Using Statistical Analysis to Calculate the Size of Biodiversity Literature

Alicia Esquivel

NDSR Resident, Chicago Botanic Garden

June 6, 2017 | Inaugural Digital Data in Biodiversity Research Conference

Share your thoughts on social media using #BHLib #BHLNDSR

BHL
Biodiversity Heritage Library

Inspiring Discovery through Free Access to Biodiversity Knowledge

10 years of inspiring discovery

through free & open access

to biodiversity literature & archives

from the **15th-21st centuries**

Mission

> The Biodiversity Heritage Library improves research methodology by collaboratively making biodiversity literature openly available to the world as part of a global biodiversity community.

BHL
Biodiversity Heritage Library

NDSR

NDSR's mission is "to build a dedicated community of professionals who will advance our nation's capabilities in managing, preserving, and making accessible the digital record of human achievement."

MISSOURI BOTANICAL GARDEN

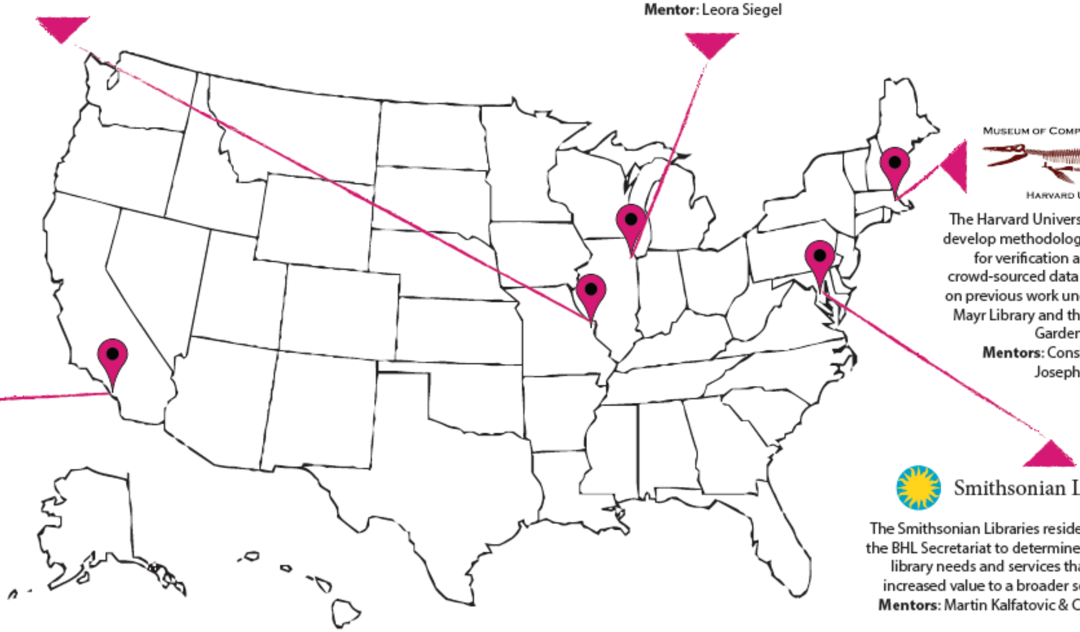
The Missouri Botanical Garden resident will explore user interface modifications to the BHL portal to enable image searching, browsing, and display.
Mentors: Douglas Holland & Trish Rose-Sandler

CHICAGO BOTANIC GARDEN

The Chicago Botanic Garden resident will address content and gap analysis, reviewing the domain of biodiversity literature underpinning the field of biodiversity, estimating the amount of that literature in the public domain, and exploring methodologies to visualize digital collections.
Mentor: Leora Siegel

NATURAL HISTORY MUSEUM
LOS ANGELES COUNTY

The Natural History Museum Los Angeles County resident will consult with BHL partners such as DPLA and Europeana to determine how BHL data works in these large-scale digital libraries and categorize high value tools and services.
Mentor: Richard Hulser



MUSEUM OF COMPARATIVE ZOOLOGY
HARVARD UNIVERSITY

The Harvard University MCZ resident will develop methodologies and propose tools for verification and integration of crowd-sourced data corrections, building on previous work undertaken by the Ernst Mayr Library and the Missouri Botanical Garden (MBG).
Mentors: Constance Rinaldo & Joseph deVeer

Smithsonian Libraries

The Smithsonian Libraries resident will work with the BHL Secretariat to determine additional digital library needs and services that will provide increased value to a broader set of BHL users.
Mentors: Martin Kalfatovic & Carolyn Sheffield

INSTITUTE of Museum and Library SERVICES

BHL
Biodiversity Heritage Library

Content Analysis

What's in BHL?

119, 682 titles
201,703 volumes
51,917,032 pages
3,732,986 species

*as of 5/12/17

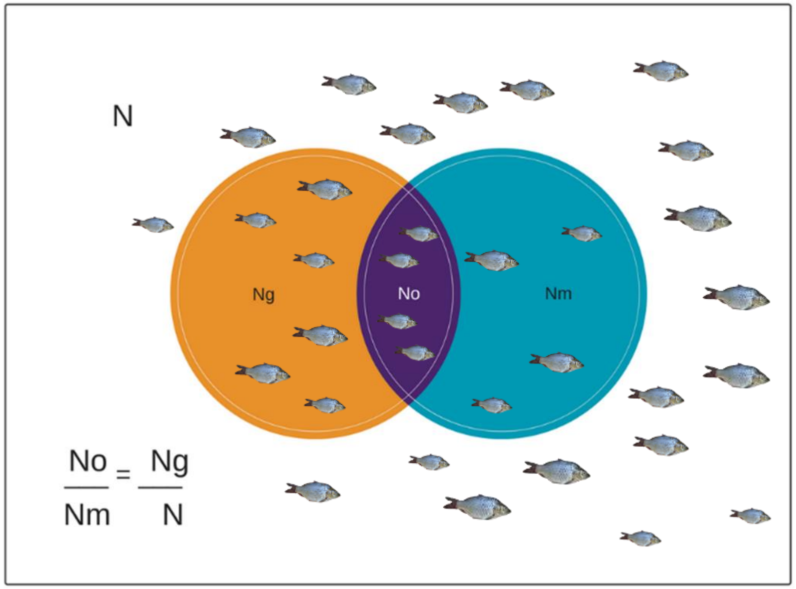
What's not in BHL?

Everything else

Biodiversity Heritage Library

Capture-Mark-Recapture (CMR)

N= unknown population
Ng= first capture
Nm= second capture
No= recaptured



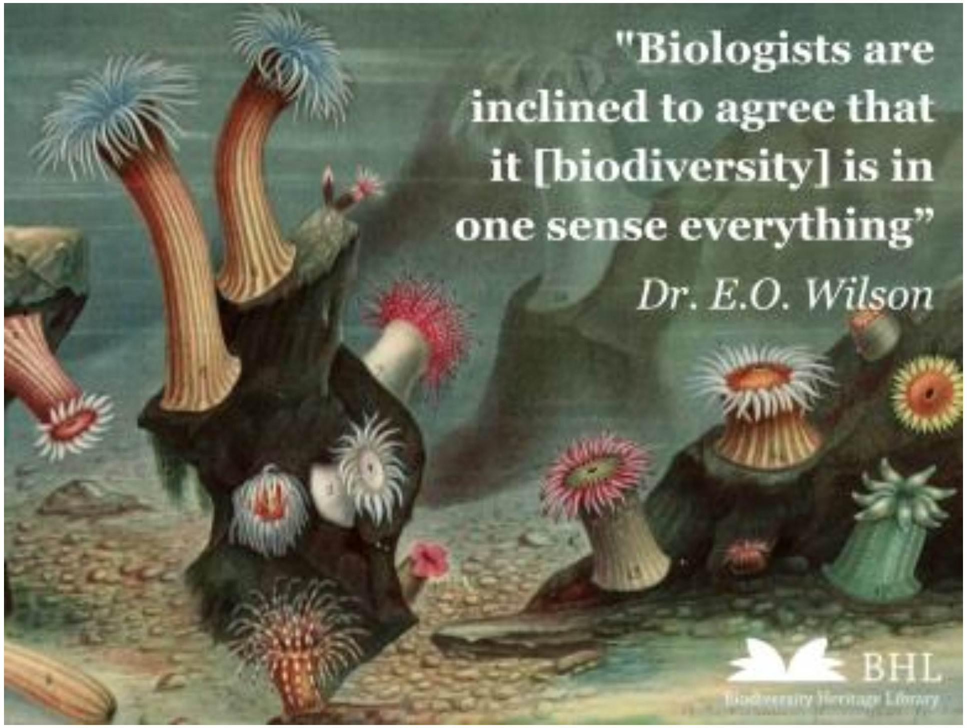
$$\frac{No}{Nm} = \frac{Ng}{N}$$

CMR in Lit Reviews

- Kastner, M., S. E. Straus, K. A. McKibbon, and C. H. Goldsmith (2009). "The Capture-mark-recapture Technique Can Be Used as a Stopping Rule When Searching in Systematic Reviews." Journal of Clinical Epidemiology.
- Lane, D., Dykeman, J., Ferri, M., Goldsmith, C., Stelfox, H. (2013). “Capture-mark-recapture as a tool for estimating the number of articles available for systematic reviews in critical care medicine.” Journal of Critical Care.
- Khabsa M, Giles CL (2014) “The Number of Scholarly Documents on the Public Web.” PLOS ONE 9(5): e93949. https://doi.org/10.1371/journal.pone.0093949
- Ariño, AH.(2010) “Approaches to estimating the universe of natural history collections data.” Biodiversity Informatics, 7: 81–92.

CMR for BHL

1. Define area of study
2. Determine how to search for organism
3. Determine which statistical model to use



"Biologists are inclined to agree that it [biodiversity] is in one sense everything”

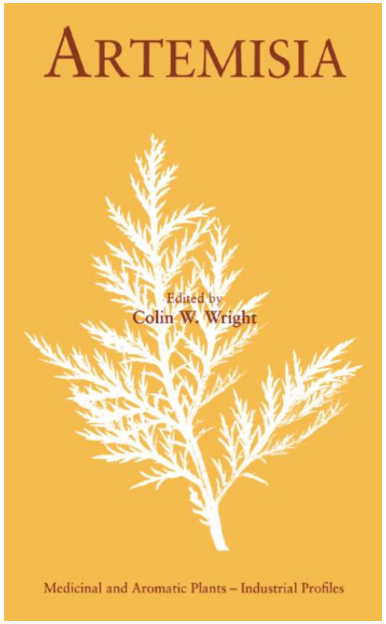
Dr. E.O. Wilson

BHL
Biodiversity Heritage Library

CMR for BHL

1. Define area of study
2. Determine how to search for organism
3. Determine which statistical model to use

Artemisia frigida © 1998 Gary A. Monroe from Encyclopedia of Life



ARTEMISIA

Edited by
Colin W. Wright

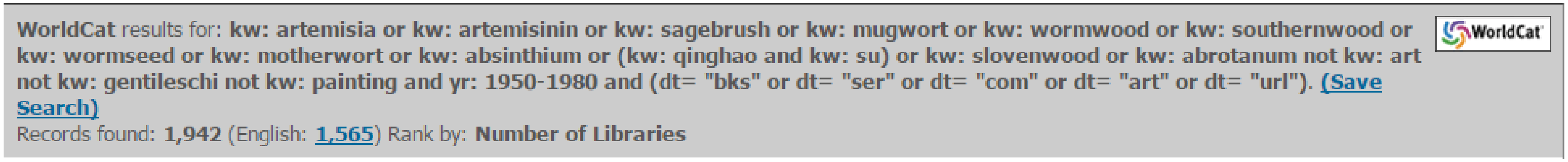
Medicinal and Aromatic Plants – Industrial Profiles

Artemisia vulgaris © 2008 Zoya Akulova from Encyclopedia of Life

BHL
Biodiversity Heritage Library

CMR for BHL

1. Define area of study
2. Determine how to search for organism
3. Determine which statistical model to use

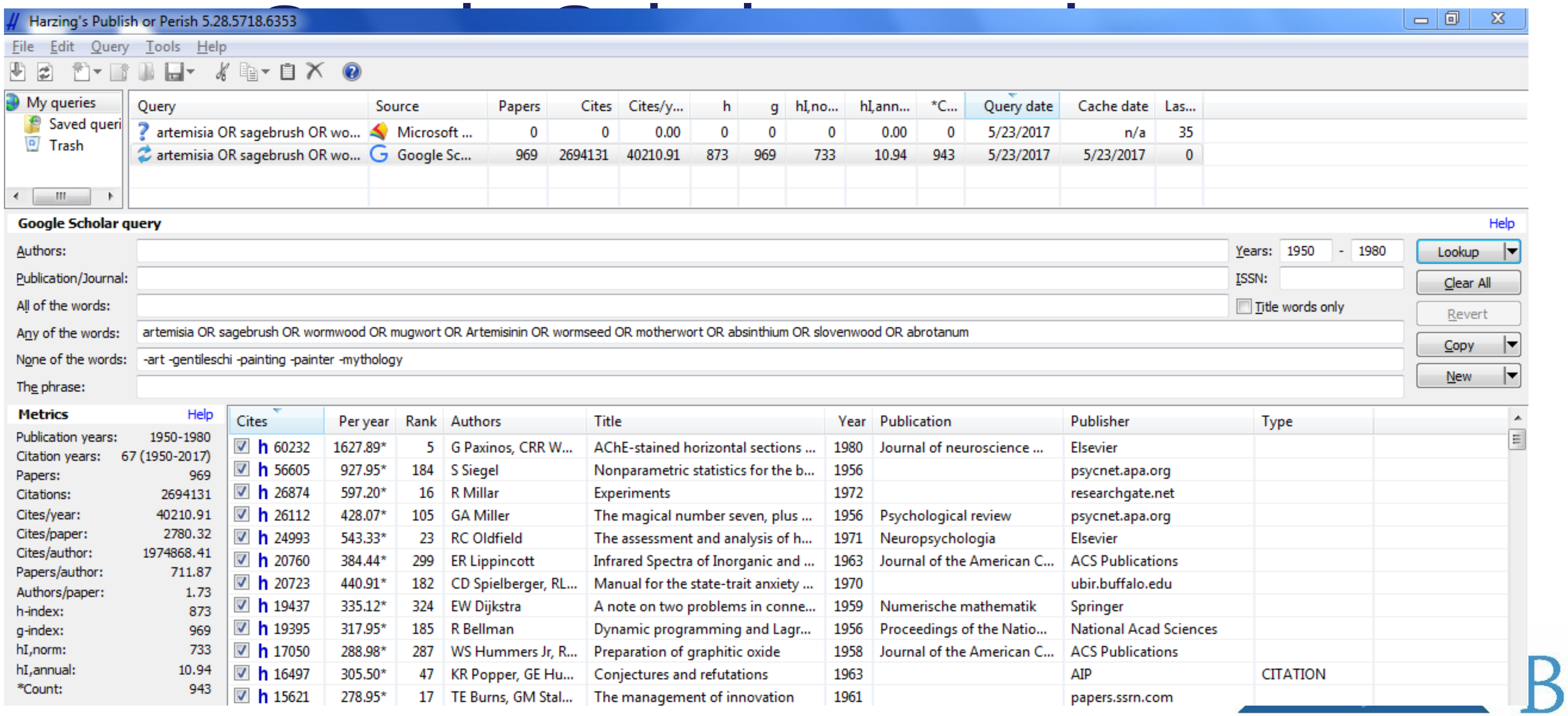


WorldCat results for: **kw: artemisia or kw: artemisinin or kw: sagebrush or kw: mugwort or kw: wormwood or kw: southernwood or kw: wormseed or kw: motherwort or kw: absinthium or (kw: qinghao and kw: su) or kw: slovenwood or kw: abrotanum not kw: art not kw: gentileschi not kw: painting and yr: 1950-1980 and (dt= "bks" or dt= "ser" or dt= "com" or dt= "art" or dt= "url"). (Save Search)**

Records found: **1,942** (English: **1,565**) Rank by: **Number of Libraries**

CMR for BHL

1. Define area of study
2. Determine how to search for organism
3. Determine which statistical model to use



Query	Source	Papers	Cites	Cites/y...	h	g	hI,no...	hI,ann...	*C...	Query date	Cache date	Las...
artemisia OR sagebrush OR wo...	Microsoft ...	0	0	0.00	0	0	0	0.00	0	5/23/2017	n/a	35
artemisia OR sagebrush OR wo...	Google Sc...	969	2694131	40210.91	873	969	733	10.94	943	5/23/2017	5/23/2017	0

Authors:

Years: 1950 - 1980

Publication/Journal:

ISSN:

All of the words:

□ Title words only

Any of the words: artemisia OR sagebrush OR wormwood OR mugwort OR Artemisinin OR wormseed OR motherwort OR absinthium OR slovenwood OR abrotanum

None of the words: -art -gentileschi -painting -painter -mythology

The phrase:

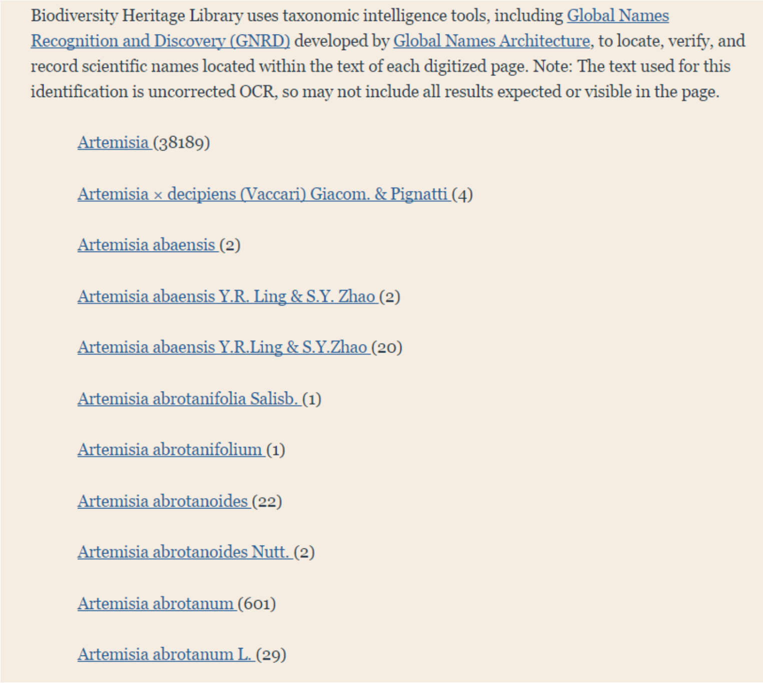
Publication years:	1950-1980
Citation years:	67 (1950-2017)
Papers:	969
Citations:	2694131
Cites/year:	40210.91
Cites/paper:	2780.32
Cites/author:	1974868.41
Papers/author:	711.87
Authors/paper:	1.73
h-index:	873
g-index:	969
hI,norm:	733
hI,annual:	10.94
*Count:	943

Cites	Per year	Rank	Authors	Title	Year	Publication	Publisher	Type
☑ h 60232	1627.89*	5	G Paxinos, CRR W...	AChE-stained horizontal sections ...	1980	Journal of neuroscience ...	Elsevier	
☑ h 56605	927.95*	184	S Siegel	Nonparametric statistics for the b...	1956		psycnet.apa.org	
☑ h 26874	597.20*	16	R Millar	Experiments	1972		researchgate.net	
☑ h 26112	428.07*	105	GA Miller	The magical number seven, plus ...	1956	Psychological review	psycnet.apa.org	
☑ h 24993	543.33*	23	RC Oldfield	The assessment and analysis of h...	1971	Neuropsychologia	Elsevier	
☑ h 20760	384.44*	299	ER Lippincott	Infrared Spectra of Inorganic and ...	1963	Journal of the American C...	ACS Publications	
☑ h 20723	440.91*	182	CD Spielberger, RL...	Manual for the state-trait anxiety ...	1970		ubir.buffalo.edu	
☑ h 19437	335.12*	324	EW Dijkstra	A note on two problems in conne...	1959	Numerische mathematik	Springer	
☑ h 19395	317.95*	185	R Bellman	Dynamic programming and Lagr...	1956	Proceedings of the Natio...	National Acad Sciences	
☑ h 17050	288.98*	287	WS Hummers Jr, R...	Preparation of graphitic oxide	1958	Journal of the American C...	ACS Publications	
☑ h 16497	305.50*	47	KR Popper, GE Hu...	Conjectures and refutations	1963		AIP	CITATION
☑ h 15621	278.95*	17	TE Burns, GM Stal...	The management of innovation	1961		papers.ssrn.com	

BHL
Biodiversity Heritage Library

CMR for BHL

1. Define area of study
2. Determine how to search for organism
3. Determine which statistical model to use



Biodiversity Heritage Library uses taxonomic intelligence tools, including Global Names Recognition and Discovery (GNRD) developed by Global Names Architecture, to locate, verify, and record scientific names located within the text of each digitized page. Note: The text used for this identification is uncorrected OCR, so may not include all results expected or visible in the page.

- Artemisia (38189)
- Artemisia × decipiens (Vaccari) Giacom. & Pignatti (4)
- Artemisia abaensis (2)
- Artemisia abaensis Y.R. Ling & S.Y. Zhao (2)
- Artemisia abaensis Y.R.Ling & S.Y.Zhao (20)
- Artemisia abrotanifolia Salisb. (1)
- Artemisia abrotanifolium (1)
- Artemisia abrotanoides (22)
- Artemisia abrotanoides Nutt. (2)
- Artemisia abrotanum (601)
- Artemisia abrotanum L. (29)

BHL
Biodiversity Heritage Library

CMR for BHL

1. Define area of study
2. Determine how to search for organism
3. Determine which statistical model to use

Compare bibliographies and apply:

- Lincoln Petersen
- Seber probabilistic
- Poisson regression

BHL
Biodiversity Heritage Library

Thank You!

Questions?

Alicia Esquivel

6/6/17 | Inaugural Digital Data in
Biodiversity Research Conference

Stay Connected with BHL!

Follow @BioDivLibrary on social media

Join our Mailing List: library.si.edu/bhl-newsletter-signup

BHL
Biodiversity Heritage Library