Calbug
Digitization California Arthropod Collections
Jun Ying Lim
University of California, Berkeley
Dept. Integrative Biology
Calbug

• Crowd sourcing via Notes from Nature
• Transcript reconciliation (R / Python)
• Initial filters
• Normalization using reference lists
• Bulk clean up to meet database requirements
• Bulk upload into Essig's SQL databases
  – Low quality (Novel collector names, fields that fail to normalize, low transcript)
    • Manual curation of aberrant transcripts
  – High quality
Testing

• Tested some specimens that have been manually inputted into databased
• Standardized fields very useful
• UI design potentially very useful to reduce error
Essig Database

- Public
- Introduced some sanity check fields to accommodate Notes from Nature
  - EnteredBy:
  - Script version
  - Raw data from NfN (all transcription entries delimited within a single string)
- Locality fields later georeferenced (some using geoLOCATE)