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

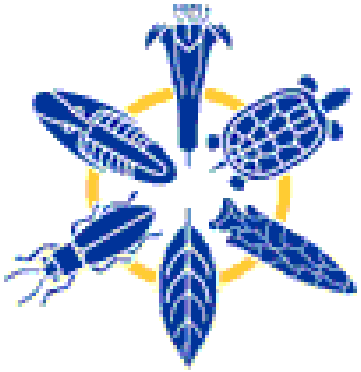
San Diego Natural History Museum

LA County Museum

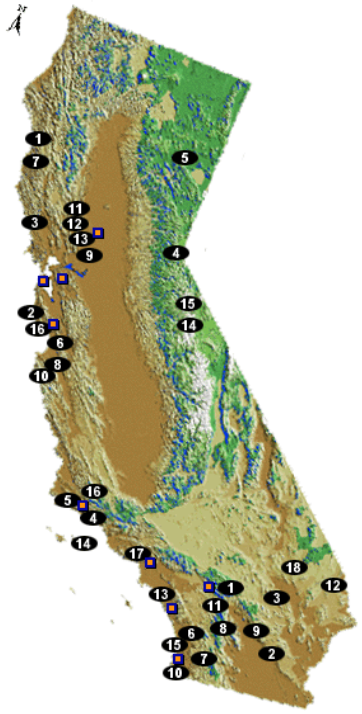
Santa Barbara Museum of Natural History

“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 Berkeley Natural History Museums (BNHM)

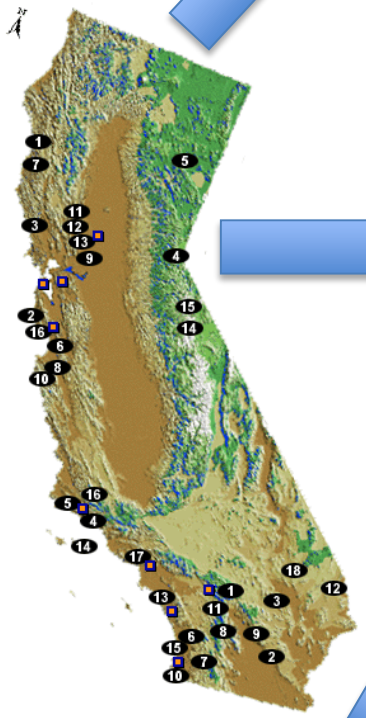


University of California Field Stations

Essig Museum of Entomology

Natural Reserve System
UNIVERSITY OF CALIFORNIA

UC NRS
(n=39)



UC ANR Research and Extension Centers
(n=9)

Jepson Herbarium

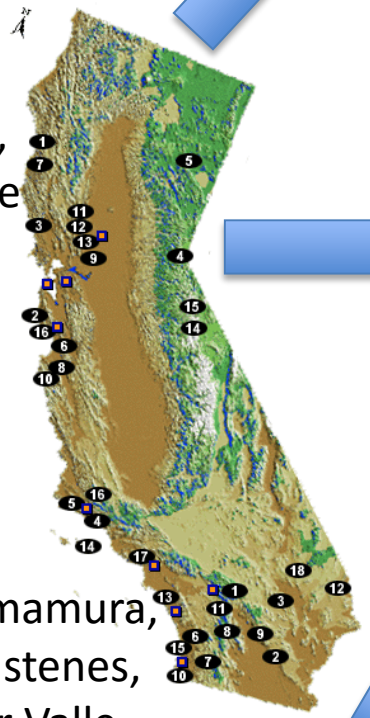
Museum of Vertebrate Zoology

UC Center for Forestry Sites
(n=4)

Essig Museum of Entomology
(Pete Oboyski & Kip WIll)

Saghen
(Faerthen Felix, Jeff Brown, Erica Krimmel)

PIs
Kip Will, Rosie Gillespie, George Roderick, Maggie Kelly



Hopland
(Bob Keiffer)

Jepson Herbarium
(Kim Kersh and Andrew Doran)

Digitizing Team
Misha Leong, Jennifer Imamura, Gracen Brilmyer, Ilona Istenes, Savannah Miller, Skyler Valle, Victoria Knorr

Museum of Vertebrate Zoology
(Chris Conroy, Carla Cicero, Michelle Koo)

Blodgett
(Frieder Schur)

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?


FIRST
AID KI
Jif
Ensatina

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

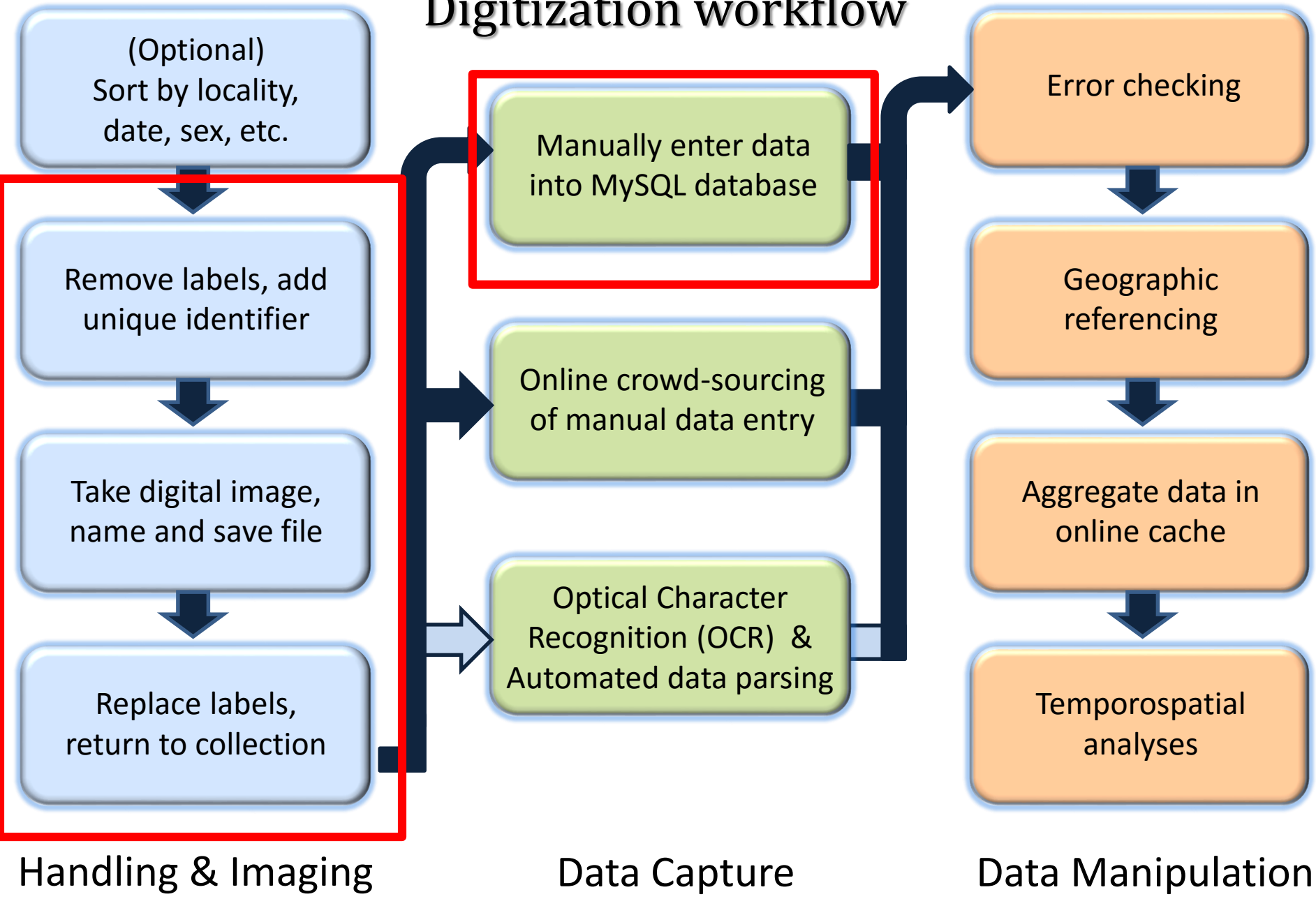
Spectrum of Options

- 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

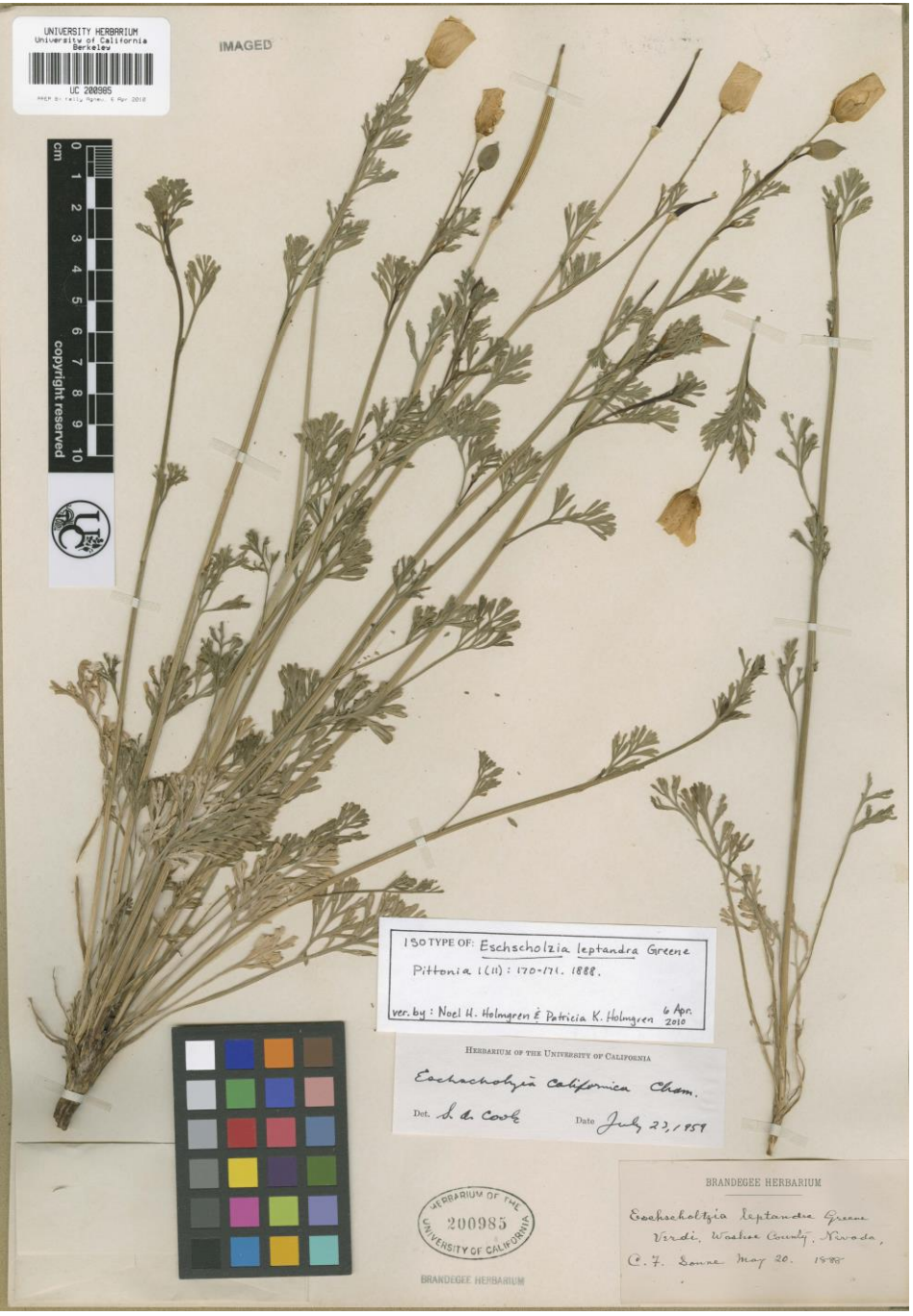


The Jepson Herbarium has an efficient in-house plant specimen digitizing set-up.

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 though the Jepson specific workflow, and were returned when completed on a following trip.

Need: connection and funding to pay for staff time.


51. HABENARIA
78 ERIOGONUM
78. POLYGONUM
79 MONOLEPIS
79 SALSOLA
86 CALYPTRIDIUM
86 CRANDRINIA
86. LEWISIA
88. LILIUM
92. ACONITUM
92. PAEONIA
NON-FLOWERING PLANTS & FUNGI — NEED TO BE ACCESSIONED & IMAGED
NEW SPECIMENS — NEED TO BE SCANNED
OLD SPECIMENS — NEED TO BE SCANNED, & ACCESSIONED
OLD SPECIMENS — NEED TO BE SCANNED & ACCESSIONED


UNIVERSITY HERBARIUM
University of California
Berkeley
UC 200985
IMAGED
ISO TYPE OF: Eschscholzia leptandra Greene
Pittonia 1(11): 170-171. 1888.
ver. by: Noel H. Holmgren & Patricia K. Holmgren 6 Apr. 2010
Herbarium of the University of California
Eschscholzia Californica Cham.
Det. D. D. Cook
Date July 23, 1959
HERBARIUM OF THE UNIVERSITY OF CALIFORNIA
200985
BRANDEGEE HERBARIUM
BRANDEGEE HERBARIUM
Eschscholtzia leptandra Greene
Verdi, Washoe County, Nevada,
C. F. Sonne May 20. 1888
copyright reserved


IMAGED
UNIVERSITY HERBARIUM
University of California
Berkeley
UC 131697
copyright reserved
Ph. tenuacetifolia
var. cinerea n. var.
type. ! Brand
C. A. PURPUS
No. 5242
FLORA OF SOUTH-EASTERN CALIFORNIA
Phacelia
APRIL–SEPTEMBER, 1897.
TYPE OF
Phacelia tenuacetifolia var. cinerea Brand
Univ. Calif. Publ. Bot. 4: 216. 1912.
ver. B. Ertter 1985
HERBARIUM OF THE UNIVERSITY OF CALIFORNIA
131697

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.

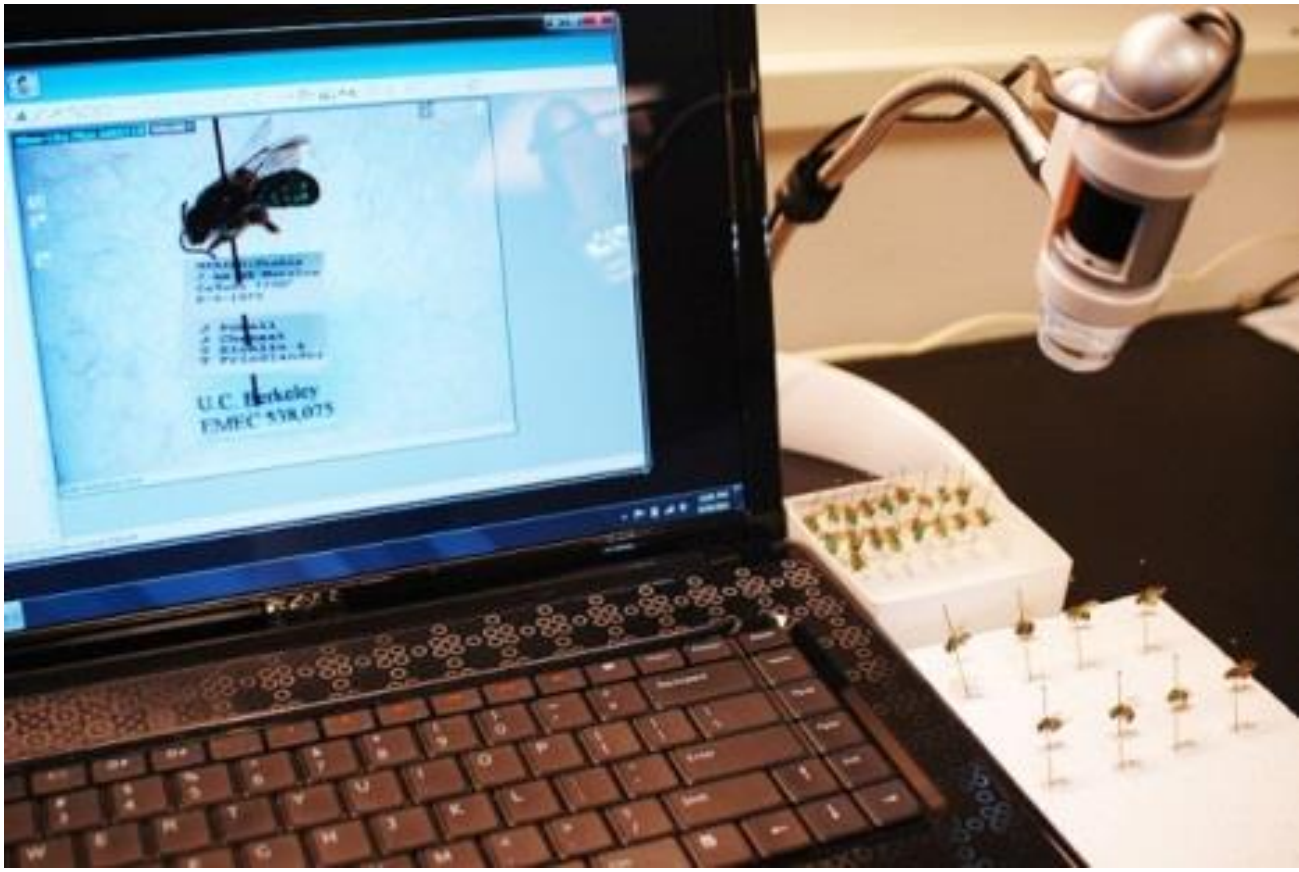
Hopland

Blodgett

Sagehen

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

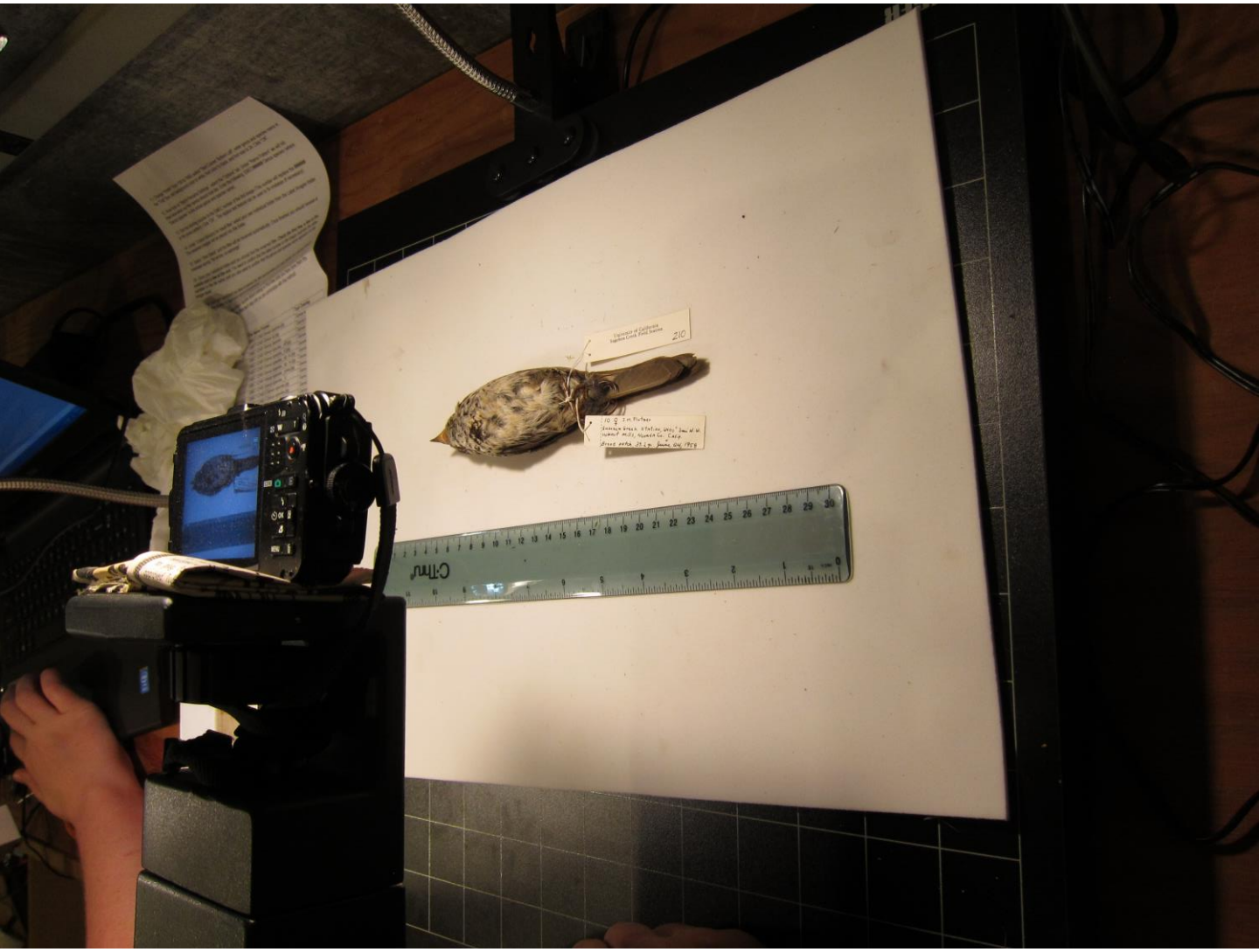



ELIZABETH LAKE
TUOL. CO., CALIF
EL. 9600'. 18 AUG 47
L.L.Jensen
Collector
UC Berkeley
EMEC
213190


UC Berkeley
EMEC
454373
3 mi.N.Termo
Lassen Co.
Calif.VIII-10-59
J.R.Powers
Collector
Hesperia
harpalus harpalus
(Edw.)


291613


U.C. Berkeley
EMEC 632.962
6 mi.W. Indio
Riverside Co.,
Calif.IV-30-49
Melilotus
E.G.Linsley
J.W.MacSwain
R.F.Smith
Podalonia 1951
luctuosa (Sm.) ♂
Det. W.D.Murray



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

OT 1731 A.S. Leopold
Sagehen Creek Field Sta., 6400', 3 mi.
NW Hobart Mills, Nevada Co., Calif.
164-37-20-0 62yr. June 27, 1971

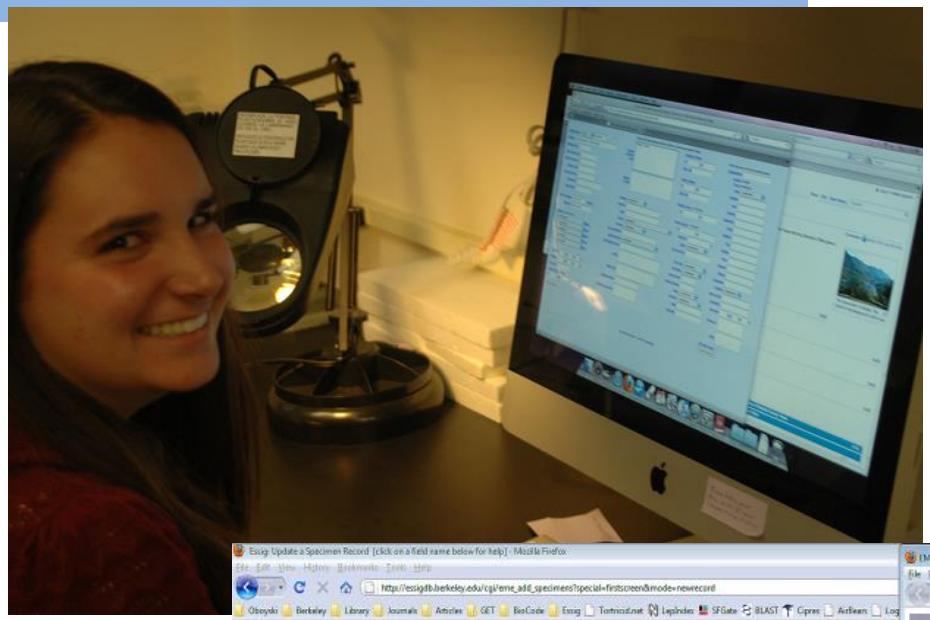
SCFS 14

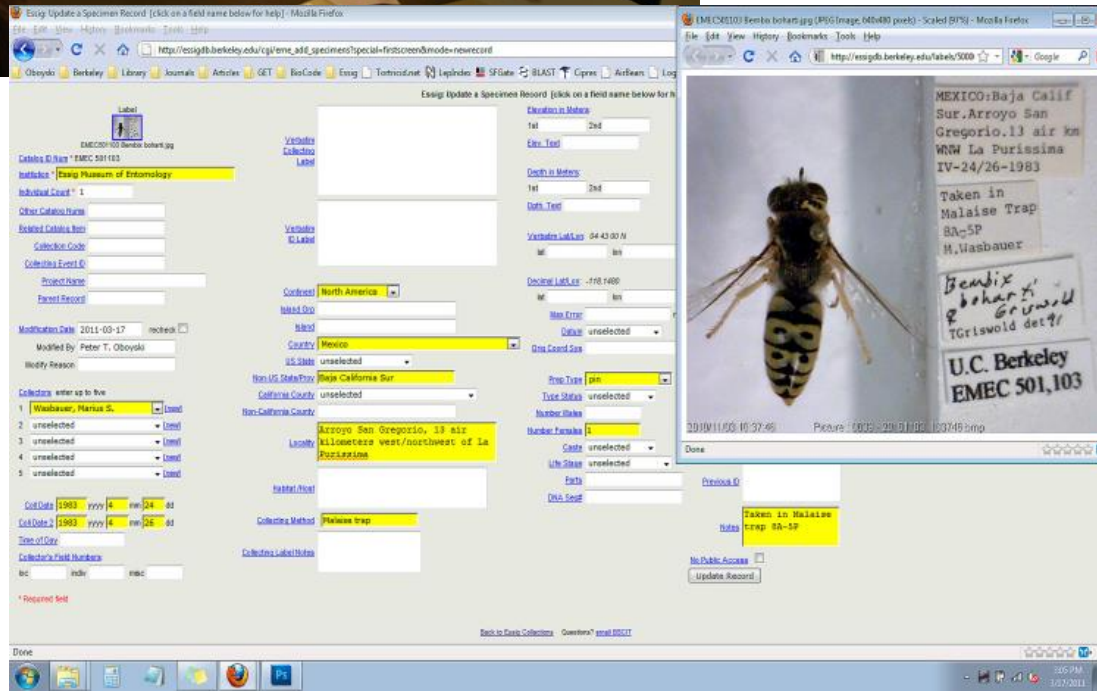
Databasing/Data Capture

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



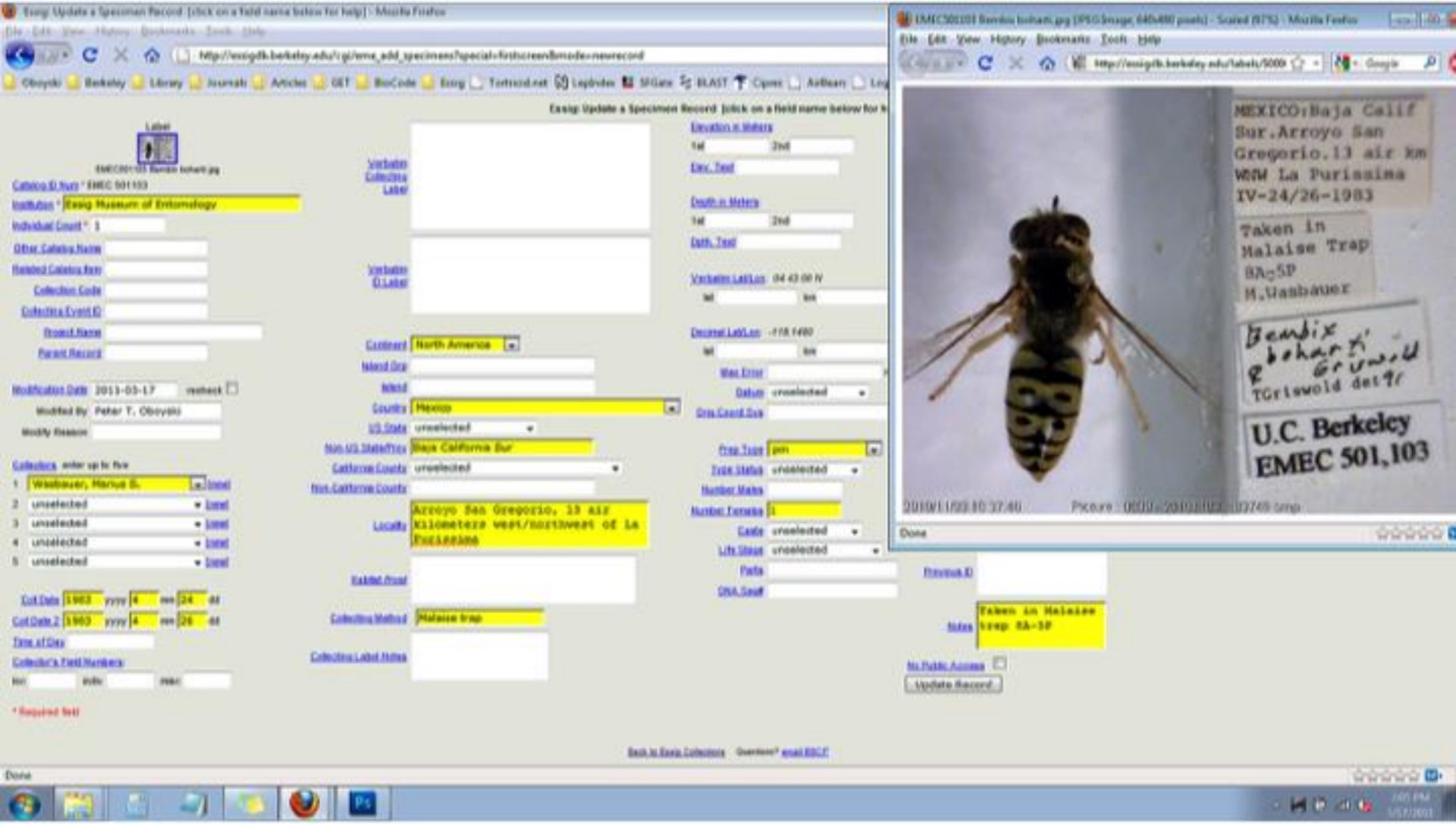


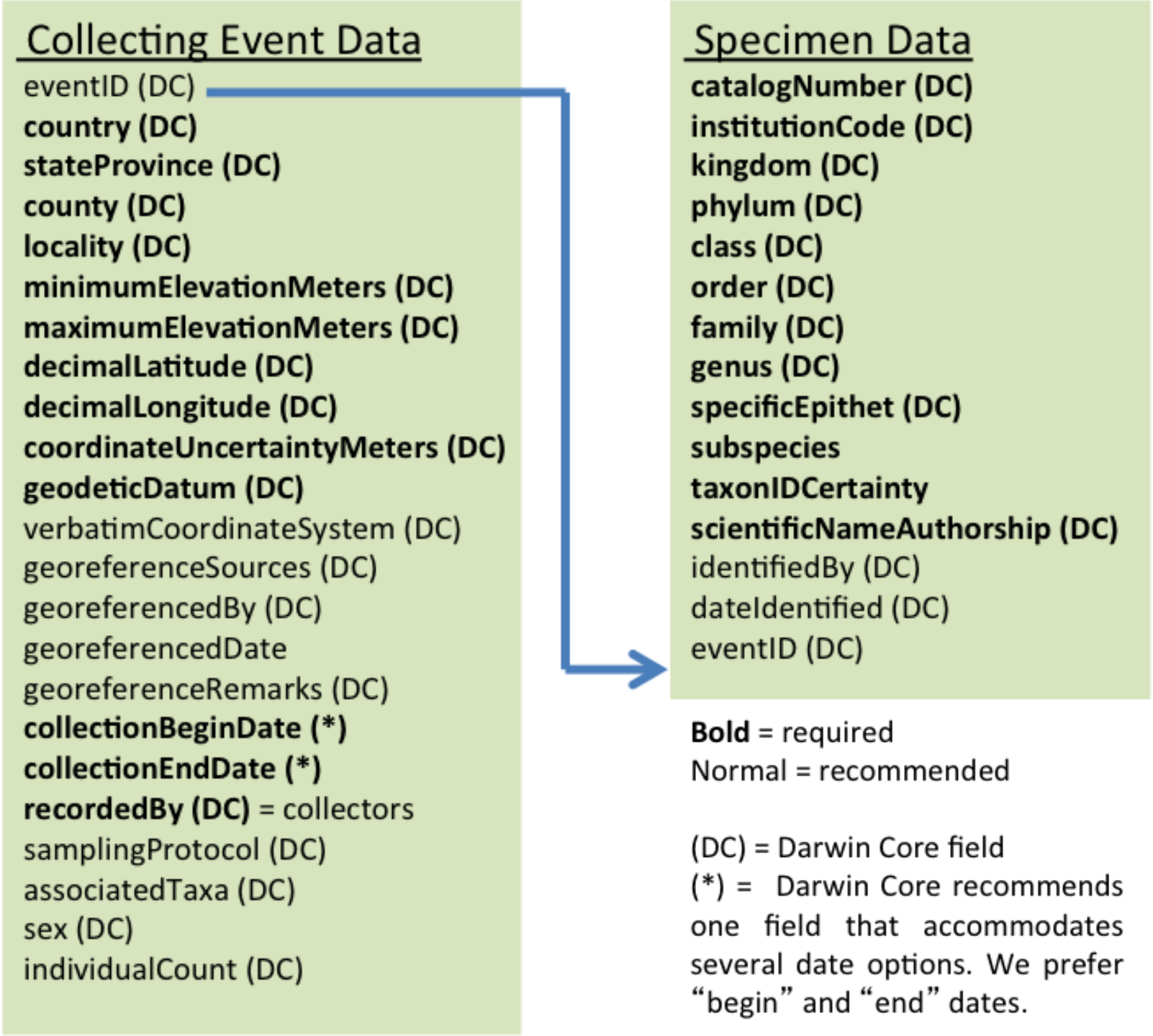
-Transcription

-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.

COLLECTION_OBJECT_ID	
ENTEREDBY	millersk
COLLECTION_CDE	Mamm
INSTITUTION_ACRONYM	BFRS
CAT_NUM	001
ACCN	
TAXON_NAME	Peromyscus l
NATURE_OF_ID	expert
ID_MADE_BY_AGENT	D.A.Airola
MADE_DATE	1977-07-14
IDENTIFICATION_REMARKS	
VERBATIM_DATE	1977-07-14
BEGAN_DATE	1977-07-14
ENDED_DATE	1977-07-14
LOCALITY_ID	
HIGHER_GEOG	
SPEC_LOCALITY	
VERBATIM_LOCALITY	Blodgett Res
ORIG_LAT_LONG_UNITS	
DEC_LAT	
DEC_LONG	
DATUM	
GEOREFERENCE_SOURCE	
GEOREFERENCE_PROTOCOL	
EVENT_ASSIGNED_BY_AGENT	
EVENT_ASSIGNED_DATE	
VERIFICATIONSTATUS	
MAX_ERROR_DISTANCE	
MAX_ERROR_UNITS	
SPECIMEN_EVENT_TYPE	
LOCALITY_REMARKS	
MAXIMUM_ELEVATION	4300
MINIMUM_ELEVATION	4300
ORIG_ELEV_UNITS	ft
COLLECTING_METHOD	
COLLECTING_SOURCE	wild caught
HABITAT	
COLLECTOR_AGENT_1	D.A. Airola
COLLECTOR_ROLE_1	collector
COLLECTOR_AGENT_2	
COLLECTOR_ROLE_2	collector
COLLECTOR_AGENT_3	
COLLECTOR_ROLE_3	
COLL_EVENT_REMARKS	
OTHER_ID_NUM_TYPE_1	collector nun
OTHER_ID_NUM_1	5
PART_NAME_1	skin
PART_CONDITION_1	unchecked
PART_BARCODE_1	
PART_LOT_COUNT_1	1
PART_DISPOSITION_1	in collection
PART_REMARK_1	
PART_NAME_2	
PART_CONDITION_2	
PART_BARCODE_2	
PART_LOT_COUNT_2	
PART_DISPOSITION_2	
PART_REMARK_2	
ATTRIBUTE_1	sex
ATTRIBUTE_VALUE_1	female
ATTRIBUTE_UNITS_1	
ATTRIBUTE_REMARKS_1	
ATTRIBUTE_DATE_1	1977-07-14
ATTRIBUTE_DET_METH_1	
ATTRIBUTE_DETERMINER_1	D.A.Airola
ATTRIBUTE_2	total length
ATTRIBUTE_VALUE_2	189
ATTRIBUTE_UNITS_2	mm
ATTRIBUTE_REMARKS_2	
ATTRIBUTE_DATE_2	1977-07-14
ATTRIBUTE_DET_METH_2	
ATTRIBUTE_DETERMINER_2	D.A. Airola
ATTRIBUTE_3	tail length
ATTRIBUTE_VALUE_3	100
ATTRIBUTE_UNITS_3	mm
ATTRIBUTE_REMARKS_3	
ATTRIBUTE_DATE_3	1977-07-14
ATTRIBUTE_DET_METH_3	
ATTRIBUTE_DETERMINER_3	D.A. Airola
ATTRIBUTE_4	hind foot wit
ATTRIBUTE_VALUE_4	24
ATTRIBUTE_UNITS_4	mm
ATTRIBUTE_REMARKS_4	
ATTRIBUTE_DATE_4	1977-07-14
ATTRIBUTE_DET_METH_4	
ATTRIBUTE_DETERMINER_4	D.A. Airola
ATTRIBUTE_5	ear from not
ATTRIBUTE_VALUE_5	18
ATTRIBUTE_UNITS_5	mm
ATTRIBUTE_REMARKS_5	
ATTRIBUTE_DATE_5	1977-07-14
ATTRIBUTE_DET_METH_5	
ATTRIBUTE_DETERMINER_5	D.A. Airola
ATTRIBUTE_6	weight
ATTRIBUTE_VALUE_6	21.5
ATTRIBUTE_UNITS_6	g
ATTRIBUTE_REMARKS_6	
ATTRIBUTE_DATE_6	1977-07-14
ATTRIBUTE_DET_METH_6	
ATTRIBUTE_DETERMINER_6	D.A. Airola
ATTRIBUTE_7	reproductive
ATTRIBUTE_VALUE_7	nulliparous
ATTRIBUTE_UNITS_7	
ATTRIBUTE_REMARKS_7	
ATTRIBUTE_DATE_7	
ATTRIBUTE_DET_METH_7	
ATTRIBUTE_DETERMINER_7	
COLL_OBJECT_REMARKS	

108 possible columns

56 columns filled out

ENTEREDBY	millersk
COLLECTION_CDE	Mamm
INSTITUTION_ACRONYM	BFRS
CAT_NUM	001
TAXON_NAME	Peromyscus l
NATURE_OF_ID	expert
ID_MADE_BY_AGENT	D.A.Airola
MADE_DATE	1977-07-14
VERBATIM_DATE	1977-07-14
BEGAN_DATE	1977-07-14
ENDED_DATE	1977-07-14
VERBATIM_LOCALITY	Blodgett Res
MAXIMUM_ELEVATION	4300
MINIMUM_ELEVATION	4300
ORIG_ELEV_UNITS	ft
COLLECTING_SOURCE	wild caught
COLLECTOR_AGENT_1	D.A. Airola
COLLECTOR_ROLE_1	collector
COLLECTOR_ROLE_2	collector
OTHER_ID_NUM_TYPE_1	collector nun
OTHER_ID_NUM_1	5
PART_NAME_1	skin
PART_CONDITION_1	unchecked
PART_LOT_COUNT_1	1
PART_DISPOSITION_1	in collection
ATTRIBUTE_1	sex
ATTRIBUTE_VALUE_1	female
ATTRIBUTE_DATE_1	1977-07-14
ATTRIBUTE_DETERMINER_1	D.A.Airola
ATTRIBUTE_2	total length
ATTRIBUTE_VALUE_2	189
ATTRIBUTE_UNITS_2	mm
ATTRIBUTE_DATE_2	1977-07-14
ATTRIBUTE_DETERMINER_2	D.A. Airola
ATTRIBUTE_3	tail length
ATTRIBUTE_VALUE_3	100
ATTRIBUTE_UNITS_3	mm
ATTRIBUTE_DATE_3	1977-07-14
ATTRIBUTE_DETERMINER_3	D.A. Airola
ATTRIBUTE_4	hind foot wit
ATTRIBUTE_VALUE_4	24
ATTRIBUTE_UNITS_4	mm
ATTRIBUTE_DATE_4	1977-07-14
ATTRIBUTE_DETERMINER_4	D.A. Airola
ATTRIBUTE_5	ear from not
ATTRIBUTE_VALUE_5	18
ATTRIBUTE_UNITS_5	mm
ATTRIBUTE_DATE_5	1977-07-14
ATTRIBUTE_DETERMINER_5	D.A. Airola
ATTRIBUTE_6	weight
ATTRIBUTE_VALUE_6	21.5
ATTRIBUTE_UNITS_6	g
ATTRIBUTE_DATE_6	1977-07-14
ATTRIBUTE_DETERMINER_6	D.A. Airola
ATTRIBUTE_7	reproductive
ATTRIBUTE_VALUE_7	nulliparous

Vertebrate specimens have many more columns to database than arthropods.

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

Some wet collections were donated to UC Berkeley's MVZ

Step 4

COLLECTION MAINTENANCE


want to cross-reference
hen which you have
ant
nples
ed by
mily
BIRD NESTS
Lodgett Forest Herbarium: Report on Species by Family (List without Notes)


CREAMY
Ensatina

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)


Tool Shed