QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

BHL, THE BIODIVERSITY HERITAGE LIBRARY:
Exposing the Taxonomic Literature

Constance Rinaldo, Ernst Mayr Library, Museum of Comparative
Zoology, Harvard University, Cambridge MA USA

http://wwwbiodiversitylibrary.org

WHAT IS THE BHL?

- Large scale digitization to provide open access to core published literature of biodiversity for scientists
- Key component of the *Encyclopedia of Life* http://www.eol.org (EOL) as conceived by E. O. Wilson
- Collaboration of ten major natural history, botanical garden & research libraries: American Museum of Natural History, Field Museum of Natural History, Harvard University (Botany Libraries & Ernst Mayr Library), Marine Biological Laboratory/Woods Hole Oceanographic Institute (MBL/WHOI), Missouri Botanical Garden (MOBOT), Natural History Museum, London, New York Botanical Garden & Royal Botanic Garden, Kew
- Collaboration with global taxonomic community: Global Biodiversity Information Facility (GBIF), International Commission on Zoological Nomenclature (ICZN), European Distributed Institute of Taxonomy, Atlas of Living Australia, Chinese Academy of Sciences, Museum fur Naturkunde der Humboldt-Universitat, BIOONE & more

WHY DO THIS NOW?

- Biodiversity is HOT; biodiversity studies need taxonomic data
- Taxonomic data are reported in general & specialized literature that may only be found in a few libraries & museums
- Current taxonomic research often relies on multiple texts & specimens more than 100 years old that are dispersed among libraries & museums around the world
- Digital technology offers an access solution to this “taxonomic impediment” that required taxonomists to travel the world to examine every specimen & paper related to an organism
- Taxonomic literature has extreme longevity thus the public domain literature is important
- Literature repatriation: most taxonomic literature is in the developed world while most biodiversity is not (Figure 1)

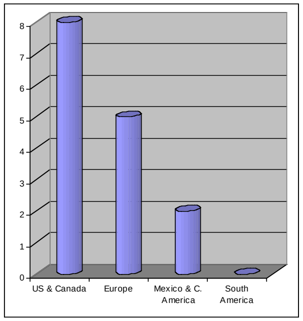
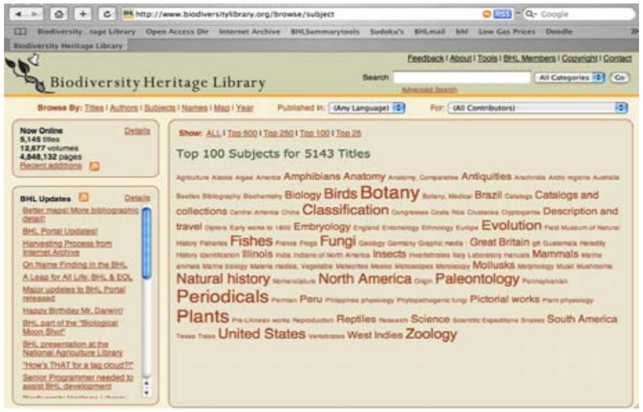


Figure 1: Distribution of copies of the *Biologia Centrali-Americana*; the copies in Central America are located in one of the twenty branches of the Smithsonian Libraries. Courtesy, Martin Kalfatovic.


http://www.biodiversitylibrary.org/browse/subject
Biodiversity Heritage Library
Open Access Dir
Internet Archive
BHLSummarytools
Feedback | About | Tools | BHL Members | Copyright | Contact
Search
All Categories
Go
Advanced Search
Browse By: Titles | Authors | Subjects | Names | Map | Year
Published In: (Any Language)
For: (All Contributors)
Details
Recent additions
Show: ALL | Top 500 | Top 250 | Top 100 | Top 25

Biodiversity Heritage Library

Now Online
5,145 titles
12,677 volumes
4,848,132 pages

BHL Updates

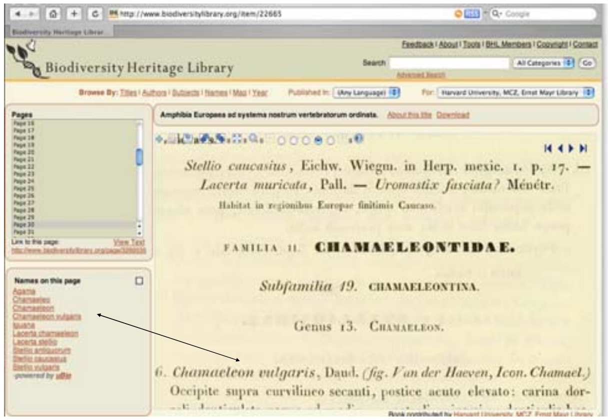
Better maps! More bibliographic detail!
BHL Portal Updates!
Harvesting Process from Internet Archive
On Name Finding in the BHL
A Leap for All Life: BHL & EOL
Major updates to BHL Portal released
Happy Birthday Mr. Darwin!
BHL part of the "Biological Moon Shot"
BHL presentation at the National Agriculture Library
"How's THAT for a tag cloud?!"
Senior Programmer needed to assist BHL development

Top 100 Subjects for 5143 Titles

Agriculture Alaska Algae America Amphibians Anatomy Anatomy, Comparative Antiquities Arachnida Arctic regions Australia Beetles Bibliography Biochemistry Biology Birds Botany Botany, Medical Brazil Catalogs Catalogs and collections Central America China Classification Congresses Costa Rica Crustacea Cryptogams Description and travel Diptera Early works to 1800 Embryology England Entomology Ethnology Europe Evolution Field Museum of Natural History Fisheries Fishes France Frogs Fungi Geology Germany Graphic media Great Britain [illegible] Guatemala Heredity History Identification Illinois India Indians of North America Insects Invertebrates Italy Laboratory manuals Mammals Marine animals Marine biology Materia medica, Vegetable Meteorites Mexico Microscopes Microscopy Mollusks Morphology Music Mushrooms Natural history Nomenclature North America Origin Paleontology Pennsylvanian Periodicals Permian Peru Philippines physiology Phytopathogenic fungi Pictorial works Plant physiology Plants Pre-Linnean works Reproduction Reptiles Research Science Scientific Expeditions Snakes South America Texas Trees United States Vertebrates West Indies Zoology

WHY A BHL PORTAL?

- Prototype developed at MOBOT as Botanicus.org & tested with scientists
- BHL Portal serves images & text files ingested from Internet Archive
- BHL Portal ingests MARCXML metadata & low resolution JPEG files; High resolution files are retrieved on the fly from IA
- Globally Unique Identifiers (GUIDs) allow links to other services such as EOL
- Taxonomic Intelligence developed at MBL/WHOI allows species name searching by users (Figure 2)
    - -TI uses sophisticated algorithm to locate name strings in the Optical Character Recognition (OCR) files that match the 9.4 million names in NameBank
    - -Iterative processing of texts increases the number of names in NameBank & the accuracy of recognition
    - -More tools are under development


http://www.biodiversitylibrary.org/item/22665
Google
Feedback | About | Tools | BHL Members | Copyright | Contact
Search
All Categories
Go
Advanced Search
Browse By: Titles | Authors | Subjects | Names | Map | Year
Published In: (Any Language)
For: Harvard University, MCZ, Ernst Mayr Library
About this title
Download
View Text

Biodiversity Heritage Library

Pages
Page 16
Page 17
Page 18
Page 19
Page 20
Page 21
Page 22
Page 23
Page 24
Page 25
Page 26
Page 27
Page 28
Page 29
Page 30
Page 31

Link to this page: http://www.biodiversitylibrary.org/page/[illegible]

Names on this page □
Agama
Chamaeleo
Chamaeleon
Chamaeleon vulgaris
Iguana
Lacerta chamaeleon
Lacerta stellio
Stellio antiquorum
Stellio caucasus
Stellio vulgaris
powered by uBio

Amphibia Europaea ad systema nostrum vertebratorum ordinata.

Stellio caucasius, Eichw. Wiegm. in Herp. mexic. 1. p. 17. — *Lacerta muricata*, Pall. — *Uromastix fasciata?* Ménétr.

Habitat in regionibus Europae finitimis Caucaso.

FAMILIA 11. **CHAMAELEONTIDAE.**

Subfamilia 19. CHAMAELEONTINA.

Genus 13. Chamaeleon.

6. *Chamaeleon vulgaris*, Daud. (*fig. Van der Haeven, Icon. Chamael.*) Occipite supra curvilineo secanti, postice acuto elevato: carina dor-

Book contributed by Harvard University, MCZ, Ernst Mayr Library

Figure 2: Taxonomic intelligence in action

WHAT ABOUT COPYRIGHT?

- Public domain literature digitized first
- Opt-in copyright model: BHL actively works with professional societies & other small publishers to integrate publications into the BHL.
- Agreements to digitize 46 titles have been signed with the BHL providing digitization at no cost to society & museum publishers with material served from BHL portal & files available to publishers
- Discussions with commercial publishers for alternative agreements

HOW?

- BHL is not a legal entity: the ten member institutions signed separate Memoranda of Agreements with the BHL
- Directors of the member libraries meet annually; an elected executive council has weekly conference calls with the BHL Program Director & Technical Director
BHL member institution staff have regular conference calls to ensure that all institutions are
- Each institution has a separate contract with Internet Archive, the digitization partner
- IA has small scanning centers in London, DC & Illinois & large centers at the Boston Public Library (thanks to the Boston Library Consortium) & the New York Public Library
- Service is provided for $.10 per page with extra charges for foldouts
- MOBOT, NYBG, Harvard & the Smithsonian have “boutique” scanning facilities to digitize oversized & unusual items
- IA provides image files & text derived from OCR
- OCLC Collection Analysis tool generated a broad look at institutional collection strengths & provided an estimate of the number of public domain materials available for immediate digitization
- Duplication is minimized using tools developed by member libraries such as a serials bidding tool, monograph de-duping tool & others
- Workflow within the libraries includes generating picklists, identifying acceptable items within the picklist, barcoding, generating packing lists,checking out books, packing books, checking in & reshelving returned books & reviewing rejected items

WHERE DO WE GO FROM HERE?

- Article-level analysis of serials using automated tools
- Further develop global partnerships & incorporate multiple languages
- Linkages to molecular, morphological & other data types
- Improved OCR for non-Roman & non-standard scripts
- Enhance connections with EOL & others
- Expand content access & tools to new audiences
- Strengthen underlying architecture
- Further develop partnerships with commercial & society publishers
- Ingestion of collections that are open access & available

BIBLIOGRAPHY

Godfray, H.C.J., B.R. Clark, I. J. Kitching, S.J. Mayo & M.J. Scoble. 2007. “The Web and the Structure of Taxonomy,” *Systematic Biology*, 56(6): 943-955.
Leary, P.R. , D. P. Remson, C.N. Norton, D.J. Patterson & I.N. Sarkar. 2008. “uBioRSS: Tracking Taxonomic Literature Using RSS,” *Bioinformatics* 23(11): 1434-1436.
Minelli, A. 2003 “The Status of Taxonomic Literature,” *Trends in Ecology and Evolution* 18(2):75-78.
Sarkar, I.N., R. Schenk & C.N. Norton. 2008. “Exploring Historical Trends Using Taxonomic Name Data,” *BMC Evolutionary Biology* 8:144.

Funding for the BHL has come from the MacArthur Foundation, the Sloan Foundation, the Moore Foundation & individual BHL member institutions, including Harvard University. Any opinions, findings, & conclusions or recommendations expressed in this material are those of the author & do not necessarily reflect the views of funding agencies.

Please refer questions to: Connie Rinaldo,Secretary for the BHL, crinaldo@oeb.harvard.edu.