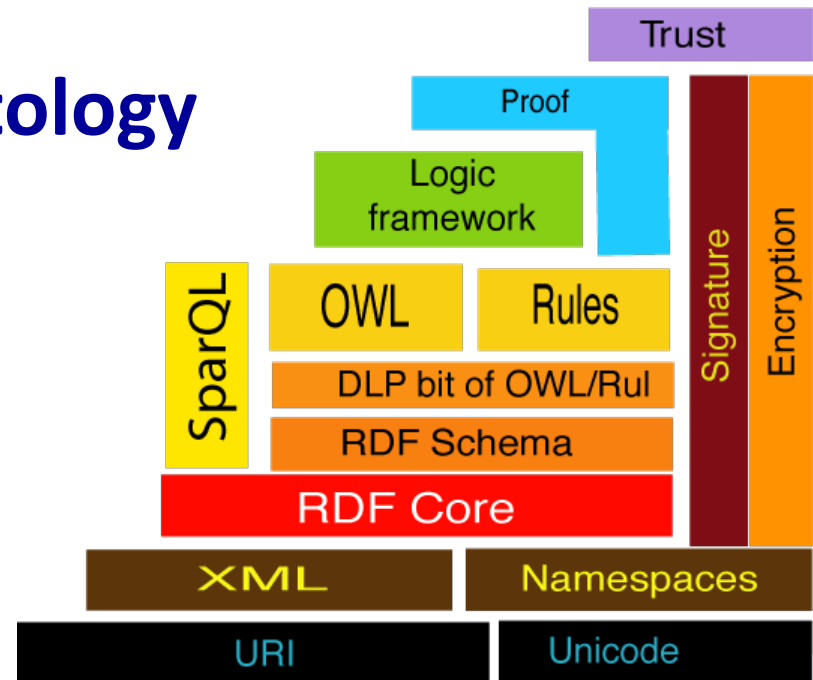


Constructing a Coleoptera Anatomy Ontology – how and why



Nico M. Franz

School of Life Sciences, Arizona State University

Third International Tenebrionoidea Symposium

August 7-8, 2013 – Arizona State University, AZ

The challenge (and promise) in a nutshell *

"The largest contribution to *Eleodes* taxonomy was a monograph on the U.S. and Baja California species by Blaisdell (1909)."

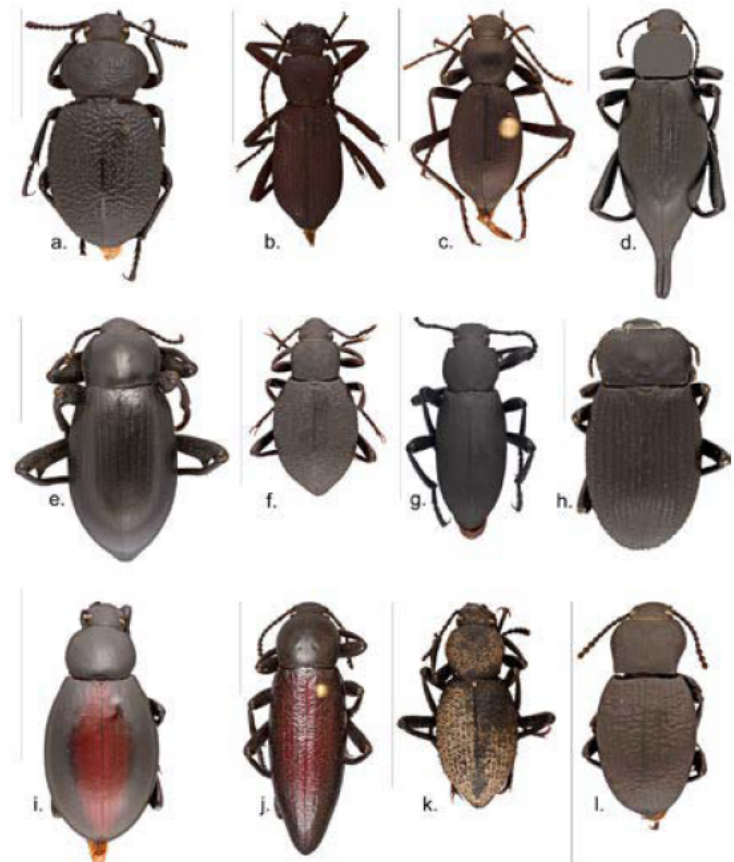
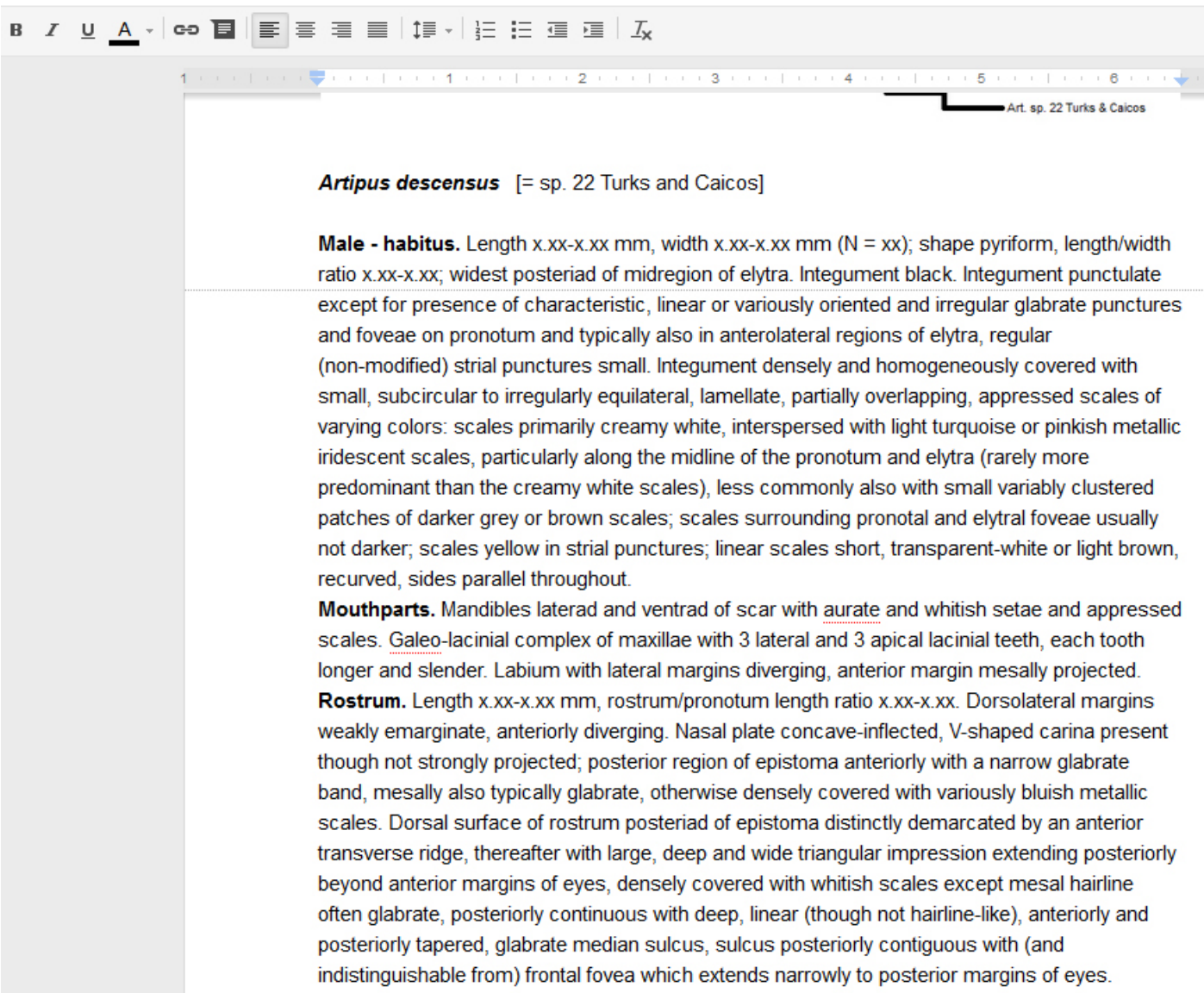


Figure 1. Dorsal habitus of *Eleodes* species: (a) *E. (Blapyllis) scabrosus*; (b) *E. (Caverneleodes) rugosifrons*; (c) *E. (Caverneleodes) wynnei*; (d) *E. (Eleodes) eschscholtzi*; (e) *E. (Eleodes) lorentensis*; (f) *E. (Litheleodes) letcheri*; (g) *E. (Metablapyllis) n. sp.*; (h) *E. (Melaneleodes) pedinoides*; (i) *E. (Promus) scapularis*; (j) *E. (Steneleodes) gravidus*; (k) *E. (Tricheleodes) pilosus*; (l) *E. (incertae sedis) commatus*. Scale bars = 1 cm. Photos by A. Smith / ASU Wheeler Lab.

* Taken from the ARTS proposal.

Natural language isn't like logic, and vice-versa

Last edit was yesterday at 4:26 PM



The image shows a screenshot of a rich text editor interface. At the top, there is a toolbar with various icons for bold, italic, underline, text color, background color, bulleted list, numbered list, link, unlink, and source code. Below the toolbar is a horizontal ruler with a scale from 1 to 6. The main text area contains a taxonomic description for *Artipus descensus*. The text is formatted with bold for section headers and italics for the species name. The description includes details about the male habitus, mouthparts, and rostrum, with some terms like 'aurate' and 'galeo-lacinal' underlined in red.

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 iridescent 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.

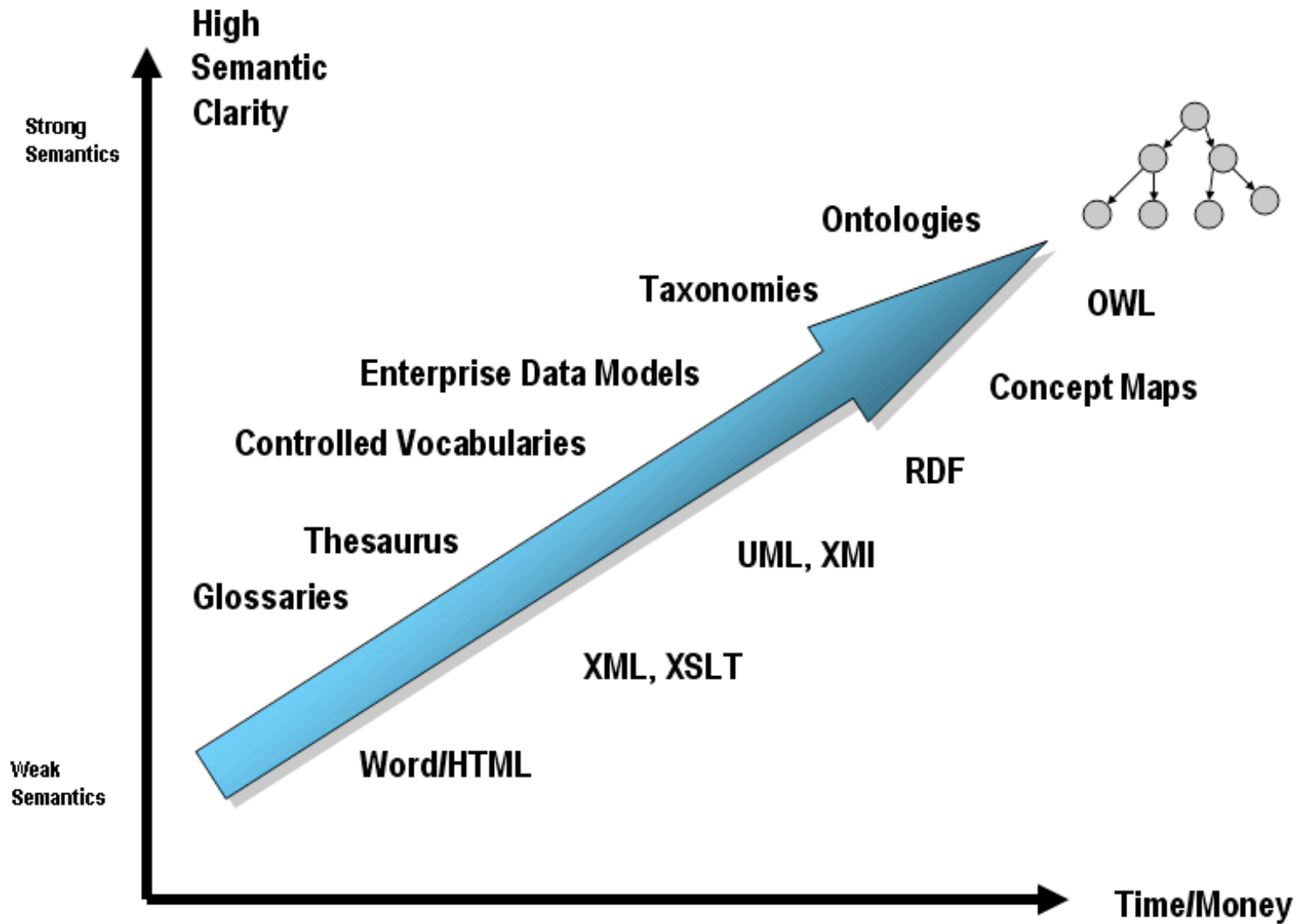
Mouthparts. Mandibles laterad and ventrad of scar with aurate and whitish setae and appressed scales. Galeo-lacinal complex of maxillae with 3 lateral and 3 apical lacinial teeth, each tooth longer and slender. Labium with lateral margins diverging, anterior margin mesally projected.

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 carina 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

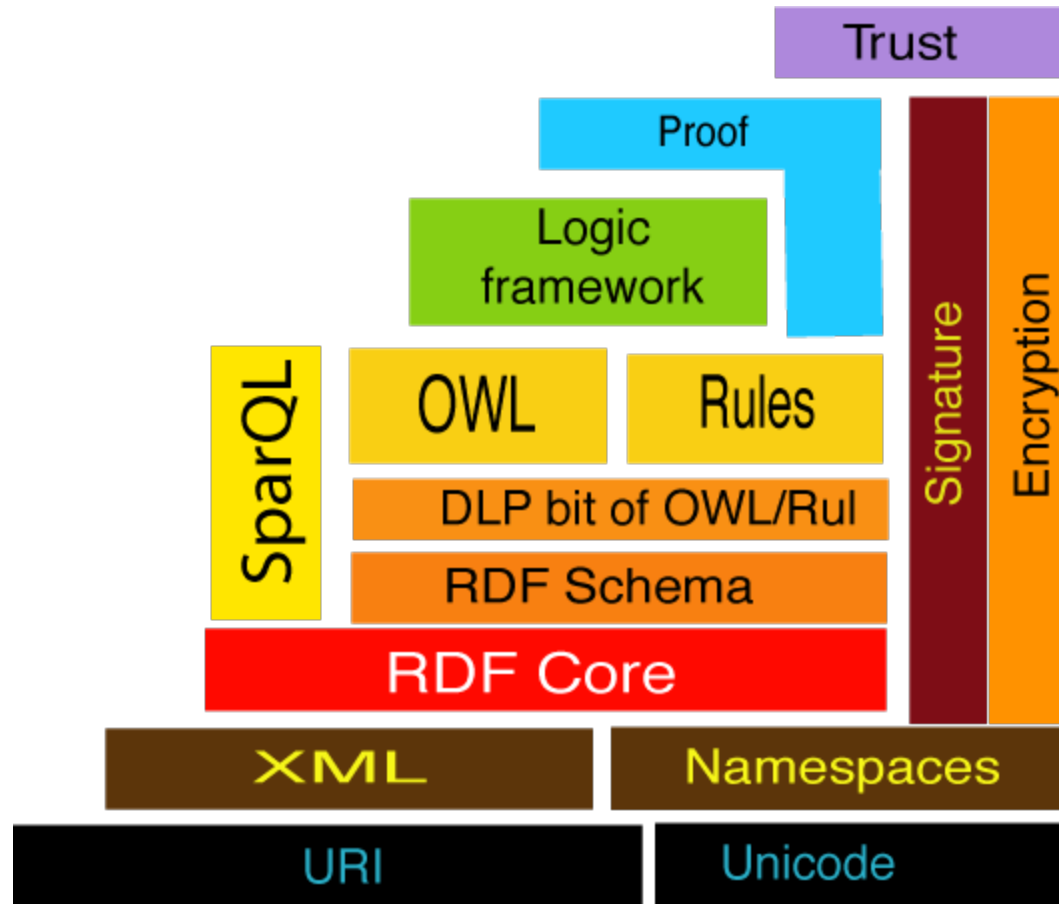
```
17  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
18  % 2. Define the variables and domains - node (= taxon names), length
19  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
20
21  node(1..n).
22  length(0..1).
23
24  % Define the domains - nodes and lengths
25  #domain node(N).
26  #domain node(N1).
27  #domain node(N2).
28  #domain length(L).
29  #domain length(L1).
30  #domain length(L2).
31
32  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
33  % 3. Generate stable models - n taxonomic trees (single root, etc.)
34  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
35
36  % Select 1 node to serve as the root of the tree
37  1{root(NN) : node(NN)}1.
38
39  % Select each node to have only a single parent
40  1{edge(N,NN) : node(NN)}1 <- node(N) & not root(N).
41
42  % Assign rankings = path lengths of 1 to the edges (counting paths); if there is an edge then rank = 1
43  path(N,N1,1) <- edge(N,N1).
44
45  % Define cumulative paths from child to grandparent, etc. (nodes N to N1 to N2, lengths L1 + L2)
46  path(N,N2,L1+L2) <- path(N,N1,L1) & path(N1,N2,L2).
47
48  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
49  % 4. Universal constraints - root is path-less, no cyclical trees
50  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
51
52  % Prohibit cyclical networks
53  <- path(N,N,L).
54
55  % All nodes must have a path of some length to the root
56  path(N,N1) <- path(N,N1,L).
57  <- not path(N,N1) & root(N1) & N!=N1.
58
59  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
60  % 5. Additional (optional) constraints
61  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
62
63  % Constraint 1 - specify that a particular taxon (number) is always the root
64  % Taxon 1 is specified as the root
65  <- not root(1).
```

Where science is heading – "ladder" of controlled vocabularies



⇒ We will spend more time explaining ourselves to computers.

Advancing from natural language to (logic-proven) trust



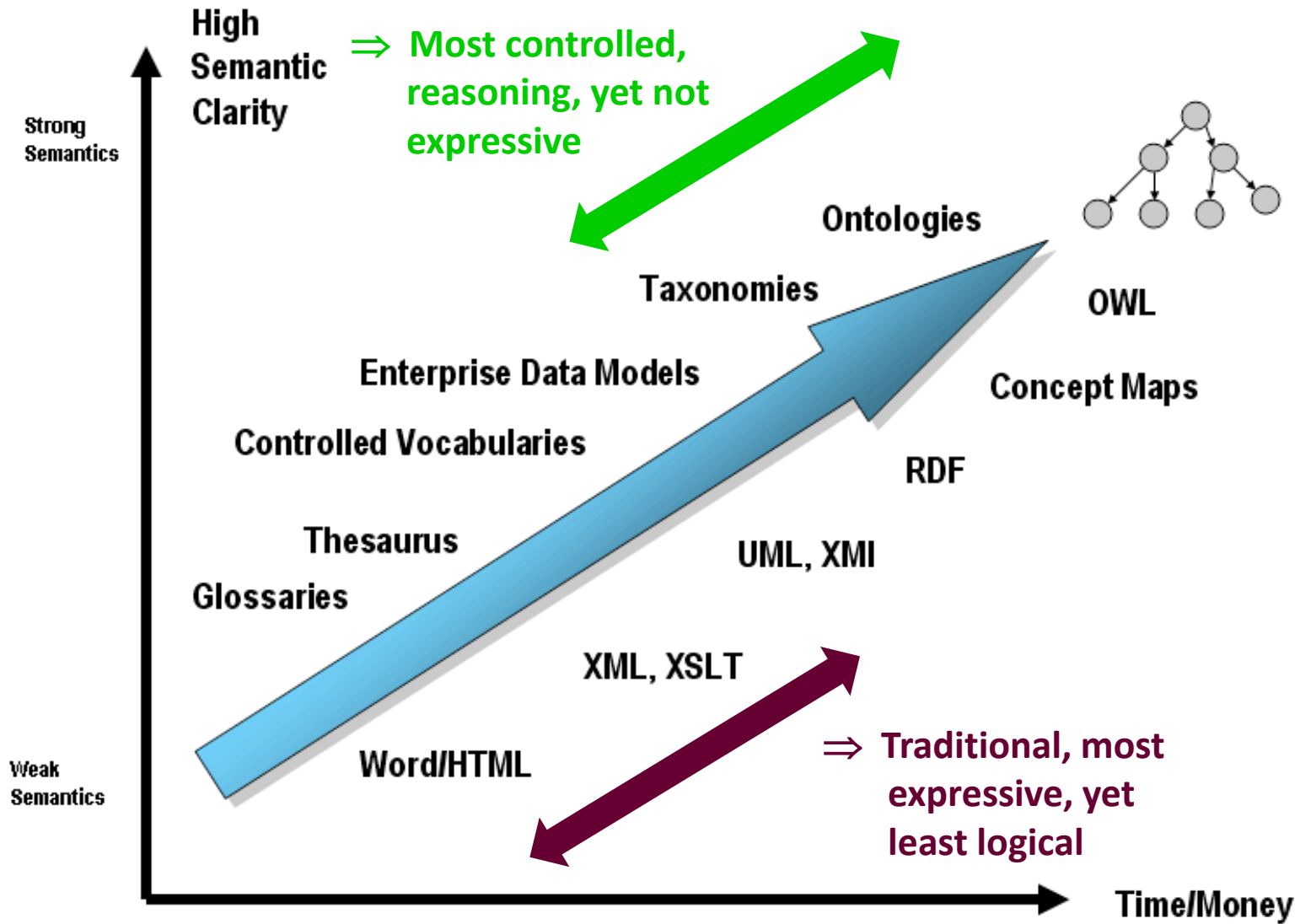
Published taxonomic information
.doc, .xls, .jpg, .pdf, .html (etc. ⇒ natural language)

Overview of upper-level ontology domains

RELATION TO TIME	CONTINUANT				OCCURRENT
	INDEPENDENT		DEPENDENT		
GRANULARITY					
ORGAN & ORGANISM	Organism (NCBI Taxonomy)	Anatomical Entity (FMA, CARO)	Organ Function (FMP, CPRO)	Phenotypic Quality (PaTO) (HAO, etc.)	Biological Process (GO)
CELL & CELLULAR COMPONENT	Cell (CL)	Cellular Component (FMA, GO)	Cellular Function (GO)		
MOLECULE	Molecule (ChEBI, SO, RnaO, PrO)		Molecular Function (GO)		Molecular Process (GO)

⇒ Ontology domains that are *specific* to taxonomy remain underdeveloped.

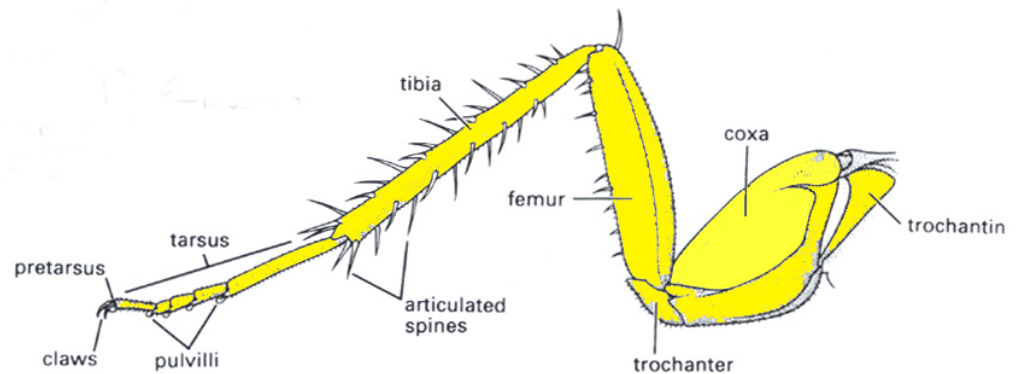
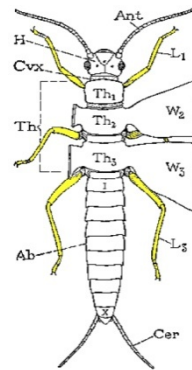
Interim message – controlled language: neat, though at what cost?



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?

1. Hymenoptera Anatomy Ontology **Glossary** – <http://portal.hymao.org/>
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?*"