

# Advances in Biodiversity Discovery and Data Quality Improvement

The VertNet Team  
Vertebrate Digitization Workshop  
Ithaca, NY  
May 3-7, 2015



# What is VertNet?

1. **Community** of Data Publishers and Users
2. **Cloud-based platform** for access to millions of global biodiversity *occurrence* records (specimens, observations, media)
3. **Services** for publishing data, improving data quality, tracking issues and feedback, and reporting usage statistics

Fast, easy, scalable, and cheap!

# Simple Darwin Core

*The standard for sharing data*

169 fields...

<i>Who</i>	Collectors, Preparators
<i>What</i>	Taxonomy, Preparations, Types
<i>Where</i>	Geography, Coordinates
<i>When</i>	Dates
<i>About</i>	Sex, Age, etc.

# Our Community: The Vertebrate Networks



*Est. 1999, 2004*

Mammal  
Networked  
Information  
System

**MaNIS**

*Est. 2001*

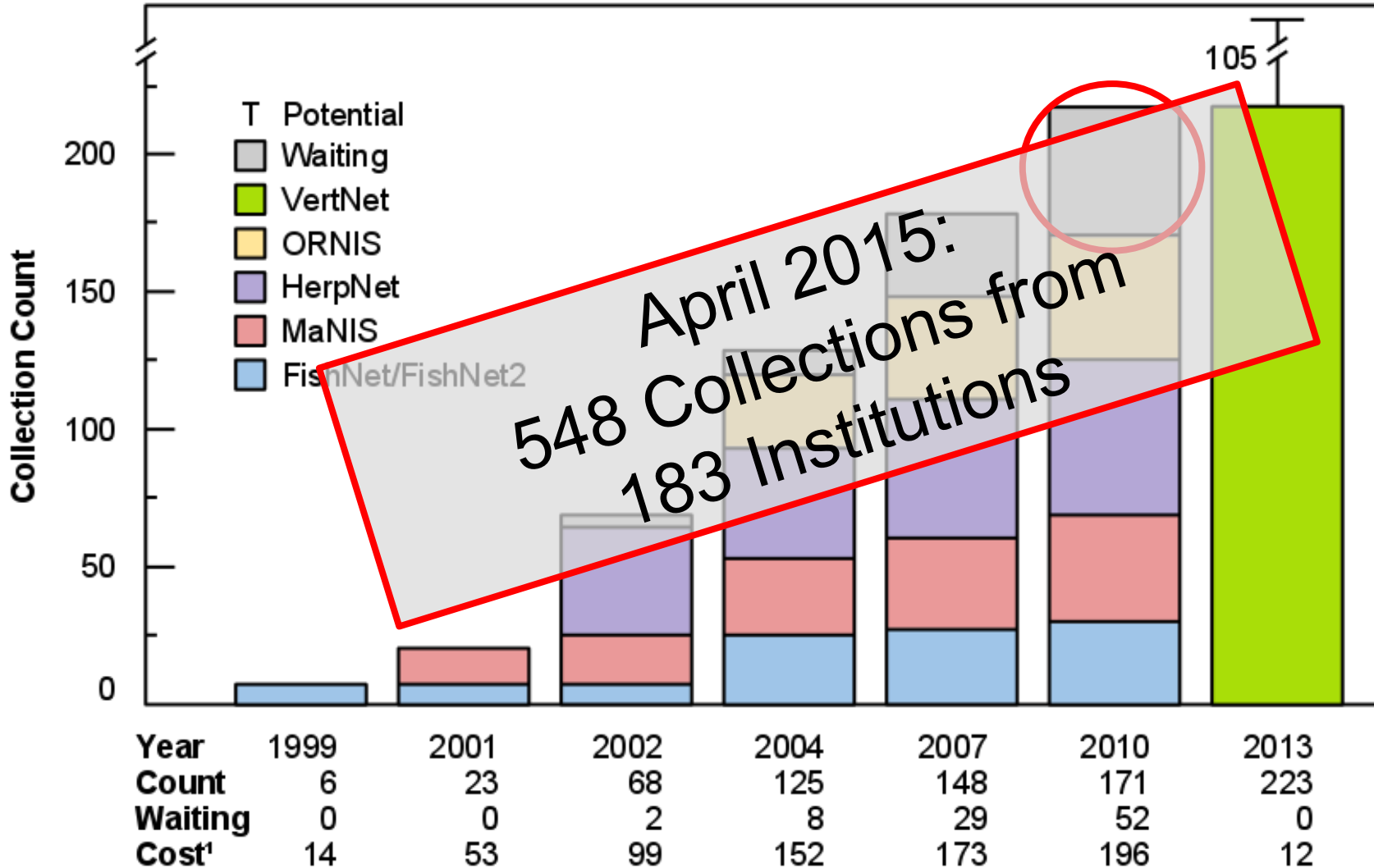


*Est. 2002*



*Est. 2004*

# Vertebrate Network Growth



<sup>1</sup> Total Operating Cost per Year in USD (x1000)

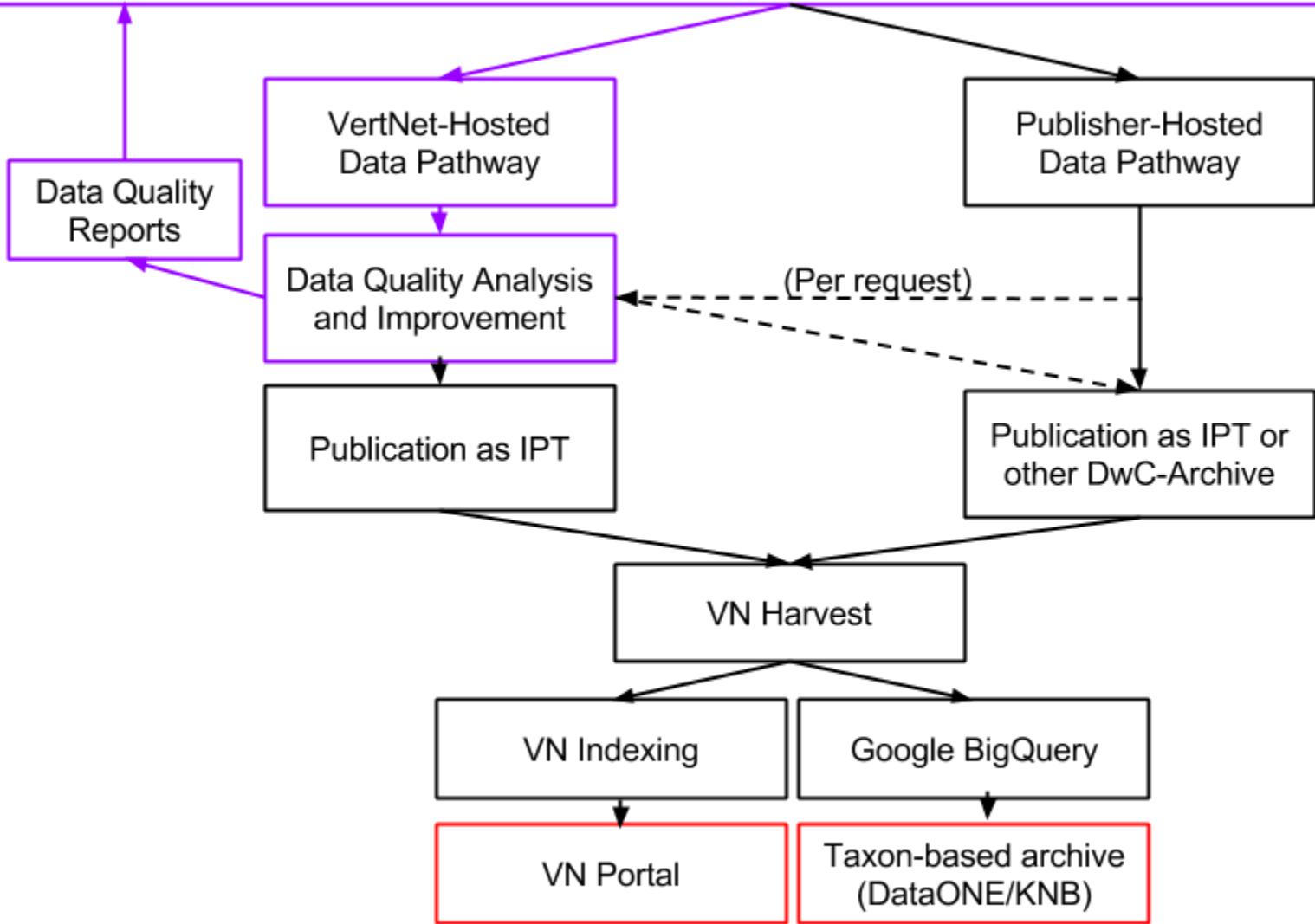
# Integrated Publishing Toolkit (IPT)

Free open source software tool used to  
publish and share biodiversity datasets

*Darwin Core Archives of Data  
Collection Metadata*

<http://gbif.org/ipt>  
<http://ipt.vertnet.org>

Source databases/collections  
(Arctos, Specify, KEmu, Brahms, Symbiota, Filemaker, Excel, etc...)



Source databases/collections  
(Arctos, Specify, KEmu, Brahms, Symbiota, Filemaker, Excel, etc...)

VertNet-Hosted  
Data Pathway

Publisher-Hosted  
Data Pathway

Data Quality Analysis  
and Improvement

(Per request)

Publication as IPT

Publication as IPT or  
other DwC-Archive

VN Harvest

Other aggregators (e.  
g., GBIF, iDigBio)

Issue Tracking/  
Feedback/GitHub

VN Indexing

Google BigQuery

VN Portal

Taxon-based archive  
(DataONE/KNB)

Data Users



# Data Published through VertNet

- 166,254,927 million occurrence records – including 150M eBird records – published to GBIF (32%) and harvested by BISON (68%)
- 17,423,087 records from 265 collections in VertNet Portal
- 9,283,225+ records published to iDigBio (34%)
- 827,615 non-vertebrate records hosted on the VertNet IPT



# The VertNet Data Portal



# It begins with a word...



## Search VertNet

[Need Help?](#) →

full text record search...



[Table](#) [Map](#)

0 results

[Download](#)

...or two or three...



# Focus your search...



## Search VertNet

[Need Help?](#)

Find records with... ×

These filters  Is fossil  Has tissues  Has media  Has type status  Is mappable  In circle on map

**All these words**

**Exact phrase**

**Any of these words**

**None of these words**

**Record type**

**Darwin Core terms**

<input type="text" value="InstitutionCode"/>	<input type="text" value="CatalogNumber"/>	<input type="text" value="RecordedBy"/>
<input type="text" value="Class"/>	<input type="text" value="Genus"/>	<input type="text" value="SpecificEpithet"/>
<input type="text" value="Country"/>	<input type="text" value="StateProvince"/>	<input type="text" value="County"/>

**Years ranging from**  to



# Still confused? We can help!

[Search](#)[Publishers](#)[Join](#)[Resources](#)[About](#)[Feedback](#)[Login](#)

Search VertNet

[Need Help?](#)

full text record search...

[More search options](#)



When in doubt...

## RESOURCES

### HELP & HOW TO

This is where you'll find a variety of guides, tutorials, and FAQs for publishing data and using the **VertNet Portal**. Don't forget to check **Training Workshops** and **Publications & Videos** for other descriptive and educational materials. We haven't forgotten about **developers** either. If all else fails, **contact us** with your questions.

[Publishers' Help](#)

[Portal Help](#)

[Portal FAQ](#)

**A VertNet Guide to...**  


[GitHub Reference & Set Up for Data Issue Tracking](#)

**A VertNet Guide to...**  


[Copyright and Licenses for Dataset Publication](#)

**A VertNet Guide to...**  


[Usage Reporting for Data Publishers](#)

# Results



## Search VertNet

Need Help? →

Callipepla californica

More search options



Table

Map

1-100 of 4,159

Download


Pro tip! Click on a table row to see the full occurrence details.





Identification	Taxonomy	Location	Year	Map	Media
CRCM Birds 98-3639	Aves: Callipepla californica	United States, Washington, Whitman: Pullman, Whitman Co., Washington	1996	📍	
DMNH Bird 27839	Aves: Callipepla californica brunnescens	United States, California, Santa Clara: Isabel Valley	1972	📍	
DMNH Bird 35208	Aves: Callipepla californica	United States, California, Santa Cruz: Watsonville, Rancho San Andreas	1973	📍	
DMNH Bird 44329	Aves: Callipepla californica brunnescens	United States, California, Santa Clara: Isabel Valley	1974	📍	
DMNH Bird 44330	Aves: Callipepla californica	United States, California, Santa Clara: Isabel Valley	1974	📍	



# Results – with Media

The **Cornell** Lab of Ornithology  **Macaulay Library** [Donate Now](#)

Search recordings by catalog number or species:

All Media   

[Home](#) [Find](#) [Order](#) [Guides](#) [About](#) | [Sign in](#)

ML 107045



[LINKS AND EMBED »](#)



[Add to Bin](#)

[Add to Cart](#)

California Quail Background Species (2)

## California Quail -- *Callipepla californica*

[BROWSE SPECIES »](#) [MORE VIDEO »](#) [MORE AUDIO »](#)

Age/Sex	Identification	Solicitation	Behavior	Note
1 Adult Male	100%	Natural	call	
	Sight and Sound		advertise	

# Results Detail – Darwin Core

## DMNH Bird 27839

Darwin Core

Rights

Spatial qua



Google

### Term

InstitutionID

CollectionID

InstitutionCode

CollectionCode

BasisOfRecord

OccurrenceID

DMNH

Bird

PreservedSpecimen

a51c2e17-c591-11e3-acd2-03413f76937f

Darwin Core

Rights

All Terms

RecordLevel

Occurrence

Event

Location

Identification

Taxon

Google

Submit data issue



Map Satellite

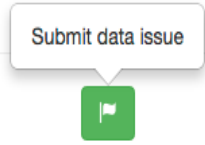


Map data ©2014 Google, INEGI Terms of Use



# Results Detail – Rights

## DMNH Bird 27839 *Callipepla californica brunnescens*



Darwin Core **Rights** Spatial quality Index

### Rights

Delaware Museum of Natural History (DMNH) data records may be used by individual researchers or research groups, but they may not be repackaged, resold, or redistributed in any form without the express written consent of a curatorial staff member of the DMNH. If any of these records are used in an analysis or report, the provenance of the original data must be acknowledged and the DMNH notified. The DMNH and its staff are not responsible for damages, injury or loss due to the use of these data.

### Citation

Delaware Museum of Natural History (DMNH)

### Organisation

Delaware Museum of Natural History

### Resource

DMNH Birds

### Contact

Jean L. Woods, Ph.D. <[jwoods@delmnh.org](mailto:jwoods@delmnh.org)>

### Published

Thu Sep 11 00:00:00 UTC 2014

### Source URL

[http://ipt.vertnet.org:8080/ipt/resource.do?r=dmnh\\_birds](http://ipt.vertnet.org:8080/ipt/resource.do?r=dmnh_birds)

# Other Ways to Access VertNet Data

VertNet IPT: Darwin Core Archives

<http://ipt.vertnet.org/>

VertNet API (Application Program Interface)

<https://github.com/VertNet/webapp/wiki/Introduction-to-the-VertNet-API>

rVertNet: R wrapper to the VertNet API

<http://cran.r-project.org/web/packages/rvertnet/index.html>

DataONE/KNB (Knowledge Network for Biocomplexity): Data Repository

Taxon-based Archives of Fish, Birds, Mammals, Reptiles, Amphibians

<https://knb.ecoinformatics.org>

# Data Quality: Cleaning the Community's Dirty Laundry



# Sex



## Distinct Values:

“Male”	189
“Female”	184
Unresolved	331

From subset of 2.7M records

# Preparations



Distinct preparations: 10,712

Distinct terms: 2,969

Distinct tissues: 26 (tissue, liver, higado, +t,...)

“vodka!00,” “surf,” “stst01”

From subset of 2.7M records

# VertNet Data Migrators

- Toolkit for working with VertNet data
- Customized to each original data source
- Converts original data into Darwin Core for upload to Integrated Publishing Toolkit resource
- Improves consistency and completeness (e.g., classification, geography, sex...)
- Provides reports back to original data source

# Sample Migration - Country

Australia	AU	1	1
'Australia'	AU	0	1
Australia (South ?)	AU	0	1
Australia			
	<b>A</b>	<b>B</b>	<b>C</b>
	<b>D</b>	<b>E</b>	
Australia	1	catalogNumber	verbatimCountry
Australia	2	A-35860	USA
Australia	3	A-35861	USA
'Australia'	4	R-10424	Belgian Congo
Australia	5	A-16728	USA
Australia	6	R-160643	USA
Australia	7	A-33990	USA
Australia	8	A-37236	USA
Australia	9	A-37237	USA
Australia	10	A-37652	USA
Australia	11	A-33765	[Tanzania]
Australia	12	A-33766	[Tanzania]
Australia	13	A-33767	[Tanzania]
Australia	14	A-33768	[Tanzania]
Australia	15	A-33769	[Tanzania]
Australia	16	A-33770	[Tanzania]
Australia	17	A-33771	[Tanzania]
Australia	18	A-15292	USA
Australia	19	A-34524	USA
Australia	20	A-34525	USA
Australia	21	A-33543	[Tanzania]
Australia	22	R-32878	E Madagascar
Australia	23	R-32568	USA
Australia	24	A-37400	USA

# Filter for "Tissues"



## Search VertNet

[Need Help?](#)

Find records with...

These filters

- Is fossil  **Has tissues**  Has media  Has type status  Is mappable  In circle on map

All these words

mvz grinnell

Exact phrase

"butorides virescens"

Any of these words

tissue OR preserved

None of these words

NOT bufo NOT california

Record type

Any type

Darwin Core terms

InstitutionCode

CatalogNumber

RecordedBy

Class

Genus

SpecificEpithet

Country

StateProvince

County

Years ranging from

1888

to

1922





# Results Detail – Spatial Quality

Darwin Core 

Rights

Spatial quality

Index



Warning: Some validations could not be performed. Check below.

## Data completeness

Are coordinates present?	✓	Yes
Is the country value present?	✓	Yes
Are both coordinates 0 (zero)?	✓	No
Do coordinates have three or more decimal figures?	✓	Yes
Do coordinates have datum?	✓	Yes

## Data inconsistencies

Are coordinates within specified country? <sup>1</sup>	✓	Yes
Distance outside of specified country (in degrees) <sup>1</sup>	✓	0
Distance outside of species range map (in degrees) <sup>1</sup>	⊘	Could not be assessed

## Data Errors

Is latitude between 90 and -90?	✓	Yes
Is longitude between 180 and -180?	✓	Yes
Are coordinates transposed? <sup>1</sup>	✓	No
Is latitude hemisphere correct? <sup>1</sup>	✓	Yes
Is longitude hemisphere correct? <sup>1</sup>	✓	Yes

<sup>1</sup>Assessed with [Map Of Life validation tools](#)

# Our Community...

## Challenges

- Knowledge gaps
- Varied experiences
- Highly variable data
- Limited resources

## Solutions

- Communication
- Training
- Issue Tracking
- Licensing
- Behaving!



# VertNet “Help Desk” Services



Meet 'em where they are

# Report for Institution Code MVZ, Resource mvz\_bird

Usage stats for March, 2015, generated on 2015/04/02

Number of searches that retrieved data from the resource: 348

Number of records retrieved in searches: 6081

List of queries that retrieved data from the resource:

Query	Times
Dinemelli	1
lamprotornis caudatus	1
Chordeiles minor	1
melanerpes formicivorus	1
ramphastos vitellius	1
Cotinga	1
catalognumber:1500000000	1
type:specimen specimen	1
pipra	1
Trogon rufus	2
"Ficedula narcissus"	2
mallard	1
Phegornis mitchellii	1
collin county	1
colinus nigrogularis	1
country:"Trinidad"	1
class:aves recorder	1
colinus cristatus leucotis year:1950 year:2015	1
stateprovince:"alaska" specific epithet:acuminata class:aves	1
country:"Panama" (passeriformes OR picidae OR cuculiformes)	1
tricholaema	1
Redstart in institutions and names	1

## Cumulative values for this year

Number of search events that retrieved data from the resource: 3371

Total number of records retrieved from searches: 72381

Number of download events that retrieved data from the resource: 183

Total number of records downloaded: 68902

## Cumulative values since the beginning

Number of search events that retrieved data from the resource: 3964

Total number of records retrieved from searches: 82935

Number of download events that retrieved data from the resource: 665

Total number of records downloaded: 489735

# Training



**40+ Workshops**

**VertNet Guides**

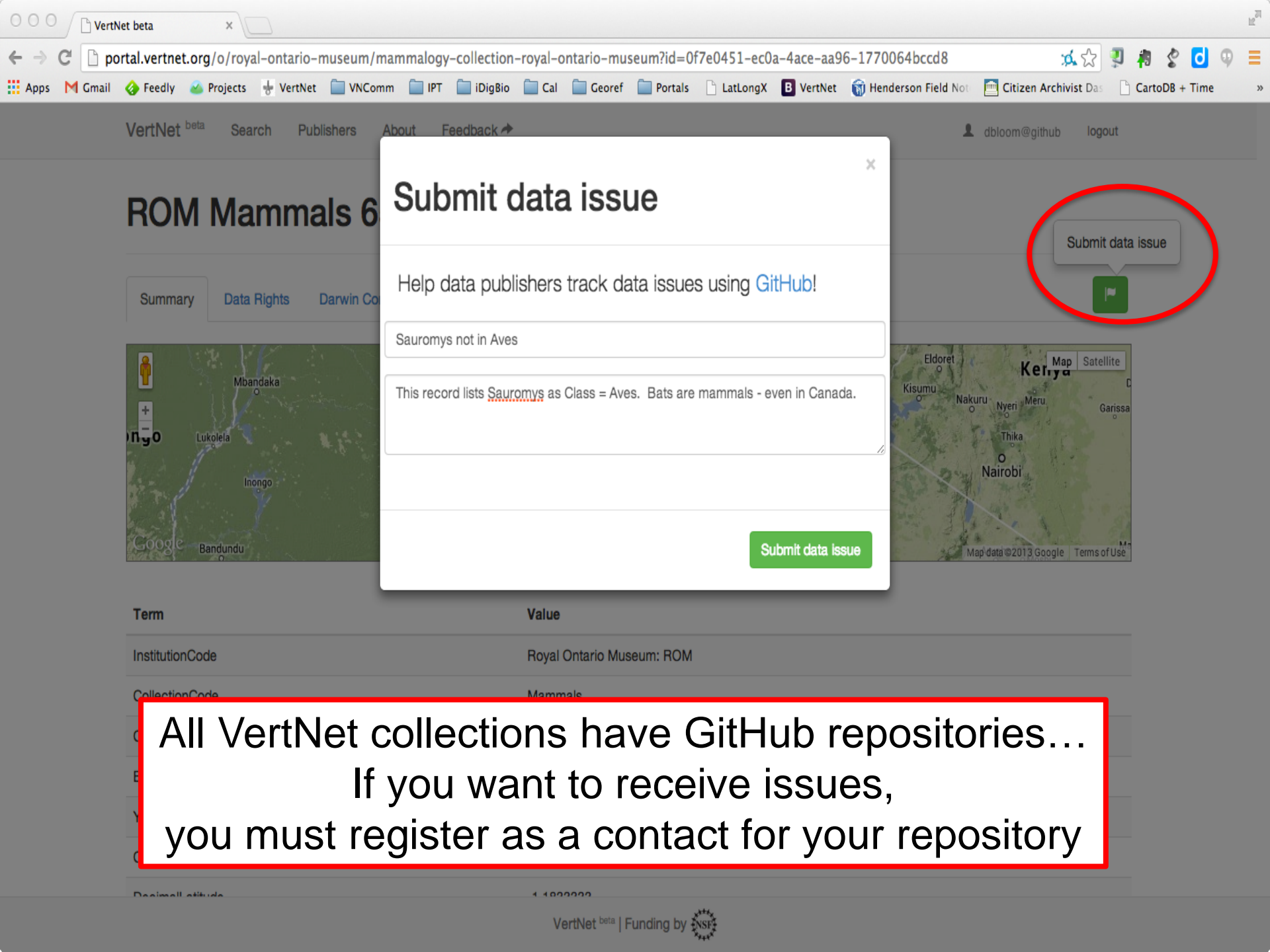
**Impromptu Instruction**

<http://www.vertnet.org/resources/workshops.html>

# Issue Tracking

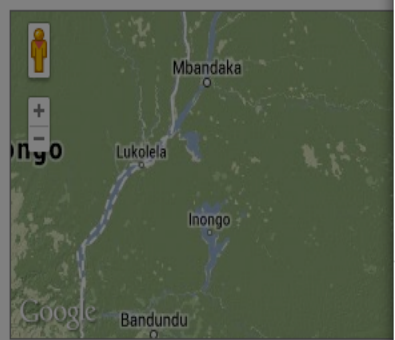


<http://www.vertnet.org/resources/issuetrackingguide.html>



# ROM Mammals 6

Summary Data Rights Darwin Co



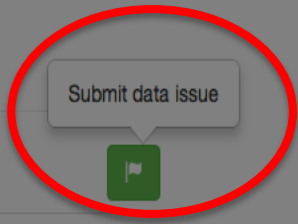
## Submit data issue

Help data publishers track data issues using [GitHub!](#)

Sauromys not in Aves

This record lists Sauromys as Class = Aves. Bats are mammals - even in Canada.

Submit data issue



All VertNet collections have GitHub repositories...  
If you want to receive issues,  
you must register as a contact for your repository

Term	Value
InstitutionCode	Royal Ontario Museum: ROM
CollectionCode	Mammals
DecimalLatitude	1.182222

Browse Issues Milestones [New Issue](#)  
[Back to issue list](#) Issue #3

dbloom opened this issue a month ago [Edit](#)

## [ROM Mammals 65045] SAUROMYS PETROPHILUS - SAUROMYS Test Issue #2

No one is assigned No milestone

This is a test. The issue of Sauromys being nested in Class:Aves has been addressed: [#1 \(comment\)](#)

Brad: pls, confirm you have received a notice of this issue and then close it.

thanks  
VertNet

.....

You can [view the original detail page](#) on VertNet. Here are the original record contents:

Term	Value
CollectionID	urn:lsid:biocol.org:col:34909
InstitutionCode	Royal Ontario Museum: ROM
CollectionCode	Mammals
DatasetName	Mammalogy
BasisOfRecord	PreservedSpecimen

**Closed**  
1 comment

Labels

Issue navigation sidebar: code icon, info icon, search icon, list icon, chart icon, link icon, close icon



Browse Issues Milestones

New Issue

- Everyone's Issues 5
- Assigned to you 0
- Created by you 0
- Mentioning you 0

No milestone selected

- Labels
- bug 0
  - duplicate 0
  - enhancement 0
  - invalid 0
  - question 0
  - wontfix 0

Manage Labels

New label  
New label name

5 Open 0 Closed Sort: Newest

<input type="checkbox"/>	Close	Label	Assignee	Milestone
<input type="checkbox"/>				
<input type="checkbox"/>				
<input type="checkbox"/>				
<input type="checkbox"/>				
<input type="checkbox"/>				
<input type="checkbox"/>				

Keyboard shortcuts available

# Licensing

## What CAN and SHOULD be protected?

<http://www.vertnet.org/resources/datalicensingguide.html>

### RESOURCES

INTRODUCTION

WHAT WE BELIEVE

WHY ARE RIGHTS AND  
LICENSES NEEDED?

WHAT CAN BE PROTECTED BY  
COPYRIGHT AND LICENSING?

COMPILATIONS AND  
COPYRIGHT

SWE

TH

VI

I'VE SELECTED A LICENSE,  
WHERE DOES IT GO?

FOOTNOTES AND OTHER  
RESOURCES

[RETURN TO PUBLISHERS' HELP](#)

#### GUIDE TO COPYRIGHT AND LICENSES FOR DATASET PUBLICATION

DOWNLOAD [PDF VERSION](#) OF THE GUIDE

#### INTRODUCTION

This quick guide is intended to give data publishers and users of data published to the VertNet portal, as well as other biodiversity data portals, an understanding of the uses and limits of copyright, licensing, and waivers for biodiversity data and datasets. This is NOT a legal document or contract. This IS a well-researched guide intended to provide you with a clear

License is not required for publication in VertNet, but...  
we encourage CC0  
or other machine-readable licenses

opinion in keeping with their home institutions. Data users may also need to seek answers to questions of use with local legal professionals, although we are happy to help in any way possible.

In keeping with VertNet's primary goals and beliefs, this guide is intended to make things as simple and clear as possible, but dealing with these issues can be complicated. The more caveats we/you add, the more complicated things can become. So, buckle up and drive on in.

# Community Norms for Data Use and Publication



Right Decision  
Wrong Decision

Behave responsibly!

<http://www.vertnet.org/resources/norms.html>



The future of VertNet

NSF funding ends in fall 2015, but the VertNet team remains committed to sustaining VertNet and open access data...

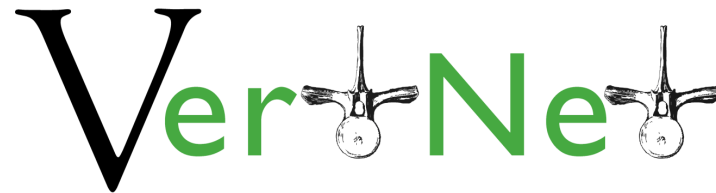
What does this mean?

- New model: non-profit, LLC, iDigBio
- Reduced services, at least temporarily
- Possible contract charges for services
- Expanded funding opportunities

How you can help: if you are writing a collections digitization grant, consider \$\$ for VertNet support to help publish your data

*“Information is only of value if you give it to people with the ability to do something with it.”*

*- Gen. Stanley McChrystal*



<http://www.vertnet.org>

*If you have questions, contact:*

*Carla Cicero, VertNet PI*

[ccicero@Berkeley.edu](mailto:ccicero@Berkeley.edu)

*Dave Bloom, VertNet Coordinator*

[dbloom@vertnet.org](mailto:dbloom@vertnet.org)