

# Adding Value to Biodiversity Images through Community Annotation in the Morphbank System

Greg Riccardi  
*College of Information*  
Florida State University  
Tallahassee, FL 32306-2100 USA  
01-850-644-2869  
Riccardi@ci.fsu.edu

Morphbank, an on-line collection of museum-quality biological images, is an NSF funded project designed to facilitate the on-line collaboration of biologists around the world. Our primary goal is to aid in the collection and management of images that are useful in phylogenetic research.

Morphbank users are actively collaborating on the creation of information that represents the associations among images and related biodiversity data objects. This demo will explain the Morphbank annotation tool and data models and give examples of how users create structured information in the system. Schematized annotation provides biologists with a flexible framework to create semantically-rich annotations using their own data models.

The demo will include access to the Morphbank image repository, image annotation tools and association mapping tools. We will demonstrate the scientific process that uses the annotation tools to create semantically rich associations. Users will be given access to the full capabilities of the working system.

## 1. Morphbank Objects

Morphbank is an open Web repository of images that serves the biodiversity research community. It is currently being used to document specimens in natural history collections, to voucher DNA sequence data, and to share research results in disciplines such as taxonomy, morphometrics, comparative anatomy, and phylogenetics.

Morphbank contains information about organisms. Each image in the system is associated with one or more specimens. Each specimen is a representation of information about an organism. Specimens are in turn associated with localities, contributors, taxonomic concepts

The initial requirements analysis for the Morphbank system identified several challenges in discovering and creating information about images and their related objects.

- Finding images and specimens associated with a specific species and genus,
- Finding and recording information about that image and its related objects, and
- The discovery and recording of ad-hoc associations among the various objects.

Discovering and recording ad-hoc data is the most problematic. It is particularly difficult to find ways that users can record associations among objects.

As long as data is well formatted and constrained to the database schema, finding and retrieving it is simple. However, as we've discovered, there is no practical limit to the amount of information a scientist may wish to store with a particular specimen. Most of the knowledge is contained in the memory of these scientists or in hand written notebooks. Although it is recognized that manual annotation is expensive and time consuming it is nevertheless still essential in documenting collaborative knowledge in biological systems [2]. Translating and storing this knowledge in a searchable form is the challenge.

### 1.1 Morphbank Objects

Each object in the Morphbank system is uniquely identified and includes a set of standard fields that assist us in cataloging its location and type, the identification of the user who added the object, the date and time of creation, an optional description of the object, and the last time the object was modified. These attributes allow anyone accessing Morphbank sufficient information to find and catalog data and associate related objects. Each object is externally identified by a Life Science Identifier (LSID) [13].

### 1.2 Morphbank Object Relationships

Since each Morphbank object is uniquely identified, any object can be the target of a stored reference. A single column within a Morphbank table holding a foreign key may refer to several an object of any type. Thus a collection object can be heterogeneous. For instance, an annotation object may define an association among images, specimens, locations, users, or even other annotations.

This flexibility allows for the creation of complex collections of objects that can be shared with other users of the Morphbank system. Although there are a series of predefined relationships in Morphbank, the use of unique identifiers allows users to define an unrestricted set of complex relationships of objects within the confines of the system.

## 2. Morphbank Object Annotation

A variety of annotation technologies allow users to add value to images by creating associations between those images, text and other digital objects. Morphbank takes this one step further by making the associations into first class objects that can themselves be annotated and associated with other objects. Morphbank also allows associations to take on specific semantic characteristics that constrain their meaning and thereby improve searching and understanding.

<sup>1</sup> Supported by NSF contract DBI-0446224, 2005-2008  
WWW 2007, May 8--12, 2007, Banff, Canada.

### 3. Biological Annotation Requirements

A problem of biodiversity annotation is that biologists have increased the number of specimens they can gather but have not increased their ability to catalog, identify, and study them. Collaborations still include the exchange of physical specimens and the manual annotations of the images using indexed cards and paper documents. At the functional level, many users have developed their own specific but proprietary solution to this problem. Through the use of Morphbank and a Web based annotation tool, we can solve most if not all of these problems.

#### 3.1 Morphbank Object Annotation

A variety of annotation technologies allow users to add value to images by creating associations between those images, text and other digital objects. Morphbank takes this one step further by making the associations into first class objects that can themselves be annotated and associated with other objects. Morphbank also allows associations to take on specific semantic characteristics that constrain their meaning and thereby improve searching and understanding.

Image annotation is available in a variety of image management Web sites. The simplest annotations are found in systems that support attaching tags to images and other media. Flickr.com and YouTube.com, e.g., allow users to add text attributes (tags) to images and use those tags to support searching. FotoTagger.com, among others, goes a step further and allows the tags to be attached to specific locations on images.

Blogging is another form of image annotation in which text passages are linked to images, Web pages and other digital objects. A blog entry creates an associate between its own text and the linked objects.

Annotea.org supports the creation of RDF attributes for image tags. These attributes can be used to provide search inference capabilities for users of image repositories.

Another annotation strategy involves the development of laboratory notebooks such as those under development at the United States Department of Energy, National Collaboratories under the guidance of Dr. Jim Myers [11]. These middle-ware products present researchers, applications, problem-solving environments (PSE), and software agents with a layered set of application services that provide a finite set of capabilities for the creation and management of meta-data, the definition of semantic

relationships between data objects, and the development of electronic research records [10]. Users are able to record associations between digital objects across and among projects.

Morphbank seeks to combine these ideas by allowing incorporating an extensible annotation type system and by systematically expanding the scope of associations by including any objects referenced by globally unique IDs (GUID).

Morphbank was designed to allow users to take advantage of Web service products to gain access to the data by conforming to industry practices and standards but maintain the ontology of the original data. Users will browse or search the Web site for Morphbank objects using a variety of tools provided through the Web site.

### 4. Demonstrations

The demonstration of the Morphbank system will allow participants to collaborate by adding annotations to a collection of images of tropical fish in order to reach consensus on the similarity and diversity of their markings and colorings. Each participant will be able to identify the various features on the fish images.

Participants will use the Morphbank tools to compare the image annotations. Each person will be able to record their evaluations of these annotations.

If all goes well, the participants will reach consensus on the feature identification. Experts from outside of the workshop will add their own evaluations.

#### Full list of Authors:

*Greg Riccardi, Andrew Deans, David Gaitros, Neelima Jammingumpula, Corinne Jorgensen, Peter Jorgensen, Austin Mast, Karolina Maneva-Jakimoska, Debbie Paul, Fredrik Ronquist, Katja Seltman, Steven Winner*

*College of Information and College of Computational Science*

*Florida State University, riccardi@ci.fsu.edu*