

# Data Quality and Issue Tracking

iDigBio Summit 2013

Tallahassee, Florida

# What is VertNet?

Access to loads of biodiversity data via a searchable cloud-based platform on the web.

Fast  
Easy  
Scalable  
Cheap



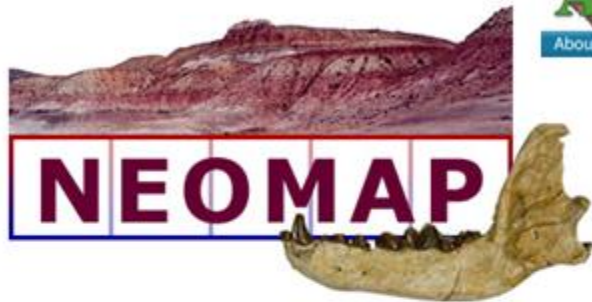
# VertNet builds on existing vertebrate networks



Data are published to:



# VertNet Collaborations



# Technical Challenges

- Cloud-based aggregation
  - Caching and Indexing
  - Search and Downloads
    - Cost

# Social Challenges

- Knowledge gaps
- Varied experience
- Limited resources

# Great Progress So Far

Complete: 51 institutions (160 collections)  
108 Data Resources

In Progress: 29 institutions (85 collections)  
Waiting: 87 institutions (174 collections)

2,068,859,726 Records Served  
June '12-June '13



Da

ges

To complete  
w

as possible  
es.



VerNe

# Standardized Vocabularies

DwC Fields	# Standard Terms	# Discovered Terms*
verbatimSex	6	710
verbatimTypeStatus	8	1861
verbatimCountry	238	4962
verbatimLifeStage	64	12721

# Non-Standardized Vocabularies

DwC Fields	# Useful Terms	# Discovered Terms*
Preparations	26	10712

\* ~2.7 million records reviewed



# Issue Tracking

## Why?

Allow data consumers to:

- Report questionable or incorrect data
- Seek clarification and ask questions

## How?

Users can communicate via:

- Posted Email addresses
- VertNet portal Feedback form
- **New! GitHub Issue Tracking**



# GitHub

**Built for developers...**

**co-opted just for you!**

- Public, project-based code repository
- Automatic notification services
- Interactive discussion
- Public URLs for easy sharing
- Long-term archiving
- Subscription opportunities
- Easy administration (we'll help!)



# How It Works

## **What we do:**

- Create an Organization
- Create repositories
- Assign administrators
- Provide support online and offline
- Monitor Issues

## **What you do:**

- Get a GitHub username
- Respond to issues (or not)

# Inside GitHub

The screenshot displays the GitHub web interface for the organization 'mvz-vertnet'. The browser's address bar shows the URL <https://github.com/organizations/mvz-vertnet>. The page features a navigation bar with links to 'Explore', 'Gist', 'Blog', and 'Help'. Below the navigation bar, there are tabs for 'News Feed', 'Pull Requests', 'Issues', and 'Teams'. The 'News Feed' tab is active, showing a list of recent activity. On the right side, there is a sidebar titled 'Repositories (8)' with a search bar and a list of repositories. Two red arrows are overlaid on the image: one points from the word 'Activity' to the 'News Feed' section, and another points from the word 'Repositories' to the 'Repositories (8)' sidebar.

**Activity**

**Repositories**

**News Feed**

- 15 days ago **eightysteale** closed issue **mvz-vertnet/mvz-bird#6** [MVZ Bird 86208] Turdus rufitorques - test
- 15 days ago **eightysteale** opened issue **mvz-vertnet/mvz-bird#6** [MVZ Bird 86208] Turdus rufitorques - test
- 16 days ago **ccicero** commented on issue **mvz-vertnet/mvz-egg#4** OK, thanks! On Sat, Nov 2, 2013 at 9:26 AM, Aaron Steele [eightysteale@gmail.com](mailto:eightysteale@gmail.com) wrote: sorry this came from my dev machine (not production) which ...
- 16 days ago **ccicero** commented on issue **mvz-vertnet/mvz-egg#4** sorry this came from my dev machine (not production) which is why the link didn't work. On Sat, Nov 2, 2013 at 9:10 AM, Carla Cicero [ccicero@berkel...](mailto:ccicero@berkel...)
- 16 days ago **ccicero** commented on issue **mvz-vertnet/mvz-egg#4** when I click to view original detail page on VertNet, I get an error: Oops! Google Chrome could not connect to localhost:8080 http://localhost:8080...
- 17 days ago **eightysteale** opened issue **mvz-vertnet/mvz-bird#5** [MVZ Bird 86208] Turdus rufitorques - test

**Repositories (8)**

- mvz-vertnet/mvz-bird**
- mvz-vertnet/mvzobs-bird**
- mvz-vertnet/mvz-egg**
- mvz-vertnet/mvz-herp**
- mvz-vertnet/mvzobs-herp**
- mvz-vertnet/mvz-hild**
- mvz-vertnet/mvz-mammal**
- mvz-vertnet/mvzobs-mammal**

# Flagging An Issue

VertNet beta

portal.vertnet.org/o/royal-ontario-museum/mammalogy-collection-royal-ontario-museum?id=Of7e0451-ec0a-4ace-aa96-1770064bccd8


Apps Gmail Feedly Projects VertNet VNComm IPT iDigBio Cal Georef Portals LatLongX VertNet Henderson Field Not Citizen Archivist Das CartoDB + Time

VertNet beta Search Publishers About Feedback

dbloom@github logout

## ROM Mammals 65044 SAUROMYS PETROPHILUS

Summary Data Rights Darwin Core



Submit data issue

Term	Value
InstitutionCode	Royal Ontario Museum: ROM
CollectionCode	Mammals
CatalogNumber	65044
BasisOfRecord	PreservedSpecimen
Year	1965
Country	BOTSWANA
DecimalLatitude	1.182222

VertNet beta | Funding by NSF

# Flagging an Issue

VertNet beta

portal.vertnet.org/o/royal-ontario-museum/mammalogy-collection-royal-ontario-museum?id=0f7e0451-ec0a-4ace-aa96-1770064bccd8

Apps Gmail Feedly Projects VertNet VNComm IPT iDigBio Cal Georef Portals LatLongX VertNet Henderson Field Not Citizen Archivist Das CartoDB + Time

VertNet beta Search Publishers About Feedback

dbloom@github logout

## ROM Mammals 6

Summary Data Rights Darwin Co

Sauromys not in Aves

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

Submit data issue

Submit data issue

Term	Value
InstitutionCode	Royal Ontario Museum: ROM
CollectionCode	Mammals
CatalogNumber	65044
BasisOfRecord	PreservedSpecimen
Year	1965
Country	BOTSWANA
DecimalLatitude	1.182222

VertNet beta | Funding by NSF

# Issues in GitHub

The screenshot shows the GitHub interface for the repository `mvz-vertnet / mvz-herp`. The page displays a list of 5 open issues, sorted by newest. The issues are:

- #5: [MVZ Herp 195816] *Batrachoseps wrighti* - Test: Look *Batrachoseps* in Oregon! (Opened by atrox10 19 days ago)
- #4: [MVZ Herp 195816] *Batrachoseps wrighti* - Test: Look *Batrachoseps* in Oregon! (Opened by atrox10 19 days ago)
- #3: [MVZ Herp 195816] *Batrachoseps wrighti* - Test: Look *Batrachoseps* in Oregon! (Opened by atrox10 19 days ago)
- #2: [MVZ Herp 65979] *Hyla eximia* - see if CArol gets this (Opened by atrox10 a month ago)
- #1: [MVZ Herp 15073] *Elgaria coerulea* - wow it's in oregon (Opened by mikoo a month ago)

The left sidebar shows filters for 'Everyone's Issues' (5), 'Assigned to you' (0), 'Created by you' (0), and 'Mentioning you' (0). It also includes a 'Labels' section with filters like 'bug', 'duplicate', 'enhancement', 'invalid', 'question', and 'wontfix'. A 'New Issue' button is visible in the top right.

A red arrow points to the word 'Issues' in the bottom right corner of the page.

# Issues in GitHub

Public **rom-vertnet / mammalogyspecimens** Unwatch 6 Star 1 Fork 0

Browse Issues Milestones [New Issue](#)

[Back to issue list](#) Issue #3

**Closed** 1 comment Labels

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

Please kick the tires!  
<http://portal.vertnet.org>

