Strategies for Integrating Source Materials and Paleobiology Collections

Patricia Holroyd
Key Source Material Metadata

Common Types of Source Material Metadata

• Who Made It?
• When Was It Made?
• What Is It?
• How Much of It Is There?

Key Biodiversity Metadata

• How does it relate to a specimen and to a particular time and place in earth history (i.e., a georeferenced point and a time of occurrence)?
• For paleontology, time of occurrence is in specialized vocabulary (in extended Darwin Core)
Discipline-Specific Metadata Issues

Collecting info is not always explicitly related to the organism’s age or environment

Use locality numbers to relate multiple specimens to age/environmental data
Organizing Data Tied To Occurrences

Many generations and types of documents tied to Locality V3310

Locality V3310
Camel jaw
No collector listed
Common Source Materials

Field Notes
Maps
Museum Records
Scientific Artwork
Stratigraphic sections
“Gray literature” reports

Personal papers & correspondence
All types of photographic media

Many Kinds of Media
Different Storage Needs
Differing ways to organize

And increasingly “born digital” data (e.g., CT data, GPS logs, email)
Organizing Source Materials in “Functional Units”

Groupings based on

• common uses or points of reference in existing catalogs and curatorial work flows
  - Check-box for supplemental locality files on locality cards

• Media type/storage requirements
  (i.e., maps in a map case)
The geographic location of an occurrence should be the most stable piece of information, so it is a practical way to associate related data that may have few other apparent commonalities.

Most entries tied to specific localities or geologic units
Most “Misc.” is correspondence related to finds in other museums or still in field
Source Material Challenges + Decision-making

• Secure and conserve source materials
  o Do you limit access to the source itself?
  o Do you capture metadata and/or the source itself?
• Generate and organize useful data on source materials
  o How do you document the relationship between archival objects and collections (e.g., finding aids, databases)?
  o How do you organize the source material to reflect those relationships?
• Facilitate discovery
  o Draw on user information to guide data capture and generation
Ways To Relate These Data: UCMP example

Collection data stored in two tables:
Specimen & Locality

318,056 specimens & 90,585 localities

Images stored in CalPhotos, an external image database
Approx. 6000 labels
1700 archives images
3800 specimen images

MySQL tables with Perl-scripted web interface
Integrating source metadata + existing database structures

Problem:
Often impractical to develop new programming

One solution:
We just extended data tables and added Archon record number hyperlink
Source Metadata Record

Use to identify materials, key disciplinary data, standard subject heading
Records can be associated with multiple links to other resources.

Source Material
Metadata

Link that generates live SQL query on specimen database

UCMP UC Museum of Paleontology Specimens

Number of matches: 3176
next 100

• Download your results (tab-delimited text file with .xls file extension, 3176 lines, file size =427.3 K)
• Map specimens with a US county

QUERY: SELECT from ucmp WHERE spec_archives like "%46%" ORDER BY cast(specno as unsigned integer)

Click on the Spec # to see the full specimen record

<table>
<thead>
<tr>
<th>Spec #</th>
<th>Collection</th>
<th>Class</th>
<th>Genus species ssp</th>
<th>Other Name</th>
<th>Type Status</th>
<th>Period</th>
<th>Epoch</th>
<th>Loc ID#</th>
<th>Locality Name</th>
<th>County</th>
<th>State/Prov</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>76901</td>
<td>Vertebrates</td>
<td>Mammalia</td>
<td>Merychys elegans</td>
<td></td>
<td>Tertiary</td>
<td>Miocene</td>
<td>V6465</td>
<td>Boron</td>
<td>Kern County</td>
<td>California</td>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>76902</td>
<td>Vertebrates</td>
<td>Mammalia</td>
<td>Merychys elegans</td>
<td></td>
<td>Tertiary</td>
<td>Miocene</td>
<td>V6465</td>
<td>Boron</td>
<td>Kern County</td>
<td>California</td>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>76903</td>
<td>Vertebrates</td>
<td>Mammalia</td>
<td>Merychys elegans</td>
<td></td>
<td>Tertiary</td>
<td>Miocene</td>
<td>V6465</td>
<td>Boron</td>
<td>Kern County</td>
<td>California</td>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>76904</td>
<td>Vertebrates</td>
<td>Mammalia</td>
<td>Merychys elegans</td>
<td></td>
<td>Tertiary</td>
<td>Miocene</td>
<td>V6465</td>
<td>Boron</td>
<td>Kern County</td>
<td>California</td>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>76905</td>
<td>Vertebrates</td>
<td>Mammalia</td>
<td>Merychys elegans</td>
<td></td>
<td>Tertiary</td>
<td>Miocene</td>
<td>V6465</td>
<td>Boron</td>
<td>Kern County</td>
<td>California</td>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>76906</td>
<td>Vertebrates</td>
<td>Mammalia</td>
<td>Merychys elegans</td>
<td></td>
<td>Tertiary</td>
<td>Miocene</td>
<td>V6465</td>
<td>Boron</td>
<td>Kern County</td>
<td>California</td>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>76907</td>
<td>Vertebrates</td>
<td>Mammalia</td>
<td>Oxydactylus</td>
<td></td>
<td>Tertiary</td>
<td>Miocene</td>
<td>V6465</td>
<td>Boron</td>
<td>Kern County</td>
<td>California</td>
<td>United States</td>
<td></td>
</tr>
<tr>
<td>76908</td>
<td>Vertebrates</td>
<td>Mammalia</td>
<td>Aleomeryx occidentalis</td>
<td></td>
<td>Tertiary</td>
<td>Miocene</td>
<td>V6465</td>
<td>Boron</td>
<td>Kern County</td>
<td>California</td>
<td>United States</td>
<td></td>
</tr>
</tbody>
</table>
Locality data sent to map display

Source Material Metadata

Live query result

Individual specimen record

Clickable Google map of Archibald collecting
Workflow for UCMP CLIR project

1. Discovery – We knew we had many documents tied to both collectors and localities and then we found more
2. Prioritization – Field Notes 1st; supplemental locality files 2nd; personal papers 3rd based on frequency of use + funding + staging of work space and people
3. Concurrent conservation (rehousing) and data extraction/data entry
4. Entry of explicit links (data associations) in related in-house databases

Future Possible Workflow

1. Discovery – Find more
2. Prioritization – materials near end of lifespan
3. Digital Reproduction of prioritized items
4. Sustainability - Integrating source material data extraction and capture into standard cataloging workflows and increasing granularity
Acknowledgments & Thanks

CLIR Hidden Collections Program and NSF-BRC
Also: Keck Foundation, Bureau of Land Management, US Geological Survey, and National Park Service, CalTrans

UCMP Archive Team: Dave Smith, Joyce Gross, Mark Goodwin, Bill Clemens, Judy Scotchmoor, Susumu Tomiya, Sarah Werning, Elizabeth Ferrer, Lucy Chang, Kaitlin Maguire, Jeff Benca, Jessica Jedvaj, Dana Riess, Elizabeth Childs, Amanda Burtt, Alexandra Schwarz, Carly Tribull, Cindy Kok, Kathryn Rieck, Kevin Phan, Julie Ding, Min-Sun Jun, and Anoop Mannan, Dave Strauss, Doug Clarke, Claire Englander, Erica Clites, Diane Erwin, Ken Finger

Many colleagues throughout the UC-Berkeley Natural History Museums and library system and the iDigBio community