Why ontologies are important for understanding morphological images

Greg Riccardi and Austin Mast
Florida State University
Topics

• Morphbank overview
• Images and Research Work Processes
• Ontology Integration

riccardi@ci.fsu.edu
Early Exposure to Biodiversity Informatics

- The Friendly Book by Margaret Wise Brown
  ill. Garth Williams ©1954
- Includes
  - Morphology
  - Behavior
  - Ecology
  - Taxonomy
- Current version is missing 4 pages
  - I like bugs
  - I like seeds
  - I like fish

riccardi@ci.fsu.edu
Early exposure to biodiversity info

From *Friendly Stories* by Margaret Wise Brown
"I like seeds" has disappeared

Your search - margaret wise brown "i like seeds" - did not match any documents.

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.
- Try fewer keywords.

riccardi@ci.fsu.edu
Topics

• Morphbank overview
• Images and Research Work Processes
• Ontology Integration
Morphbank Project

- Repository of images of organisms
  - 220,000 images so far
  - Each image has a context:
    - Specimen, taxon, locality, specimen part, view angle, etc.

- Repository of information related to the images
  - Specimens, localities, users, groups, taxa, annotations, collections
  - Owner, group, date, permissions
  - Unique identity for each object

- All objects are connected through collections and associations (aka annotations)
Why is Morphbank Successful?

• Create value for users
  – Actively adapt system to needs of users
  – Give users credit for contribution

• Support for contributors
  – Hire people who like people
  – Make sure that anyone who wants to give us content is made welcome
  – Help contributors to understand metadata standards and the value of data
Sorting Images by Observed Features

- Geranium 'Brookside'
- Geranium 'Blue Sunrise'
- Geranium 'Nimbus'
- Geranium 'Brookside'
- Geranium 'Himalayense'
- Geranium 'Orion'
- Geranium 'Johnson's Blue'
- Geranium 'Johnson's Album'
- Geranium phaeum 'In'
- Geranium macrorrhizum 'In'
- Geranium macrorrhizum 'Va'
- Geranium maculatum
- Geranium maculatum 'Album'
- Geranium nodosum 'Svelte'
- Geranium macrorrhizum
- Geranium macrorrhizum 'In'

NSF
Defining a Character and its States

State: 0
Elongate
- Uynneohorax undulatus
- Callochrysa marmorata
- Elopysana

State: 1
Broad and flat
- Pardachirus pavo

State: 2
Laterally compressed subrectangular
- Asaromorbis intermedius
- Forcipiger longirostris
- Chaetodon fasciatus
- Chaetodon nitens
- Pemacanthus semicirculatus

State: 3
Laterally compressed ovoid
- Carangoides hedianensis
- Chromis arnoldi
- Sphyraena

Undefined
- Amphiprion bicinctus
Linking to Morphbank Images

http://www.morphbank.net/?id=111266&imgType=jpeg
Upload of Images and Metadata

• Web site for small number of images

• Bulk insertion
  – Organize metadata
    • Present in excel or xml
  – Organize images
    • Attach to metadata
  – Send info to Morphbank
    • Via ftp and/or service interface
Other People’s Software

• NIMBY (not in my backyard)
  Software
    – I’d love to use your software as long as I don’t have to maintain it.
    – Invest in standards that allow other people’s software to enhance my infrastructure
Live Examples From Service API

• Interface with image tools
  – mediars

• Interface with RSS Readers
  – Find by keyword
  – Find recent changes

• Interface with Google maps
  – Georss

• Import and export with XML
Controlled Vocabulary in Morphbank

• Each image has a specimen and a view
  – Specimen has Darwin Core fields
    • Taxon, sex, form, preparation type, ...
  – View has fields that describe what you see
    • specimen part, view angle, sex, imaging technique, imaging preparation type

• Field values create a controlled vocabulary for describing image content

• Search for keywords within vocabulary
  – geranium (taxon)
  – leaf (specimen part)
  – adaxial (view angle)
Topics

• Morphbank overview
• Images and Research Work Processes
• Ontology Integration
Morphology Work Processes

• What did Linnaeus do?
  – Collected specimens
  – Examined and described features
  – Created illustrations with annotations
  – Catalogued features and separated specimens

• What has changed?
  – How has the process been improved?
  – What are new work processes?
Morphology Publication Example

Head, anterior view, female

1. [Modified from R1] Shape of ventral part of clypeus: (a) rounded, broadly projecting over mandibles (Fig. R8); (b) not projecting from cranial margin or slightly and narrowly projecting medially (Fig. R7). (CI = 0.20, RI = 0.67)
2. Shape of projecting ventral clypeal margin: (a) straight or rounded (not illustrated); (b) with a median incision (Fig. R8). (CI = 0.29, RI = 0.29)
3. [= R2] (CI = 1.00, RI = 1.00)
4. Direction of clypeo-pleurostomal lines: (a) ventrally diverging (Fig. R8); (b) ventrally converging (Fig. R7a). (CI = 0.33, RI = 0.57)
5. [= R3] (CI = 0.17, RI = 0.38)

<table>
<thead>
<tr>
<th>Character</th>
<th>1</th>
<th>6</th>
<th>11</th>
<th>16</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syneropus</td>
<td>1</td>
<td>b–b–b</td>
<td>aaaaa</td>
<td>aaaca</td>
<td>cbbac</td>
</tr>
<tr>
<td>Pencillus</td>
<td>2</td>
<td>b–b–b</td>
<td>aaaaa</td>
<td>aaaca</td>
<td>cbbac</td>
</tr>
<tr>
<td>Ceroptus</td>
<td>3</td>
<td>b–b–b</td>
<td>aaaba</td>
<td>aaaca</td>
<td>cbac</td>
</tr>
<tr>
<td>Synophromorpha</td>
<td>4</td>
<td>b–b–a</td>
<td>aaaba</td>
<td>aaacb</td>
<td>–baab</td>
</tr>
<tr>
<td>Xestophanes</td>
<td>5</td>
<td>b–b–a</td>
<td>aaaba</td>
<td>aaabb</td>
<td>–ba–b</td>
</tr>
<tr>
<td>Diastrophus</td>
<td>6</td>
<td>abb–b</td>
<td>ababa</td>
<td>aaacb</td>
<td>–ba–c</td>
</tr>
<tr>
<td>Gonaspis</td>
<td>7</td>
<td>b–b–b</td>
<td>aaaba</td>
<td>aaabb</td>
<td>–ba–c</td>
</tr>
<tr>
<td>Liposthemes</td>
<td>8</td>
<td>abaaa</td>
<td>adaaa</td>
<td>ababa</td>
<td>ccaca</td>
</tr>
<tr>
<td>Ai. kerner</td>
<td>9</td>
<td>aaaaab</td>
<td>aacaa</td>
<td>aabaa</td>
<td>b????</td>
</tr>
<tr>
<td>Cecconia</td>
<td>10</td>
<td>b–aba</td>
<td>baaa</td>
<td>aaaba</td>
<td>a????</td>
</tr>
<tr>
<td>Antistrophus</td>
<td>11</td>
<td>b–aba</td>
<td>bcaaa</td>
<td>a?aba</td>
<td>ababa</td>
</tr>
<tr>
<td>Rhodus</td>
<td>12</td>
<td>b–aba</td>
<td>aaaaa</td>
<td>aaaaa</td>
<td>a????</td>
</tr>
<tr>
<td>Hedickiana</td>
<td>13</td>
<td>b–aba</td>
<td>baaa</td>
<td>aaaa</td>
<td>a????</td>
</tr>
</tbody>
</table>
CTOL Images

- Apparent work process
  - Collect
  - Image
  - Annotate
CTOL in Morphbank

Teleost anatomy and development

Version 1.78

cerotobranchial 5 tooth

ID: TAO:0000694

Definition:
The tips of the ceratobranchial 5 teeth form before the ceratobranchial itself (3.0 mm NL). As they grow ventrally, they meet the perichondrally ossified surface of ceratobranchial 5 and become firmly ankylosed to the bone. Ceratobranchial teeth are the only teeth present in zebrafish and other cypriniforms.

Database References: ZFIN

Develops From: dental organ

Part Of: pharynx
cerotobranchial 5 bone

Exact Synonym: "pharyngeal teeth"

Is A: tooth
Spider ATOL Workflow

- **Spider ATOL:** 35000 images organized according to genus and anatomy
- **Purpose of images**
  - Illustrate the matrix
  - Or populate the matrix with values, according to information in images
- **Workflow:**
  - Define “standard views” based on presentation of features in the literature
  - Define characters and associate with standard views
  - Acquire an image for each view, for each genus
  - Add images to matrix of genus x character
  - Add values to matrix with assistance from images
Online Work Process

• Collect and digitize specimens
• Catalog metadata and images
  – Add to online digital repositories ASAP
  – Protect digital information from outsiders
• Annotate images
• Create phenome descriptions
  – Extend ontology
  – Annotate anatomy
• Share online among collaborators

riccardi@ci.fsu.edu
Advantages of Online Workflow

- Illustration of
  - matrix, phenomes, morphology
- Integrity of work process/provenance
- Active sharing among collaborators
- Smooth transition to publication and archiving
- Fitness for use: Bob Morris quote:
  - We are coming to love the phrase "fitness for use". In a sense, it should be the only thing that matters for science. In the end, science is about falsification of hypotheses. You have a model and you have some data. Your question is always of the form "Does data D falsify hypothesis H and was it actually data fit for that purpose?"
Topics

• Morphbank overview
• Images and Research Work Processes
• Ontology Integration
  – The next project for Morphbank
Morphontify

• Specify (Beach, U. Kansas)
  – Specimen management
  – Desktop tool for specimen metadata management

• Morphster and Ontobrowser (Miranker, U. Texas)
  – Ontology Management for Phylogenetics
  – Extension of ontology to incorporate annotations

• Integration of Specify, Morphster and Morphbank
  – Searching for images using ontology terms
  – Linking images and other digital objects to ontology terms
  – Access to information from any user interface
Welcome to Specify 6

The Specify Software Project is delighted to announce the availability of Specify 6, a museum and herbarium collections data management system for Windows, Mac OS X, and Linux. Specify 6 processes specimen information for computerizing holdings, for tracking collection management transactions, and for mobilizing species occurrence data to the internet. Specify 6 was released on April 10, 2009, it is free and open source licensed.

Specify 6 has an intuitive user interface and highly-customizable data forms aimed at streamlining routine collections data tasks while preparing and validating collection information for research analysis. Specify has numerous features including robust support for paleontological data, field notebooks, file attachments, GUIDs, hierarchical storage locations, data uploads through the Specify Workbench and Excel, repository agreements, accession logging, conservation treatments, collection object containers, along with numerous additional functions. Specify 6 supports the use of record sets for various types of processing, such as georeferencing with GEOLocate, label and report printing, and importing and exporting, and Specify's data model now handles all institutional collections within a single database for simplified administration.
Morphster Project

• Dan Miranker at U. Texas at Austin
  – Ferner Cilloniz and other students
  – NSF funding

• Morphster is an ontology management system
  – Desktop application
  – Import and transform various ontology representations
  – Image annotation

• Ontobrowser is a Web site
  – Browse ontology terms
  – Illustrate ontology terms with images

riccardi@ci.fsu.edu
Illustration of Ontology

• Associate feature that can be seen in an image with ontology terms that describe the image
  – Area of interest in the image
  – Terms that describe anatomy, shape, etc.

• Replace
  – The (un) controlled vocabulary of Morphbank
  – With the controlled vocabulary of Morphster

• Resulting system is
  – Better for users because it is illustrated
  – Better for harvesters because it is precise

riccardi@ci.fsu.edu
Ontobrowser with Morphbank

**Herrerasaurus**

**Anatomy Ontology**
- **Skeleton**
  - tooth
  - maxilla
- **Jaw**
  - Naris
  - prefrontal
  - PosteriorProcessOfDentary
- **Openings**
  - opisthotic
  - postorbital
  - lacinal
  - Mandibular_fenestra
  - Infratemporal_fenestra
  - surangular
  - squamosal
- **Axial_Skeleton**
  - quadratojugal
  - Antorbital_fenestra

**Anatomy Illustration**
- **Name**: maxilla
- **Definition**: No definition
- **Comments**: No comments
- **Relationships**:
  - maxilla [Subclass of] Skeleton
  - maxilla [Subclass of] Axial_Skeleton
  - maxilla [Subclass of] Cranial
  - maxilla [Subclass of] Skull

**Found Morphbank Images**
- **Exemplar Images**
- **Anatomy Selected Image**
- **Similar Images**
Try Ontobrowser


- An example search:
  - Select ontology Herrerasaurus
  - Click the "Show Advance" button "Enable-> Morphbank"
  - Go to the "Term Keyword Search" on the right and search for maxilla.