THE FLORIDA STATE UNIVERSITY

COMPUTER SCIENCE DEPARTMENT

UPDATED
April 2006

REQUIREMENTS AND RECOMMENDATIONS FOR THE MORPHBANK WEB BASED PROJECT

AUTHORS AND CONTRIBUTORS

Fredrik Ronquist

David Gaitros

Austin Mast

Neelima Jammigumpula

Wilfredo Blanco

Matt Buffington

Maureen O'Leary

Seth Kaufmann

Andy Deans

Greg Riccard

Shayne Steele

Johan Liljeblad

Susanne Schulmeister

Michael Sharkey.

TABLE OF CONTENTS

INTRODUCTION

Purpose

This document provides for the recommended requirements, specifications, and designs for the MorphBank web site. The document is a report on the initial review of the MorphBank project and recommendations of web site layout and design, data, graphics, annotations, software, hardware, operations and maintenance to fully implement the MorphBank web site. The recommendations provided in this document are for all parties involved in the implementation and development of MorphBank, more specifically for the Hymenoptera working group.

Scope

MorphBank is a web-based project designed to provide biological information on the Hymenoptera order to the user community. MorphBank is a repository of data on Hymenoptera specimens on five categories: 1) the specimen, 2) collection data, 3) taxonomy, 4) graphics and 5) graphical annotations, and has the capability to search, browse, update and add on the many data fields maintained in MorphBank. MorphBank also maintains many views of graphical images of the specimens maintained in the database and a new feature of MorphBank is annotating the images. This new aspect will give the user the ability to document on a graphical image of a specimen, in fact, many

researchers would have the capability of annotating images, and therefore would provide an excellent means of communication and documentation on Hymenoptera specimens.

As MorphBank is web-based, the information maintained in MorphBank can be updated from a worldwide community of biological researchers which will enhance the research by providing a comparative morphological study on the Hymenoptera order and increase documentation and communication in the biological community and possibly used to discover new species. In addition, MorphBank will be accessible to all interested users who have access to the Internet and would promote new interests in the biological sciences.

The MorphBank web site has been initially designed to maintain and provide information for phylogenetic and biodiversity research for certain species, initially within the Hymenoptera order, and once successful, the web site can be mirrored for other orders.

Definitions, acronyms and abbreviations

CSS	Cascading Style Sheet
DS	Database Schema and Low Level Software Library
ESRI	Environmental Systems Research Institute
GAT	Annotation Technology
GIF	Graphics Interchange Format
HTML	Hypertext Markup Language
IEEE	Institute of Electrical and Electronics Engineers
JPEG	Joint Photographic Experts Group (image format)

MBP	MorphBank Project
OS	Operational Specifications
PHP	PHP Hypertext Preprocessor
SID	Specimen Image Database
SQL	Structured Query Language
SSC	Server and Software Configuration
TC	Trademark and Copyright
WSRD	Web Site Requirements and Design
XML	Extensible Markup Language

References

IEEE Std 830-1998. IEEE Recommended Practice for Software Requirements Specifications.

IEEE Std 2001-2002. IEEE Recommended Practice for Internet Practices – Web Page Engineering – Internet./Extranet Applications. IEEE Computer Society.

Trademark/Servicemark Application, Principal Register. 2003. US Department of Commerce, Washington. 9 July 2004 <http://teas.uspto.gov/>.

U.S. Copyright. 2000. Library of Congress, Washington. 15 July 2004 <http://www.copyright.gov/>.

Ronquist, Fredrik. E-mail from David Kopicki. July 9, 2004.

The INHS Insect Collection Database. Illinois Natural History Survey. <http://ctap.inhs.uiuc.edu/Insect/>.

College of Agricultural & Natural Resources. Department of Entomology and Applied Ecology. <http://ag.udel.edu/departments/ento/insectdb.htm>.

Stilleye. <http://www.stilleye.com>.

Hymenoptera. Department of Entomology at University of Nebraska at Lincoln. <http://www.ucmp.berkeley.edu/arthropoda/uniramia/hymenoptera.html>.

The International Plant Names Index. <http://www.ipni.org/index.html>.

Arctos. University of Alaska. <http://arctos.database.museum/SpecimenSearch.cfm>.

Glish. <http://www.glish.com/>.

A List Apart (ala). A List Apart Magazine. <http://www.alistapart.com/>.

The Paleobiology Database. National Center for Ecological Analysis and Synthesis. <http://www.paleodb.org/>.

BioImage Database. 21 July 2004. ORIEL Project – Online Research Information Environment for the Life Sciences. <http://www.bioimage.org>.

Computer Hardware: Information about Computer Tape Drives. 22 July 2004. <http://www.computerhope.com/help/tape.htm>.

Exabite Tape Backup Systems. 22 July 2004. <http://www.penguincomputing.com/products/peripherals/exabyte.php>.

Gaitros, David. Personal interview. 7 July 2004.

Gaitros, David. Personal Interview. 20 July 2004.

Gaitros, David. "Representation of Ontology Annotation Information in Grid Computing." Prospectus submitted to Department of Computer Science, Florida State University. 30 March 2004.

Ichthyology Web Resources. 21 July 2004.
http://www2.biology.ualberta.ca/jackson.hp/IWR/Specimen_Database/index.php.

Ittoolbox Knowledge Base. 15 July 2004. <http://database.ittoolbox.com>.

Jammigumpula, Neelima. Email message to Joe Barrett. 12 July 2004.

LaCie: AIT U&I Tape Drive. 22 July 2004.
<http://www.lacie.com/products/product.htm?id=10046>.

MorphBank. 21 July 2004. <http://www.morphbank.com>.

Omni Site Security. 21 July 2004. <http://www.omni-secure.com>.

Oracle8(TM) Server Backup and Recovery Guide Release 8.0. June 1997. Oracle
Corporation. 16 July 2004.
<http://www.rohan.sdsu.edu/doc/oracle/server803/A54640_01/toc.htm>.

Rycroft, S.D. "Searchable Online Image Database." Masters dissertation submitted to
University of Glasgow. September 2003.

Specimen Image Database (SID). 21 July 2004. <http://darwin.zoology.gla.ac.uk/~sid>.

The Tree of Life Project. 21 July 2004. <http://www.tolweb.org/>.

Zawodny, Jeremy. Jeremy Zawodny's MySQL Stuff.
16 July 2004. <http://jeremy.zawodny.com/mysql/>.

PHP Authentication and Authorization tutorials:

Fuecks, Harry. The PHP Anthology Volume 2, Chapter 1 – Access Control. 2 February
2004. <http://www.sitepoint.com/article/anthology-2-1-access-control>.

Meloni, Julie. Authenticate and Track Users with PHP. 2 February 2000.
<http://webmonkey.wired.com/webmonkey/00/05/index2a.html>.

Meloni, Julie. PHP-Based User Authentication. 23 May 2000.

<http://www.zend.com/tut/authentication.php>.

PHP, MySQL and Authentication 101.

<http://www.devarticles.com/c/a/MySQL/PHP-MySQL-and-Authentication-101>.

Yank, Kevin. Managing Users with PHP Sessions and MySQL. 14 June 2003.

<http://www.sitepoint.com/article/users-php-sessions-mysql>.

Apache Web Server Authentication and Authorization tutorials:

Apache Tutorial: .htaccess files. Apache HTTP Server Documentation Project.

<http://www.weberblog.com/Manuals/Apache2.0.x/howto/htaccess.html>.

Authentication, Authorization, and Access Control – Apache HTTP Server Version 1.3.

Apache HTTP Server Documentation Project.

<http://httpd.apache.org/docs/howto/auth.html>.

Authentication, Authorization, and Access Control – Apache HTTP Server Version 2.0.

Apache HTTP Server Documentation Project.

<http://www.weberblog.com/Manuals/Apache2.0.x/howto/auth.html>.

Lopez, Daniel. Restricting Website Access with Apache 2. 8 November 2002.

<http://www.samspublishing.com/articles/article.asp?p=29967>.

Overview

Since the MorphBank project (MBP) is a very large project, it has therefore been divided into six sub-projects: Web Site Requirements and Design (WSRD), Database Schema and Low Level Software Library (DS), Graphical Annotation Technology (GAT), Operational Specifications (OS), Trademark and Copyright (TC), and Server and

Software Configuration (SSC). Each project team has provided documentation in this SRS on the required tasks assigned to that team.

The documentation requirements cover the overall functionality of the web site, including the layout, in addition to the features of searching, browsing, and viewing of biological data, and general information web pages, such as publications, and news. The requirements section also covers the operations and maintenance functions of the MorphBank web site, such as access control, backup, mirror site, and hardware configuration. All of the above mentioned requirements have been defined in Section 3, Specific Requirements and is followed by the recommended specifications and designs, if applicable, for the MorphBank system.

Further research has been performed by the project teams of SSC, GAT, DS and TC and the team research and recommendations are provided as appendices to this document. Appendix A provides a report covering the hardware and software recommendations to run the MorphBank web site. The GAT group has researched many applications that could be used to provide efficient graphical annotations and the results of their findings are located in Appendix B. The DS group has researched and documented the many possible features on the database interactions both with the web site and the database and provided their findings at Appendix C. Lastly, the TC group have provided their sample trademarks in Appendix D.

OVERALL DESCRIPTION

Project perspective

The MorphBank website project is a self-contained project, and will run independently from other products, except the interface with the Internet. Other related projects currently in use are the Specimen Image Database (SID), Ichthyology Web Resources, BioImage Database and Tree of Life Project.

SID is very similar to the current MorphBank, except SID has a few additional features. SID is designed to access a distributed database, while MorphBank's database is currently centralized. SID can deliver Web Services so that other sites may link to their images, contains a Java applet to upload multiple images at once to the web server, can use the Glasgow Taxonomy Name server to verify that an image's taxon name is valid, and the software can be downloaded for free so that other organizations may setup a similar server on their own. The web site contains a lot of information on how SID was implemented.

The IWR Specimen Database has data on fish specimens in permanent collections worldwide. The list of users who are allowed to submit and edit data in the database is rather small right now. Searching for data can be done at a "low-level," in that searching can be done with SQL statements and data is mainly available in the form of text.

The BioImage Database project has the goal of providing a free image database of high-quality scientific images of biological specimens, with a well-structured searchable metadata, similar to MorphBank. Images can be two or three-dimensional, with no limit in format. Access will soon be available through E-BioSci (http://www.e-biosci.org/).

Images may not necessarily reside on the BioImage server, except for thumbnail images and full metadata.

The Tree of Life Web Project's goals are broader than the previous website mentioned. The intended audience ranges from elementary students to researchers. Its goals include: provide a uniform framework to publish information about the evolutionary history and characteristics of all groups of organisms, present a modern scientific view of the evolutionary tree, aid learning about biological diversity, provide a life-wide database and searching system about characteristics of organisms, and provide a means to find taxon-specific information. Most of the contributions are submitted by invited authors and the website has the look and feel of an online Encyclopedia.

Software interfaces

At this initial stage of the recommended requirements for the MorphBank database, there are only a few known software interfaces. The web pages will be written in HTML/XHTML and JavaScript. The applet was written in Java, PHP is used to communicate with SQL, and XML is used to work with the data in the database. The GAT group has recommended that Inote be used as the application for the graphical annotations.

Project functions

As mentioned above, MorphBank is a database maintaining biological information on the Hymenoptera order. MorphBank is web-based and will therefore be able to reach a large biological research community. Some of the functions that MorphBank will provide are:

- provides graphical annotations on specimen images;
- provides searching capabilities by taxonomy, specimen, annotation/image, and advanced searching;
- provides browsing throughout the MorphBank database;
- provides a detailed data fields of the specimens;
- provides still images of the specimens;
- provides the ability to add annotations to graphical images of specimens;
- provides the ability to link to any page on the web site without backtracking;
- provides for a publication schema;
- provides for a list of annotations on a specimen;
- provides for user login for updating annotations; and
- provides for the ability to download images.

The above are some of the primary functions that MorphBank will be able to handle. Please see Section 2.6 for additional features that could be future functionalities of the MorphBank project.

User characteristics

The primary target user community for MorphBank is the biological research community. The biological research user community will be the users responsible for providing new specimens and updating the existing MorphBank. A biological researcher would more than likely have a doctoral degree and have a vast amount of experience in the phylogenetics.

be of interest is added user-defined reports. It is recommended that a query a user repeatedly requests from MorphBank is stored in either the web server or possibly on the users hard drive. This way the user could simply click on the report and obtain the new data, and gives the system the flexibility to work with a variety of researchers. Other possible report(s) to be considered are charts and statistics based on the biological data maintained in MorphBank. Lastly, the search features should be enhanced to provide for all boolean criteria.

Future considerations for updating and adding to the MorphBank database should be implemented as a link between current data and new contributions. It is recommended that a "Contribute" button or pull down menu item that will give the user access to a web page for submission of images. Similarly, the Login and Registration links on the top of every page shall be changed to the current user's name in the format of: Title/First Name/Last Name/ and Logout.

Another possible consideration for future flexibility of the system is to have a search results page that is dynamic and would display the fields according to the data fields searched. For example, if a part of the search was on the minimum elevation, then the minimum elevation field should display as a part of the search result.

Interfaces to other applications/projects should be implemented, such as a taxonomic name server, so that taxonomic names describing the images can be verified. Other automated error-checking software may be developed so that less time will be needed for manual quality control.

Lastly, future considerations for detailed audit functions should be implemented. This would include data updates and login attempts, both successful and unsuccessful, to be monitored and logged for security and quality control purposes.

SPECIFIC REQUIREMENTS

MorphBank Requirements

Web Site Functional Requirements

Uniform software language

The MorphBank web site shall be coded in a uniform software language in order to keep the web site accessible to as many users as possible. In addition, the web site should be written in a software language that is in use today by multiple browsers and can adapt to changes to the web. 3.2.1.1, 3.2.1.2

Universal language

The MorphBank web site shall be written in a universal language, since MorphBank will be used worldwide. It will need to be written in an universal language accepted by the majority of the world. 3.2.1.4

Creation data for web pages

MorphBank shall have all web pages documented with initiation information in order to maintain the life cycle information and work on archiving in the future. 3.2.1.3

Support multiple browsers

MorphBank shall be designed for cross-browser support because it will not be known what browsers a user may prefer and increase versatility. 3.2.1.5

Format of the web site

MorphBank shall have a consistent format to the web site. 3.2.1.3

Image file format

MorphBank shall have all images stored in a small format in order to keep data size down to a minimum and improve the capability of downloading images, and an universally accepted format. This feature should also assist with sending images via e-mail. 3.2.2.13.1

Measurement system

MorphBank shall be written in a universal measurement system. 3.2.1.4

Enabling/disabling features

MorphBank shall have certain features enabled and/or disabled depending upon user level. 3.2.1.8

Web Site Visual Requirements

Menus

MorphBank shall provide for the ability to link among all pages on the web site in order to achieve easy navigation through the web site. The menu shall be accessible on every page. 3.2.2.4

Display of the security access rights

MorphBank shall have security levels for the various types of users as described under Access Control. Considerations should be made for all possible user accessibility, and a web page for login of the user. 3.2.1.6, 3.2.1.9, 3.2.1.10

Registration

MorphBank shall be able to register various types of users. A registration page should be put in place to handle this feature, and should be drawn from the user schema. 3.2.1.7

Browsing capability

MorphBank shall have a means to browse through the data for those users who are not familiar with the technical biological data, or just wish to browse through the records maintained in MorphBank, there should be a user friendly browse feature. 3.2.2.8

Searching capabilities

MorphBank shall provide searching capabilities by specimen, taxonomy, and annotation. Searching should be made available through text input and where available through populated data in MorphBank. Search options should be efficient for users. 3.2.2.9, 3.2.2.10, 3.2.2.11, 3.2.2.12

Search results page

MorphBank shall provide a generic search results page that will have an overview of the specimens in list format with small thumbnail photos. Enough data to get a user's attention to proceed with the Specimen detail page.

Home page

MorphBank shall provide a home page that is appealing to the eye and simplistic in terms navigation. 3.2.2.1

Help

MorphBank shall provide a help feature that will provide the user with documentation on how to traverse throughout the web site, in addition to specific details of the fields and functionality of the web site. 3.2.2.3

Search page features

MorphBank shall provide search pages that are easy for users in selecting fields that they desire to query on. 3.2.2.9, 3.2.2.10, 3.2.2.11, 3.2.2.12

Taxonomy searching

MorphBank shall provide for the option of searching for specimens through taxonomy identifiers. In addition to providing more detail on the searches by adding several search field options. 3.2.2.9

Specimen searching

MorphBank shall provide for the option of searching for the specimens through identifiers most frequently used. In addition to providing more detail on the searches by adding several search field options. 3.2.2.11

Annotation searching

MorphBank shall provide options for searching through the annotations for a specific image on a specimen. 3.2.2.10

Advanced searching

MorphBank shall provide for an advanced searching option not provided by the taxonomy, specimen, and annotation. The advanced search option should allow for all possible data fields. 3.2.2.12

General information web page

MorphBank shall provide a general information web page to include the history and contact information. 3.2.2.1

Specimen data page

MorphBank shall provide for a specimen data page that contains all of the data fields for that particular specimen, in addition to graphical image. The specimen data page should define the specimen thoroughly. The web page should provide an enlarged image of the specimen. 3.2.2.13

Specimen annotation list

MorphBank shall provide for a current list of the annotations for a specimen. 3.2.2.13.5

Image of the specimen

MorphBank shall provide a still image of the specimen on the specimen data page. Format should be a universal format and small in size for future download capabilities. 3.2.2.13.1

Annotation display page

MorphBank shall provide for a link to the Annotation display page from the specimen data page. 3.2.2.13.6

News section

MorphBank shall provide for a news section to keep users of the MorphBank web site apprised of current news both from a functional standpoint and interesting biological discoveries. 3.2.2.3

Publications web page

The publications section should maintain all pertinent data on publications written about MorphBank or pertinent to the Hymenoptera order. 3.2.2.6

Access Control Requirements

Purpose of access control

Access control will be important in this project to ensure the integrity and availability of the website's content. The integrity of the MorphBank website data is important, as it will be used for scientific research. The data needs to be accurate and not tampered with. Except for brief intervals for maintenance, the website should be available for use 24 hours a day and every day of the year. Confidentiality will not be as important, as the website's content will be freely available to anyone with access to the World Wide Web. It will not be necessary to encrypt data (except logins) over the network, web server, or database. Access Control will help control actions of the website users, and prevent them from corrupting the data either purposely or accidentally. 3.2.3

Types of users

Different user access levels are needed, since permissions related to updates to the web site content and layout should be restricted. Everyone should be allowed to search and browse through the site and view images, image descriptions, and image annotations. A registered user should be a scientist actively researching in the field pertaining to the type of images they want to submit. Registered users can submit data, which is later quality controlled before it is made publicly available. Registered users should be able to make annotations to any image, not just images that they personally submitted. Another

type of user is needed to quality control data that has been submitted, but is not yet available for public viewing. Finally, website administrators need additional privileges in order to control the content and layout of the website. 3.2.3.1

Login

Users should use the same login method, whether they are registered users, super users or administrators. The user should input their username and password, then this should be verified by a program on the web server. 3.2.3.2

Passwords and usernames

Users should select their own username, while passwords will be initially assigned. Users should be able to change their password, and reclaim their password if they forget what it is. Passwords should be strong so that they are not easily guessed or broken by an automated dictionary attack. Usernames and passwords should be kept secret, so login information should not be passed over the network in the clear. 3.2.3.3

Revoking privileges

Users' permissions should be removed for misuse. User education should be detailed and precise so that users know what is allowed on the site. If misuse is suspected, then a panel should decide if a user's permissions should be revoked, after gathering all necessary information including an explanation from the accused. Misuse will likely be rare from registered users. 3.2.3.4

Registration

An individual or a few individuals are needed to determine if someone is qualified to become a registered user. The potential registered user needs to state what research

their data submissions are related to. A registered user needs to be active in the research field related to their data submissions. Users should be able to register within five business days. 3.2.3.5

Quality control

Images and image descriptions need to conform to a certain format, decided on by an order's controlling panel (for example, Hymenoptera). A group of people is needed to manually quality control data that has been submitted from the registered users. Quality control should be partially done automatically, so that manual quality control can be done quickly. A taxonomy name server should be used to verify that taxonomic names are valid in a submission. There should be separate quality control groups for every order, such as hymenoptera. The size of the quality control group depends on the workload, which will depend on the complexity of the required format, rate of data submission, and the amount of workload reduced by automated quality control programs. The leaders of an order, such as Hymenoptera, should certify quality control checkers. Certified checkers should be able to quality control an image submission on their own, and with a minimal amount of time. It is not necessary for a panel to meet (in person or virtually) in order to quality control image submissions. Manual quality control can be done individually. Proper user education and checking of registered users' qualifications will go a long way towards reducing the amount of quality control that is needed. 3.2.3.6

Monitoring

Actions of users should be monitored and logged, including updates, submissions and successful and unsuccessful login attempts. This will help maintain the security and

quality control of the site. Users should be notified that their actions will be monitored. 3.2.3.7

Access control implementation

Access control can be implemented four ways (at least): Java application running on the web server, third-party software, Apache web server modules, or PHP scripts. Implementation should be inexpensive, less than $50 per month. It also should be flexible and easy to implement – the administrators should have the expertise to implement the software and it should be compatible with existing software. 3.2.3.8

Security Requirements

Security

Physical security will be important to prevent unauthorized access to hardware. Intrusion detection is needed to detect changes to important files and unauthorized access to the local network. An updated virus checker is needed to detect and delete viruses, worms, and Trojans. Vulnerability assessments will needed occasionally to prevent future exploits. The operating system should be patched as often as necessary. 3.2.4, 3.2.4.1, 3.2.4.2, 3.2.4.3

Server and Database Requirements

Server and database configuration

The main database server is connected to a web server and is only allowed to communicate to that web server. The web server's main function is to store data submissions. Once the submissions have been approved, they are sent to the database

server. The database server updates itself and then sends the changes to the web server, which successively sends those changes to the mirror sites.3.2.5

Server Backup Requirements

Backups to database

A database that is available 24/7 needs to be able to withstand many different "threats." This explains why the database should be backed up. Human errors, disasters at the location, intentional actions by other users, or equipment problems can cause loss of all data. In the event of any possible danger to the database, a fresh "hot" copy needs to be available right away. There should be a log that can enable point-in-time recovery, in the event that a restore is needed due to error. There are many different software solutions to this problem. 3.2.6

Mirror Site Requirements

Purpose of mirror systems

Mirror Systems are an important part to web based systems today. Mirrors provide a system the ability to serve more customers at a given time in fast and efficient manner, which is not possible with a stand-alone system. Mirror systems prevent bottlenecks in a system when system load is very high and provide services to the users when the main system is at its maximum capacity or offline for repairs or maintenance. The mirrors also provide an offsite backup copy of a system's data in times of catastrophic failure at the main site so the data is not lost. 3.2.7.1, 3.2.7.2, 3.2.7.3, 3.2.7.4, 3.2.7.5, 3.2.7.6

Mirror site updates

- All data broadcast by mirror sites should be identical.
- Mirror sites should be synchronized on at least a weekly basis.
- A standard time for synchronization of mirror sites should be decided.
- Mirror sites should not be synchronized simultaneously. Mirror sites should be synchronized sequentially to avoid all mirrors being offline at the same time.
- The process of synchronization should be automated and need no interaction by a system administrator except for the initial setup.
- Automation of mirror site synchronization should be controlled by a scheduling utility and a shell script that will ensure identical file structures on all mirror sites.
- A utility should be used that only updates files that have been changed or deleted from the file system on the main site as to reduce transmission time.3.2.7.7

MorphBank Specifications

Web Site Functional Specifications

HTML

The web site will be coded in HTML 4.0 or higher, in order to use standard basic HTML and XHTML tags currently in use today.

XML

Version 1.0 to be used in conjunction with the GAT.

Style sheet

The web site shall adhere to CSS 2.1 regulations and validations in accordance with W3 standards. The standard format is: a) center page header, b) navigation

drop-down box, c) memo text, d) 2 vertical columns, e) copyright information at bottom of page, f) times new roman, color black, g) gray overtones. The style.css is used for the basic formatting tools implemented via XML and style2.css is used for the navigation and drop down boxes.

General web page specifics

The web site shall be written in English, currently the most known language, and the primary language spoken in the U.S., origination of the web site. Also, the web site shall use the SI system for all measurements. Lastly, ever web page shall have the following general information (to be used for life cycle purposes in the future): 1) creation date, 2) last update date and time, 3) creator/administrator, 4) archival date, if applicable, and 5) special notes.

Browsers

Recommended Microsoft Internet Explorer 5.5 or higher. Currently reviewing Mozilla and Netscape for future use.

Login

Login page will contain the first and last name of the user, in addition to e-mail address and password and will be in compliance with Access Control specifications. 3.2.3.2 The e-mail address and password will be submitted on an encrypted line for avoidance of packet retrieval by third party.

Registration

There will be a registration page which will contain general registration information: Last name, first name, affiliation, address 1, address 2 (multi-line address),

city, state, zip, country, country zip, e-mail, and reminder notice (for forgotten passwords). This feature will be in compliance with Access Control specifications.

3.2.3.5

Enabling/disabling – login features

This feature will toggle the availability of the "Add Annotation" button featured on the Specimen Data page. Administrative access (access level 4) will be achieved through a direct link and will NOT be linked to from the site itself.

Password creation

Initially the system will create a password for the user.

Password change

On the login page, there is an option for the user to change his or her password.

Web Site Visual Specifications

Home Page

The MorphBank home page shall feature basic information on the current functionality of the MorphBank System as well as the three (3) most current news articles. A navigation menu as well as access to login and registration pages will be easily accessible.

About MorphBank/contact information

The web page entitled About MorphBank/Contact will contain general text information under two sub-titles: 1) About MorphBank and 2) Contact Information. Under the header of About MorphBank will be general text information on the history of MorphBank, how to contribute to MorphBank. Under the Contact information header

there will be a list of contacts and each contact will include full name, affiliation, and e-mail addresses. Web site to be maintained by the web site administrator.

Help

A quickly accessible link that brings up a new window featuring functionality information for the MorphBank system and will be accessible from every page. The help pages will also provide details specifications for each data field.

Menus

A navigation menu shall be visible and easily accessible on every page throughout the MorphBank website. The menu shall be created with JavaScript Version 1.1 or higher, implemented using Cascading Style Sheets and XHTML Version 1.0 or higher (for implementation). The menu should be multi-tiered, featuring a main menu and sub-menus that are accessed when the mouse pointer is placed over an applicable main menu item. The following is the recommended format for the menu, and where each item leads when clicked on:

Search

Taxonomy — Linking to the Taxonomy Search Page

Annotation — Linking to the Annotation Search Page

Specimen — Linking to the Specimen/Image Search Page

Advanced — Linking to the Advanced Search Page

Browse

Linking directly to the Browse Page

Archive

News — Linking to the News Archive Page

Publications — Linking to the Publications Archive Page

About

About Us — Linking to the About MorphBank and Brief History Page

Credits — Linking to the List of Contributors Page

Home

Linking back to the home page.

News

The news about MorphBank will be maintained on the home page. Only the three (3) most current news articles shall be maintained by the site administrator. The news articles shall be short pieces of information that inform the user of current updates and activity on the site. These updates should feature a title and the date posted, both in bold, a descriptive yet short log of information, and the name of the individual, in italics, of the individual responsible for submission.

Publications

Publications web page will be simple and have the following fields available for view: a) primary author, b) all secondary authors, c) date published, d) publication type (journal, book, etc), e) name of publication, and f) full text view of the publication.

Search Results

The search results should display in order of highest percent match to the least, if two items match with the same percent they will be ordered by their specimen ID. Every

page of the search results should display the number of matches according to the selection in the search page. On the top of the page there should be a display telling the user which results are being viewed. There should also be links to the next and previous results pages. On the bottom of the page there should be the same bar as the top with one addition, there should be a link to go to the top of the page. Each data item returned should be in the following format:

- Specimen ID in brackets.
- The Order/Family/Genus/Species, with the species in italics.
- Body Part.
- Developmental Stage.
- Sex.
- To the left of the image there should be some text telling the user how many annotations were made on that image, example: "7 Annotations".
- A small version of the image should be displayed on the right of all the data displayed.

Browsing options

The Browsing page will be divided into two parts. The top half will be the tree structure and the bottom half will be a replica of the search results page. There will also be a button that allows the user to turn on/off the search results so the user can browse the tree he/she reaches a point where they would like to see what is in the database for that section.

Search Tree

The root of the tree will be the Order Hymenoptera with all the Families within it coming out has branches. The initial page will come up with just this section of the tree. The branches will be scrolling if need be. The tree will only be populated with the K P C O F G S that are currently in the database. Has you click on the K P C O F G S in the tree it becomes the root with the next level as the branches coming out of it. The lowest level of the tree will be the individual specimen.

Search Results

This will be on the bottom half of the browse page. It will look identical to the main search results page. It will display a search based on the last branch clicked on. So if you click on a Family under Hymenoptera it will show all the data in the database that is in that Family. The search results would be treated the same as on the main search results page.

Taxonomy search page

The taxonomy search page will provide several pertinent fields for a user to search on; these field are provided below with a short description.

Keyword — text box to search on a keyword that is related to the specimen, such as preparation type, location, etc. Field can be used to search on any field within the data on a specimen.

Taxon name — text box for searching on taxon name.

Sex — check box for M=male, F=female, H=Hermaphrodite, I=Indeterminate (examined but could not be determined), U=unknown (not examined), and T=transitional (between sexes).

Stage — drop-down box for searching on development/growth stages; will be populated with current data in MorphBank.

Date identified — two drop-down boxes for a range of dates to be input, one a start date and the second an end date; date format is DD/MM/YYYY.

Identifiers — text box for searching a specific identifier.

Annotation/image search page

The annotation/image search page will provide the user with the ability to search through the many annotations on a specific specimen. The fields available for searching on are provided below:

Collector ID — drop-down box to search on the user ID to obtain the person who collected the specimen; will be populated with data from the existing MorphBank database.

Annotation type — drop-down box to search on the types of annotation, such as graph, data, image (i.e., JPG, TIF, GIF), etc. This data will be populated from the existing MorphBank database.

Date submission — two text field boxes for searching a range of dates of submitted annotations. The first will be the start date and the second the end date; format is MM/DD/YYYY.

Body part — drop-down box for searching the body parts of the specimen and populated from existing data in MorphBank.

View angle — drop-down box for searching the view angle of the specimen and populated from the existing data in MorphBank.

Technique — text box for searching the technique of the annotation.

Accession number — drop-down box for searching the number of entries in the MorphBank database.

Annotation content — text field for searching a word or a string for the comment section of the annotations.

Specimen search page

The specimen search page will contain combo boxes and check boxes to enhance the ease of searching. The following are the fields to be displayed on the specimen search page and provide for searching options:

Collection number — first field on the page, since most frequently used search option. Label "Collection Number" is located on the left of a text only input field for searching the collection number of a specific specimen; unique alphanumeric.

Collector name — label "Collector Name" is located on the left of a text box for searching the person who collected specimen or taking the observation.

Sex — label "Sex" is located on the right of the check boxes for check box for M=male, F=female, H=Hermaphrodite, I=Indeterminate (examined but could not be determined), U=unknown (not examined), and T=transitional (between sexes).

Stage — label “Stage” is located on left of a combo box, populated from the database with all possible stages available.

Locality — label “Locality” is located on left of a text box and will provide search capabilities on where the specimen is collected.

Country — label “Country” is located on left of a combo box with a list of possible countries where specimen is collected; format using ISO3166-1 two letter codes.

Continent/ocean — label “Continent/Ocean” is located on left of a text box for searching on continent or ocean where a specimen is collected.

Collection dates — label “Collection Date” is located on the left of two date fields one start date and one end date to view specimens collected within a certain time frame; dates are formatted MM/DD/YYYY.

Advanced search page

The advanced search page is provided for users who will require more advanced searching, not provided on the taxonomy, specimen or annotation search pages. In fact these search pages have been combined to provide the ability to search on many features stored on a specimen.

Keyword — text box field for input of the Kingdom, Phylum, Class, Order Family, Genus, or Species name.

Sex — check box for M=male, F=female, H=Hermaphrodite, I=Indeterminate (examined but could not be determined), U=unknown (not examined), and T=transitional (between sexes).

Taxon name — text box for searching on taxon name.

Stage — drop-down box for searching on development stages; will be populated with current data in MorphBank.

Identified by — text box for search on the person who identified the specimen,

Date identified — two drop-down boxes for a range of dates to be input, one a start date and the second an end date; date format is DD/MM/YYYY.

Identifiers — text box for searching a specific identifier.

User — text box for searching for annotations by researcher who has created the annotation.

Annotation type — drop-down box to search on the types of annotation, such as graph, data, image (i.e., JPG, TIF, GIF), etc. This data will be populated from the existing MorphBank database.

Date identified — two drop-down boxes for a range of dates to be input, one a start date and the second an end date; date format is DD/MM/YYYY.

Body part — drop-down box for searching by body parts; this field is populated by the current entries in the MorphBank.

View angle — drop-down box for searching by view angle; this field is populated by the current entries in the MorphBank.

Technique — text box for searching the technique.

Accession number — text box for searching the accession number.

Annotation content — text field (4 x 40) for searching a string within the notes portion of the annotation.

Collector — text box for searching the name of the collector.

Preparation type — text box for searching a specific type of preparation (i.e., skin, slide, etc.)

Development stage — drop-down box for searching developmental stages populated with current MorphBank data.

Scientific name — text box for searching the scientific name of a specimen.

Sub-species — text box for searching the sub-species epithet name of the specimen.

Notes — text field (4 x 40) for searching a string for the notes section of a specimen.

Institution code — drop-down box for searching the institution codes currently populated by the MorphBank database.

Collection code — text box for searching the collection code defined as the collection within the institution; unique alphanumeric.

Field number — drop-down box for searching the field numbers currently in the MorphBank database.

Date identified — two drop-down boxes for a range of dates to be input, one a start date and the second an end date; date format is DD/MM/YYYY.

Time collected — two drop-down boxes for a range of times to be input, one a start time and the second an end time; military time and format is HH:MM:SS.

Continent/ocean — drop-down box for searching the continent/ocean where specimen collected.

Country — drop-down box for searching the country where specimen collected; format using ISO3166-1 two letter codes.

State/province — drop-down box for searching the state or province where the specimen collected.

County — drop-down box for searching the county where the specimen collected.

Locality — text box for searching the locality where specimen collected.

Latitude — text box for searching the latitude where the specimen collected; format is decimals.

Longitude — text box for searching the longitude where the specimen was collected; format is decimals.

Minimum elevation — text box for searching the minimum elevation where the specimen was collected; SI system format and elevation above (positive) or below (negative) sea level.

Maximum elevation — text box for searching the maximum elevation where the specimen was collected; SI system format and elevation above (positive) or below (negative) sea level.

Minimum depth — text box for searching the minimum depth where the specimen was collected above (negative) or below (positive) sea level; SI system in meters.

Maximum depth — text box for searching the maximum depth where the specimen was collected above (negative) or below (positive) sea level; SI system in meters.

Specimen data page

The specimen data page will contain all of the data fields on a specimen and maintained in the MorphBank database, excluding database functional fields. The web page has five sections: 1) image, 2) taxonomy, 3) specimen information, 4) collection data, and 5) annotation list.

Image

The image will be in JPEG format; the image is a magnified view of the specimen.

Taxonomy

All data fields under the taxonomy schema will be shown in this section. The Kingdom, Phylum, Class, Order, Form, Genus, and Species.

Specimen information

All data fields to be displayed on the physiological characteristics of the specimen.

Collection data

All data fields to be displayed on the geographical aspects on where a specimen was found.

Annotation list

All data fields to be displayed showing a list of the current annotations for the specimen.

Annotation link

A button to be located below the annotation list that will open the create annotation page if the user has update rights, otherwise an error message will be displayed.

Access Control Specifications

Types of users

Four types of users of the website are needed:

1. Unregistered user – can search and browse through the website for images, image descriptions, and annotations. An unregistered user can apply to become a registered user. An unregistered user does not login to the site. Other types of users have the same permissions as an unregistered user, until they successfully login to the site.

2. Registered user – has all of the permissions of an unregistered user, plus a registered user is allowed to submit images along with the associated metadata. A

registered user is allowed to submit annotations to any image. Annotations are not manually quality controlled.

3. Super user – has all of the permissions of a registered user, plus is allowed to view submitted images and their associated metadata in a protected directory. The super user is allowed to execute commands that will send the submitted data over to a different directory for public viewing, after the submitted data has been manually quality controlled. A super user is also allowed to quality control submitted data. A super user is certified as a quality control checker by an order's (e.g. Hymenoptera) controlling panel.

4. Administrator – has all of the permissions of a super user, plus is allowed to access additional directories containing commands and configuration data used to control the content and layout of the website.

Login

Users will click on a link or hypertext, which will enable them to get a login dialog box. The login box will prompt for a username and password, and there will be a button to click on to submit the username/password combination. The username and password will be limited to ten characters or less each, so more characters are ignored. A program will check the combination against a user table to check if it is valid. The user table will also contain the group (registered user, super user, or administrator) that the user belongs to, in order to determine permissions. If the combination is valid, the user is allowed to proceed. A user gets three login chances before their account is locked out for a total of five minutes. The login dialog box is then closed. This is implemented by ignoring all login attempts from that particular IP address for five minutes. After the five

minutes, the login process can be started over. If a particular login attempt is unsuccessful, the output to the screen should be similar to "Username or Password incorrect. Try again." If a third login attempt is unsuccessful, the output to the screen should be similar to "Account locked out for five minutes. Try back again then."

Passwords and usernames

Users select their own usernames when they apply to become a registered user. Usernames will be in lower-case letters and numbers. Passwords will be randomly generated for new users. These will be e-mailed to the new users. After that, the user can change their password. Passwords should be strong – containing at least eight characters, and a combination of upper-case letters, lower-case letters, numbers and special characters. When a password is changed by a user, a program should automatically check to make sure that it passes as a strong password – cannot be a previous password, must be unique among all of the users' passwords, have at least eight characters, and contain upper and lower-case letters, numbers and special characters. When changing their password, a user should re-enter their old password, then enter their new password twice. If a password is forgotten, the user can request that their password be e-mailed to them. Do not use secret questions for the user to reclaim a forgotten password, as the answers can be easily guessed. Since username and password combinations should be kept private, HTTP message digest authentication should be used to obtain the username and password of a user. Basic authentication only uses very weak encryption – base64 encoding. Usernames and passwords should be stored in an encrypted state on the database or server. This encrypted file data should also be stored in a protected file for

additional security. Only the administrator should have read permission on the files containing passwords.

Revoking privileges

Registered users should have their permissions revoked for misuse, if they frequently and blatantly disregard recommended data formats. Slight deviations from standard formats should be allowed, however. It is possible that the user is unaware of all proper procedures – in this case more user education may be needed. Also, careless mistakes are possible. Misuse should be proven intentional for a user to have permissions taken away. A panel should be formed to revoke the registration of a user if necessary. Revocation of another's privileges should not be left up to a single individual. In an emergency, there may not be time to form a review panel, in the case that a user is purposely attacking the integrity of the database. This situation will likely be rare. It is possible that the attacking user is actually someone else who has compromised a registered user's account.

Registration

For a person to become a registered user, they must be active in the research field related to their specimen submissions. They should be published in the research area. Exceptions can be made for graduate students that have not been published yet. Users should only submit data for the family or families that are related to their research. Exceptions can be made for host species in a parasitic relationship. Users should be able to receive their registration information (username/password combination, information on how to submit data to the website) within five business days. The following information

is required in the registration process: desired username, first and last name, institutions they are affiliated with, address, e-mail, phone number, type of images they wish to submit, current research fields, name of one or more recent publications, and education.

Quality control

Users need to be educated on exactly what is considered acceptable behavior and procedures for uploading and modifying images, metadata, and annotations. Proper screening of potential users and user education will cut down on the amount of quality control needed. On the average, a user should have to resubmit an image and its metadata less than once, after it has passed automated quality control checking. Forms for the submission of images and annotations should be designed carefully to guide the registered users on the proper format of data. Most of the fields (greater than 50%) in the submission forms should contain text to select (such as drop down boxes or toggle buttons), instead of allowing the submitter to type in a free-form answer. The forms should state what fields are mandatory. Error checking should be automated before the data is submitted, preferably either with a Java applet or JavaScript. Annotation submissions should only allow specific body parts to be annotated, for example head, antenna, etc. Submission forms for annotations should also include a list of characteristics for a family or genus. Experienced biologists should be selected to manually review images and their metadata, to make sure that images and metadata are in the proper format, such as making sure that a particular view is allowed, and no required metadata is missing. Manual quality control is not needed for annotations. The quality control process for an image only needs to be done by one person – a consensus is not

needed. A taxonomy name server should be used when an image and its metadata is submitted to ensure that the proper names are used. Adjustments will be needed for new species or genuses that are not listed in the taxonomy name server. A submitter should e-mail the website's administrator when an image's metadata is rejected for this reason, so that the taxonomy name service can be updated, assuming that the original submission was accurate. Manual quality control should be minimal and objective – it should take less than 5 minutes per specimen image and associated metadata, if it is a known species or genus. The website should state that users are encouraged to e-mail the administrator with any errors found on the website. Submitters should be given an objective rating based on how many images they have got accepted by the website, and how many errors that have committed. This will both motivate users to submit high quality data and also help the super users evaluate individual submissions.

Monitoring

The actions of all users of the website should be monitored for security purposes and to determine improper behavior. Updates to images and annotations should be logged, with the userID and time of update included in the log. On the website, users should be notified that their actions are monitored. Logs should be written to a protected directory, to help prevent attackers from overwriting the log. Logs should be archived to a separate hard drive, tape drive or some other storage device once a week, or whenever backups are done. Registered users should be able to see their past history of changes, including image and image description submissions and annotations.

Access control implementation

The only third-party software found that fit the requirements was OSS v3.0 by Omni-Secure (http://www.omni-secure.com). It is $497 for 5 domain licenses for 5 years; fewer licenses for less time are cheaper. It appears to have a large amount of functionality. Java applications do not appear to be used often for web server authentication and authorization – mainly for Java applets and stand-alone applications it appears. The recommendation is to use Apache web server modules for basic functionality, since they seem to be easy to implement for authentication and authorization. For additional functionality and flexibility, PHP scripts should be used along with the Apache modules. See the reference list for the URLs of tutorials for authentication and authorization in Apache and PHP. Apache modules can implement digest authentication.

Security Specifications

Physical security

The server is to be stored in a room inside of the school where it can be administered and be physically accessible. The room will be locked and the administrator will have access at all times. Case locks will be used to ensure that no one can open up the server and steal components from it. The mirrors need the same physical security.

Software implementation

Use of Tripwire — monitors your system for changes to important files. The software creates an initial database blueprint from the existing important files, and then runs periodically to check for inconsistencies. Alerts are generated when a file has been

modified. The blueprint will need to be updated when the MorphBank database is updated. Red Hat 7.1 includes Tripwire as an optional package during install.

Use of Chkrootkit — shell script that scans system for known exploits, trojan commands, and worms. Chkrootkit can be found at http://www.chkrootkit.org.

Use of NESSUS — will search and locate vulnerabilities on your system by actively trying to perform known exploits against it. If a vulnerability is found, the program will make recommendations about upgrades, configuration changes, and where to find patches. The software gives explanation as to why it finds services vulnerable. NESSUS is available at http://Nessus.org.

Use of Astaro Security Linux — includes firewall, intrusion protection, URL filtering, and a VPN gateway. Features include packet inspection, deep packet filtering, application-level intrusion detection, content filtering, virus detection for email traffic and web traffic. www.astaro.com.

Calibrate Firewall — calibrate firewall on database server (write-only) to only accept transactions from the mirror sites. This will be based on IP address.

Unix configurations

Make sure all latest OS and patches are installed. This should be taken care of biweekly. (http://www.redhat.com/security/)

Regularly review user list and eliminate unnecessary accounts.

Disable ability to log on remotely as root. (enable the "CONSOLE" line in /etc/default/login and to disable use of ftp by root by adding "root" to /etc/ftpusers)

Display legal notice upon login. It is generally believed that login warning banners will help when prosecuting unauthorized users. The following commands can be used:

- echo "Message here." >> /etc/motd
- echo "Message here." >> /etc/issue
- echo "Message here." >> /etc/default/telnetd
- echo "Message here." >> /etc/default/ftpd

Mount the file system as read-only. Mounting file systems containing system binaries read=only makes malicious or accidental changes to the system more difficult.

Server and Database Specifications

Refer to the Server Configuration White Paper Appendix A 4.1.

Backup Specifications

Backups failures to database

The following is a list of possible failures that can occur:

- **Process failure**: This type of failure occurs when handling a program statement.
- **Instance failure:** This type of failure can be caused by hardware problems such as power outages and software problems such as operating system crashes. In this type of failure, the data in the buffers of the system global area is not written to the data files.
- **User or application error:** When a user deletes a file that is still required, it can create the need for the system to be backed up to an exact certain point in time (just before the delete occurred).

- **Media (disk) failure:** This error can occur in reading or writing to a file that is required to operate the database. This error can potentially cause the loss of all files on a disk drive. (Data files, log files, and control files are affected in this case).

Backup types

The following is a list of different types of backups:

- Whole database backups - most commonly used by systems administrators.
    - Consistent whole database backup
    - Inconsistent whole database backup
- Table space backups
- Datafile backups
- Control file backups
- Archive log backups
    - Consistent whole database backup: datafiles and control files are consistent to the same point in time.
    - Inconsistent whole database backup: portions of the disk are being written to while the backup is occurring. This method is beneficial to systems that run 24/7.

Proposed database backup software

- Backup/Restore/Archive- By *Datalink*
- BrightStor Enterprise Backup- By *Computer Associates*
- Global Data Manager- By *VERITAS*

- HyperTape Network Management Suite- By *BridgeHead*
- NetWorker- By *Legato Systems, Inc.*
- SnapManager- By *Network Appliance, Inc.*
- Time Navagator- By *Atempo*

The software should enable the system to have a journal or log of each backup, equipped with the time and date of each event. This will enable the exact point-in-time recovery that would be potentially essential to keeping the system running. For example, if a user "accidentally" deletes a file or record that is supposed to be retained in the system, a recovery of the system's status at the previous backup should be possible. For these reasons, it is proposed that new additions to the database be backed up quite frequently, i.e. every 1 hour, 2 hours, etc., on a faster data drive (Internal/External hard drive, DVD backup, etc.). Since storing the new additions that occur each day would require quite less memory than that of the entire database, a hard drive or DVD drive would be quite efficient. Depending on how large the MorphBank system becomes and how frequent updates of records occur, the entire system should be backed up to a tape drive once or twice a week.

Proposed backup hardware solutions

Tape Drive Backup System: Tape drives are very large drives that enable the user to backup very large quantities of data and are still cost efficient. They are very easily maintained and used, however the only disadvantage is that memory access speeds are slower. The following is a list of some common tape backup systems to consider:

- The 8mm Tape Drive is faster than its DAT counterpart with transfer rates up to 6MB/sec. Very much like the tape used in a video camera.
- The Digital Audio Tape (or DAT**)** is a slower solution used for backing up Network/ Computer information.
- The Digital Linear Tape (or DLT) is very cost efficient, fast, and reliable. Also known for their efficiency in backing up networks, these drives stream data across a single stationary head.
- Last but not least, the first tape drive standard invented in 1983 is the QIC (Quarter Ink Cartridge). It is a magnetic data tape storage.

Exabite Systems makes a popular tape backup system. This company has designed a system that supports up to **80GB** of storage per tape with a **6MB/Sec** transfer rate. The high end drive made by this company is **$2,933.00**, ranging down to about **$900.00**. The accessories, i.e. tape cartridges, cleaning cartridges, etc. range from roughly **$17.00** to **$88.00**. Another part of the system is the Rackmount, which is priced at roughly **$2,900.00**.

LaCie is another company that has a design for the backup system. For a system that has a **35GB** storage capacity with a **4MB/Sec** transfer rate, not to mention almost all the accessories included, the price is **$799.00**. This is much cheaper than the **Exabite** systems, although with a sacrifice in capacity and transfer rate. Their larger system has a **50GB** storage capacity with a **6MB/Sec** transfer rate and all accessories included runs for

$1,199.00. The cartridge access time on this system is **35 s,** where in the cheaper system it is **28 s**. Thus, more storage, faster transfer rate, more time to access the cartridge.

Proposed storage of backup hardware

The backup hardware, i.e. the tape drive system should be stored in a separate facility other than the actual database. This is because in the event of the database itself being destroyed, it would be a tragedy to have the backup destroyed as well. The room where it is stored should provide shelter from any sort of fire, i.e. fire-proof vault, and any sort of water damage, i.e. emergency sprinklers. Storage should not be an unbearable issue, since the drives are no bigger than the size of a filing cabinet (...Max). In the event that the hardware was to be stored in the same room, or a normal room other than the fire-proof vault solution, it should at the very least be concealed from ceiling emergency sprinklers, soft drinks, etc.

Mirror Site Specifications

Mirror system hardware

The hardware for the mirror systems will be off the shelf computer hardware. The mirror will consist of one box running all the applications. The system should have a storage space of at least 100 gigabytes. The mirror box should be compatible with a Linux / UNIX based operating system. Hardware performance will vary depending on the expected load to the mirror site but computational speed should be similar the main site to keep response time consistent. Backup and RAID systems for the mirror systems will not be needed since data backup units are on the main site. Mirror sites will be read only sites serving users who wish to retrieve data from the data base. Adding Annotations,

making modifications, or adding new data items will only be handled by the main MorphBank system.

Mirror system software

The MorphBank system will consist of a Linux based operation system (Fedora Linux). An Apache Web server will supply web services and the database engine will be a My SQL database. Plug-ins like tomcat will enable PHP to be used to build a web interface to the My SQL database.

New mirror registration

The initial MorphBank site must have at least one mirror when the system is deployed to handle users' request for data since the main system will be read only. As traffic to the system increases or other organizations decide to volunteer in hosting a MorphBank mirror, the new mirror system will be integrated to current operational system. When the current MorphBank staff approves the volunteering organization, the new mirror's administrator will be given access to the MorphBank's data and will be registered on the main site with rights to copy all the system's data. The new mirror will receive a copy of the database on a large storage device by mail when ever possible to prevent heavy bandwidth losses in transmitting large amount of data through the network. With the initial image of the system the new mirror can simply synchronize with the main server and be up to date with less movement of data through the Internet. The last thing needed to make the new mirror system available is to register the mirror's IP address with the DNS (Domain Name Server) of MorphBank so routing can take place.

The goal being that the main site will route the user to the mirror closest to their geographical location for better performance.

Mirror maintenance

MorphBank's mirror system will include data verification process, software updates, and hardware inspection (cleaning, check connections). The mirror system specifications must comply with current standards. Mirrors sites will not be required to be backup or RAIDED since data is securely kept at the main site. Mirror updates such as new data items will be transmitted via the web during times when traffic to that specific mirror is lowest. The updates to a mirror will be increment-only (changes to the main database will only be sent).

Mirror user interaction

MorphBank will consist of several computer systems interacting with each other with purpose of serving scientist around the world. User interaction with the MorphBank system should be transparent to the user. When submissions or data modifications are submitted to the mirror, the mirror will route the submissions to the main site for processing.

Mirror recover procedures

The main system will ping (checking that they are up) the mirrors every couple of hours. If a mirror does not respond to the ping from the main site, the routing of customers to that mirror will not occur. If the mirror has a storage device failure a new copy of the database will be sent to the site to restore the data of the mirror. Once a mirror is back online and updated, routing to the mirror will resume.

Updates to mirror sites

- Updates shall be made to the mirror sites from the main server on a weekly basis.
- SQL Journaling shall be enabled on the main database server at the FSU site.
- Every Friday a copy of the journal will be sent to each of the mirror sites and played to bring the mirror databases up to date.
- Every Friday morning a shell script shall be used to update the web content of the mirror sites to be up-to-date with the main server.
- Each mirror site shall be updated sequentially and not simultaneously.
- The update script shall utilize the standard unix/linux rsync utility.
- The updating shall be automated by using the standard unix/linux scheduling utility cron.

What is rsync?

Rsync is an open source utility that provides fast incremental file transfer. Rsync is freely available under the GNU General Public License. Rsync's specialty is efficiently synchronizing file trees across a network. It is available at http://rsync.samba.org/ .

How to use rsync

Here is an example of how to update a file tree to a remote machine using a secure shell connection.

- rsync -a -e ssh source/
- username@remotemachine.com:/path/to/destination/

If a file was originally in both source/ and destination/ and you delete it from source/, you probably want it to be deleted from destination/ on the next rsync. However,

the default behavior is to leave the copy at destination/ in place. Assuming you want rsync to delete any file from destination/ that is not in source/, you'll need to use the --delete flag.

What is cron?

Cron is the name of program that enables unix users to execute commands or scripts (groups of commands) automatically at a specified time/date. The cron utility comes standard with the Fedora Core 2 distribution that will be installed on the main MorphBank server. It can be used for anything. For example, you could add a cron job that automatically copies a MySQL database to a separate location on your site as a backup.

How to use cron.

Cron stores it's entries in the crontab (cron table) file. This is generally located in your /etc directory.

If you want a program or shell to run every day at 3:29 AM you would add the following line to your /etc/crontab file:

29 3 * * * root updatescript.sh >> /dev/null 2>&1

This specific example would run the updatescript.sh script as a root user at 3:29 AM every day without any output to the terminal.

Below is a table of what each field of a line in the crontab file does.

Field	Meaning
1	Minute (0-59)
2	Hour (2-24)
3	Day of month (1-31)

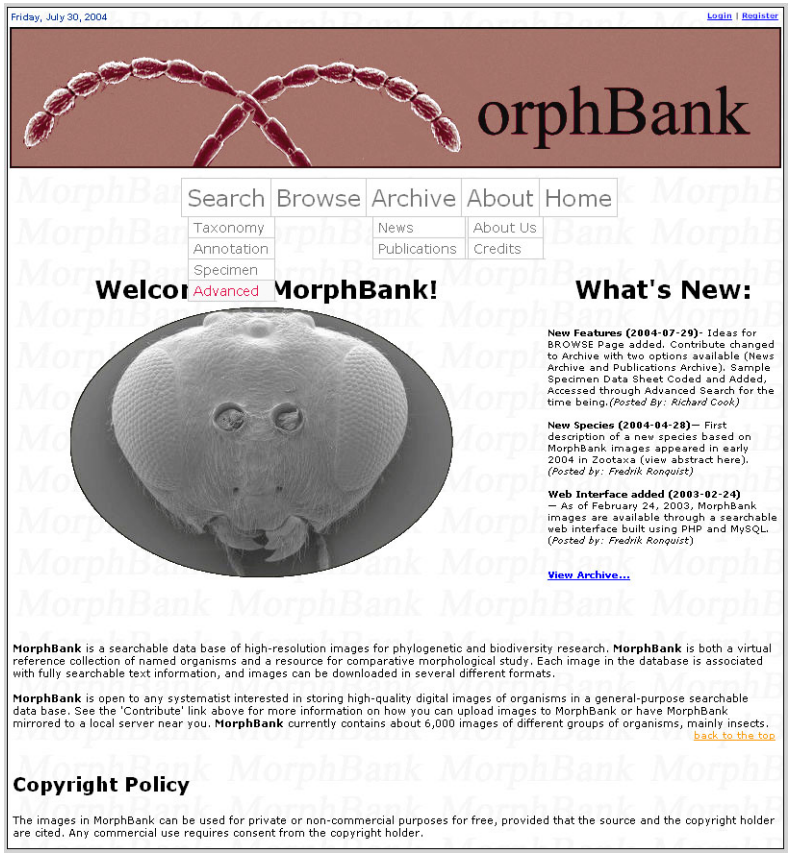
4	Month (1-12, Jan, Feb, etc)
5	Day of week (0-6) 0 = Sunday, 1 = Monday etc or Sun, Mon, etc)
6	User that the command will run as
7	Command to execute

Each user on your system can have their own crontab which would be stored in /var/spool/cron/. To edit a user's crontab entry, simply log into your system for that particular user and type crontab -e. After you've added all your entries you can use the command crontab -l to list them. If you wanted to remove your crontab file you could run crontab -r to delete it. To edit a user's crontab file as root you can run crontab -e -u username

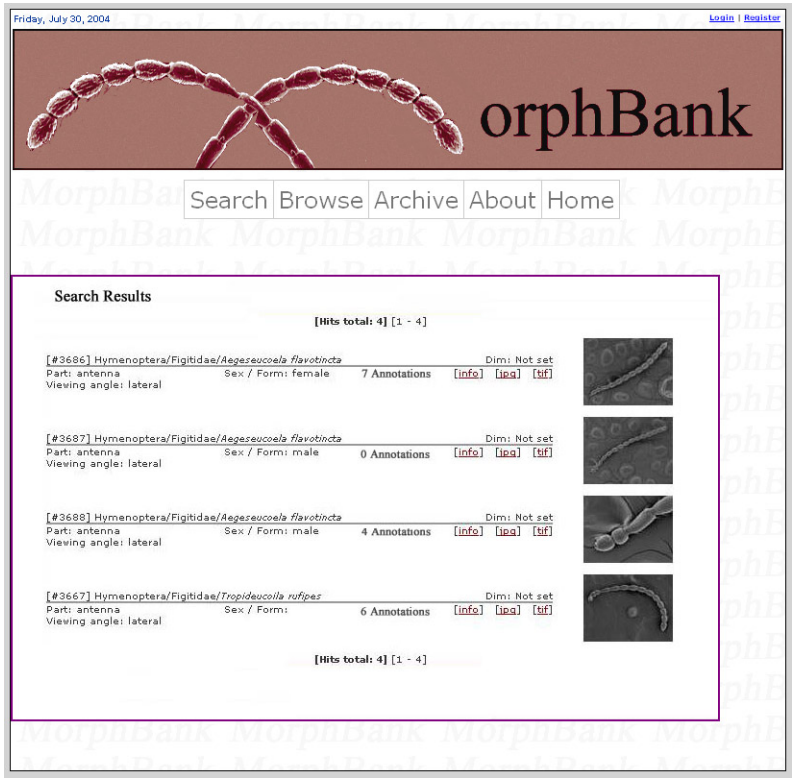
Cron is very configurable and is a great tool for every system administrator to automate tasks.

Web Site Visual Design Recommendations

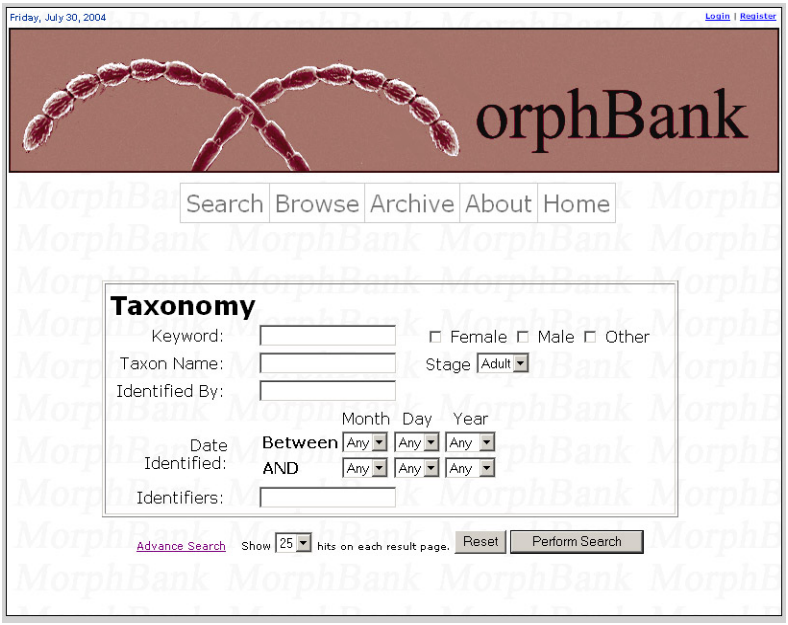
Home Page


Friday, July 30, 2004
Login | Register
orphBank
Search
Browse
Archive
About
Home
Taxonomy
Annotation
Specimen
Advanced
News
Publications
About Us
Credits
MorphBank!
What's New:
New Features (2004-07-29)- Ideas for BROWSE Page added. Contribute changed to Archive with two options available (News Archive and Publications Archive). Sample Specimen Data Sheet Coded and Added, Accessed through Advanced Search for the time being.(Posted By: Richard Cook)
New Species (2004-04-28)— First description of a new species based on MorphBank images appeared in early 2004 in Zootaxa (view abstract here). (Posted by: Fredrik Ronquist)
Web Interface added (2003-02-24) — As of February 24, 2003, MorphBank images are available through a searchable web interface built using PHP and MySQL. (Posted by: Fredrik Ronquist)
View Archive...
MorphBank is a searchable data base of high-resolution images for phylogenetic and biodiversity research. MorphBank is both a virtual reference collection of named organisms and a resource for comparative morphological study. Each image in the database is associated with fully searchable text information, and images can be downloaded in several different formats.
MorphBank is open to any systematist interested in storing high-quality digital images of organisms in a general-purpose searchable data base. See the 'Contribute' link above for more information on how you can upload images to MorphBank or have MorphBank mirrored to a local server near you. MorphBank currently contains about 6,000 images of different groups of organisms, mainly insects.
back to the top
Copyright Policy
The images in MorphBank can be used for private or non-commercial purposes for free, provided that the source and the copyright holder are cited. Any commercial use requires consent from the copyright holder.

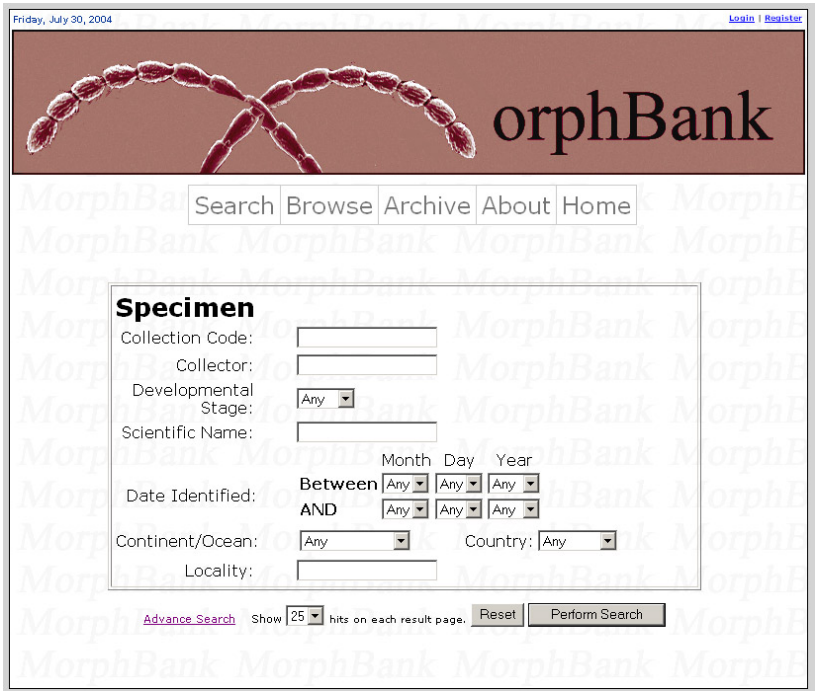
Search Results Page



Taxonomy Search Page

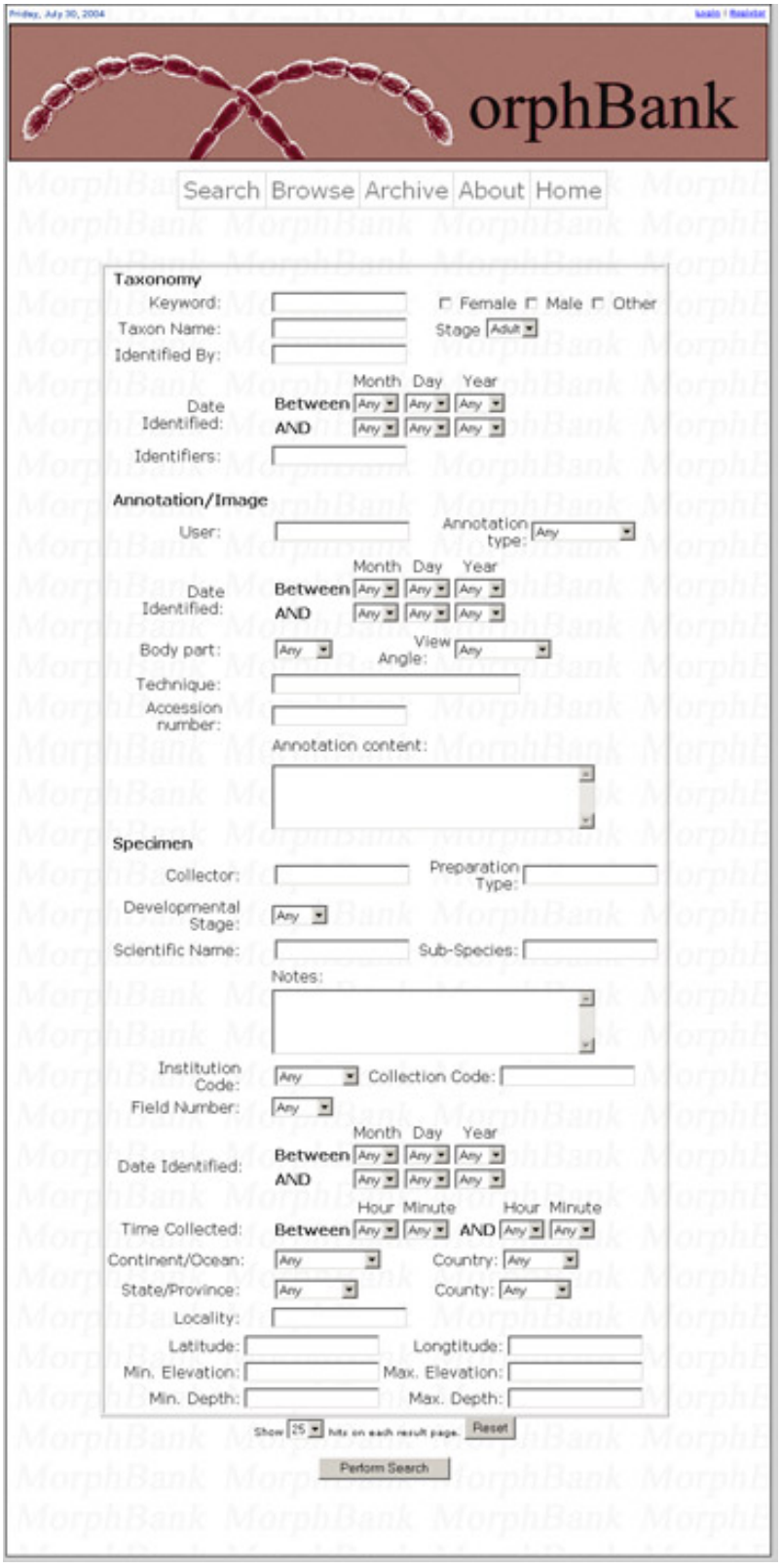


Specimen Search Page



Annotation/Image Search Page

Advanced Search Page


Friday, July 30, 2004
orphBank
Search Browse Archive About Home
Taxonomy
Keyword:
Female Male Other
Taxon Name:
Stage Adult
Identified By:
Month Day Year
Date Identified:
Between Any Any Any
AND Any Any Any
Identifiers:
Annotation/Image
User:
Annotation type: Any
Month Day Year
Date Identified:
Between Any Any Any
AND Any Any Any
Body part: Any
View Angle: Any
Technique:
Accession number:
Annotation content:
Specimen
Collector:
Preparation Type:
Developmental Stage: Any
Scientific Name:
Sub-Species:
Notes:
Institution Code: Any
Collection Code:
Field Number: Any
Month Day Year
Date Identified:
Between Any Any Any
AND Any Any Any
Hour Minute Hour Minute
Time Collected:
Between Any Any AND Any Any
Continent/Ocean: Any
Country: Any
State/Province: Any
County: Any
Locality:
Latitude:
Longtitude:
Min. Elevation:
Max. Elevation:
Min. Depth:
Max. Depth:
Show 25 hits on each result page.
Reset
Perform Search

Specimen Data Page

Friday, July 30, 2004 Login | Register

orphBank

Sample Specimen Data Page

Annotation #1
By: Dr. Jay Gould
Rating: 2/5

Annotation #1
By: David Gaitros
Rating: 4/5

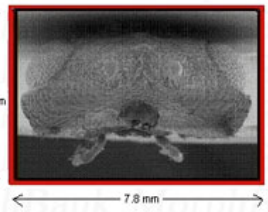
Annotation #1
By: Dean Philip
Rating: 5/5

Annotation #1
By: Dr. Ronquest
Rating: 5/5

Annotation #1
By: Dr. Melvin Shayma
Rating: 1/5

Add Annotation

Submitted By: Not available
Submission Date: 2001-06-07
Copyright Holder: Johan Liljeblad
Copyright Contact: N/A
View Number: 02
Body Part: Head
Viewing Angle: Anterior
Technique: SEM
Preparation: Ammonia cleaning, airdried, gold coated
Magnification: 206
Dimension (px): 1424x1067
Resolution (px/inch): 151



Taxonomy

Kingdom:
Phylum: Arthropoda
Class: Insecta
Order: Hymenoptera
Family: Cynipidae
Genus: Acraspis
Species: erinacei
Form: Parthenogentic female

Specimen

Scientific Name:
Sub-Species:
Sex: *Unknown*
Development Stage: Larva
Preparation Type: Gold

Collection Data

Time Collected: 9:35 am
Continent/Ocean: Africa
Country: Zaire
Locality: Jungle
Latitude: 145.6
Longitude: 67.8

Minimum Elevation: 100 ft. Above Sea Level
Maximum Elevation: 500 ft. Above Sea Level
Minimum Depth:
Maximum Depth:

Collector Name: Dean, Philip
Institution Code: 23
Collection Code: 54
Field Number: 78
Date Collected: 12-5-1998