquiring new cabinets, drawers, unit trays
- Entomological IDs and overhaul
- Updated species lists
- Transferring specimens from formalin to ethanol
Moving forward...

• An enhanced collection is easier to use and build off of
• Now know what each field station has
• Improved linkages to main campus
• Developing protocols for people who are doing research at the field station
• Helping get people to come to (and contribute) collection
• Involvement of the local community (imaging, databasing, BioBlitz)