

Biodiversity Heritage Library



Nancy E. Gwinn
Smithsonian Institution Libraries
March 24, 2008



Encyclopedia of Life

- ◆ Major project to create a single Web page for every known species (1.8 million!)
- ◆ Total funding will reach at least **\$50M**
- ◆ EOL needs the literature underpinning in the BHL project
- ◆ BHL now key partner in EOL project
- ◆ EOL launched on 9th May, 2007
 - First 30,000 pages presented at TED conference Feb 27, 2008

Biodiversity
Heritage
Library



Synthesis Center
Field Museum

Informatics
Marine Biological
Laboratory &
MOBOT

Secretariat
Smithsonian

Education &
Outreach
Smithsonian/Harvard

Serine Molecule

Encyclopedia of Life

The screenshot shows the EOL website interface for the species *Opisthothylax immaculatus*. The page includes a navigation bar at the top with links for 'HOME', 'ABOUT EOL', 'LANGUAGE: US EN', and 'ABOUT EOL'. Below the navigation bar is the EOL logo and the species name 'Opisthothylax immaculatus' with the common name 'GRAY-EYED FROG'. The page features a main image of the frog on a green leaf, a sidebar with 'IMAGES', 'CHARTS', 'MAPS', 'VIDEOS', 'RSS FEED', and 'EXPERT CHAT', and a central 'OVERVIEW' section. The 'OVERVIEW' section contains a detailed description of the frog, its status, and a list of sources and references. On the right side, there is an 'EVOLUTIONARY CLASSIFICATIONS' section showing a phylogenetic tree with 'CLASS Amphibia' and 'ORDER Anura'. Below that is a 'CONTRIBUTE TO EOL' section with a list of related species and a 'RELATED ARTICLES' section with a list of species names.

Opisthothylax immaculatus
GRAY-EYED FROG

IUCN CONSERVATION STATUS: LEAST CONCERN

LAST POSTING: 04/20/2007

IMAGES | CHARTS | MAPS | VIDEOS | RSS FEED | EXPERT CHAT

EVOLUTIONARY CLASSIFICATIONS

CLASS Amphibia

ORDER Anura

CONTRIBUTE TO EOL

1. *Urosaurus maritimus* (Images)

2. *Mercops aploaster* (Content)

3. *Urosaurus maritimus* (Maps)

4. *Oryza sativa* (Charts)

5. *Kiwa hirsuta* (Images)

RELATED ARTICLES

SPECIES | HABITAT | DIET

[Xenopus laevis](#)

[Bufo marinus](#)

[Rana catesbeiana](#)

[Rana pipiens](#)

[Rana temporaria](#)

[Rana sylvatica](#)

[Amphytoma maculatum](#)

[Bufo bufo](#)

“The launch of the Encyclopedia of Life will have a profound and creative effect in science... this effort will lay out new directions for research in **Every branch of biology**”
E.O. Wilson

[MY EOL \(885\)](#) | [MY PREFERENCES](#) | [LANGUAGE, USER](#) | [ABOUT EOL](#)



Encyclopedia of Life



Hello Max!
If you are not "Max", please [click](#) or [register](#)

BIOGEO	Filter
Coastal	(109,733)
Marine	(239,888)
Polar	(12,837)

GEOGRAPHY	Filter
Arctic	(5,729)
Antarctica	(8,858)
North America	(235,549)

TYPE OF ORGANISM	Filter
Animals	(1,276,000)
Fungi	(64,000)
Plants	(265,000)

SEARCH

Life Form Environment

Ursus maritimus

POLAR BEAR

IUCN CONSERVATION STATUS: **VULNERABLE**

LAST POSTING: 04/02/2007

[IMAGES](#) | [CHARTS](#) | [MAPS](#) | [VIDEOS](#) | [PODCASTS](#) | [RSS FEED](#) | [EXPERT CHAT](#)



IMAGES

UP Fish and Wildlife Services



← BACK 1 2 NEXT →

FIELD NOTES

Polar Bear and Cubs

EVOLUTIONARY

EOL CLASSIFICATIONS



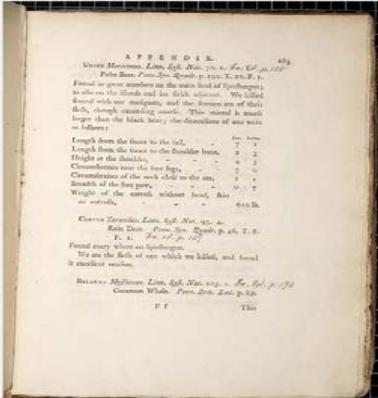
[HOME](#) | [ABOUT](#)

TABLE OF CONTENTS

- [Overview](#)
- [Research Data](#)
- [Nomenclature History](#)
- [Physiology and Biochemistry](#)
- [Specimen Information](#)
- [Genetics](#)
- [Molecular Biology](#)
- [Physical Description](#)
- [Original Description](#)
- [Geographic Range](#)
- [Habitat](#)
- [Reproduction](#)
- [Lifespan & Causes of Death](#)
- [Behavior](#)
- [Diet & Eating Habits](#)
- [Predation](#)
- [Ecosystem Roles](#)
- [Economic Relevance to Humans](#)
- [Conservation Status](#)
- [Conservation Measures](#)
- [Classification / Taxonomic Notes](#)
- [Links to Other Resources](#)
- [Biodiversity Heritage Library](#)**
- [Literature References](#)

BIODIVERSITY HERITAGE LIBRARY

STATUS: UPDATED 05/07/2007 | COMMENTS: [ORIGINAL DESCRIPTION](#)



Chicago, 1778: 185

Milgram, Constantine John Plover, Baron, 1744-1792

A Voyage Towards the North Pole Undertaken By His Majesty's Command, 1776

London: Printed by W. Benger and J. Alcock, for J. Norton ... 1776

vol. 207, (2) p. 14. Hatched before it [plate: B., map: plate 25 on, etc.]

CONTRIBUTE TO EOL

Add content to already viewed pages

- [Ursus maritimus \(Images\)](#)
- [Merops apiaster \(Context\)](#)
- [Ursus maritimus \(Images\)](#)
- [Onycha asavia \(Context\)](#)
- [Kleia hirsuta \(Images\)](#)

[MENU](#)

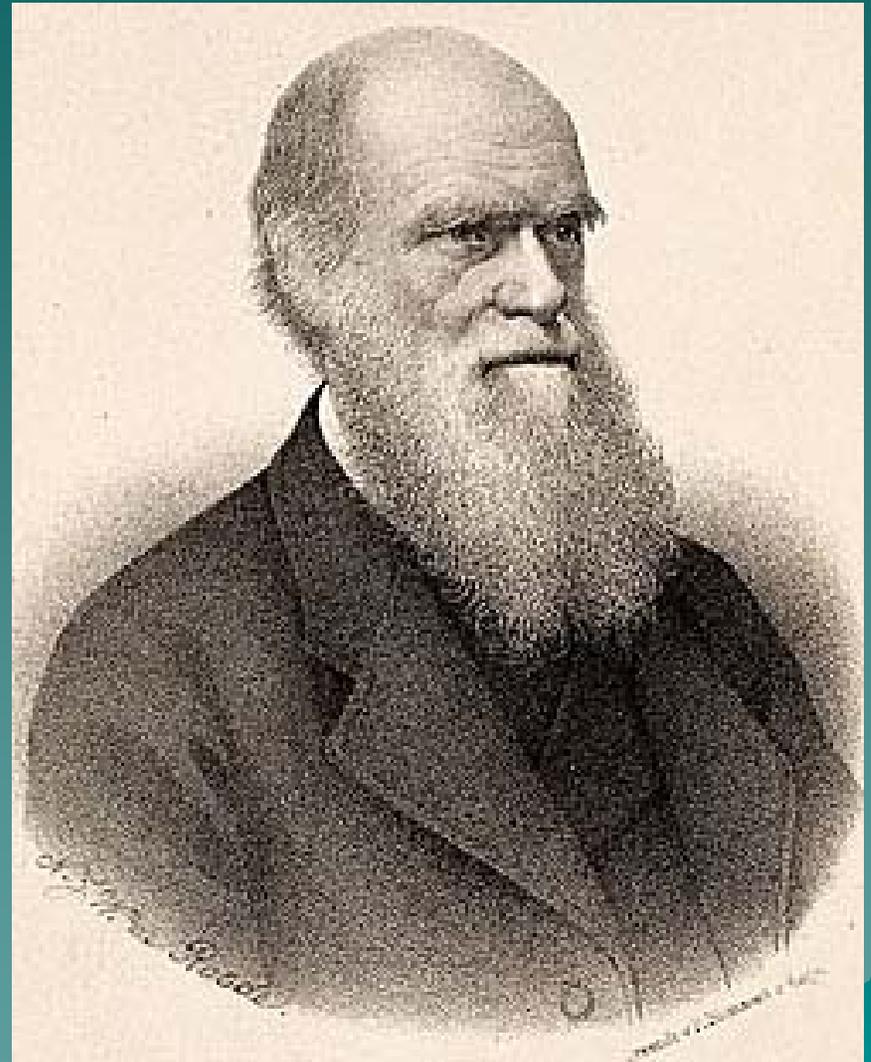
RELATED ARTICLES

SPECIES	HISTORY	DATA
Ursus americanus		
Ursus arctos		
Ursus thibetanus		
Puma hirsuta		
Cynanthus chrysatus		
Elanoides forficatus		
Odobenus rosmarus		
Alouatta palliata		

“The cultivation of natural science cannot be efficiently carried on without reference to an extensive library.”

Charles Darwin, et al
(1847)

Darwin, C. R. et al. 1847. *Copy of Memorial to the First Lord of the Treasury [Lord John Russell], respecting the Management of the British Museum. Parliamentary Papers, Accounts and Papers 1847*, paper number (268), volume XXXIV.253 (13 April): 1-3. [Complete Works of Charles Darwin Online]



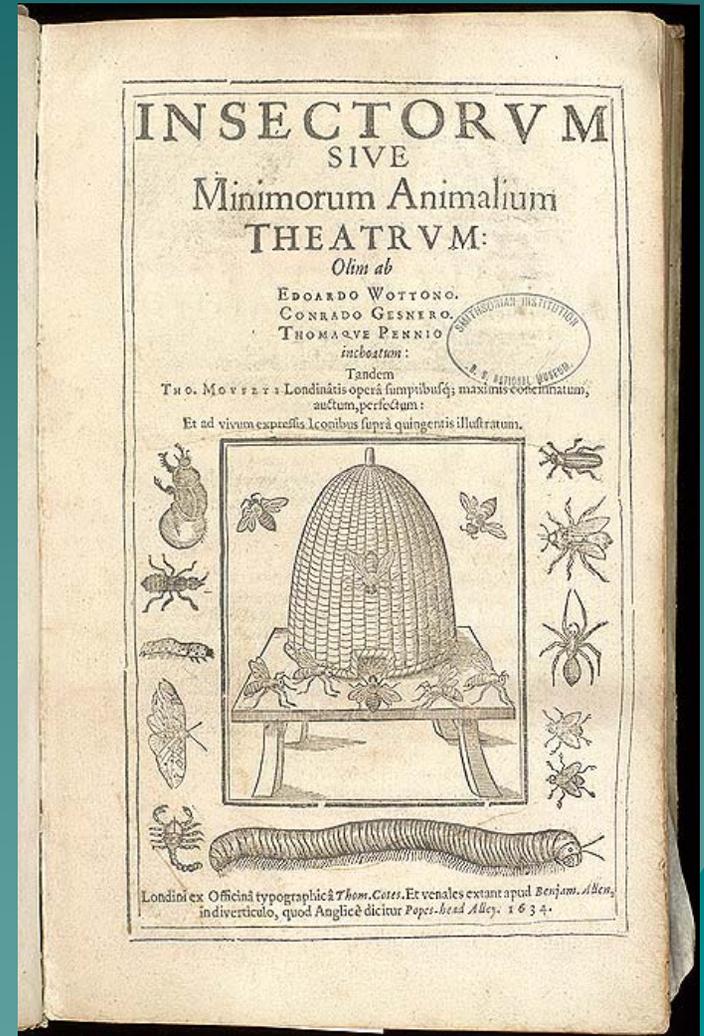
Taxonomic Literature

The cited half-life of publications in taxonomy is longer than in any other scientific discipline

* * *

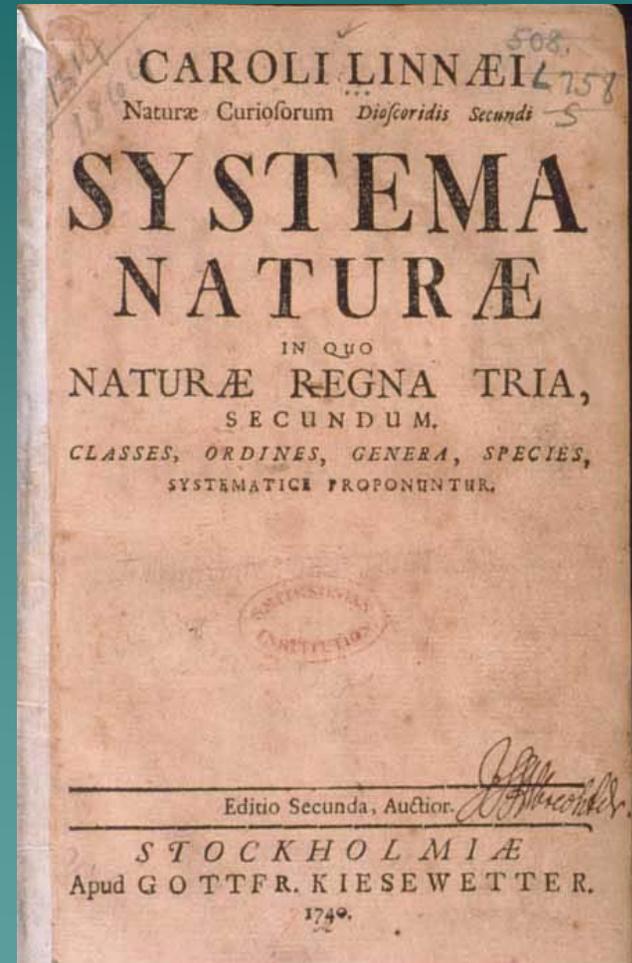
The decay rate is longer than in any scientific discipline

~ *Macro-economic case for open access*
Tom Moritz



Taxonomic Literature

- Over 250 years of systematic description of life
- *Systema naturae* (10th ed. 1758) by Carl von Linné



Taxonomic Literature

- Taxonomic descriptions must be published for the name to be valid
- Publications must be available to the public through trusted sources
- Libraries have been the traditional place





Biodiversity
Heritage
Library





Mission:

Provide Open Access to Biodiversity Literature

Goals:

Digitize the core published literature on biodiversity and put on the Web

Agree on approaches with the global taxonomic community, rights holders and others

How big is the Biodiversity domain?

- ◆ Over 5.4 million books dating back to 1469
- ◆ 800,000 monographs
- ◆ 40,000 journal titles (12,500 current)
- ◆ 50% pre-1923



BHL MEMBERS

Museums

- Field Museum (Chicago)
- Natural History Museum (London)
- Smithsonian Institution Libraries (Secretariat)
- American Museum of Natural History (New York)

Botanical Gardens

- Missouri Botanical Garden
- New York Botanical Garden
- Royal Botanic Gardens, Kew

University Libraries

- Botany Libraries, Harvard University
- Ernst Meyer Library of the Museum of Comparative Zoology
Harvard University

Research Institute Library

- Marine Biological Laboratory / Woods Hole Oceanographic
Institution Library

All signed MOU's



Other Members Coming

University of Illinois, Urbana-Champaign
(contributing member)

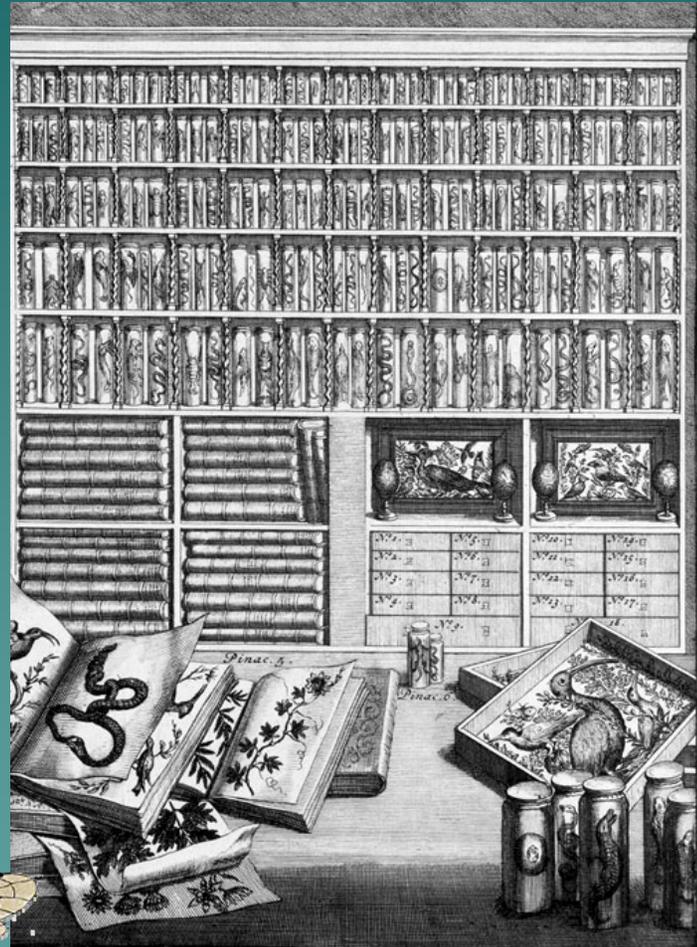


International discussions promising

- ◆ Positive discussions have already taken place with the Chinese Academy of Sciences
- ◆ Australian Government likely to fund scanning as part of Atlas of Australian Life
- ◆ EU has no funding budgets – exploration at national level in Netherlands, Germany, Spain
- ◆ Talks with Malaysia

BHL Collections

- 1.3 million catalogue records
- 73% are monographs (remainder are serials at title-level)
- 63% is English language material
- The next most popular language (9%) is German
- About 30% of material was published before 1923



Why now?



- ◆ Cost low – 10-19 cents a page
- ◆ Other projects funded recently – BL/Microsoft /Google big ten
- ◆ Tractable, well-defined scientific domain
- ◆ Taxonomic information has exceptionally longevity
- ◆ Supports GBIF and other international initiatives

Where are we now?

- ◆ Key partner of Encyclopedia of Life
- ◆ Working Groups have agreed **technical plan, metadata standards and image standards**
- ◆ **Internet Archive**

The Internet Archive

- 501(c)(3) organization
- Dedicated to “Universal Access to Human Knowledge”
- Founder of the Open Content Alliance
- Provides:
 - Mass scanning
 - Archival storage of files
 - Image processing
 - Technology development



‘Scribe’ scanners installed in NHM-
London, NYC, Boston, Washington,
Illinois



Washington, DC:

- 1 Scribe machine at Smithsonian Libraries
- 10 Scribe facility at Library of Congress with Fedlink (operational Spring 2008)



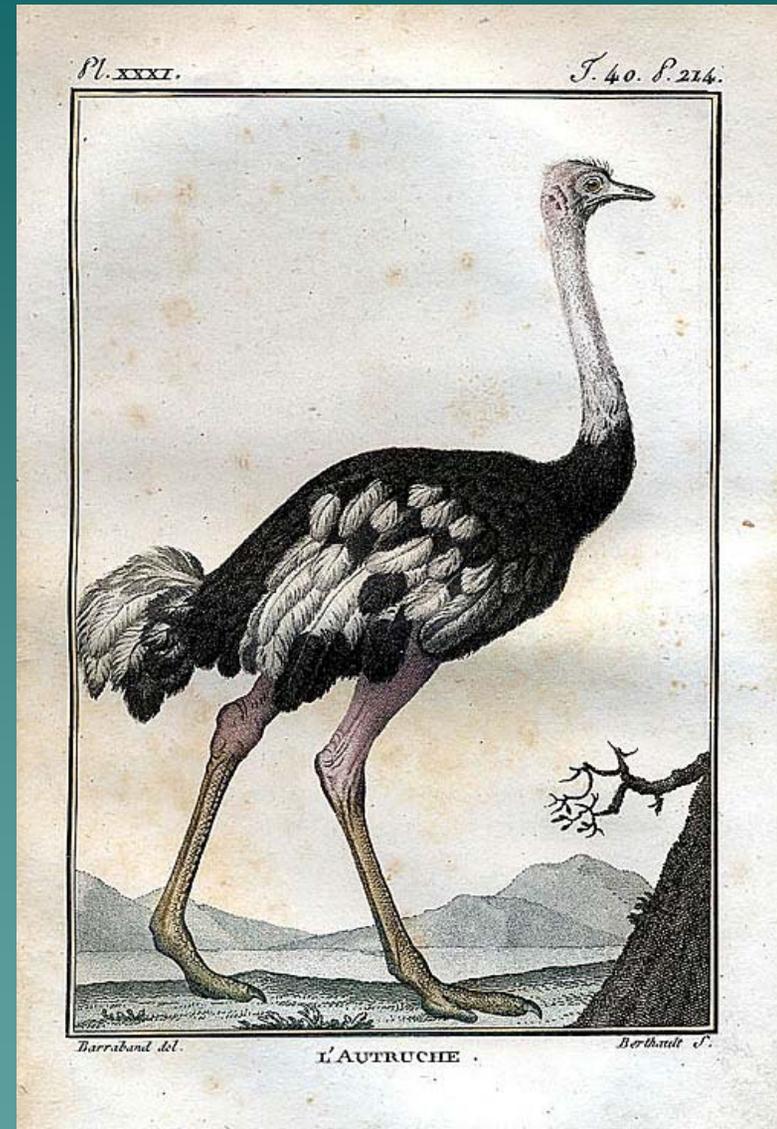
Status

- ◆ 10,000 volumes scanned
- ◆ Close to 4 million pages
- ◆ Portal up and running with 7,000 vols.



“All accumulated information of a species is tied to a scientific name, a name that serves as a link between what has been learned in the past and what we today add to the body of knowledge.”

~ Grimaldi & Engel, 2005,
Evolution of the Insects

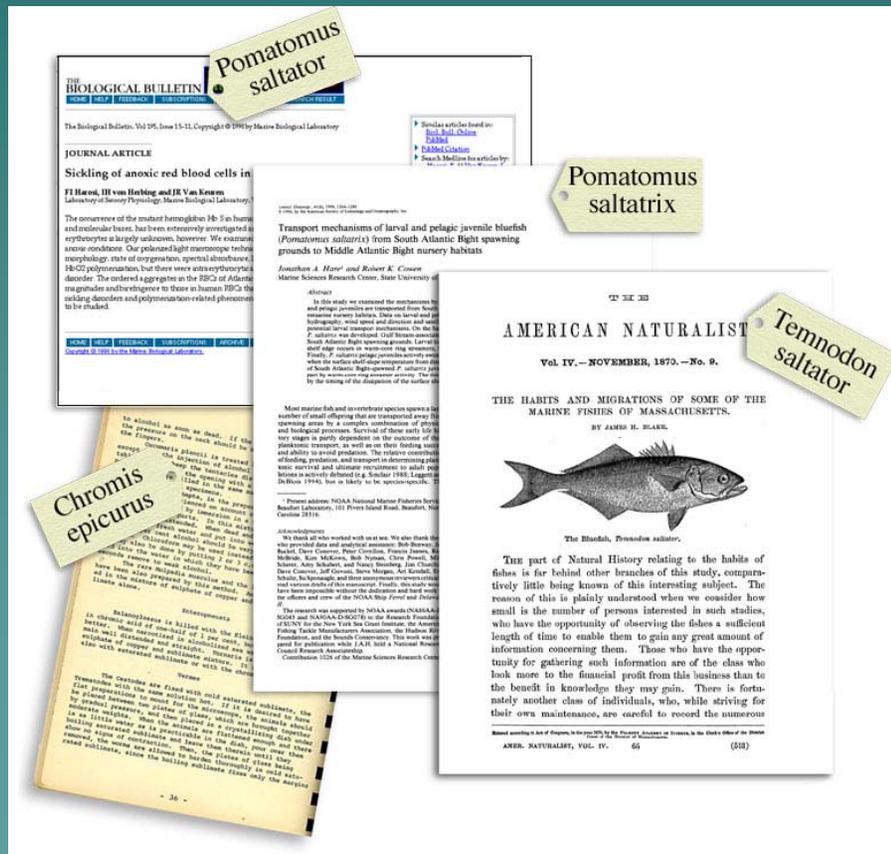


- ◆ Information about named groups (taxa) of organisms (taxon-related information)
- ◆ Extends back at least 1000 years
- ◆ Books, journals, surveys
- ◆ Museum specimens, herbaria
- ◆ In many languages and is distributed



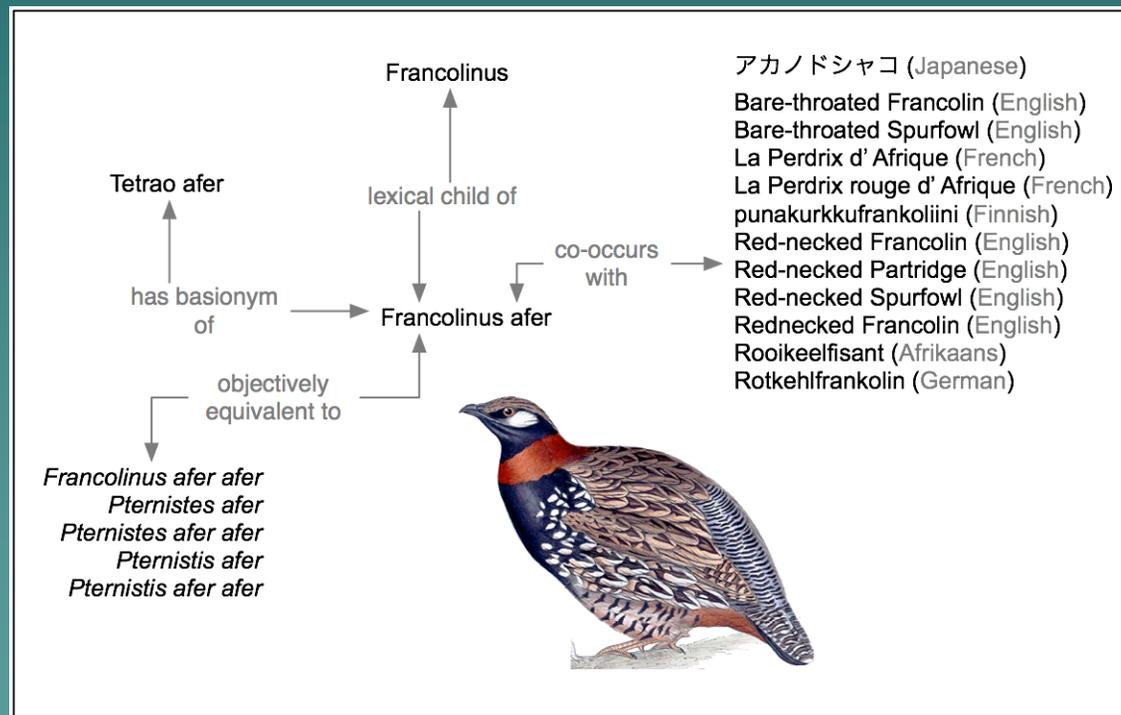
From T.E. Glover, The Fishes of Southwestern Japan, c.1870

The challenge for contemporary DIGITAL libraries



Goal:
Use one name to
find the content
for all names

Reconciliation – linking alternative names for the same organism



A query initiated with any name, can be expanded to all names and will unify data associated with each

But what about

- ◆ Difficult (impossible?) to repurpose much of the material
- ◆ Quality of images often questionable
- ◆ Sketchy / inaccurate bibliographic data

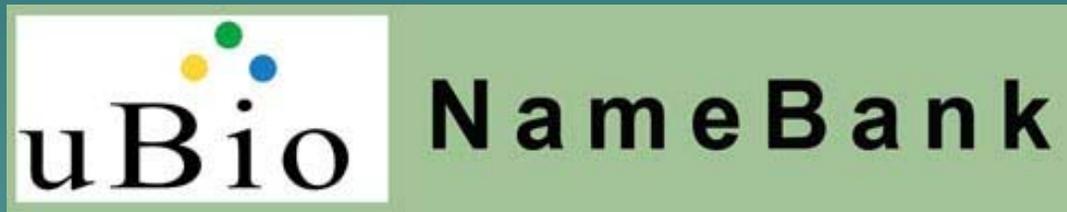


What makes this project different ?

TAXONOMIC INTELLIGENCE



Taxonomic intelligence is the inclusion of taxonomic practices, skills and knowledge within informatics services to manage information about organisms



ClassificationBank

Established at the Marine Biological
Laboratory/Woods Hole Oceanographic Institute

Taxonomic Intelligence

- ◆ 10.7 million name strings in NameBank
- ◆ Uses sophisticated algorithm (TaxonGrab) to locate likely name strings in OCR text
- ◆ Processing of BHL texts will both increase the number of name strings in NameBank and increase the accuracy of name string recognition



Now Online [Details](#)
 3,006 titles
 7,602 volumes
 2,920,990 pages
 Recent additions 

BHL Updates  [Details](#)

- [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 tad cloud?"](#)
- [Senior Programmer needed to assist BHL development](#)
- [Biodiversity Heritage Library - Europe](#)
- [Eggplant Leafroller Moth reared on Potatoes](#)
- ["An amazing service..."](#)
- [BHL Name Services v.1.0 released](#)
- ["Discovered Bibliographies" through Natural Language Processing algorithms](#)
- [Three Hundred Years of Linnaean Taxonomy](#)
- [Flora, Fauna, and Fine Books](#)
- [BHL Portal, an early review](#)
- [More BHL material online](#)
- [BHL Presentation at LITA National Forum](#)
- [BHL members attend Open Content Alliance in San Francisco](#)
- [New name finding functionality](#)
- [2 Million pages!](#)
- [BHL members gather at Missouri Botanical Garden](#)
- [Scanning starts at](#)

Show: [ALL](#) | [Top 500](#) | [Top 250](#) | [Top 100](#) | [Top 25](#)

Top 100 Subjects for 3006 Titles

19th century Africa Algae America Amphibians Anatomy *Anatomy, Comparative* Antarctica Antiquities Arachnida Arctic regions Australia Bats Beetles Belize Bibliography Biology
 Birds Borneo **Botany** Brazil Catalogs Catalogs and collections Central America Chicago China **Classification** Colombia Compositae Copper engraving Costa Rica
 Crustacea Description and travel Dictionaries Diptera Early works to 1800 Embryology England Ethnology **Evolution** Field Museum of Natural History Fisheries **Fishes** France Frogs Fungi
 Geology Germany gmppo Graphic media : Great Britain *gt* Guatemala Heredity Histology History Illinois India Indians of North America Insects Invertebrates Laboratory manuals Lizards
Mammals Marine animals Meteorites MEXICO Microscopes Microscopy Mollusks Morphology Musci **Natural history** Nomenclature **North America**
 Paleontology Parasites Pennsylvanian **Periodicals** Permian **Peru** Philippines physiology Pictorial works Plant physiology **Plants** Pre-Linnean works
 Protozoa Reptiles Science Scientific Expeditions Snakes South America Texas Uncolored United States Venezuela Vertebrates West Indies **Zoology**

<http://www.biodiversitylibrary.org/Default.aspx>

Page Delivery

http://biodiversitylibrary.org/item/bca_04_00_00

Biodiversity Heritage Library *Prototype* Search Biodiversity Heritage Library Search

Browse By: [Titles](#) | [Authors](#) | [Subject](#) | [Map](#) | [Year](#) [About](#) | [BHL Members](#) | [Copyright](#) | [Contact](#)

Biologia Centrali-Americana : [About this title](#) [Download \(63 MB PDF\)](#)

Biol. Centr. Am. *Reptilia, Tab. 42.*



Tab 21
Tab 22
Tab 23
Tab 24
Tab 25
Tab 26
Tab 27
Tab 28
Tab 29
Tab 30
Tab 31
Tab 32
Tab 33
Tab 34
Tab 35
Tab 36
Tab 37
Tab 38
Tab 39
Tab 40
Tab 41
Tab 42

[Link to this page](#) [View Text](#)

Names on this page
[SPILOTES](#)
-powered by [uBio](#)

Book contributed by [Smithsonian Institution Libraries](#)

Taxonomic Intelligence

The screenshot displays a web browser window with the URL http://biodiversitylibrary.org/item/bca_04_00_00. The page is titled "Biodiversity Heritage Library" and "Prototype". The main content area shows a page from "Biologia Centrali-Americana : Reptilia" with the page number "180". The text on the page includes the name "Epicrates, Wagler" and a list of species: "1. Epicrates", "Boa cenchria, Linnaeus, Proc. Zool. Acclimat. Soc. Lond. ii. p. 385, 1802", "Epicrates cenchria", "Epicrates maurus", "Cliftia fusca", "Epicarsius cupreus", "Epicrates cupreus", "CORALLUS", "Corallus", "Corallus cooki", "Corallus cookii", "Corallus hortulanus", "C[orallus] cooki", and "C[orallus] hortulanus".

The sidebar on the left contains a "Pages" list with "Page 180" selected. Below it, a "Names on this page" box lists the following taxa: OPHIDIA EPICRATES, Epicrates, Epicrates cenchria, Boa cenchria, Epicrates cenchria, Epicrates maurus, Cliftia fusca, Epicarsius cupreus, Epicrates cupreus, CORALLUS, Corallus, Corallus cooki, Corallus cookii, Corallus hortulanus, C[orallus] cooki, and C[orallus] hortulanus. The sidebar also includes a "Link to this page" button and a "View Text" button.

The right sidebar contains a "Library" search bar, a "Search" button, and links for "Members", "Copyright", and "Contact".

At the bottom of the page, there is a footer that reads "1856, p. 96, t. 2." and "Smithsonian Institution Libraries".

The taxonomic intelligence sidebar is highlighted with a red box and contains the following text:

Names on this page
OPHIDIA EPICRATES
[Epicrates](#)
[Epicrates cenchria](#)
[Boa cenchria](#)
[Epicrates cenchria](#)
[Epicrates maurus](#)
Cliftia fusca
Epicarsius cupreus
[Epicrates cupreus](#)
CORALLUS
Corallus
[Corallus cooki](#)
[Corallus cookii](#)
[Corallus hortulanus](#)
C[orallus] cooki
C[orallus] hortulanus
-powered by [uBio](#)

Publishers & Permissions

- Seek permissions from copyright holders of journals
- Opt in Copyright Model: The BHL will actively work with professional societies and associations to integrate their publications into the BHL in a way that serves the societies' missions and goals
- BHL will digitize learned society backfiles and mount them through the BHL Portal at no cost.
- Will provide a set of files to the publishers for reuse as they see fit



Successes

- 49 signed permissions
- *Malachologia* the most recent
- *Entomological News*
- *Journal of Hymenoptera Research*
- *Herpetological Review*
- California Academy of Sciences
- BioOne



Funding

- ◆ Initial \$3 million from John D. and Catherine T. MacArthur Foundation
- ◆ Gordon Moore Foundation
- ◆ Proposals to IMLS, NSF
- ◆ Individual members (Harvard, Smithsonian, NY Botanical Garden)

Challenges

- ◆ Experience confirms project will work
- ◆ Sustainable platform
- ◆ Ability to scan fold-outs, over-sized volumes
- ◆ Time to access pages slow
- ◆ Mirror sites
- ◆ How to represent results to users?
 - 2.9 million pages in BHL portal
 - 14.7 mill. Name occurrences using Taxon Finder
 - One search can yield 19,000 occurrences of single name

LINKS

- Biodiversity Heritage Library
<http://www.biodiversitylibrary.org/>
- Biodiversity Heritage Library Blog
<http://biodiversitylibrary.blogspot.com>
- Encyclopedia of Life
<http://www.eol.org/>
- Smithsonian Institution Libraries
<http://www.sil.si.edu/>
- Universal Biological Indexer and Organizer
<http://www.ubio.org/>
- Biologia Centrali-Americana
<http://www.sil.si.edu/digitalcollections/bca/>

