

# Words, words, words, you fishmonger Using Wikidata to Reconcile Taxonomies

Richard Littauer  
Te Herenga Waka Victoria University of Wellington  
<http://richard.social>

# International Commission on Zoological Nomenclature

 Search...

 All

 Case

 SEARCH

[HOME](#)   [ABOUT](#)   [THE CODE](#)   [CASES](#)   [NOMENCLURAL NOTES](#)   [OFFICIAL LISTS AND INDEXES](#)   [BULLETIN OF ZOOLOGICAL NOMENCLATURE](#)   [SUBSCRIPTION AND ACCESS](#)   [ICZN PUBLICATIONS](#)  
[ZOOBANK](#)   [LIST OF AVAILABLE NAMES](#)   [FAQS](#)   [MAKE A DONATION](#)

Home » [The Code](#) » The Code Online

[Contents](#) | [Search](#) | [Glossary](#) | [Help](#)

[Open in new window](#)

## CONTENTS

- [Home](#)
- [Explanatory Note on The Code](#)
- [Preface to the Fourth Edition](#)
- + [Introduction](#)
- [Preamble](#)
- + [Chapter 1: Zoological nomenclature](#)
- + [Chapter 2: The number of words in the scientific names of animals](#)
- + [Chapter 3: Criteria of publication](#)
- + [Chapter 4: Criteria of availability](#)
- + [Chapter 5: Date of publication](#)
- + [Chapter 6: Validity of names and nomenclatural acts](#)
- + [Chapter 7: Formation and treatment of names](#)

## INTERNATIONAL CODE OF ZOOLOGICAL NOMENCLATURE

### Fourth Edition

[Incorporating **A.** Declaration 45, Addition of Recommendations to Article 73 and of the term "specimen preserved" to the Glossary, **B.** Declaration 44, amendments of Article 74.7.3, with effect from 31 December 1999, and **C.** the Amendment on e-publication, amendments to Articles 8, 9, 10, 21 and 78, with effect from 1 January 2012]

*adopted by the  
International Union of Biological Sciences*

The provisions of this Code supersede those of the previous editions with effect from 1 January 2000



ISBN 0 85301 006 4

The author of this Code is the International Commission on Zoological Nomenclature

# International Commission on Zoological Nomenclature

 Search... All     Case**SEARCH**[HOME](#)[ABOUT](#)[THE CODE](#)[CASES](#)[NOMENCLATURAL NOTES](#)[OFFICIAL LISTS AND INDEXES](#)[BULLETIN OF ZOOLOGICAL NOMENCLATURE](#)[SUBSCRIPTION AND ACCESS](#)[ICZN PUBLICATIONS](#)[ZOOBANK](#)[LIST OF AVAILABLE NAMES](#)[FAQS](#)[MAKE A DONATION](#)[Home](#) » [The Code](#) » The Code Online[Contents](#) | [Search](#) | [Glossary](#) | [Help](#)

## CONTENTS

- [Home](#)
- [Explanatory Note on The Code](#)
- [Preface to the Fourth Edition](#)
- [Introduction](#)
- [Preamble](#)
- [Chapter 1: Zoological nomenclature](#)
- [Chapter 2: The number of words in the scientific names of animals](#)
- [Chapter 3: Criteria of publication](#)
- [Chapter 4: Criteria of availability](#)
- [Chapter 5: Date of publication](#)
- [Chapter 6: Validity of names and nomenclatural acts](#)
- [Chapter 7: Formation and treatment of names](#)

## INTERNATIONAL COD

### Fourth Edition

[Incorporating [A. Declaration 45](#), [Addition of 44](#), [amendments of Article 74.7.3](#), with effect [21](#) and [78](#), with effect from 1 January 2012]

*adopted by the  
International Union of Biological Sciences*

The provisions of this Code supersede those o



ISBN 0 85301 006 4

The author of this Code is the International Commission on Zoological Nomenclature

## "The Code":

- French and English
- 90 Articles
- Glossary
- Examples (which don't count)

# CAROLI LINNÆI

## EQUITIS DE STELLA POLARI,

ARCHIATRI REGII, MED. & BOTAN. PROFESS. UPSAL.  
ACAD. UPSAL. HOLMENS. PETROPOL. BEROL. IMPER.  
LOND. MONSPEL. TOLOS. FLORENT. SOC.

# SYSTEMA NATURÆ

PER

## REGNA TRIA NATURÆ,

## SECUNDUM

CLASSES, ORDINES,  
GENERA, SPECIES,

GUM

*CHARACTERIBUS, DIFFERENTIIS,  
SYNONYMIS, LOCIS.*

TOMUS L. Sach Ekstrand

EDITIO DECIMA, REFORMATA. *M. Wallberg*

Cum Privilegio S:æ R:æ M:tis Sveciæ

HOLMIÆ.

IMPENSIS DIRECT. LAURENTII SALVII

1758.

## Article 3: The beginning

- Starts with Linnaeus 1758, on Jan. 1.
  - But the Code doesn't specify which calendar.



## Article 11.2 Latin alphabet

*Haematopus  
unicolor*

And not:

እናወያኑኩኝ ብሔርና



## Article 11.3 Derivation

Doesn't have to be Latin.

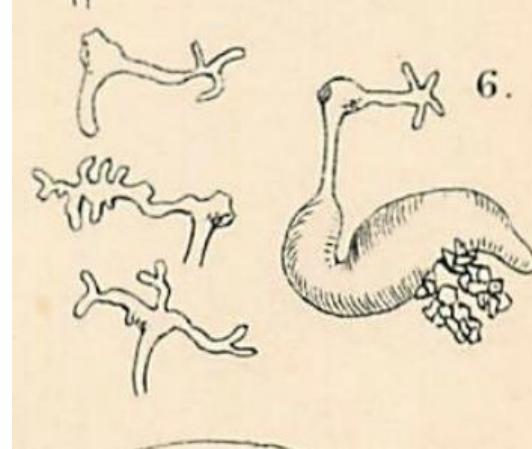
Cf. *Manu antiquus* Marples, 1946



## Article 11.9 Species names

Can be:

- an adjective in the nominative singular
- a noun in the nominative singular
- a noun in the genitive case, or
- an adjective in the genitive case representing related species



# Article 23 Principle of priority

First means first.



Gould, public domain

Huia

## ► †*Heteralocha acutirostris* (Gould)

*Neomorpha acutirostris* Gould, 1837: *Synop. Birds Australia* 1: pl. 11 & text – North Island.

*Neomorpha crassirostris* Gould, 1837: *Synop. Birds Australia* 1: pl. 11 & text – North Island.

*Neomorpha Gouldii* G.R. Gray, 1841: *List Gen. Birds* (2nd edition): 15. Unnecessary *nomen novum* for *Neomorpha acutirostris* Gould, 1837.

*Heteralocha Gouldi* (G.R. Gray); Cabanis 1851, *Mus. Heineanum* 1: 218. Unjustified emendation.

*Heteralocha gouldi* (G.R. Gray); G.R. Gray 1862, *Ibis* 4: 217. Unjustified emendation.

*Heteralocha acutirostris* (Gould); Buller 1872 (Apr.), *History of the Birds of N.Z.*, 1st edition (part 1): 63.

*Heteralocha auctirostris* (Gould); Enys 1876, *Trans. Proc. N.Z. Inst.* 8: 204. Misspelling.

## Article 30.1 Gender of genus-group names

If Latin, Latin.

If Greek, Greek.

If Greek but looks like Latin, like Latin.

If disclaimed, not Greek or Latin.

If “common or variable”, masculine.

If ends in *-ops*, masculine.



## Article 30.2 Gender of genus-group names

If in “modern European language”,  
the gender of that word.

Else, as noted by author.

Or by combination.

Or masculine;  
unless ends in *-a* (feminine);  
*-um*, *-on*, or *-u* (neuter).



## Article 31.2 Agreement

Species must agree with genus.

31.2.1: But not compound names.

31.2.2: Unclear N or Adj:

- Did the author note it?
- Is usage decisive?
- Can it be regarded as either?



# Some unclear names from NZ I've found

*Megalapteryx didinus, Euryapteryx curtus, Euryapteryx curtus curtus, Dendrocygna, Cygnus atratus, Cygnus sumnerensis chathamicus, Branta, Numida, Callipepla, Callipepla californica, Synoicus, Ptilinopus, Eudynamys, Aegotheles, Hirundapus, Hirundapus caudacutus, Hirundapus caudacutus caudacutus, Apus, Apus pacificus, Apus pacificus pacificus, Zapornia pusilla, Tribonyx hodgenorum, Haematopus, Pluvialis dominicus, Numenius hudsonicus, Arenaria, Calidris, Calidris acuminata, Calidris ferruginea, Calidris alpina, Calidris melanotos, Calidris pusilla, Phalaropus, Phalaropus fulicarius, Phalaropus lobatus, Anous albivittus, Gygis alba, Larus dominicanus dominicanus antipodum, Onychoprion fuscatus serratus, Onychoprion anaethetus melanopterus, Chlidonias leucopterus, Chlidonias hybridus, Chlidonias hybridus javanicus, Aptenodytes patagonicus, Pygoscelis papua taeniata, Diomedea, Diomedea antipodensis, Diomedea antipodensis antipodensis, Thalassarche, Thalassarche impavida, Hydrobates leucorhous, Thalassoica, Daption capense, Ardenna pacifica chlororhyncha, Pelecanoides georgicus, Pelecanoides georgicus georgicus, Morus, Sula leucogaster plotus, Bubulcus ibis coromandus, Ardea intermedia plumifera, Botaurus poiciloptilus, Aquila, Tyto alba delicatula, Athene, Eurystomus orientalis pacificus, Todiramphus, Todiramphus sanctus, Cacatua galerita, Platycercus eximius, Cyanoramphus, Cyanoramphus novaezelандiae chathamensis, Anthornis, Anthornis melanura obscura, Anthochaera carunculata, Gerygone, Gerygone albofrontata, Mohoua, Artamus cyanopterus, Gymnorhina, Petroica, Petroica macrocephala, Petroica macrocephala macrocephala, Poodytes punctatus stewartianus, Passer domesticus, Acanthis, Acanthis flammea, Anas, Oreortyx, Coturnix coromandelica, Lophura nycthemera, Lagopus, Lagopus lagopus scotica, Pluvialis apricaria, Passer montanus, Alectoris rufa*

# The Code has some issues

Here's a cute parrot again.

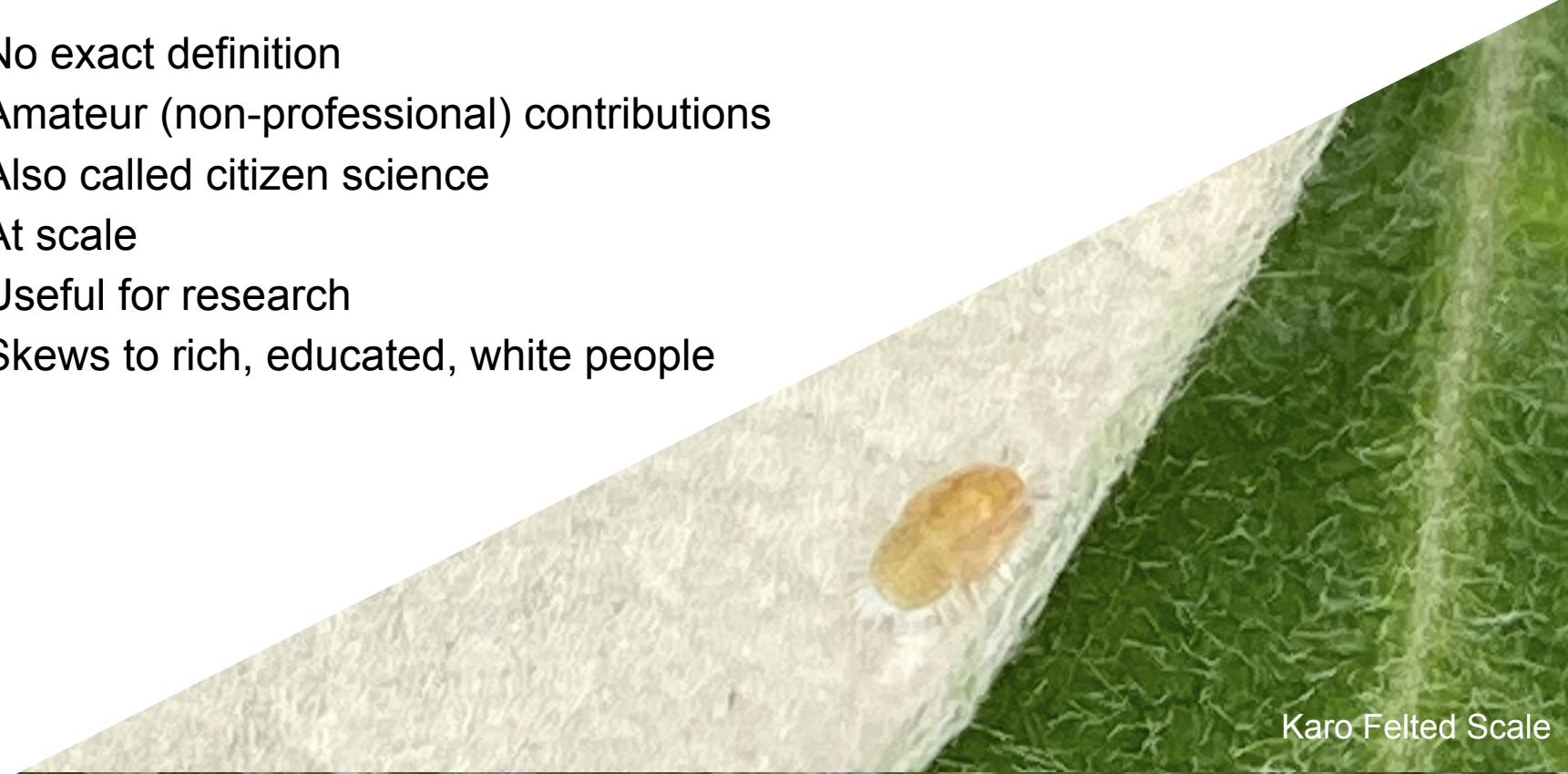
## ► *Strigops habroptila* G.R. Gray

*Strigops habroptilus* G.R. Gray, 1845: *Gen. Bird.*  
*Strigopsis habroptilus* (G.R. Gray); Bonaparte 1850  
*Strigops greyii* G.R. Gray, 1862: *Ibis* 4: 230 – So  
*Stringops (Strigops) habroptilus* (G.R. Gray); Poey 1874  
*Stringops habroptilus* G.R. Gray; Reichenow 1881  
*Stringops Greyi* G.R. Gray; Reichenow 1881, *Journ. für Ornith.*  
*Strigops habroptilus habroptilus* G.R. Gray; Mathews 1912  
*Strigops habroptilus innominatus* Mathews & Irene 1912  
*Strigops habroptilus parsonsi* Mathews & Iredale 1912  
*Strigops habroptila* G.R. Gray; Dickinson 2003,



# Community Science

- No exact definition
- Amateur (non-professional) contributions
- Also called citizen science
- At scale
- Useful for research
- Skews to rich, educated, white people



Karo Felted Scale

# SCIENCE

- Silver Fern, collected by Banks and Solander
- C. 1769-1770, Aotearoa
- Specimen in Te Papa



Ponga

# SCIENCE

- Silver Fern, photographed by Littauer
- Nov 14, 2022, Ohakune
- <https://www.inaturalist.org/observations/141999533>

Still science.



# SCIENCE

eBird New Zealand CHECKLIST S122400012

Mon 14 Nov 2022 7:06 PM

Edit date and effort

Tongariro NP -- Maungawhero Forest Walk Ruapehu District County, Manawatu-Wanganui, New Zealand

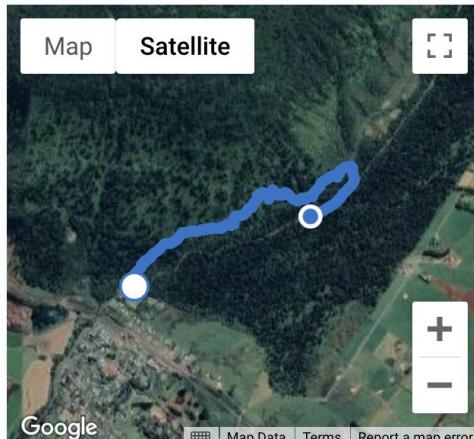
Edit location

Richard Littauer



Traveling Incomplete

2 1 hr, 3 min 1.33 mi



eBird Mobile Tracks

Submitted from eBird for iOS, version 2.13.20

Edit comments



8

Species Observed 16 individuals

1 New Zealand Pigeon *Hemiphaga novaeseelandiae*

1 Long-tailed Koel *Urodynamis taitensis*

1 Tui *Prosthemadera novaeseelandiae*

2 Gray Gerygone *Gerygone igata*

5 Tomtit *Petroica macrocephala*

3 Silvereye *Zosterops lateralis*

1 Song Thrush *Turdus philomelos* \*

2 Eurasian Blackbird *Turdus merula* \*

# eBird

- Birds.
- Different protocols for birding
- Audio and photo uploads
- For science.



Takahē

eBird Stats WORLD

10,826

Species

97,753,170

Complete Checklists

1,034,384

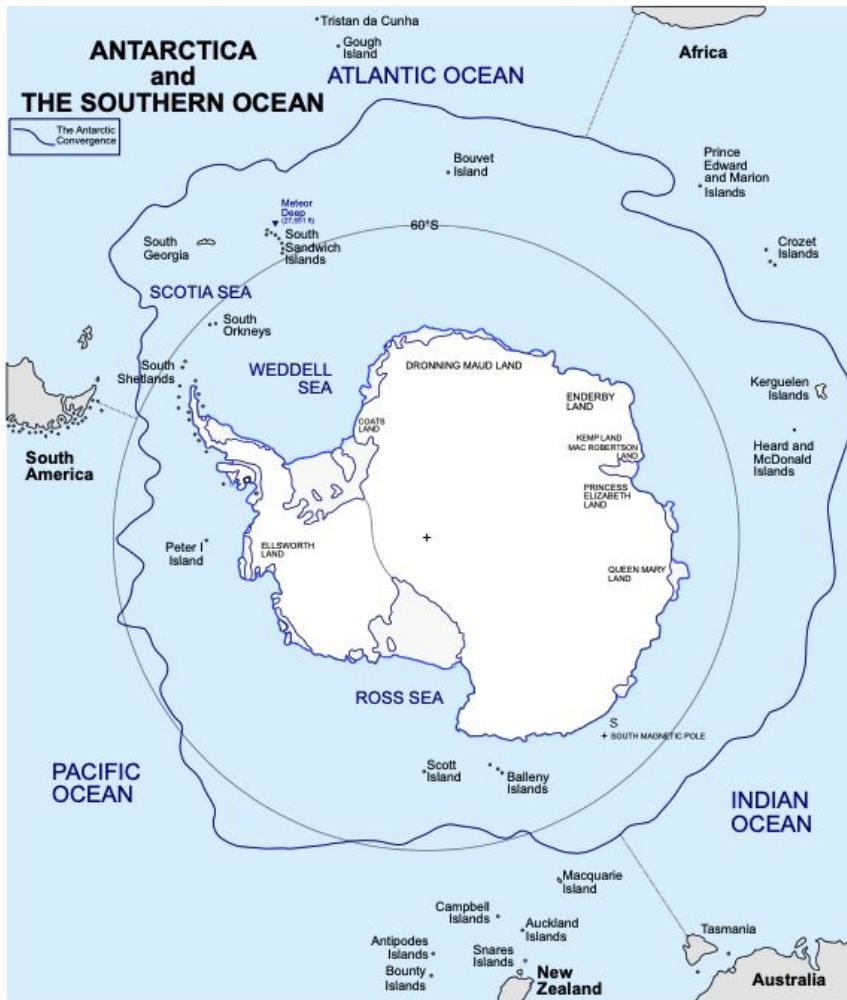
eBirders

# iNaturalist

- All momo
- Photos or audio
- "Research grade"
- Comments
- 250m observations
- 3m observers



Korimako, Manuka, & Rō



# Community science to track Bird Flu in the Subantarctic?

- Part of my PhD work
- HPAI is present in South Georgia, but not yet in New Zealand
- Let's compare data!

# Different datasets

**Table 1.** Summary of datasets used in this study. \* Auckland, Campbell, the Snares (\*\* ∩ Subantarctic Is.), and Macquarie Islands are treated as geographic subsets of the Subantarctic Islands region.

	All (Distinct)	Crozet	Falkland Is.	SGSSI	Kerguelen I.	Marion & PEI	Auckland Is.*	Campbell I.*	The Snares**	Macquarie I.*	Subantarctic Is.
Latitude	–	-46.33	-51.80	-54.42	-49.24	-46.90	-50.74	-52.55	-48.02	-54.63	-54.50
Longitude	–	51.55	-59.30	-36.56	69.45	37.77	166.12	169.15	166.60	158.70	164.00
Radius	–	500	300	500	250	500	250	250	250	250	800
<b>Observations</b>											
eBird	319,652	1,132	182,225	68,393	5,644	11,226	15,668	9,075	10,558	13,580	51,032
iNat	27,184	460	10,392	6,088	1,209	806	3,340	1,809	2,210	1,039	8,229
GBIF	6,498	1,481	405	2,495	1,640	14	51	98	30	200	463
RNBWS	2,171	19	603	1,121	9	73	80	64	80	70	346
ASOBS	215	0	0	0	0	0	0	0	0	61	215
All Records	355,720	3,092	193,625	78,097	8,502	12,119	19,139	11,046	12,878	14,950	60,285



Use OpenRefine?

# Main features



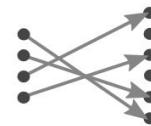
## Faceting

Drill through large datasets using facets and apply operations on filtered views of your dataset.



## Clustering

Fix inconsistencies by merging similar values thanks to powerful heuristics.



## Reconciliation

Match your dataset to external databases via reconciliation services.



## Infinite undo/redo

Rewind to any previous state of your dataset and replay your operation history on a new version of it.



## Privacy

Your data is cleaned on your machine, not in some dubious data laundering cloud.



## Wikibase

Contribute to Wikidata, the free knowledge base anyone can edit, and other Wikibase instances.

# Filtering names

OpenRefine Avilist [Permalink](#)

Facet / Filter Undo / Redo 6 / 6

Refresh Reset all Remove all

scientific\_name invert reset

genia|  case sensitive  regular expression

12 matching records (33,844 total)

Show as: rows records Show: 5 10 25 50 100 500 1000 records « first < previous

		sort_	taxon_r	protonym	order	family_na	family	famil	scientific_name	R
1	All	sort_	taxon_r	protonym	order	family_na	family	famil	scientific_name	R
2	3107.	3107	species	lotreron Eugeniae	Columbiformes	Columbidae	Pigeons, Doves	21	Pt scientific_name  ae	
3	9632.	9632	subspecies	Galbula melanogenia	Galbuliformes	Galbulidae	Jacamaras	93	Galbula ruficauda melanogenia	
4	11695.	11695	subspecies	Conurus xanthogenius	Psittaciformes	Psittacidae	African & New World Parrots	106	Eupsittula pertinax xanthogenia	
5	12098.	12098	species	Psittacula melanogenia	Psittaciformes	Psittaculidae	Old World Parrots	107	Nannopsittacus melanogenia	
6	12099.	12099	subspecies	Psittacula melanogenia	Psittaciformes	Psittaculidae	Old World Parrots	107	Nannopsittacus melanogenia melanogenia	
7	12100.	12100	subspecies	Cyclopsitta suavissima	Psittaciformes	Psittaculidae	Old World Parrots	107	Nannopsittacus melanogenia suavissimus	
8	12101.	12101	subspecies	Cydopsittacus fuscifrons	Psittaciformes	Psittaculidae	Old World Parrots	107	Nannopsittacus melanogenia fuscifrons	
9	12235.	12235	species	Eos cyanogenia	Psittaciformes	Psittaculidae	Old World Parrots	107	Trichoglossus cyanogenia	

Facet / Filter

Undo / Redo 29 / 29

12 rows

Show as: rows records Show: 5 10 25 50 rows

	All	Artist	Lifespan	Profession
1.	Anna Jóelsdóttir <input checked="" type="checkbox"/> Hübschmannov'a, Anna (11) <input checked="" type="checkbox"/> Macková, Anna (11) <input checked="" type="checkbox"/> Neborová, Anna (11) <input checked="" type="checkbox"/> Tweeddale, Anna (11) <input checked="" type="checkbox"/> Anders, Anna (11) <input checked="" type="checkbox"/> Beeck, Anna (11) <input checked="" type="checkbox"/> Heindl, Anna (11) <input checked="" type="checkbox"/> Jermolaewa, Anna (11) <input checked="" type="checkbox"/> Waser, Anna (11) <input checked="" type="checkbox"/> Matoušková, Anna (11) <input checked="" type="checkbox"/> Create new item		born 1947	contemporary artist
2.	Fridriksdóttir, Gabriela Choose new match		born 1971	painter, sculptor
3.	Helgadóttir, Gerður Choose new match		1928–1975	sculptor, stained-glass artist
4.	Gunnfríður Jónsdóttir <input checked="" type="checkbox"/> Jónsdóttir, Jóní (19) <input checked="" type="checkbox"/> Create new item	edit	1889–1968	sculptor
5.	Sveinsdóttir, Júliana Choose new match			
6.	Sigurdardóttir, Katrín Choose new match			
7.	Kristín Jónsdóttir <input checked="" type="checkbox"/> Jónsdóttir, Jóní (19) <input checked="" type="checkbox"/> Create new item			
8.	Matthiasdóttir, Louisa Choose new match			
9.	Margret the Adroit <input checked="" type="checkbox"/> Margret the Adroit (19)			

Match this Cell Match All Identical Cells Cancel

**Jónsdóttir, Jóní**  
(Icelandic artist, born 1972)

Reconciling!

## Add columns from reconciled column latin\_name

Add property

Preview

Reset

Suggested properties

- host
- hymenium type
- iNaturalist taxon ID
- main food source
- maximum viable temperature
- minimum viable temperature
- natural reservoir of
- optimum viable temperature
- parent taxon
- taxon common name
- taxon name
- taxon range map image

Western Gull	4345	Larus occidentalis	Larus
Calonectris leucomelas	4189	Calonectris leucomelas	Calonectris
Brown Pelican	4328	Pelecanus occidentalis	Pelican
Heermann's Gull	4353	Larus heermanni	Larus
Phalaropus	3957	Phalaropus	Scopidae
Calonectris leucomelas	4189	Calonectris leucomelas	Calonectris
Common Murre	4519	Uria aalge	Uria
Larus glaucescens	4399	Larus glaucescens	Larus
Tufted Puffin	4509	Fratercula cirrhata	Puffin
Western Gull	4345	Larus occidentalis	Larus
Diomedea nigripes		Diomedea nigripes	Diomedea
Bonaparte's Gull	144502	Chroicocephalus philadelphia	Chroicocephalus
Ring-billed Gull	4364	Larus delawarensis	Larus
Larus glaucescens	4399	Larus glaucescens	Larus
Sterna anaethetus		Sterna anaethetus	Sterna

Add data from Wikidata

OK

Cancel

# Automatically?

## Current script:

- Download all eBird and iNat data
- Filter for relevant species
- Hit the endpoint!

```
2  
3 # ---- Wikidata endpoints ----  
4 WIKIDATA_SPARQL_URL = "https://query.wikidata.org/sparql"  
5 WIKIDATA_RECONCILE_URL = "https://wikidata.reconcil.link/en/api"  
6 USER_AGENT = "EBirdINatWikidataReconcile/0.1 (richard@burntfen.com)"  
7  
8
```

## Wikidata reconciliation for OpenRefine

This web service can be used to align datasets to Wikidata items in OpenRefine.

Use the following URL in OpenRefine: <https://wikidata.reconcil.link/en/api>.

Replacing "en" by another language code will display items and properties in your language, when they are available.

This interface works with OpenRefine from 2.6 rc2 onwards. It is not compatible with Google Refine.

- Documentation of the protocol
- Documentation of this application for developers and Wikibase admins
- Source
- Bug tracker
- Service status

## Tutorials for OpenRefine users

- [Official documentation in English](#)
- [Official documentation in Japanese by 山本 泰智](#)
- [Tutorial by Antoine Courtin \(in French\)](#)
- [Comparison of reconciliation services for OpenRefine by Aaron Tay \(in English\)](#)
- [Using OpenRefine to Reconcile Name Entities by Karen H. \(in English\)](#)
- [Reconciling in OpenRefine - eine Einführung von histHub \(in German\)](#)

## Videos

- [Reconciliation in OpenRefine](#), screencast by Owen Stephens (in English, 24:06)
- [OpenRefine demo at WikidataCon 2017](#), live demo by Arthur P. Smith (in English, 19:52)
- [Exemple de réconciliation et d'enrichissement avec Wikidata](#) (silent screencast, 3:27) by Ettore Rizza
- [Wikidata editing with OpenRefine](#), screencast by Antonin Delneuch (in English, 3 videos of 5 minutes each)

# Automatically!

## Potential cross-dataset matches (same QID, different names)

### Q1242812 — Double-banded Plover

scientific_name	source	match_method	wikidata_rank	normalized_name	score
Anarhynchus bicinctus	both	fuzzy	species	Anarhynchus bicinctus	100.000
Anarhynchus bicinctus	only_inat	fuzzy	species	Anarhynchus bicinctus	81.000

### Q42326 — Rock Dove

scientific_name	source	match_method	wikidata_rank	normalized_name	score
Columba livia	only_inat	exact	species	Columba livia	
Columba livia (Feral Pigeon)	only_ebird	fuzzy	species	Columba livia (Feral Pigeon)	67.000

→ [Seqg12023.py compare\\_species.py](#)

Found 150 unique scientific names from NZ-WGN in eBird

Found 157 unique scientific names from iNaturalist

Comparison summary (before Wikidata, normalized names):

eBird (NZ-WGN) unique names (normalized): 150

iNat unique names (normalized): 157

In both (normalized): 105

Only eBird (normalized): 45

Only iNat (normalized): 52

Wikidata looked up 10/202 names

Wikidata looked up 20/202 names

Wikidata looked up 30/202 names

Wikidata looked up 40/202 names

Wikidata looked up 50/202 names

Wikidata looked up 60/202 names

Wikidata looked up 70/202 names

Wikidata looked up 80/202 names

Wikidata looked up 90/202 names

Wikidata looked up 100/202 names

Wikidata looked up 110/202 names

Wikidata looked up 120/202 names

Wikidata looked up 130/202 names

Wikidata looked up 140/202 names

Wikidata looked up 150/202 names

Wikidata looked up 160/202 names

Wikidata looked up 170/202 names

Wikidata looked up 180/202 names

Wikidata looked up 190/202 names

Wikidata looked up 200/202 names

Wikidata looked up 202/202 names

Wikidata reconciliation summary:

QIDs found (Overall): 181

eBird entries lacking QIDs: 19

iNat entries lacking QIDs: 3

Final CSV with Wikidata reconciliation written to: [ebird\\_inat\\_wikidata.csv](#)

Markdown review file written to: [wikidata\\_review.md](#)

# Automatically!?

- Still got some kinks
  - Doesn't deal well with hybrids - these aren't in Wikidata
  - Depends upon Wikidata being good
  - Could use Avibase?
  - What about a database of names in use at the time of the data in use?

But it's a start!



## Welcome to Avibase

Avibase is an extensive database information system about all birds of the world, containing over 54 million records about 10,000 species and 22,000 subspecies of birds, including distribution information for 20,000 regions, taxonomy, synonyms in several languages and more. This site is managed by Denis Lepage and hosted by Birds Canada, the Canadian copartner of Birdlife International. Avibase has been a work in progress since 1992 and I am now pleased to offer it as a service to the bird-watching and scientific community.

# Thank you.

Follow along? <https://www.burntfen.com/blog>

Also <https://mastodon.social/@richlitt> and  
<https://bsky.app/profile/richlitt.bsky.social> and <http://richard.social>

