

Institutional Council Meeting

BHL Technical Update

William Ulate

BHL US/UK Technical Director

Marine Biological Laboratory Woods Hole, Massachusetts May 6-7, 2013



BHL Technical Update

- Merge the BHL Australia website User Interface and the BHL-US/UK website Functionality
- The Global Names Architecture project (National Science Foundation)
- The Art of Life: Data Mining and Crowdsourcing the Identification and Description of Natural History Illustrations from the Biodiversity Heritage Library (National Endowment for the Humanities)



BHL Technical Update

• BHL AU-USome

The NSF GNA project

The NEH Art of Life project



(thanks to B.Crowley for the name)





...Notes & News from the BHL Staff

THURSDAY, JULY 14, 2011

BHL goes live in Australia with launch of BHL-Australia site

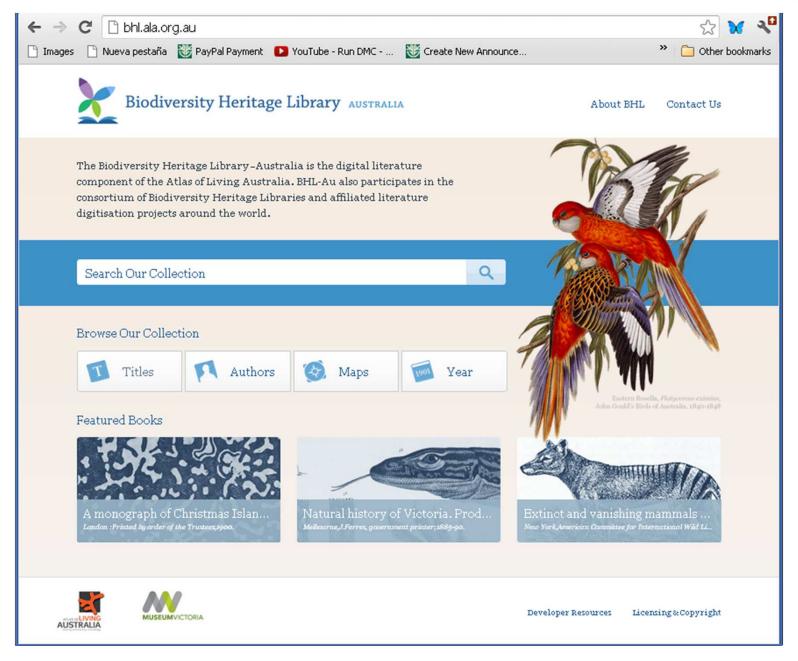


The newest addition to the BHL global family, BHL-Australia, has now launched. BHL in Australia is the digital literature component of the Atlas of Living Australia (ALA), a national project that aims to make biodiversity information more accessible and useable on line. It is a partnership between CSIRO, Australia's national science agency, the Australian natural

history collections community and the Australian Government.

The Australian node project is being managed through Museum Victoria, based in Melbourne. Dr Elycia Wallis, the BHL Australia node team lead said "We are so pleased to have an Australian BHL node up and running. Now the fun part of developing new tools and functionality can start so that we can share these back to the global BHL community." The Australian node has taken the core BHL code and reworked it with a redesigned user interface. Designer Simon O'Shea and developer Michael Mason at Museum Victoria wanted to concentrate on providing the core functionality of BHL as well as giving an Australian feel to the site design by using







2011 Usability Test

- A list of key differences between user interfaces
- Feedback from (17) users on their preferences
- Usability Test Notes and Survey Summary
- Usability Test Report



BHL AU-USome (2012)

Usability Test Report

- Names
- OCR
- Illustrations
- More information on Species
- Book Viewer
- Advanced Search



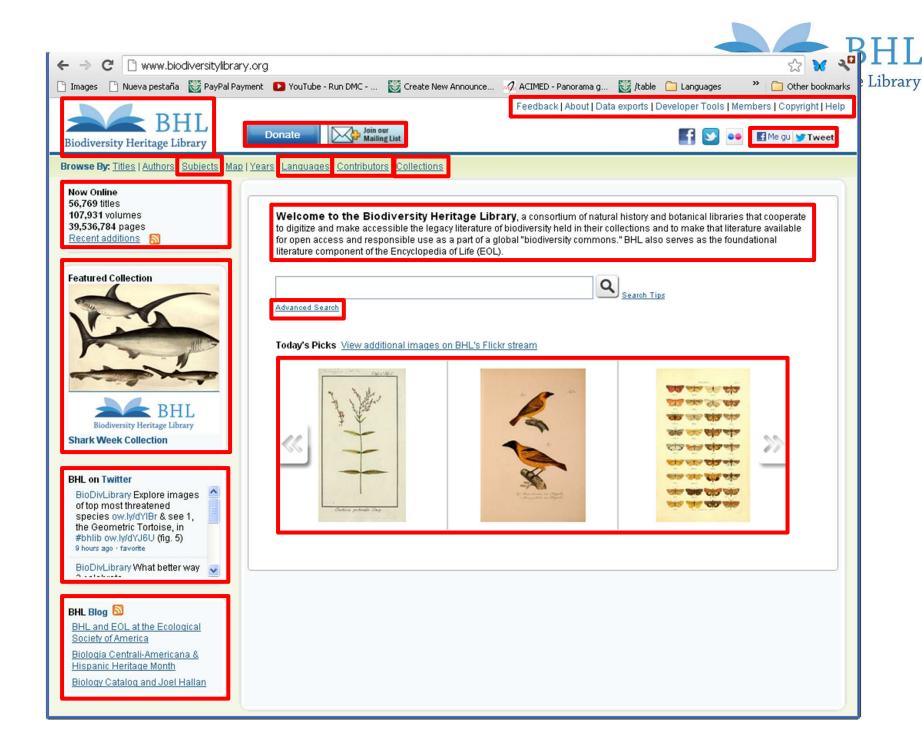
New US/UK functionality since 2011

Home page

- Featured Collections
- Browse by Collection
- Links on top right of main page
 - feedback, exports, members
- Now Online stats box on main page
- Recently Added view (never in BHL-AU)
- Twitter feed integration
- Blog integration
- Donate/Mailing List buttons
- Flickr images on home page

Other

- Social Media (Like/Tweet) buttons
- Advanced search interface
- DOIs added to bibliography page for titles
- Title variants in bibliography page for titles
- Schema.org markup added
- Darwin's Library annotation viewer
- Icon to "Add record to Mendeley library"





Article data model changes

- In the book viewer, display a list of articles contained in the book/journal being viewed. Pick one and navigate to the start page.
- Need a "landing page" for articles with article metadata.
- Link to the book viewer, an external location (a PDF in another repository), or nothing.
- Display Articles that match the search term in a new section of Search results.
- Add an option to "View Record" or "View Article".
- Browsing should include articles related, not only titles.



- Advantages
 - AU and US share IDs across portals!
 - Model has been kept synchronized
- Disadvantages
 - US Code modified since AU launched
 - AU didn't incorporate certain functionality



Proposed Timeline

1. DESIGN PHASE – Aug. 6 –24.

Input: Specifications/descriptions/notes about new features as soon as possible. *Outcome:* A full suite of designs, that incorporate comments from the 2011 usability survey and incorporating any new features that you have planned or are already building.

2. COMMENT AND RESPONSE PHASE – Aug. 24 – Sep. 14.

Input: Comments to the suites of designs.

Outcome: Comments and responses to comments

3. FINAL SIGN OFF – Sep. 21

Output: The full suite of designs signed off by September 21, packaged, files transferred across to MOBOT.

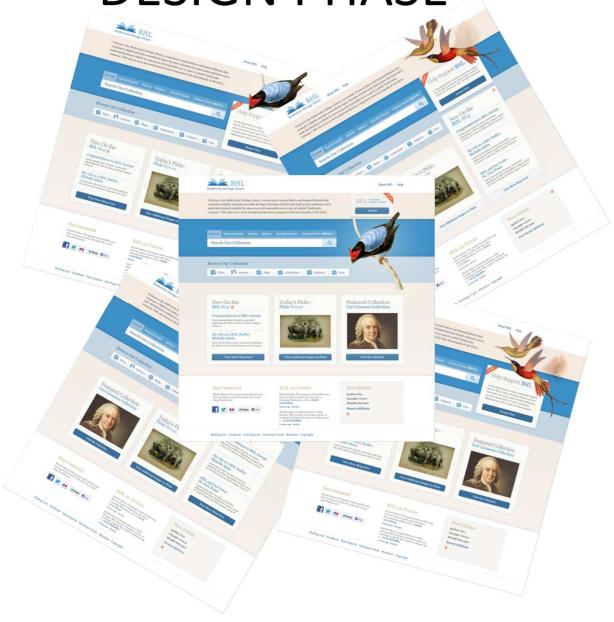
4. Simon Sherrin VISIT TO Saint Louis, MO – Oct. 1 - 22.

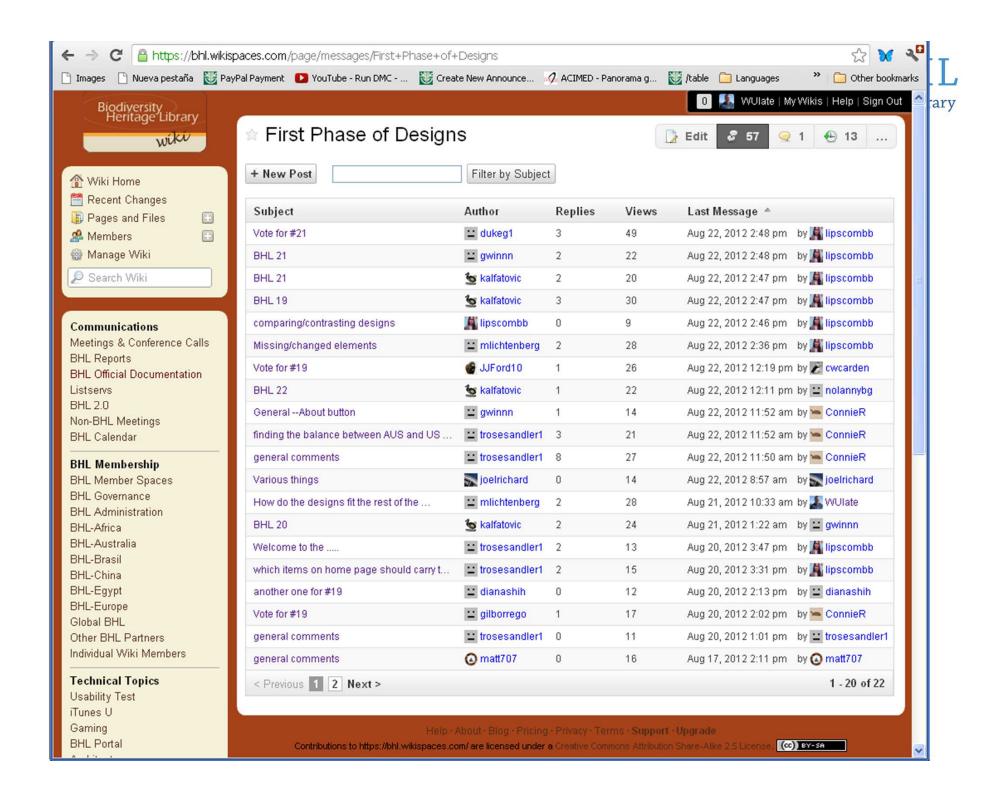
Output: BHL US codebase adapted to incorporate new designs.

5. NEW DESIGN IN PRODUCTION - . Mar. 28

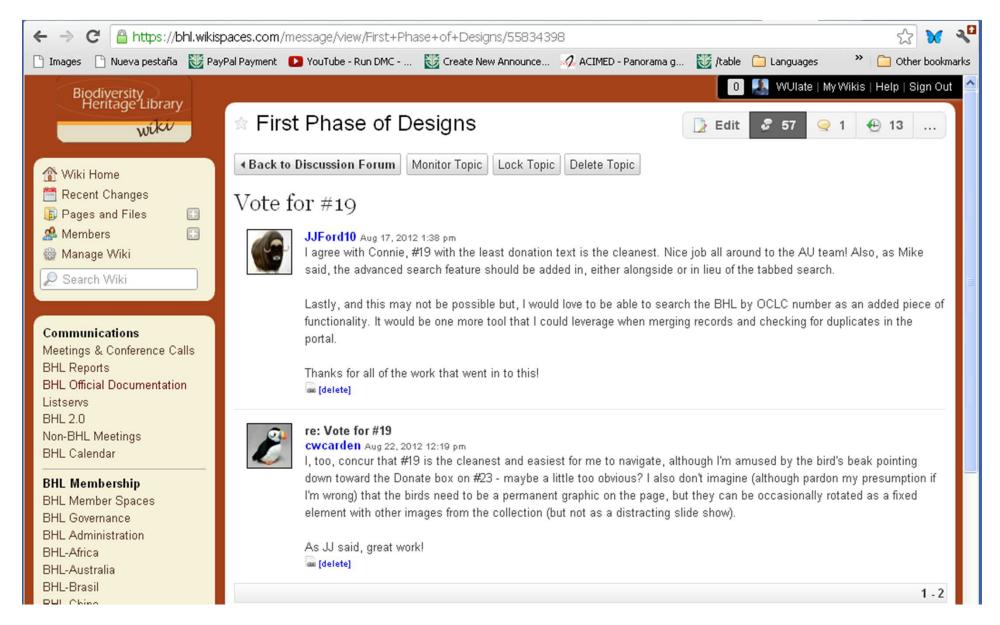


DESIGN PHASE











About BHL Help



Welcome to the Biodiversity Heritage Library, a consortium of natural history and botanical libraries that cooperate to digitize and make accessible the legacy literature of biodiversity held in their collections and to make that literature available for open access and responsible use as a part of a global "biodiversity commons." BHL also serves as the foundational literature component of the Encyclopedia of Life (EOL).

Search

Browse Our Collection By:







Collection

Q

Help Support BHL

Loren ipsum dolor sit amet, consectetur adipisc ing elit. Sed eu enim nec libero tempor varius et varius nisi. Sed cursus odio vel nulla scelerisque at Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed eu enim nec libero tempor

New On the BHL Blog

Congratulations to BHL-Europe

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed ou enim nor libero tempor varios et

My Life as a BHL Staffer: Michelle Abeln

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed eu enim nec libero tempor varios et varios

Today's Picks -Flickr Stream



Featured Collection Carl Linnaeus Collection



Visit the Collection

Stay Connected

Keep up to date with the latest news via these social media platforms.









Join Our Mailing List

Sign up for the BHL Newsletter and get the latest updates.



BHL on Twitter

BioDivLibrary The gorgon head starfish may have as many as 5000 arm tips! See a stunning Bustration of it in *bhbb ow.ly/bbtah

8 hours ago - fermille

BioDivLibrary Congratulations to BHL-Europe: Bill-Europe is hosting a series of moetings in Serlin this week to celebrate the s...bit.ly/MyOMRs.

13 hours ago - famorite

BioDivLibrary Maps that visually demonstrate vertebrate biodiversity concentrations ow.ly/dxmXh Wonder if there's something similar 4 plants?

11 hours ago - favorite

Now Online

55,800 titles

105,956 volumes 38,958,799 pages

Recent Additions



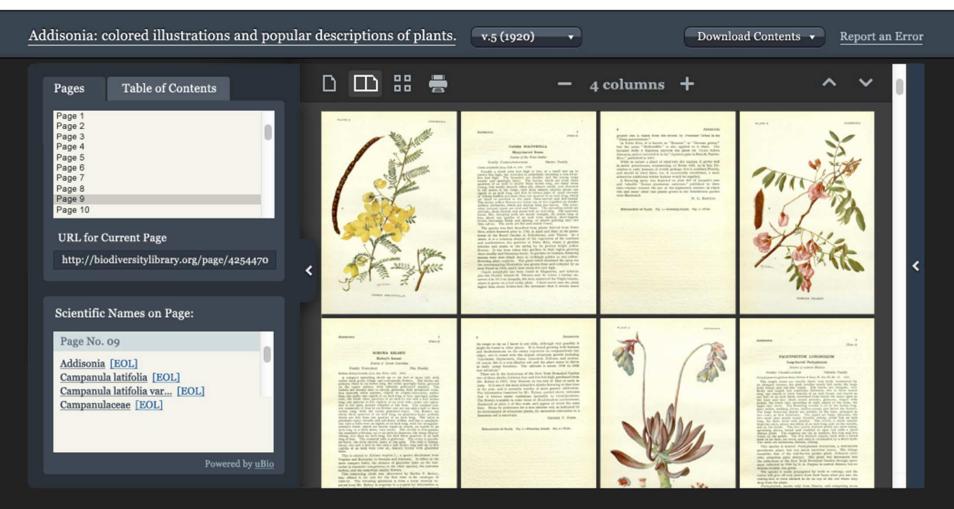
ADVANCED SEARCH



About BHL Help

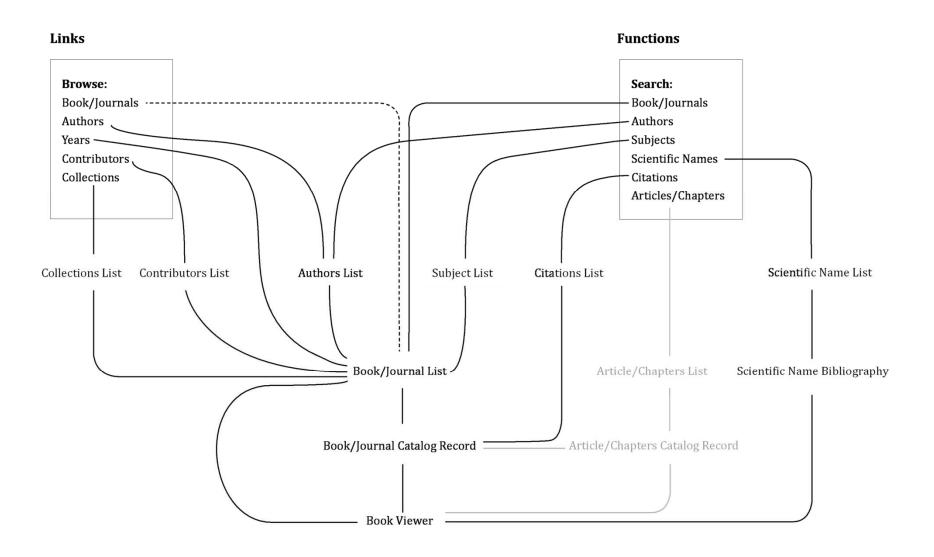


Search



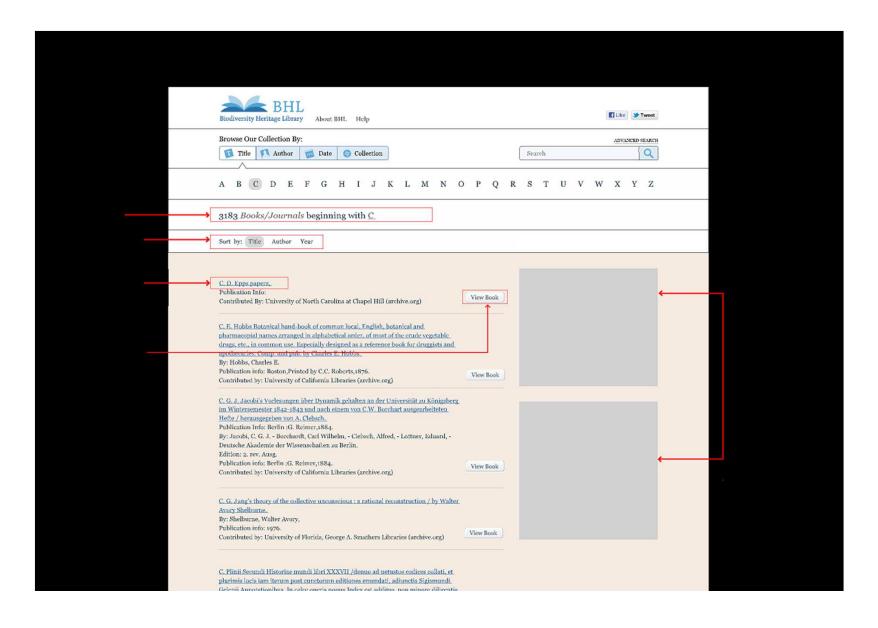


Relations between links and functions





Wireframes of Different Layers



Wireframes of Different Layers





Portal



About BHL

Help



Welcome to the Biodiversity Heritage Library, a consortium of natural history and botanical libraries that cooperate to digitize and make accessible the legacy literature of biodiversity held in their collections and to make that literature available for open access and responsible use as a part of a global "biodiversity commons." BHL also serves as the foundational literature component of the Encyclopedia of Life (EOL).

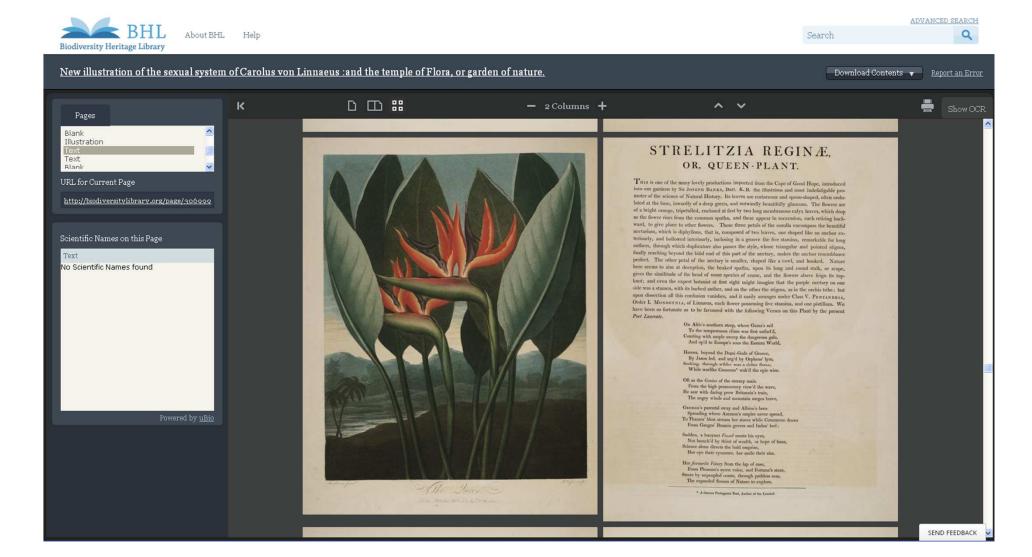


New on the BHL Blog

Today's Picks Flickr Stream Featured Collection Charles Darwin's Library



Book Viewer

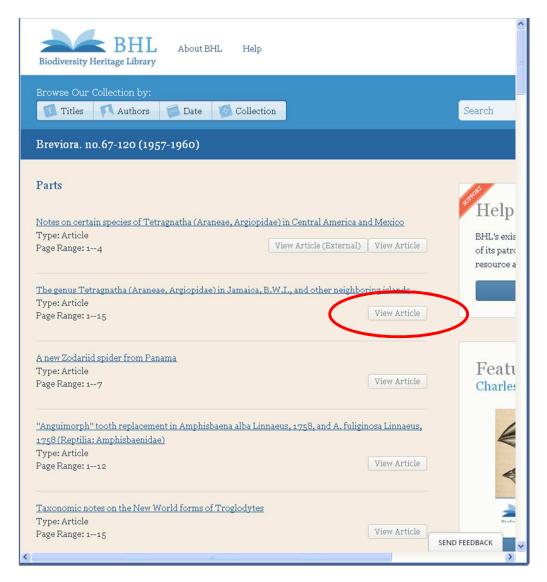


Treatment-level metadata Article-level metadata Chapter-level metadata Part-level metadata

- Disambiguating and *locating* structural components in the corpus
- Done by automated and crowdsourced means
 - Thanks Rod Page! Welcome others!
- Greatly increases semantic value of the dataset
- Addressing important makes data addressable and thus linkable

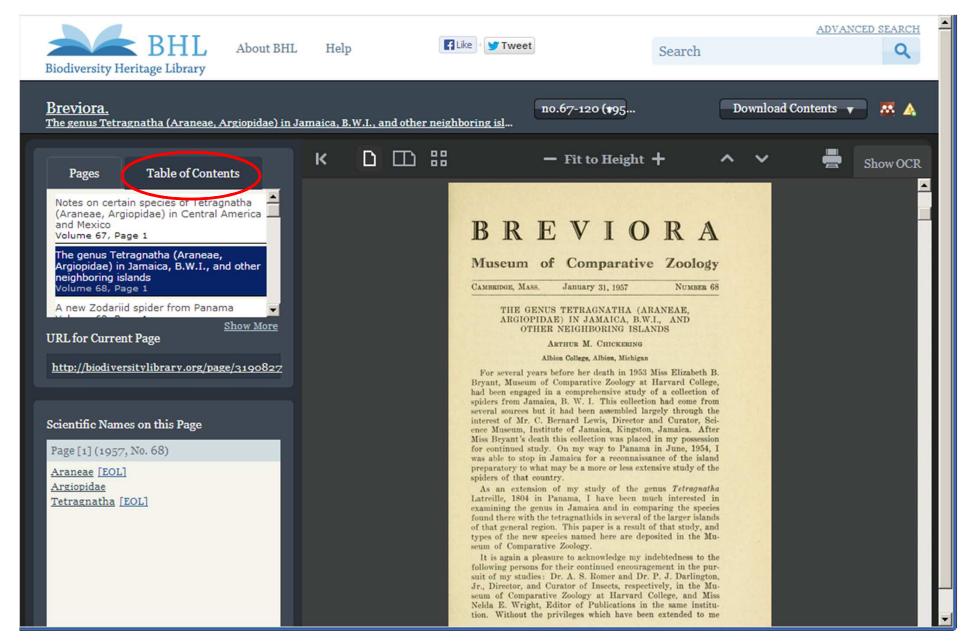


Articles in the BHL UI





Articles



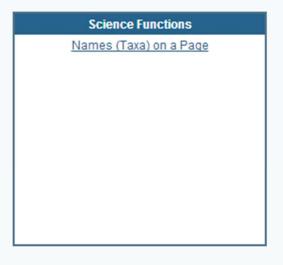


Articles

Dashboard







Library Stats		?	
	All Records	Active	
Titles	61956	59349	
Items	114985	113233	
Pages	41397934	40786634	
segments	92520	92520	
Items w/ Segments	7735	7602	
Names	151877692	151249651	
Unique Names	29313460	29117883	
Verified Names	9327768	9276202	
Show EOL Names			
Show Pages with EOL Names			
Expanded Library Stats			



PDF Generation Stats				
<u>Status</u>	# Of PDFs			
Generated	150401			
Error	36			
Declined	103			
<u>Expanded P</u>	<u>DF Stats</u>			

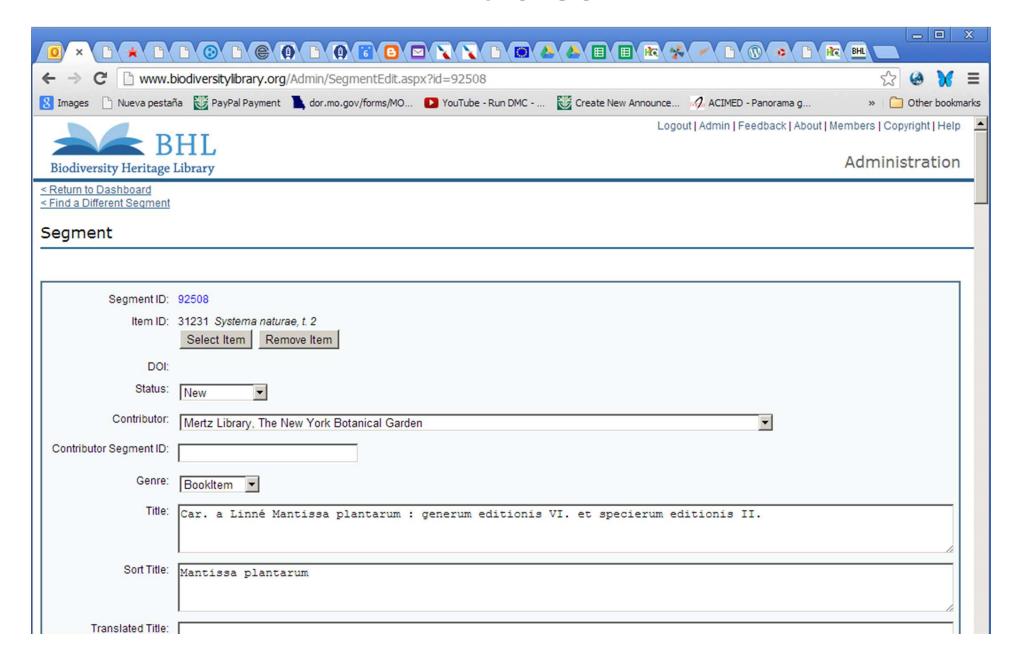
IA Harvoet State

PioStor Harvoet State

DOI Accignment State

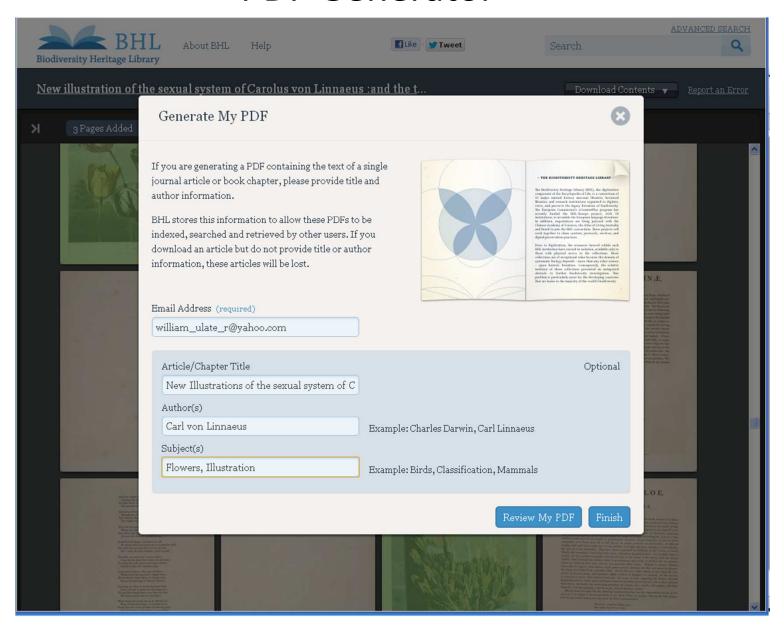


Articles





PDF Generator





Topics with the TAG

- 1. Making Tabs more prominent & choose the first one with data.
- 2. Reinstate the "Contributing Library" in the book viewer.
- 3. Too many Volumes...
- 4. Lists are too long.
- 5. Assign DOIs to articles.
- 6. Citations that link to no content.
- 7. PDFs are too big



NSF GNA Project

Citation Services in BHL



- Fulfill the role of the global repository for bibliographic citations relating to biodiversity.
- Provide an open environment for sharing and disseminating citations that suit taxonomists (series, volumes, articles, pages, and treatments.).
- Coming from multiple sources, raw citations are not standardized and reconciliation services must be provided to map variant forms together.



- The Citebank platform contains the aggregated bibliographies of BHL, other digital libraries, publishers, institutional repositories, and contributed bibliographies from specialist groups.
- Citebank is built from the open-source publishing framework Drupal and includes community-authored components for the management of bibliographic citations.



Where are we?

- Articles
 - Extend BHL data model to store article metadata
 - Build process to harvest data from BioStor
- Create user interfaces for adding article metadata and associated files
 - Define functional requirements
 - Define process flow for adding article metadata and associated files
 - Implement UI changes
- Change BHL UI to accommodate article search
- Change BHL UI to accommodate article display (TOC)



Where are we?

Link Out

- Extend BHL data model to link out to titles and items in other systems
- Create user interfaces for adding links out to titles & items in other systems
- Adjust BHL web display to show links out to titles & items in other systems

Name-finding Improvements

- Enhance name finding algorithms
- Review changes to BHL data model to accommodate enhancements
- Review changes to BHL UI to accommodate enhancements



Where are we?

- Citation reconciliation
 - Augment existing BHL APIs to return article metadata and associated files
 - Respond to requests for improvements
 from ZooBank & IPNI & Index Fungorum



Where are we going?

- Citation Reconciliation
 - merging citations with the same title; (for example, an author publishing the next segment with the same title or the same title over and over)
- Crowdsource corrections and contributions
 - Consider expanding security (login user accounts)

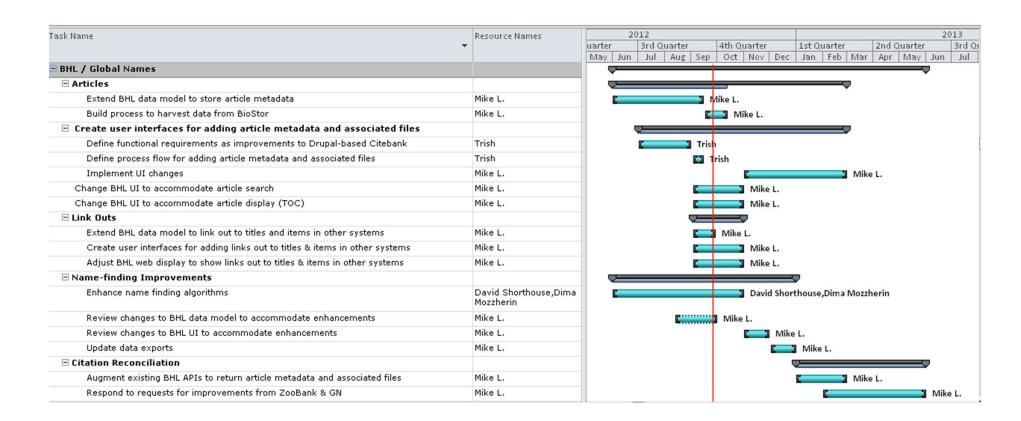


Functional requirements for a citation repository

- IMPORTING (Administrator)
- IMPORTING (General User)
- RECORD CREATION (General User)
- RECORD EDITING (General User)
- USER MANAGEMENT (Administrator)
- BROWSE (General User)
- CITATION TYPES
- OAI HARVESTING
- SPECIFICATIONS FOR DATA PROVIDERS PAGE
- CONTRIBUTORS PAGE
- REPORTING
- GLOBAL UPDATES (Administrator)
- RELATIONSHIPS BETWEEN CONTENT FILES AND CITATIONS
- FIELDS



Global Names



One year no-cost extension requested



The Art of Life

Data Mining and Crowdsourcing the Identification and Description of Natural History Illustrations from the Biodiversity Heritage Library









What is Art of Life?

- Grant given to Missouri Botanical Garden in St Louis, to work with Indianapolis Museum of Art and University of Colorado Boulder.
- Funded by National Endowment for the Humanities
- With support of staff from BHL (SIL)
- Runs May 2012-April 2014











5 Primary Objectives of Art of Life

Objective 1: Define an appropriate metadata schema for natural history illustrations

Objective 2: Build software tools to automatically identify illustrations in the BHL corpus

Objective 3: Enhance existing tools to enable the initial sorting, viewing, and editing of these identified visual resources.

Objective 4: Integrate tagging applications to enable a community of users to edit descriptive metadata for the illustrations

Objective 5: Integrate the descriptive metadata generated by users back into BHL portal both for access and preservation



The Art of Life

Data Mining and Crowdsourcing the Identification and Description of Natural History Illustrations from the Biodiversity Heritage Library



Title Amandava formosa

Type Illustrations

Date Publication: 1898

Agent Publisher: Arthur G. Butler (1844-1925)
Illustrator: F.W. Frohawk (1861-1946)

,

Description A pair of finches with green and yellow bodies resting on reeds

Subjects Scientific name: Amandava formosa (Latham, 1790)

Vernacular Name: Green Avadavat or Green Munia

Accepted Name: Estrilda formosa

Birds, finches

Inscriptions bottom center: Green Amaduvade Waxbill (Stictospiza formosa)

Source Butler, Arthur Gardiner. Foreign finches in captivity. HullBrumby and Clarke,

limited, 1889 (2nd edition). This image comes from the Biodiversity Heritage Library,

and is available online at biodiversitylibrary.org/page/17195895

Rights Public domain

ART OF LIFE workflow diagram

Purpose: Identify BHL pages with illustrations

Purpose: Classify illustrations into broad types in order to push them to other platforms for more detailed description

Purpose: Allow public & specialists to further describe illustrations

Purpose: Share described illustrations more widely both within BHL portal and beyond

EXTRACT

Users:

- · BHL Staff
- **Functionality:**
- · If there's a page with only 1 illustration generate 1 image and 1 md record; if there's a page with more than 1 illustration then generate multiple images and multiple metadata records?
- Metadata should include both citation for book or journal and page url
- · Provide initial category based on % of page covered by image, color histogram and compression algorithm
- · Quick load of pages
- · Provide coordinates of bounding box per image found
- · Extract images only found maintaining association to page id
- · View all pages of a book (left pane) side by side with images found by algorithm (right pane) in order to assess its accuracy. Need easy way to mark which images are missing in left pane and which are false positives in right pane.
- · Modifiable by BHL staff

Platform:

TBD

Developers:

IMA- role: develop algorithm, develop assessment UI

MOBOT- role:validate algorithm, provide processing environment for final run

CLASSIFY

Users:

· BHL Staff

Functionality:

- · Zoom + / -
- Quick load of pages
- Add metadata to
- · 1 page
- · all pages in a book or journal -search across books & journals (e.g. see all illus tagged as "map")
- · identify which pages should be further described
- · Go look at original page
- · View and sort by: size, book, page type, sequence
- · Push metadata (pagetype+) from classifier to BHL portal
- Push metadata & images to descriptive platforms (i.e. flickr, commons)

Platform:

TBD

Developers:

MOBOT- role:further develop paginator to serve as classfier

DESCRIBE

Users:

General public & Specialists

Functionality:

- · Show 1 illustration
- · Show title (if available) and description
- · Show Source & URL
- · Show other metadata through machine tags
- · Allow for tag addition by users
- Need simple form for entering descriptive data

Platform:

- -Wikimedia Commons (specialists) -Flickr (General public)
- **Developers:**

MOBOT- role:develop data entry form for wikimedia commons or develop standalone application for specialists describing images?

SHARE

Users:

- · EOL
- · ARTStor
- ·iTunesU

Functionality:

- · Use binomial tags for EOL automatic consumption (could be several by illustration?)
- -merge metadata from Flickr & Commons
- -change model so illus md can be stored
- -change portal UI to display md -develop export output for users

Platform:

BHL DB

Developers:

MOBOT- role:develop additional ways of sharing and packaging data via OAl-PMH, data exports, APIs, etc.





Current status of Art of Life

- Development of the algorithm is about 90% complete and will be done by May 2013
- Draft schema for describing natural history illustrations available for public review http://tinyurl.com/9hm7nsb
- Classifier tool in progress





Current status of Art of Life

- Algorithm due to significant staff changes at the Indianapolis Museum of Art the algorithm work was delayed this spring but we are now in the final stages of the algorithm work where we are identifying which of the 4 algorithms that were developed are performing most effectively on the test set. We are working the staff at MBL Woods Hole to work out procedures for running the algorithm across the entire corpus.
- Classifier Joel Richard has modified the Macaw tool that he developed for paginating and it will be used by BHL staff to do some basic classification of page images (e.g. photos, drawings, maps, etc) before the pages are sent to tagging environments such as flickr and Wikimedia commons





Current status of Art of Life

- Held first Advisory Board meeting in January which was very successful and received good feedback.
- Members include:
 - Doug Holland, Director, Missouri Botanical Garden Library
 - Dr. Hong Cui, Assistant Professor, University of Arizona
 - Dr. David Kohn, Director and General Editor, Darwin Manuscripts Project, American Museum of Natural History
 - Charles Miller, Chief Information Officer, Missouri Botanical Garden
 - Nancy Gwinn, Director, Smithsonian Institution Libraries
 - Robert Guralnick, Associate Professor at the University of Colorado at Boulder
 - Betty Smocovitis, Professor of Zoology and History at the University of Florida



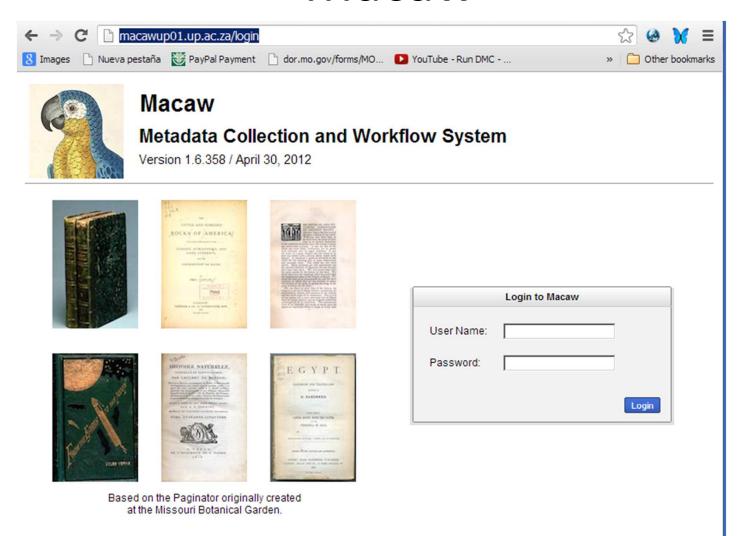


Presentations on Art of Life

- Biodiversity Informatics Standards Annual Conference, Beijing, China. Oct 2012, <u>The Art of Life Schema</u>: <u>describing and providing access to natural history illustrations from the Biodiversity Heritage Library (BHL)</u>, William Ulate, Trish Rose-Sandler, Gaurav Vaidya, Robert Guralnick
- Museums and the Web conference, Portland, OR Apr 2013, More than just a pretty
 picture: improving the discoverability of illustrations in the Biodiversity Heritage Library
 (BHL), Gilbert Borrego, Grace Costantino, Bianca Crowley, Kyle Jaebker, Trish Rose-Sandler
- Visual Resources Association conference, Providence, RI, Apr 2013, <u>A Case Study of The Art of Life: Data Mining and Crowdsourcing the Identification and Description of Natural History Illustrations from the Biodiversity Heritage Library</u>, Trish Rose-Sandler
- St Louis Regional Library Network (SLRLN) Tech Expo, St Louis MO Mar 2013, <u>The Art of Life: Data mining and crowdsourcing the identification an description of natural history illustrations from the Biodiversity Heritage Library</u> Trish Rose-Sandler



Macaw



http://macawup01.up.ac.za



Viewing Activity



Statistics

Summary

- 0 items ready to be scanned.
- 1 items in progress.
- 0 reviewed and ready to share.
- 0 being exported or verified.
- 0 items completed.
- 0 items archived.
- 0 items have errors.

290 pages scanned total.

0d 12h 48m average scanning time.

Disk Usage (GB)



Pages Per Day

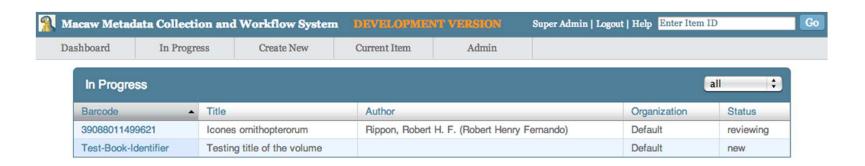


Total Pages Scanned





Viewing Activity



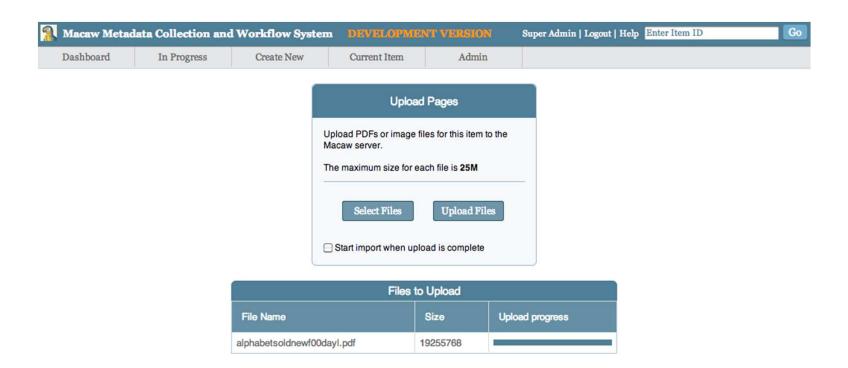


Loading Activity

Macaw Metad	lata Collection and	l Workflow System	DEVELOPME	NT VERSION	Super Admin Logout Help	Enter Item ID	Go
Dashboard	In Progress	Create New	Current Item	Admin			
	Identifier:		Add	Item			
	Organization: Needs QA: Copyright:		Default This item will be reviewed for Quality Assurance Not in Copyright				
	Creative Committee	nons: (none)	\$				
	marc_xml						
				w Row 💿			



Uploading images via browser





Uploading images via browser



testered a blacked as the south a factor of

/var/www/joelrichard.com/macaw/docs/incoming/Test-Book-Identifier

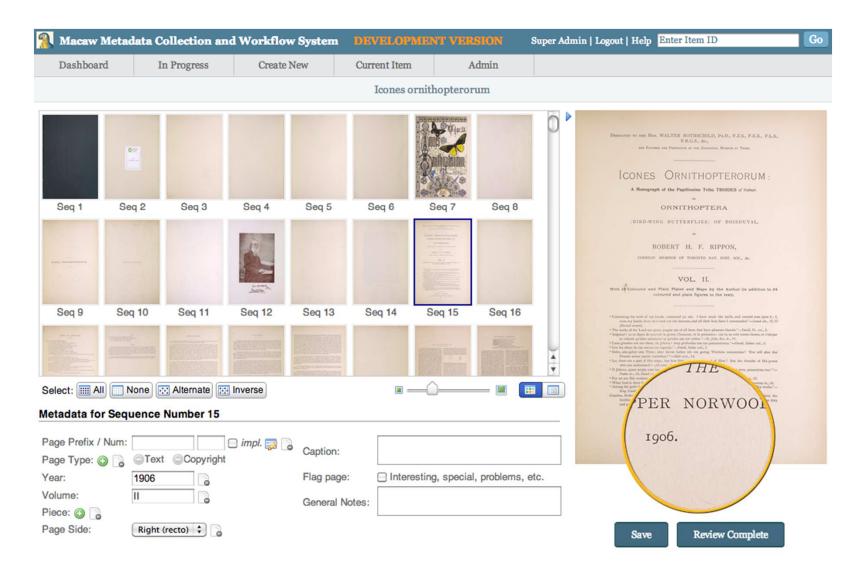
Start the import:



While the scans will be copied to the server, you can monitor the progress on the right.

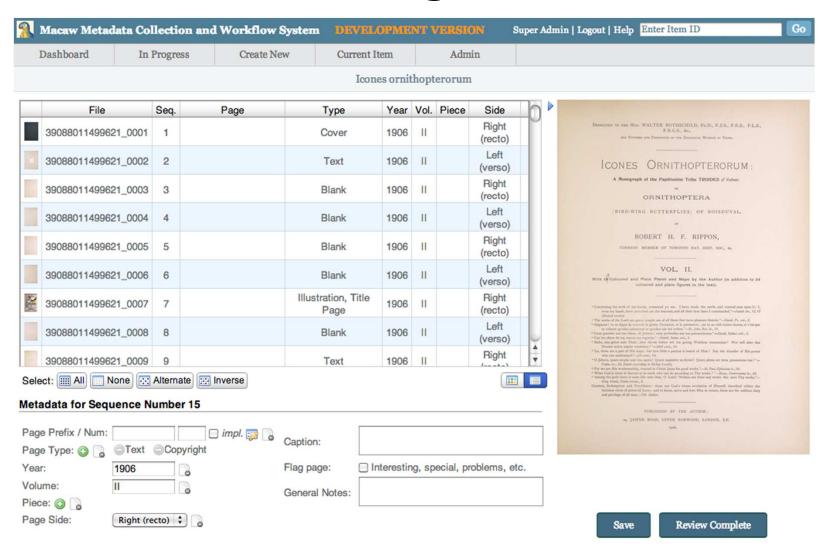


Reviewing Metadata





Reviewing Metadata





 Need to get set up with an account at IA first

Account at IA needs access to the biodiversity collection

 Uploading of completed items is done via scheduled job or the command line



Institutional Council Meeting

Thank you

William Ulate

BHL US/UK Technical Director William.Ulate@mobot.org

Marine Biological Laboratory Woods Hole, Massachusetts May 6-7, 2013