

# Who Has Time for Biological Collections Data Quality Feedback? Maybe a Community Can Help

**Teresa J. Mayfield-Meyer**  
**Collections Assistant, Museum of Southwestern Biology**  
**Treasurer, Arctos Consortium**

**SPNHC/TDWG 2018**



**Website:** <http://arctosdb.org>  
**Search:** <http://arctos.database.museum>

# What do biodiversity collection managers do all day?



**What do biodiversity collection managers do all day?**

**DIGITIZE AND MANAGE DATA**

# Stuff happens when you digitize...

# COBOL

# dBASE

# LOTUS123

# ACCESS

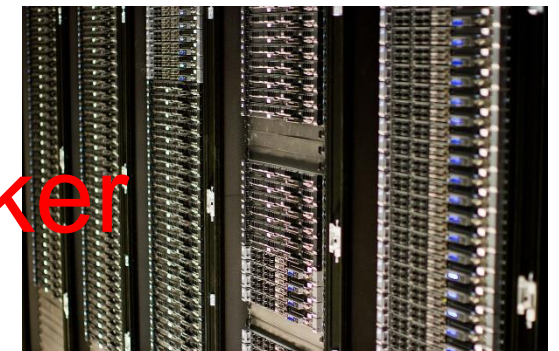
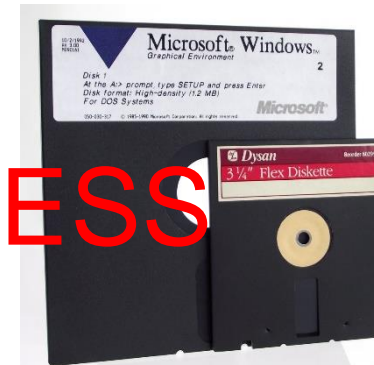
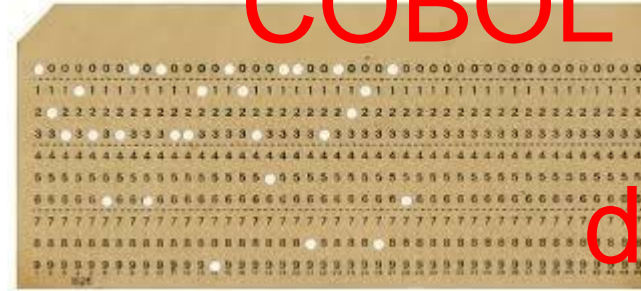
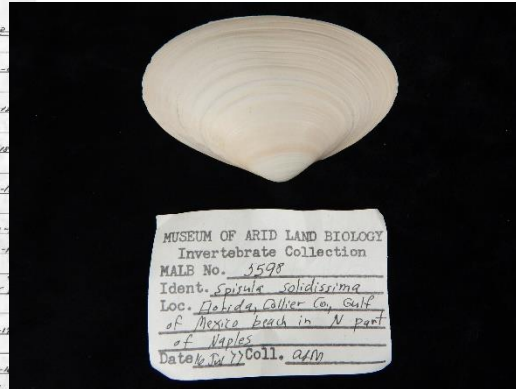
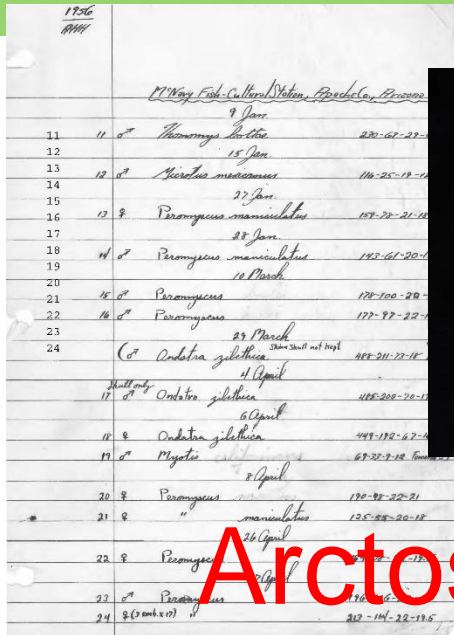
# FileMaker

# Cloud Computing

# Arctos

# EMu

# Specify Symbiota





# No Excuses (Not)

*U.S. Navy Fish-Culture Station, Proctor's*

11	♂	<i>Thomomys bottae</i>	9 Jan.	230-67-27-6
12	♂	<i>Microtus merriami</i>	15 Jan.	116-25-19-18
13	♂	<i>Peromyscus maniculatus</i>	27 Jan.	157-28-21-18
14	♀	<i>Peromyscus maniculatus</i>	28 Jan.	143-61-20-18.5
15	♂	<i>Peromyscus maniculatus</i>	10 March	178-100-29-18
16	♂	<i>Peromyscus</i>		177-97-22-18.5
17	♂	<i>Peromyscus</i>	29 March	488-211-73-18
18	♂	<i>Peromyscus</i>	4 April	485-209-70-17
19	♂	<i>Ondatra zibethica</i>	6 April	449-192-67-16.5
20	♀	<i>Ondatra zibethica</i>		69-32-9-12 (Femur 24)
21	♂	<i>Myotis</i>	8 April	190-98-22-21
22	♀	<i>Peromyscus maniculatus</i>	26 April	125-85-20-18
23	♀	<i>Peromyscus</i>	27 April	164-78-22-19.5
24	♂	<i>Peromyscus</i>		196-106-22-20
	♀ (3 emb. x 12)			213-114-22-19.5

MUSEUM OF ARID LAND BIOLOGY  
Invertebrate Collection  
MALB No. 5598  
Ident. *Spirula solidissima*  
Loc. of Hobida, Collier Co., Gulf  
of Mexico beach in N part  
Date 16 Jul 77 Coll. aLM

25 JUL 1977  
MEXICO CHIHUAHUA  
Colonia Juarez  
5000' 5 mi NW  
GSForbes, collector



# Who has time for surveys?

iDigBio data quality flags feedback

[SUMMARY](#) → [DESIGN SURVEY](#) → [PREVIEW & SCORE](#) → [COLLECT RESPONSES](#) → [ANALYZE RESULTS](#)

## Survey Design

### iDigBio data quality flags feedback

Created on 2/18/2018

10

QUESTIONS

1

PAGES

## Responses and Status

TOTAL RESPONSES

3

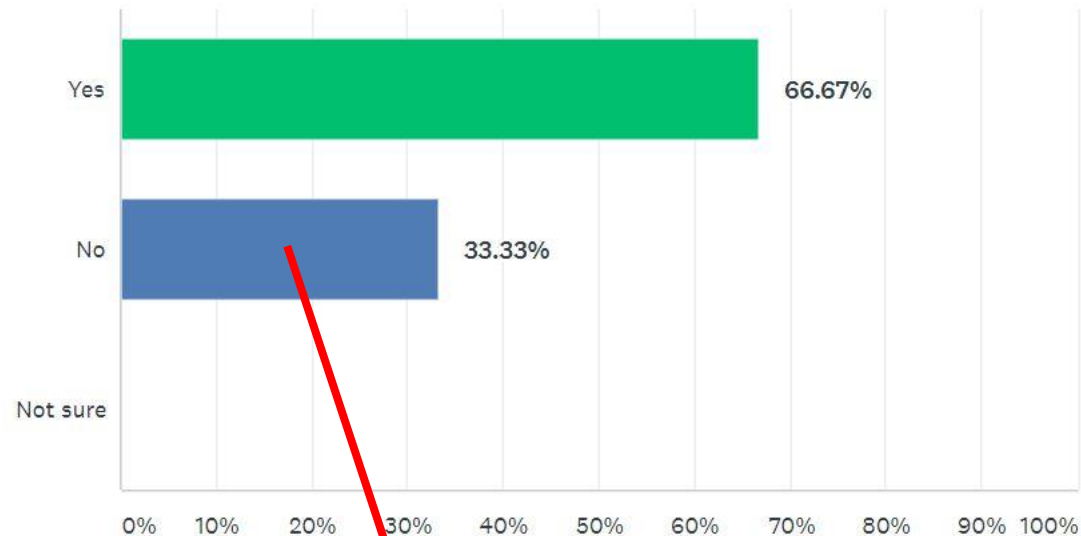
### Collectors

OPEN

# Data Publishing

Have you published your collection data to iDigBio?

Answered: 3 Skipped: 0

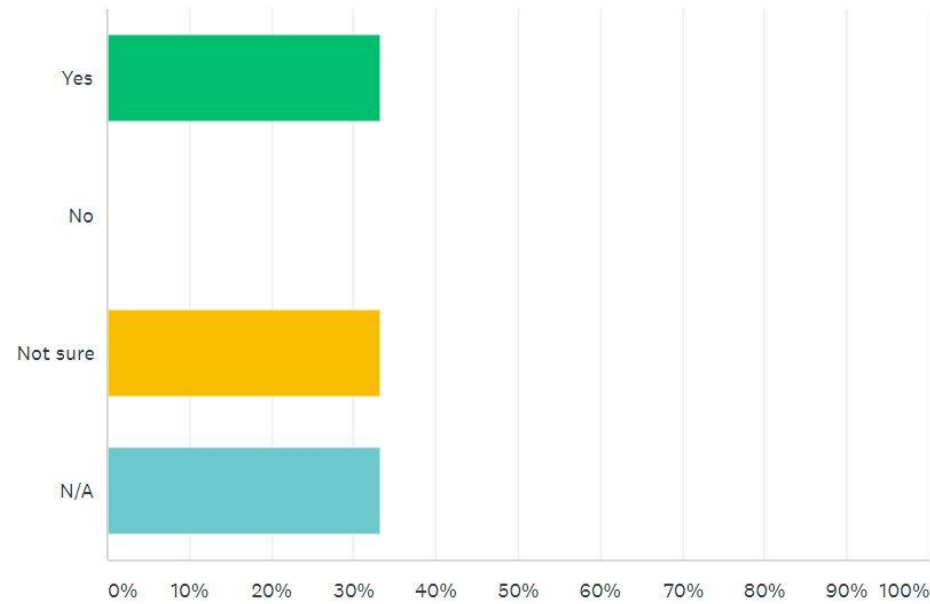


I'm not even sure what this is referencing:  
data in Arctos or data that is being migrated into  
Arctos! I have never heard this terminology until now.

# Wait, there are data flags?

Whether or not you received a data quality email from iDigBio, have you reviewed the data quality flags provided by iDigBio for your collection data?

Answered: 3 Skipped: 0





# My initial experience

## iDigBio Flags on Continent #1291



Jegelewicz opened this issue on Oct 5, 2017 · 2 comments

huge list of specimens flagged for various corrections (sigh)



Jegelewicz commented on Oct 5, 2017

Member



As my data was recently ingested by iDigBio, I received a huge list of specimens flagged for various corrections (sigh). I wanted to bring this one to the group to see if we should be paying more attention to Darwin Core, or if it is just something to let iDigBio keep "correcting" for.

Some of my specimens on islands in the Pacific, are flagged by iDigBio with "dwc\_continent\_replaced | Darwin Core Continent Corrected." one example is here:

<https://www.idigbio.org/portal/records/89015b8e-d745-430c-b846-8b250b62a>

dwc\_continent\_replaced

Is Arctos not complying with Darwin Core or is this just an artifact of iDigBio? Do we need to do anything about it or do I just need to know that these flags are not a problem? My main concern is that users of iDigBio will view our data as less reliable with flags attached.

users of iDigBio will view our data as less reliable



# My initial experience

noticed something similar in GBIF regarding dates

Darwin Core is an exchange standard; Arctos isn't "complying" with any data standards because none exist

## iDigBio Flags on Continent #1291

 Open Jegelewicz opened this issue on Oct 5, 2017 · 2 comments



Jegelewicz commented on Oct 5, 2017

Member

As my data was recently ingested by iDigBio, I received a huge list of specimens flagged for various corrections (sigh). I wanted to bring this one to the group to see if we should be paying more attention to Darwin Core, or if it is just something to let iDigBio keep "correcting" for.

Some of my specimens on islands in the Pacific, are flagged by iDigBio with "dwc\_continent\_replaced | Darwin Core Continent Corrected." one example is here:  
<https://www.idigbio.org/portal/records/89015b8e-d745-430c-b846-8b250b62afcb>

Is Arctos not complying with Darwin Core or is this just an artifact of iDigBio? Do we need to do anything about it or do I just need to know that these flags are not a problem? My main concern is that users of iDigBio will view our data as less reliable with flags attached.

I agree with your assessment: User's initial reaction to the flag will be "Arctos is broken," which is absolutely not the case.

# My initial experience

“..subjective to how data are stored in Arctos vs. other models. The subjective ones I don't think are worth our time to care about at this point”

“..some data quality flags from iDigBio that are useful because they correct objectively incorrect data”

**the DQ tests and methods iDigBio uses are in flux due to work being done in TDWG. ....**



ekrimmel commented on Oct 5, 2017

Member + 👤 ...

We've done a bit of thinking about this internally. Right now there are some data quality flags from iDigBio that are useful because they correct objectively incorrect data (like a mismatch between coordinates and country due to a missing sign). Others, like your Pacific islands example, are subjective to how data are stored in Arctos vs. other models. Many of the objective DQ tests would flag errors that we don't have because Arctos/Dusty also catches them (e.g. "April 31st is a date that doesn't exist"). The subjective ones I don't think are worth our time to care about at this point, in particular because the DQ tests and methods iDigBio uses are in flux due to work being done in TDWG.

The TDWG Biodiversity Data Quality task group has a few factions working on different aspects. One is trying to [define a framework](#) for what we even mean when we talk about data quality as a collections community. Another is getting all the aggregators, including iDigBio, to [settle on a set of the same data quality tests](#) to run on provider data and return flags for.

I don't actually think the flags are visibly negative enough to make users think "Arctos is broken." I would hope (although I guess hope is the operative word here), that people who are running analyses on or otherwise using aggregator data for something beyond browsing would notice that the flags are doing more standardizing than correcting, and that obviously different collections/databases use different but equally correct ways to say the same thing...

# My initial experience

I would hope (although I guess hope is the operative word here), that people who are running analyses on or otherwise using aggregator data for something beyond browsing would notice that the flags are doing more standardizing than correcting, and that **obviously** different collections/databases use different but equally correct ways to say the same thing...



ekrimmel commented on Oct 5, 2017

Member + 👤 ...


We've done a bit of thinking about this internally. Right now there are some data quality flags from iDigBio that are useful because they correct objectively incorrect data, like a mismatch between coordinates and country due to a missing sign. Others, like year of Pacific islands example, are subjective to how data are stored in Arctos vs. other models. Many of the objective DQ tests would flag errors that we don't have because Arctos/Dusty doesn't catch them (e.g. "April 31st" is a date that doesn't exist). The subjective ones I don't think are worth our time to care about at this point, in particular because the DQ tests and methods iDigBio uses are in flux due to work being done in TDWG.

The TDWG Biodiversity Data Quality task group has a few functions working on different aspects. One is trying to define a framework for what we even mean when we talk about data quality as a collections community. Another is getting all the aggregators, including iDigBio, to settle on a set of the same data quality tests to run on provider data and return flags for.

I don't actually think the flags are visibly negative enough to make users think "Arctos is broken." I would hope (although I guess hope is the operative word here), that people who are running analyses on or otherwise using aggregator data for something beyond browsing would notice that the flags are doing more standardizing than correcting, and that obviously different collections/databases use different but equally correct ways to say the same thing...



# One Collection's Data Flags

 **University of Texas at El Paso Vertebrate Paleo Collections**  
**University of Texas at El Paso Biodiversity Collections**

Search Portals My Stuff About/Help

Access to 23,354 records

Search Clear Form Use Last Values See results as: Specimen Records

Type: any Require Tissues? ☐

**Identifiers** [Customize](#) [Show More Options](#)

Collection: Select options

Catalog Number:

Any Identifier:

**Identification and Taxonomy** [Show More Options](#)

Any taxon, ID, common name:

**Locality** [Show More Options](#)

Any Geographic Element:  [Select on Google Map](#)

**Date/Collector** [Show More Options](#)

Help Agent Role Define

**Specimen Record** [Show More Options](#)

Part Name:  Define: Add # for exact match

**Usage** [Show More Options](#)

Basis of Citation:  Define


**Media** [Show More Options](#)

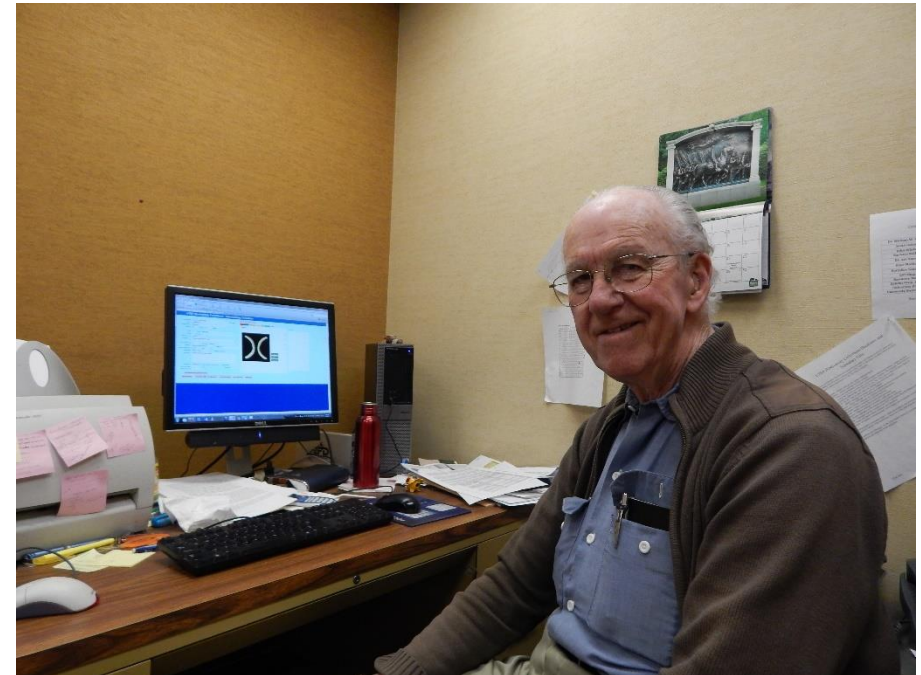
Media Type:  Define

**Relationships** [Show More Options](#)

Relationship:

Search Clear Form Use Last Values

 [Data Providers](#)  
[Report a bug or request support](#)  
[View in mobile site](#)



UTEP Fossils (Arctos)

Specimen Records: 23,354      Media Records: 258  
iDigBio Last Ingested Date: 2018-07-13

Data Corrected   Data Use   Raw

This table shows any data corrections that were performed on this recordset to improve the capabilities of iDigBio Search. The first column represents the correction performed. The last two columns represent the number and percentage of records that were corrected. A complete list of the data quality flags and their descriptions can be found [here](#). Clicking on a data flag name will take you to a search for all records with this flag in this recordset.

Flag	Records With This Flag	(%) Percent With This Flag
idigbio_isocountrycode_added	23354	100
geopoint_datum_missing	22963	98.326
dwc_datasetid_added	21141	90.524
dwc_parentnameusageid_added	21141	90.524
dwc_taxonid_added	21141	90.524
dwc_taxonomicstatus_added	21141	90.524
dwc_taxonrank_added	21141	90.524
gbif_canonicalname_added	21141	90.524
gbif_genericname_added	21141	90.524
gbif_taxon_corrected	21141	90.524
dwc_scientificnameauthorship_added	20615	88.272



last update: 2018-07-29 14:44:14.0  
globalnames score not available  
match type not available

author\_text: A. Dietr.  
display\_name: *Picea* A. Dietr.  
nomenclatural\_code: ICBN  
remark: Imported from ITIS 6 Feb 2007, added author text 7 Jul 2018 from ITIS  
scientific\_name: *Picea*  
source\_authority: ITIS  
valid\_catalog\_term\_fg: 1

Classification:

Plantae (kingdom) [ [more like this term](#) ] [ [including rank](#) ] [ [from this source](#) ]  
Tracheophyta (phylum) [ [more like this term](#) ] [ [including rank](#) ] [ [from this source](#) ]  
Pinopsida (class) [ [more like this term](#) ] [ [including rank](#) ] [ [from this source](#) ]  
Pinales (order) [ [more like this term](#) ] [ [including rank](#) ] [ [from this source](#) ]  
Pinaceae (family) [ [more like this term](#) ] [ [including rank](#) ] [ [from this source](#) ]  
Picea (genus) [ [more like this term](#) ] [ [including rank](#) ] [ [from this source](#) ]

dwc_kingdom_suspect	4	0.009
dwc_phylum_added	1	0.004



## UTEP Fossils (Arctos)

Specimen Records: 23,354

Media Records: 258

iDigBio Last Ingested Date: 2018-07-13

taxon_match_failed <i>i</i>	2274	9.737
dwc_specificepithet_added <i>i</i>	2191	9.382
dwc_specificepithet_replaced <i>i</i>	1666	7.134
dwc_family_replaced <i>i</i>	1101	4.714
dwc_infraspecificepithet_added <i>i</i>	1069	4.577
dwc_order_replaced <i>i</i>	739	3.164
dwc_genus_replaced <i>i</i>	645	2.762
dwc_family_added <i>i</i>	524	2.244
dwc_genus_added <i>i</i>	524	2.244
dwc_phylum_replaced <i>i</i>	521	2.231
dwc_class_replaced <i>i</i>	517	2.214
dwc_order_added <i>i</i>	436	1.867
geopoint_low_precision <i>i</i>	121	0.518
geopoint_datum_error <i>i</i>	100	0.428
dwc_class_added <i>i</i>	9	0.039
dwc_taxonremarks_added <i>i</i>	4	0.017
dwc_kingdom_suspect <i>i</i>	2	0.009
dwc_phylum_added <i>i</i>	1	0.004



# Is Aves not a class?

Filters Mapping Sorting Download

Add a field  Clear

Recordset: b7a79601-c07b-46d5-bd09-d4472b0d9431  
☐ Present ☐ Missing

Data Flags: dwc\_class\_replaced  
☐ Present ☐ Missing

Scientific Name: dwc:scientificName [Add EOL Synonyms](#)  
☐ Present ☐ Missing

↓ Scroll To Bottom ↓

Top 4 Taxa  
Aves  
Chiroptera  
Bivalvia sp.  
Lepisosteus sp.  
other

click or hover to wake.

3000 km  
2000 mi

Leaflet | Map data © OpenStreetMap

Total: 517

Family	Scientific Name ^	Date Collected	Country	Institution Code	Basis of Record	Columns
palinuridae	Aves	1970-06-18	United States	UTEP	FossilSpecimen	<a href="#">view</a>
palinuridae	Aves	1970-06-15	United States	UTEP	FossilSpecimen	<a href="#">view</a>

Family=palinuridae

# What the what?



University of Texas at El Paso Vertebrate Paleo Collections  
University of Texas at El Paso Biodiversity Collections

[Search](#)

[Portals](#)

[My Stuff](#)

[About/Help](#)

**UTEP:ES:4-578**

**Aves**

[<< Return to results](#)

**Dry Cave, Bison Chamber**

North America, United States, New Mexico, Eddy County

1971-- (1971-01-01 - 1971-12-31)

## Identifications

**Aves**

Animalia; Chordata; Aves; Aves

Identified by Arthur H. Harris on 1971-12-31

Nature of ID: expert

Collector(s)

Arthur H.

Parts

Part Name

Animalia > Chordata >  
Aves > Aves

## Specimen Record

[Animalia](#) > [Arthropoda](#) > [Malacostraca](#) > [Decapoda](#) > [Palinuridae](#)

**Aves**

From UTEP Fossils (Arctos)

Continent North America

Country United States

State/Province New Mexico

County/Parish Eddy County

Locality Dry Cave, Bison Chamber

Latitude 32.3723

Longitude -104.4815

Institution Code Utep

Collection Code Earth Science

Catalog Number 4-578

Collected By Collector(s): Arthur H. Harris

Animalia > Arthropoda >  
Malacostraca > Decapoda  
> Palinuridae

# The mystery deepens...



University of Texas at El Paso Vertebrate Paleo Collections  
University of Texas at El Paso Biodiversity Collections

[Search](#)[Portals](#)[My Stuff](#)[About/Help](#)

**UTEP:ES:4-578**

**Aves**

[<< Return to results](#)

**Dry Cave, Bison Chamber**

North America, United States, New Mexico, Eddy County

1971-- (1971-01-01 - 1971-12-31)

## Identifications

**Aves**

Animalia; Chordata; Aves; **Aves**

Identified by Arthur H. Harris on 1971-12-31

Nature of ID: expert

Collector(s):  
**Arthur H.**

Parts

Part Name

Data from source [Arctos](#) [ [Classifications](#) ] [ [Top](#) ]

last update: 2017-06-01 00:35:18.0  
globalnames score not available  
match type not available

display\_name: Aves  
nomenclatural\_code: ICZN  
scientific\_name: Aves  
source\_authority: ITIS  
valid\_catalog\_term\_fg: 1

Classification:

Animalia (kingdom) [ [more like this term](#) ] [ [including rank](#) ] [ [from this source](#) ]  
Chordata (phylum) [ [more like this term](#) ] [ [including rank](#) ] [ [from this source](#) ]  
Aves (class) [ [more like this term](#) ] [ [including rank](#) ] [ [from this source](#) ]

Genus = Avus

[Data](#) [Flags](#) [Raw](#)

## Taxonomy

Scientific Name	Aves
Higher Classification	Animalia; Chordata; ;
Kingdom	Animalia
Phylum	Chordata
Class	Aves
Order	decapoda
Family	palinuridae
Genus	avus
Nomenclatural Code	ICZN



# An answer, I think



University of Texas at El Paso Vertebrate Paleo Collections  
University of Texas at El Paso Biodiversity Collections

[Search](#)

[Portals](#)

[My Stuff](#)

[About/Help](#)

**UTEP:ES:4-578**

**Aves**

[<< Return to results](#)

**Dry Cave, Bison Chamber**

North America, United States, New Mexico, Eddy County

1971-- (1971-01-01 - 1971-12-31)

Identifications

**Aves**

Animalia; Chordata; Aves; Aves

Identified by Arthur H. Harris on 1971-12-31

Nature of ID: expert

Collector(s)

Arthur H.

Parts

Part Name

Animalia > Chordata > Aves

Aves

Identification=scientific name=dwc:scientificName "Aves"

Taxon rank=dwc:taxonrank "class"



# What do biodiversity collection managers do all day?



# In My Backyard

## **We need more help at home!**

- under-staffed and under-funded
- lack of experience/knowledge on the tech end
- assistance expressing the importance of the endeavor to administrators
- sustainability

# Thank You!

**Photos courtesy of:** Arctos  
Arthur H. Harris  
Museum of Obsolete Media  
Wikimedia Commons

**All opinions expressed are my own. I would like to thank**

Everyone in the **Arctos Working Group** for putting up with my multiple requests and issues on the Arctos GitHub. You all make my life so much richer!

**Dr. Art Harris** for being my iDigBio data quality flag guinea pig and partner in biological data crime.

**Deb Paul** for taking my questions seriously and making an honest attempt to answer them even while traveling.

