VertNet

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*…*

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

Our Community:
The Vertebrate Networks

Est. 1999, 2004

Est. 2001

Est. 2002

Est. 2004

Vertebrate Network Growth

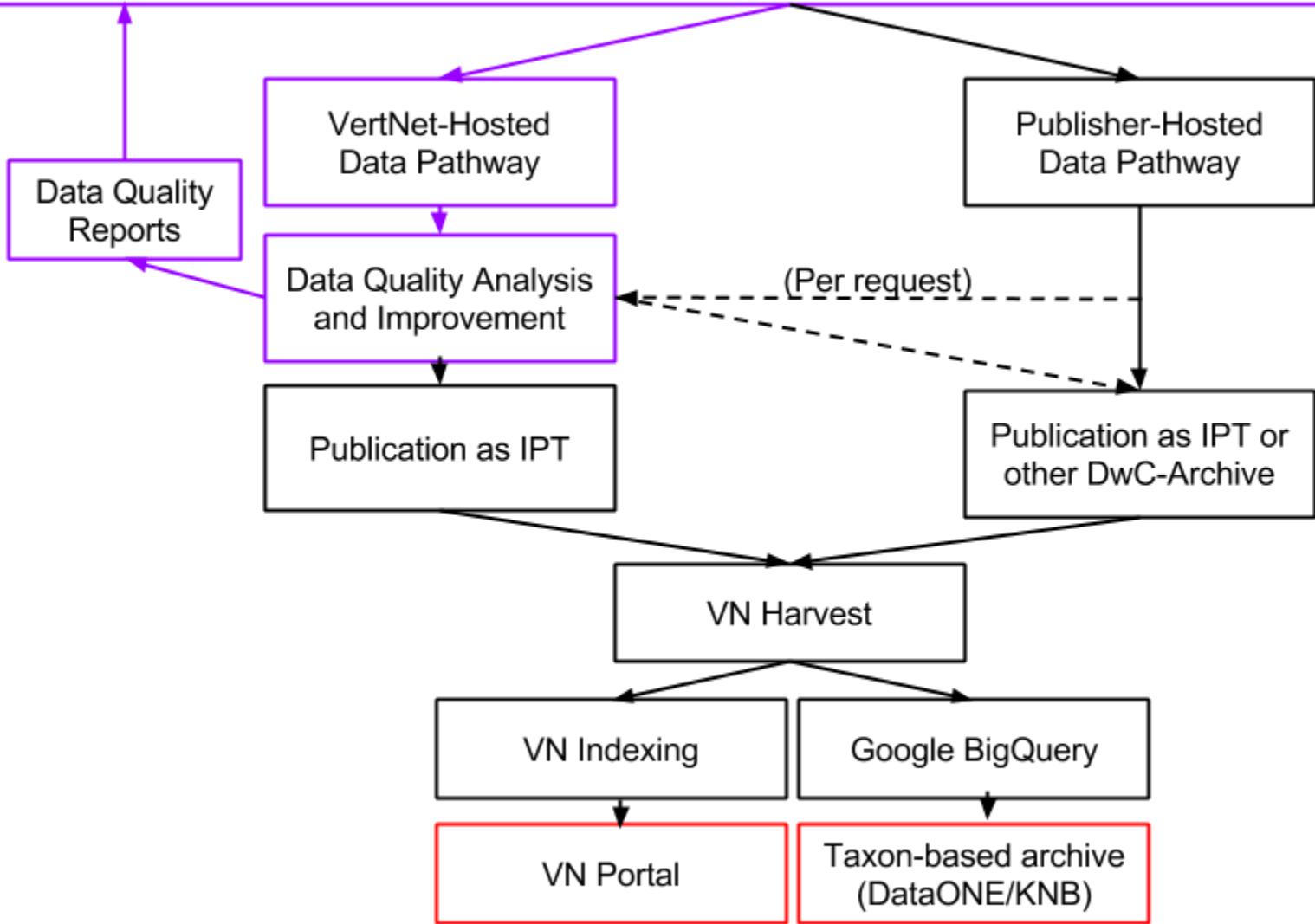April 2015:
548 Collections from
183 Institutions
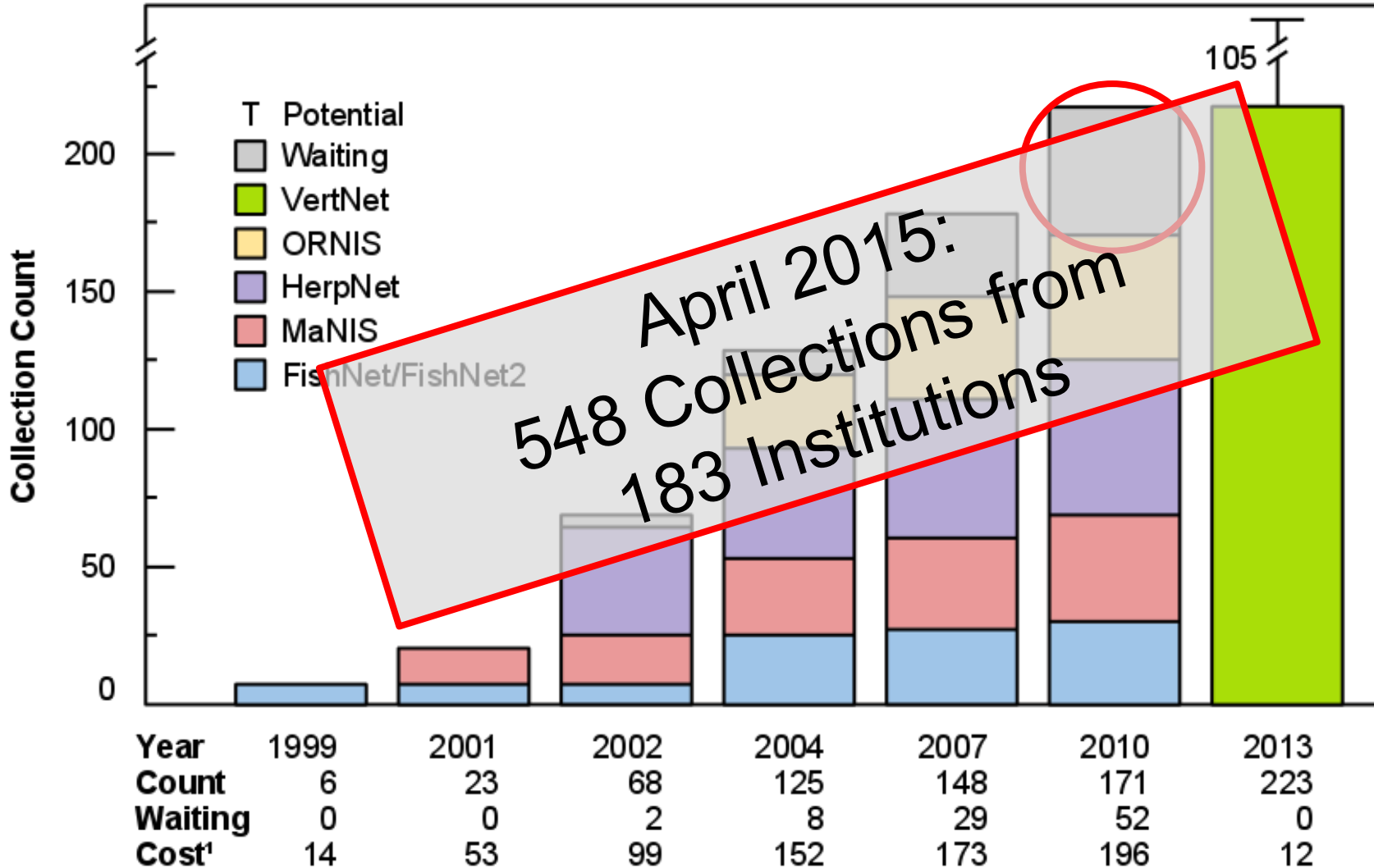
Integrated Publishing Toolkit (IPT)

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…)

Data Quality Reports

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

VN Indexing

Google BigQuery

VN Portal

Taxon-based archive (DataONE/KNB)

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

         VertNet-Hosted                          Publisher-Hosted
         Data Pathway                             Data Pathway

         Data Quality Analysis        (Per request)
         and Improvement

         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…

…or two or three...

Focus your search…

Still confused? We can help!

When in doubt...

Results

Results – with Media

Results Detail – Darwin Core

Results Detail – Rights

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			
Australia			
Australia			
'Australia			
Australia			
Australia			
Australia			
Australia			
Australia			

	A	B	C	D	E
1	catalogNumber	verbatimCountry	Country		
2	A-35860	USA	United States		
3	A-35861	USA	United States		
4	R-10424	Belgian Congo	Democratic Republic of the Congo		
5	A-16728	USA	United States		
6	R-160643	USA	United States		
7	A-33990	USA	United States		
8	A-37236	USA	United States		
9	A-37237	USA	United States		
10	A-37652	USA	United States		
11	A-33765	[Tanzania]	Tanzania		
12	A-33766	[Tanzania]	Tanzania		
13	A-33767	[Tanzania]	Tanzania		
14	A-33768	[Tanzania]	Tanzania		
15	A-33769	[Tanzania]	Tanzania		
16	A-33770	[Tanzania]	Tanzania		
17	A-33771	[Tanzania]	Tanzania		
18	A-15292	USA	United States		
19	A-34524	USA	United States		
20	A-34525	USA	United States		
21	A-33543	[Tanzania]	Tanzania		
22	R-32878	E Madagascar	Madagascar		
23	R-32568	USA	United States		
24	A-37400	USA	United States		

Filter for "Tissues"

Results Detail – Spatial Quality

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 formi	1
ramphastos vitelli	1
Cotinga	1
catalognumber:15	1
type:specimen sp	1
pipra	1
Trogon rufus	2
"Ficedula narcissi	2
mallard	1
Phegornis mitche	1
collin county	1
colinus nigrogula	1
country:"Trinidad	1
class:aves record	1
colinus cristatus l	1
stateprovince:"alaska" specificepithet:acuminata class:aves	1
country:"Panama" (passeriformes OR picidae OR cuculiformes)	1
tricholaema	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
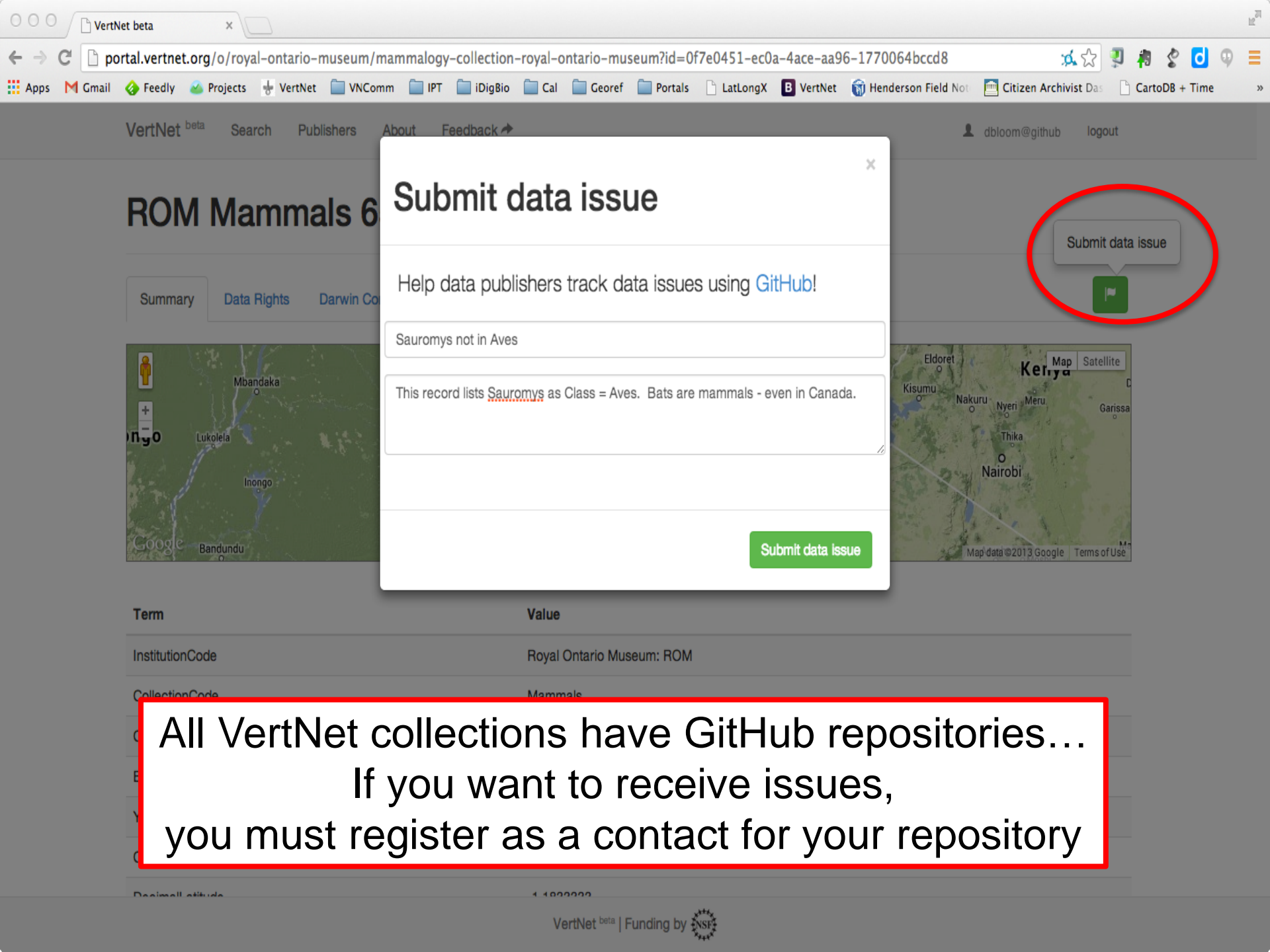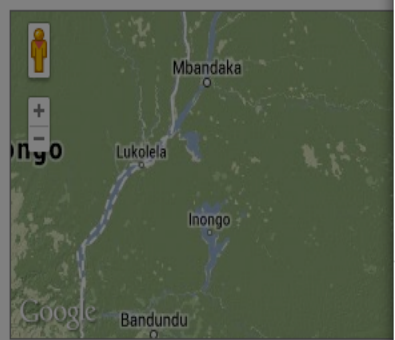
http://www.vertnet.org/resources/workshops.html

Issue Tracking

http://www.vertnet.org/resources/issuetrackingguide.html
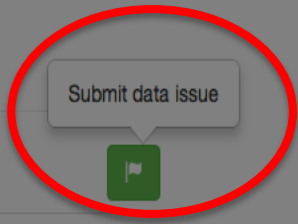
All VertNet collections have GitHub repositories…
If you want to receive issues,
you must register as a contact for your repository

This repository ▾ | Search or type a command | ? | **Explore Gist Blog Help** | 🖥 dbloom

PUBLIC | **rom-vertnet** / **mammalogyspecimens**

👁 Unwatch ▾ | 6 | ★ Star | 1 | ⑂ Fork | 0

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

Closed

1 comment

No one is assigned ⚙▾ | No milestone ⚙▾

Labels ⚙▾

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

This repository ▾   Search or type a command   ⑦   **Explore** **Gist** **Blog** **Help**   dbloom

PUBLIC   mvz-vertnet / **mvz-herp**   ⊙ Unwatch ▾ 8   ★ Star 0   ⅄ Fork 1

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 ▾

☐   Close   Label ▾   Assignee ▾   Milestone ▾

☐ ⊘ **[MVZ Herp 195816] Batrachoseps wrighti - Test: Look Batrachoseps in Oregon!**   #5
    Opened by atrox10 19 days ago

☐ ⊘ **[MVZ Herp 195816] Batrachoseps wrighti - Test: Look Batrachoseps in Oregon!**   #4
    Opened by atrox10 19 days ago

☐ ⊘ **[MVZ Herp 195816] Batrachoseps wrighti - Test: Look Batrachoseps in Oregon!**   #3
    Opened by atrox10 19 days ago

☐ ⊘ **[MVZ Herp 65979] Hyla eximia - see if CArol gets this**   #2
    Opened by atrox10 a month ago

☐ ⊘ **[MVZ Herp 15073] Elgaria coerulea - wow it's in oregon**   #1
    Opened by mkoo a month ago

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

o have

gnition on

r legal

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

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
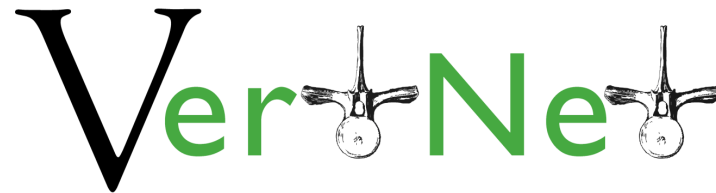
VertNet

If you have questions, contact:

Carla Cicero, VertNet PI
ccicero@Berkeley.edu

Dave Bloom, VertNet Coordinator
dbloom@vertnet.org