The Importance of Taxonomic Quality Control in Paleontological Digitization: Strategies for Increasing Fitness for Use and Trust in Aggregated Data

Katy Estes Smargiassi, Austin Hendy, Erica Krimmel, Jann Vendetti, and Lindsay Walker

Invertebrate Paleontology and Malacology

Natural History Museum of Los Angeles County
The Problem

• Responsibility for the quality of taxonomic data can be argued to belong to the
  – Data providers (who are responsible for identifications)
  – Data aggregators (who develop a unifying taxonomic backbone)
  – Downstream users (who analyze the data)
Taxonomic Quality over Quantity?

• Taxonomy of fossil specimens is fundamental to paleobiology research.

• Therefore, it is important that identifications of these specimens are as accurate and precise as possible.
Filling gaps in the LACMIP collection

<table>
<thead>
<tr>
<th>Class</th>
<th>% indetermined*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bivalvia</td>
<td>3.7</td>
</tr>
<tr>
<td>Echinoidea</td>
<td>33.6</td>
</tr>
<tr>
<td>Gastropoda</td>
<td>4.0</td>
</tr>
<tr>
<td>Malacostraca</td>
<td>91.3†</td>
</tr>
<tr>
<td>Polyplacophora</td>
<td>28.7</td>
</tr>
<tr>
<td>Scaphopoda</td>
<td>2.7</td>
</tr>
</tbody>
</table>

*Limited to data generated through the EPICC-TCN (Cenozoic only).
†Many reidentified by taxonomic expert, but awaiting updating of taxonomic dictionary
Filling gaps in the LACMIP collection

<table>
<thead>
<tr>
<th>Age</th>
<th>% indetermined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleistocene</td>
<td>5</td>
</tr>
<tr>
<td>Pliocene</td>
<td>8.2</td>
</tr>
<tr>
<td>Miocene</td>
<td>9.2</td>
</tr>
<tr>
<td>Oligocene</td>
<td>8.8</td>
</tr>
<tr>
<td>Eocene</td>
<td>12.9</td>
</tr>
<tr>
<td>Paleocene</td>
<td>10.5</td>
</tr>
</tbody>
</table>
Where to Start?

• Fossil invertebrate taxonomic groups lack the species-level compendia that aid classification of many neontological plant and animal groups.

• Existing databases provide an easy solution;
  – WoRMS: World Registry of Marine Species (taxon matching tools)
  – PBDB: Paleobiology Database

• These are being used in both the digitization process AND as the primary taxonomic backbones for data aggregators (e.g., GBIF, iDigBio).

• How well do these tools perform as a service to a major digitization effort (Eastern Pacific Invertebrate Communities of the Cenozoic-TCN)?
Expert identification

Analysis of historic labels

[Images of historic labels and people examining specimens]
Historical trends in taxon matching

- For Pleistocene age fossils, a taxon-match with WoRMS will capture about 65-85% of specimen records.
- This decreases with age of original identification.
- Very few species are extinct!
Historical trends in taxon matching

- For Pliocene-age fossils, a taxon-match with WoRMS will capture no more than 75% of
- More species are extinct!
Using WoRMS for fossil invertebrates

- Through geologic time the matches of specimen records increases as the number of extant species increase.
- Only really useful for Pliocene-age fossils onwards
Improving on a good taxonomic backbone

- Adding in the taxonomic opinions of the Paleobiology Database improves the % of matches with specimen records
Now about 40-80% of pre-Pliocene specimen records have matches with our two taxonomic resources
Improving on a good taxonomic backbone

- But, what about the gap?
- We (paleontology community) need to resolve this
Strategies for success

• Involvement of experts
  • identification of specimens
  • building taxonomic dictionaries
  • project design

• Develop taxonomic dictionaries with internal consistency

• Implement internally consistent taxonomic dictionaries when migrating to a new database or when starting fresh

• Work together to identify and ENHANCE taxonomic resources
Putting the dead to work: Late Cretaceous biogeography

Taxonomic checklists

- **Mollusca**
  - **Gastropoda**
    - **Neogastropoda**
      - **Volutidae**
        - **Drilluta**
          - *Drilluta jacksonensis* (Anderson, 1958)
        - **Konistra**
          - *Konistra biconica* (Anderson, 1958)
        - **Longoconcha**
          - *Longoconcha eunea* Saul & Squires, 2008
        - **Retipirula**
          - *Retipirula calidula* Saul & Squires, 2008
          - *Retipirula crassitesta* (Gabb, 1869)
          - *Retipirula pinguis* Saul & Squires, 2008
        - **Varens**
          - *Varens anae* Saul & Popenoe, 1993
          - *Varens formosus* Saul & Popenoe, 1993
        - **Volutoderma**
          - *Volutoderma angelica* Saul & Squires, 2008
          - *Volutoderma averillii* (Gabb, 1864)
          - *Volutoderma elderi* Saul & Squires, 2008
          - *Volutoderma magnifica* Dall, 1903
          - *Volutoderma gabbii* White, 1889
          - *Volutoderma jalama* Saul & Squires, 2008
          - *Volutoderma guernica* Saul & Squires, 2008
          - *Volutoderma santana* Packard, 1922
          - *Volutoderma suciana* Dall, 1907
          - *Volutoderma uncaea* Saul & Squires, 2008
          - *Volutoderma? antherena* Saul & Squires, 2008

Gradient analysis

Cluster analysis

Walker et al. (this meeting) – Revitalizing the Cretaceous Seas of California (CSBR)
Putting the dead to work: Plio-Pleistocene *Ocinebrina* among the collections

Wiedrick, Walter, Wetzer & Eernisse, in prep.
Putting the dead to work: Plio-Pleistocene scaphopods

Modern biodiversity

* Dentalium neohexagonum
* Dentalium agassizi
* Dentalium inversum
* Dentalium oerstedii
* Dentalium vallicollens
* Antalis pretiosa
* Rhabdus rectius
* Graptacme semipolita
* Tesseracme hancocki
* Tesseracme quadrangularis
* Fissidentalium megathyris
* Gadila aberrans
* Gadila austinclarcki
* Gadila perpusilla
* Gadila tolmei
* Cadulus californicus
* Striocadulus albicomatus
* Siphonodentalium quadrifissatum
* Compressidens stearnsii

Species distribution (past & present)

Abundance (spms)

Dentalium neohexagonum

- Modern distribution
- Pleistocene occurrences

Gadila aberrans

- Modern distribution
- Pleistocene occurrences
THANKS!

• Co-authors Austin Hendy, Erica Krimmel, Lindsay Walker, and Jann Vendetti
• Shawn Weidrick, Scott Rugh, Torey Nyborg, and Chuck Powell, for their expert identifications and research contributions
• Thanks to the many LACMIP students who cataloged thousands of specimens to make this project possible!