Constructing a Coleoptera Anatomy Ontology – how and why

Nico M. Franz
School of Life Sciences, Arizona State University

Third International Tenebrionoidea Symposium
August 7-8, 2013 – Arizona State University, AZ
"The largest contribution to *Eleodes* taxonomy was a monograph on the U.S. and Baja California species by Blaisdell (1909)."

* Taken from the ARTS proposal.
Natural language isn't like logic, and vice-versa.

**Artipus descensus**  [= sp. 22 Turks and Caicos]

**Male - habitus.** Length x.xx-x.xx mm, width x.xx-x.xx mm (N = xx); shape pyriform, length/width ratio x.xx-x.xx; widest posteriad of midregion of elytra. Integument black. Integument punctulate except for presence of characteristic, linear or variously oriented and irregular glabrate punctures and foveae on pronotum and typically also in anterolateral regions of elytra, regular (non-modified) strial punctures small. Integument densely and homogeneously covered with small, subcircular to irregularly equilateral, lamellate, partially overlapping, appressed scales of varying colors: scales primarily creamy white, interspersed with light turquoise or pinkish metallic indescent scales, particularly along the midline of the pronotum and elytra (rarely more predominant than the creamy white scales), less commonly also with small variably clustered patches of darker grey or brown scales; scales surrounding pronotal and elytral foveae usually not darker; scales yellow in strial punctures; linear scales short, transparent-white or light brown, recurved, sides parallel throughout.

**Mouthparts.** Mandibles laterad and ventrad of scar with aurate and whitish setae and appressed scales. Galeo-lacinial complex of maxillae with 3 lateral and 3 apical lacinial teeth, each tooth longer and slender. Labium with lateral margins diverging, anterior margin mesally projected.

**Rostrum.** Length x.xx-x.xx mm, rostrum/pronotum length ratio x.xx-x.xx. Dorsolateral margins weakly emarginate, anteriorly diverging. Nasal plate concave-inflected, V-shaped canina present though not strongly projected; posterior region of epistoma anteriorly with a narrow glabrate band, mesally also typically glabrate, otherwise densely covered with variously bluish metallic scales. Dorsal surface of rostrum posteriad of epistoma distinctly demarcated by an anterior transverse ridge, thereafter with large, deep and wide triangular impression extending posteriorly beyond anterior margins of eyes, densely covered with whitish scales except mesal hairline often glabrate, posteriorly continuous with deep, linear (though not hairline-like), anteriorly and posteriorly tapered, glabrate median sulcus, sulcus posteriorly contiguous with (and indistinguishable from) frontal fovea which extends narrowly to posterior margins of eyes.
Natural language isn't like logic, and vice-versa
Where science is heading – "ladder" of controlled vocabularies

⇒ We will spend more time explaining ourselves to computers.

Source: http://www.mkbergman.com/
Advancing from natural language to (logic-proven) trust

Published taxonomic information
.doc, .xls, .jpg, .pdf, .html (etc. ⇒ natural language)

Overview of upper-level ontology domains

<table>
<thead>
<tr>
<th>RELATION TO TIME</th>
<th>CONTINUANT</th>
<th>OCCURRENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRANULARITY</td>
<td>INDEPENDENT</td>
<td>DEPENDENT</td>
</tr>
<tr>
<td>ORGAN &amp; ORGANISM</td>
<td>Organism (NCBI Taxonomy)</td>
<td>Anatomical Entity (FMA, CARO)</td>
</tr>
<tr>
<td>CELL &amp; CELLULAR COMPONENT</td>
<td>Cell (CL)</td>
<td>Cellular Component (FMA, GO)</td>
</tr>
<tr>
<td>MOLECULE</td>
<td>Molecule (ChEBI, SO, RnaO, PrO)</td>
<td>Molecular Function (GO)</td>
</tr>
</tbody>
</table>

⇒ Ontology domains that are specific to taxonomy remain underdeveloped.

Interim message – controlled language: neat, though at what cost?

⇒ Most controlled, reasoning, yet not expressive

⇒ Traditional, most expressive, yet least logical
Ontologies provide structured definitions, and *some* reasoning

"Ontologies represent formalized domains of knowledge, whereby **classes are related to one another** to enable logical reasoning. For example, we have three classes: **leg, femur, and fore leg**, and these classes are related as: **femur part_of leg** and **fore leg is_a leg**.

Given a statement, perhaps read in a species description, that "**legs are yellow**" we can reason that the **fore leg must be yellow** and that the **femur must also be yellow** since they each inherit this property from their parent class: leg."

The HAO and applications – a template for Coleoptera?


2. HAO in Ontobee – http://www.ontobee.org/

3. HAO editing in OBO-Edit2 (desktop authoring tool)

4. HAO in mx (web-based editing tool) ⇒ preferred starting point for CAO
   http://mx.speciesfile.org/account/login

5. HAO in Protégé (desktop OWL editor) – reasoning = consistency checking, queries

6. Putting it all together – HAO, Phenex (EQ statements, individuals), Protégé – Smart queries
   http://sysbio.oxfordjournals.org/content/early/2013/05/31/sysbio.syt028.full.pdf+html
Prospects for a CAO, and limitations to consider

• **mx** is a nice data assembly / annotation / ontology building platform
  • requires a concerted, collaborative process to assemble CAO
  • *Eleodes* project members and collaborators can lay the ground work

• **Lots of issues remain**
  • Process is technically challenging
  • Time consuming
  • Do we *need* ontology-enabled queries of all and anything?
  • Higher-level integration with the OBO "paradigm"
  • Representing homology, apomorphy, homoplasy
  • Handling nuances, exceptions (which may have a genetic underpinning)
  • DL is good at representing EQ statements, not good at rule-based reasoning
  • Arguably it's time to reflect on the question: "what's in it *for systematics*?"