

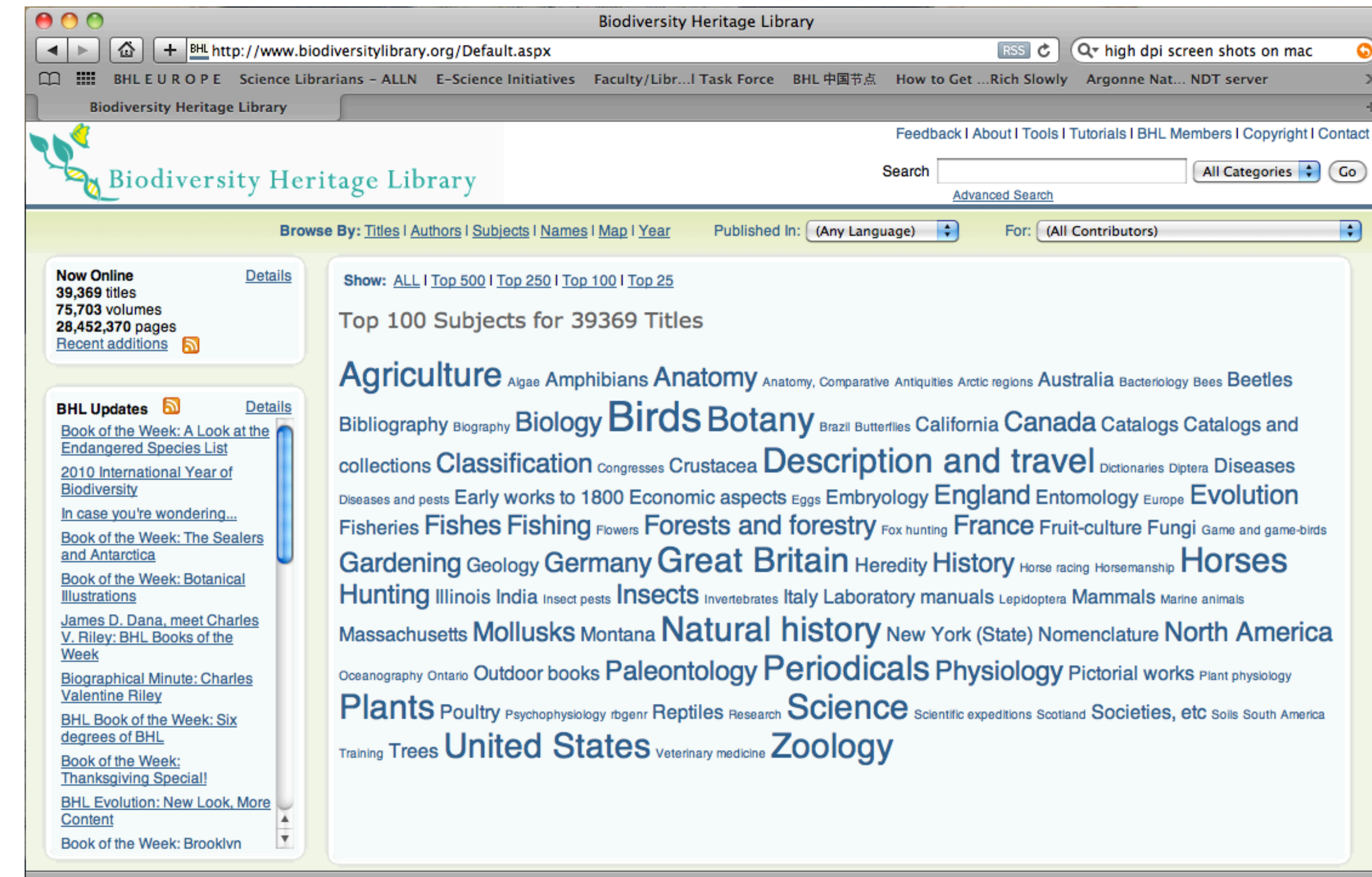


# THE GLOBAL BIODIVERSITY HERITAGE LIBRARY

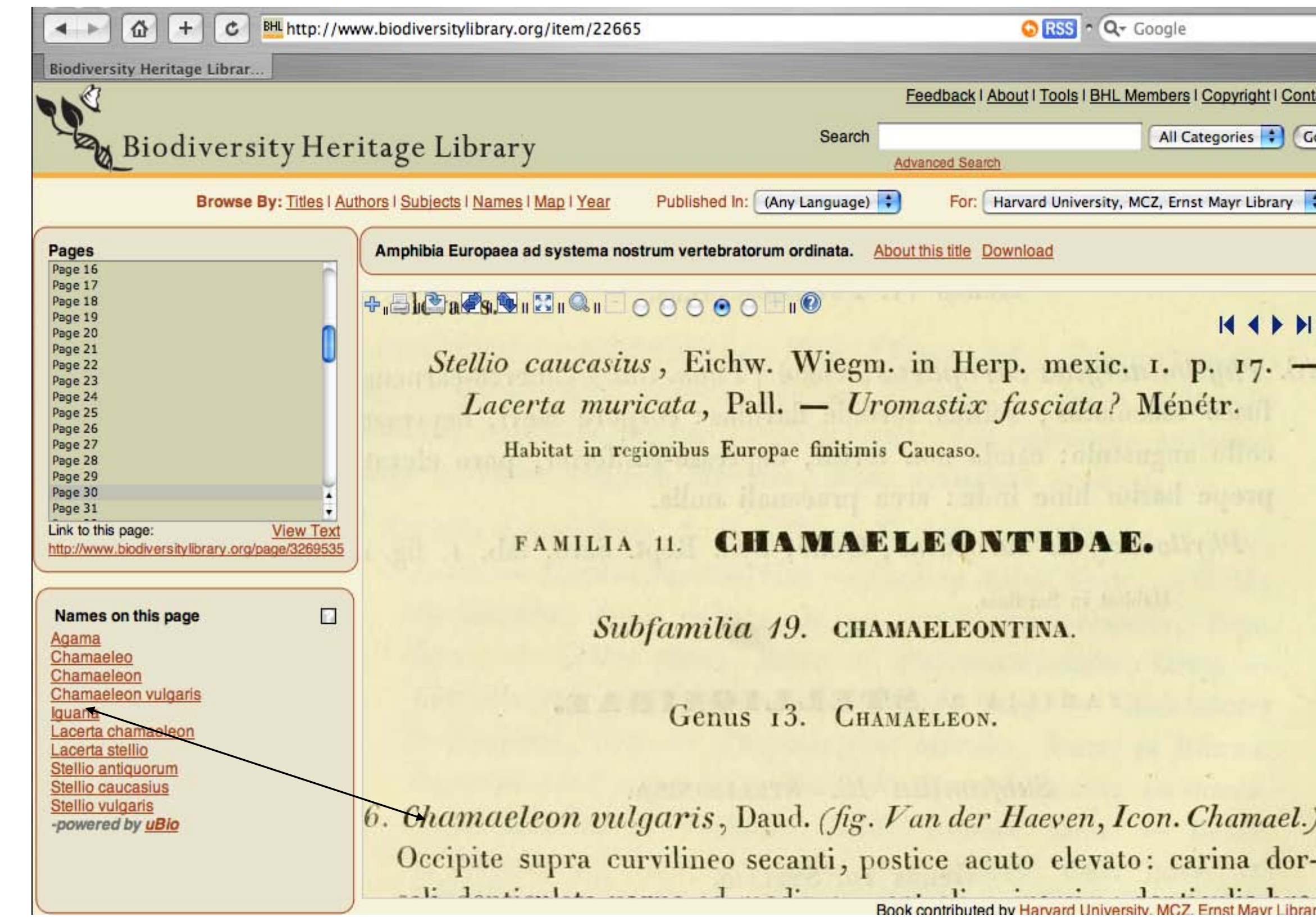
Constance Rinaldo, Ernst Mayr Library, Museum of Comparative Zoology,  
Harvard University, Cambridge, MA on behalf of the BHL



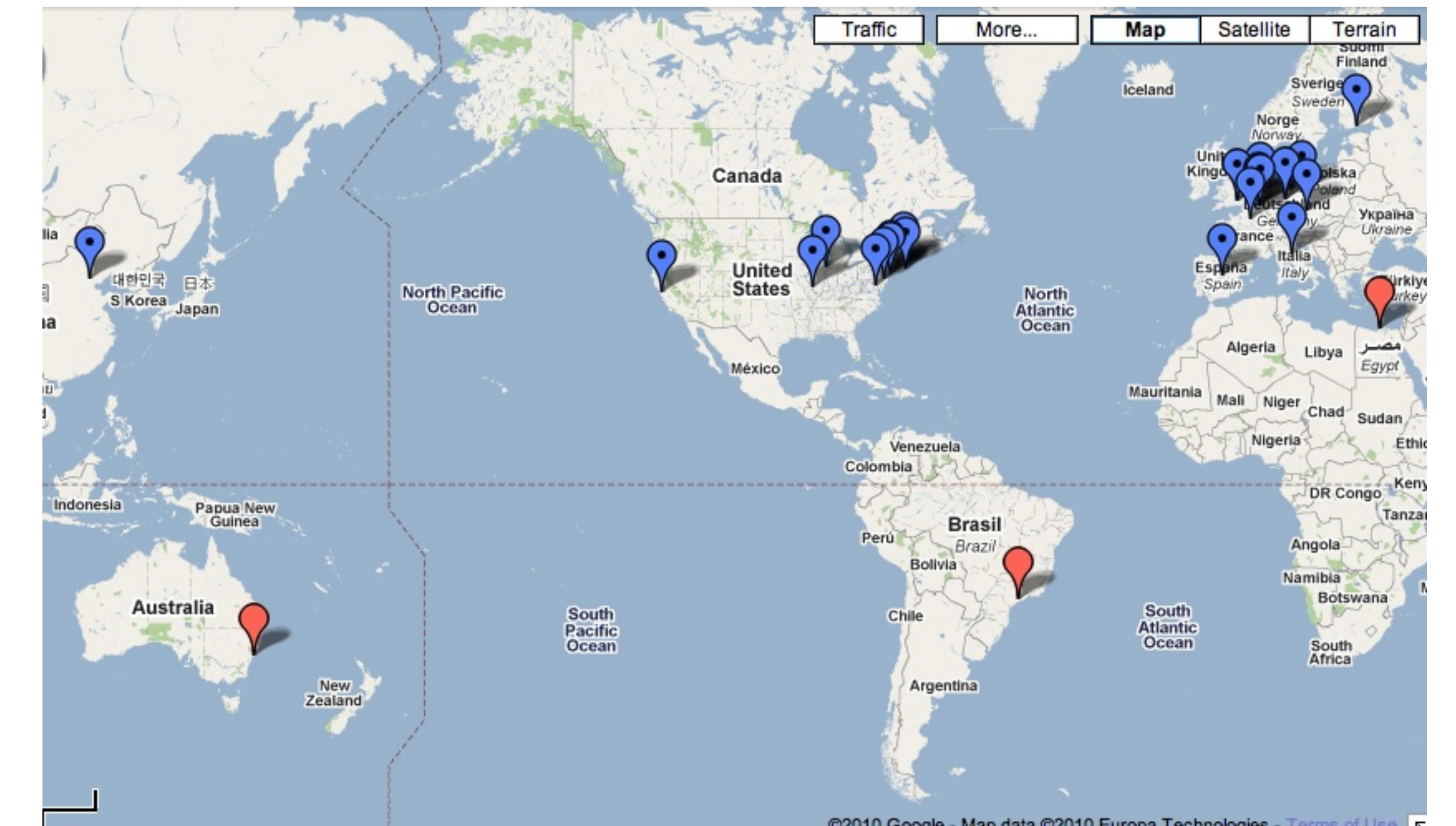
<http://www.biodiversitylibrary.org>



BHL Portal



Taxonomic intelligence in action

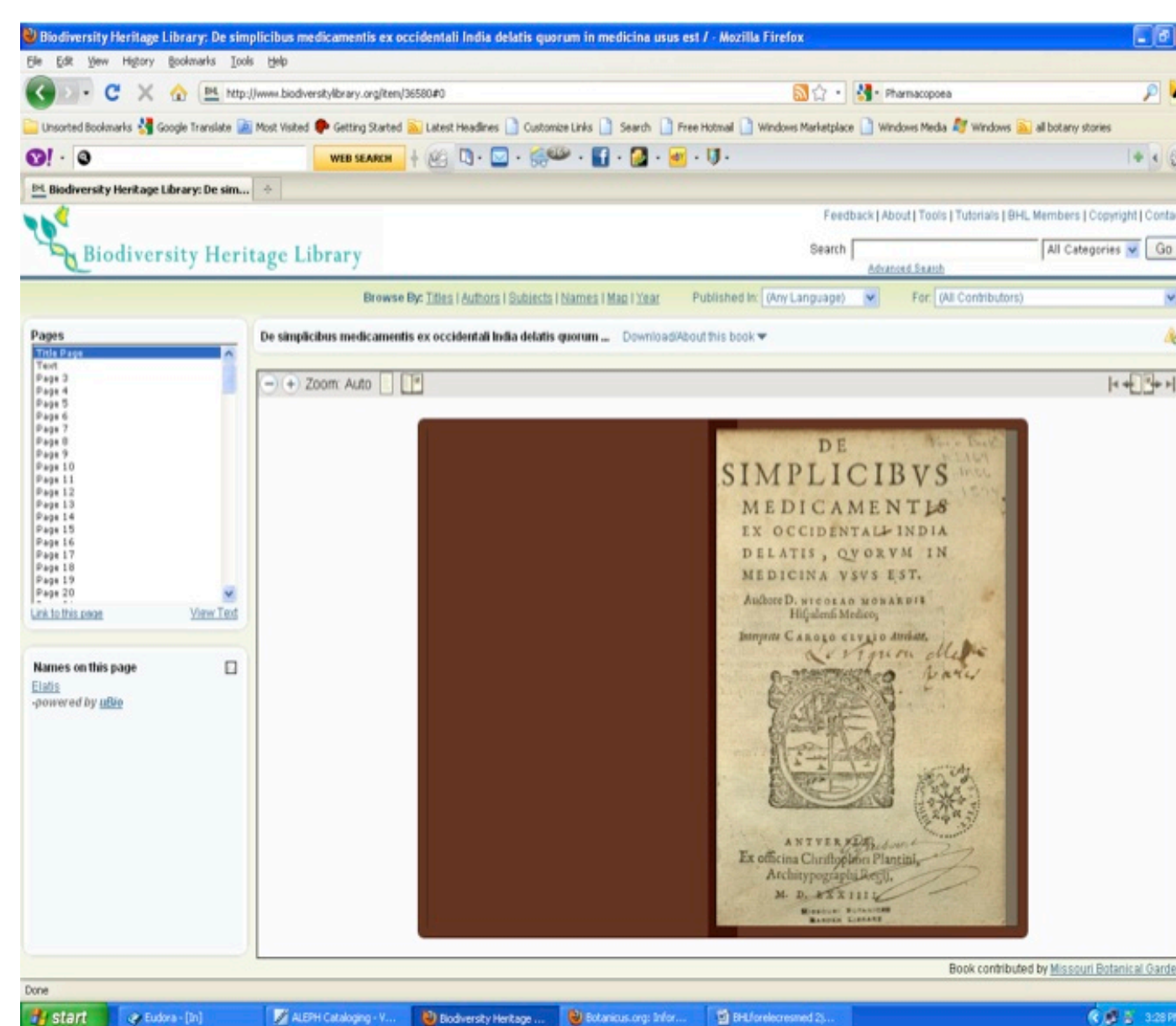


## WHAT IS THE BHL?

- 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 major natural history, botanical garden & research libraries & museums
- Collaboration with global taxonomic community: Global Biodiversity Information Facility (GBIF), International Commission on Zoological Nomenclature (ICZN), European Distributed Institute of Taxonomy, International Plant Name Index, BIONE & more

## BHL PORTAL

- Web-based entry to content and services of BHL
- Prototype developed at MOBOT as Botanicus.org & tested with scientists
- BHL Portal serves images & text files ingested from Internet Archive (IA)
- 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
  - TI uses sophisticated algorithm to locate name strings in the Optical Character Recognition (OCR) files that match the 11.1 million names in NameBank
  - Iterative processing of texts increases the number of names in NameBank & the accuracy of recognition
- PDF generator enables article-level retrieval



## WHY?

- Biodiversity studies need taxonomic data & literature
- Taxonomic data are reported in general & specialized literature that may only be 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
- Taxonomic literature is repatriated to developing countries that are rich in biodiversity but lacking research libraries

## 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

## GLOBAL BHL

- More than 22 European institutions funded by the European Union for BHL-Europe
- MOU signed with Chinese Academy of Sciences for BHL China
- MOA in process with Atlas of Living Australia
- BHL-Brazil with content from SciELO in review
- December 2009 meeting at Alexandria Library (Arabic BHL?)
- BHL Hub will contain entire scope of BHL content & services but tailored for regional-specific or language needs

## RELATED PROJECTS

- Article-level analysis of serials using automated & social tools
- BHL citation repository articles: [citebank.org](http://citebank.org)
- Incorporate multiple languages with the help of the global partners
- Linkages to molecular, morphological & other data types
- Improved OCR for non-Roman & non-standard scripts
- Medical Heritage Library through Open Knowledge Commons <https://cms.www.countway.harvard.edu/wp/?p=850>
- Expand content access & tools to new audiences
- Strengthen underlying architecture with the help of the global partners
- Ingest collections that are open access & available, including those of the global partners
- Darwin Library

## 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.
- Gwinn, N.E. & C. Rinaldo. 2008. The Biodiversity Heritage Library: sharing biodiversity literature with the world. *IFLA Journal* 35(1): 25-34.
- 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.



BHL CHINA

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 authors & do not necessarily reflect the views of funding agencies.

Please refer questions to: Connie Rinaldo, Secretary for the BHL, [crinaldo@oeb.harvard.edu](mailto:crinaldo@oeb.harvard.edu).