Mobilizing data for global change research

Patricia Holroyd & Charles Marshall
University of California-Berkeley
University of California Museum of Paleontology

- Primarily focused on research and teaching
- Fossils spanning >600 million years
- Fossils and recent specimens from worldwide in 3 storage facilities
Collections dating back to the 1890’s

A century-old backlog of variably georeferenced but uncurated specimens (i.e., no taxonomic data captured)

- Initially, locality numbers associated with specimen lots with taxonomy and specimen catalog to follow
Some of UCMP’s many partnerships

Berkeley Natural History Museums

Inter-Institutional partnerships
Focused on outreach and
Shared resource management
Partnership focused on shared discovery

Goal: Discovering the unknown through basic science by mobilizing historic and modern biological data to understand how organisms and ecological systems have and will adapt to global change

100+ faculty, staff, postdocs, and students since 2009
7 major grants
BiGCB Partnership Catalyzed Projects

- 1 of 7 major grants

- Informatics Engine for Understanding Biotic Response to Global Change

- 10 months of student support for undergraduates for mobilization of museum data for last 100,000 years of California history
How to best mobilize data in limited time frame?

- Find and empower great students
- Find low-hanging fruit
  - Attempt comprehensive curation of small collections
  - Minimal coverage of all geographic areas and taxonomic groups
- Target pilot projects for the future
Selected students with prior data entry experience
Most had no biology or museum experience
Intensive 3 week training
Empower the team
Go on vacation
Find Low Hanging Fruit

- What do we know from existing collections? UCMP data vs. published datasets
- Verify and improve site data with collaborative georeferencing
- Open cabinet. Once you touch it, process it
Georeferencing sent to grant partners in Museum of Vertebrate Zoology

Updating and refining taxonomy put aside for later

Task swapping among team members
Find and Run Pilot Projects

- Accept some things are impossible
- Take short period to get pilot data
Data mobilized

- 589 invertebrate sites updated
- 30,074 specimens cataloged and curated
  - 15,618 invertebrates
  - 14,456 vertebrates
- Cost: ~78 cents/specimen
- ~2000 person hours

- 7.7% of all cataloged UCMP specimens
- 30% of invertebrate records
- 5.6% of vertebrate records
Data mobilization and analysis

UCMP

ucmpdb.berkeley.edu
calphotos.berkeley.edu

“Ecoengine”
holos.berkeley.edu
analysis of multiple climate and biota layers
Using mobilized data to identify key areas for future work on past biodiversity

Example:
What do we know about California Pleistocene vertebrates?

<table>
<thead>
<tr>
<th>Class</th>
<th>Published</th>
<th>Cataloged</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammalia</td>
<td>504</td>
<td>24365</td>
<td>2.1</td>
</tr>
<tr>
<td>Aves</td>
<td>128</td>
<td>1669</td>
<td>7.7</td>
</tr>
<tr>
<td>Reptilia</td>
<td>58</td>
<td>1209</td>
<td>4.8</td>
</tr>
<tr>
<td>Amphibia</td>
<td>2</td>
<td>219</td>
<td>1.0</td>
</tr>
<tr>
<td>Osteichthyes</td>
<td>1</td>
<td>765</td>
<td>0.1</td>
</tr>
<tr>
<td>Chondrichthyes</td>
<td>229</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Significant mismatch in specimens, diversity, and published record

Excludes 2014-2015 publications
Seeded new and future projects

- Within UCMP, used Keck to seed 4 new projects since 2014
  - Current IMLS
  - 2 new NSF ADBC (TCN and PEN)
  - 1 pending NSF-BIO
- Multiple publications in progress

McKittrick and Maricopa: California’s Other Tarpits
Lessons Learned

- Breaking up workflows can greatly accelerate data capture
- If you touch it, process it
- Being open to innovative partnerships can open unusual opportunities
- Incorporate data mobilization into every grant
Thanks

- BiGCB and BNHM collaborators

- Team Keck
  - Verna Bowie and Sarah Hinman, Coordinators
  - Our students: Alison Young, Melissa Mast, Isaac Brazil, Molly Raven, Elyanah Posner, Eric Holt
  - Programming: Joyce Gross, Ginger Ogle, Michelle Koo

- Funders
  - W.M. Keck Foundation
  - Vice Chancellor for Research UC Berkeley