

# Getting Data Back

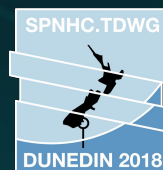
Repatriation of Augmented Information to an Institutional Repository



**Sharon Grant**

Janeen Jones  
Kate Webbink  
Pete Herbst  
Rob Zschernitz

**SPNHC/TDWG**  
2018



On the 9th of April 2010 the Field Museum received a momentous email from the ORNIS (ORnithology Network Information System) team informing them that they could now access the products of a nationwide georeferencing project; its bird collection could be, quite literally, put on the map.

On the 7th of August 2017 those data (along with the sister datasets from FISHNet (FISH NETwork) and MaNIS (Mammal Network Information System) finally made their way into the Museum's collection management system.

It's easy to get data out, why is it so hard to get it back? To make it easier, what do we need to do in terms of coordination, staffing, and/or technological resources? How can tools like data quality flags better accommodate the needs of data-providers as well as data-users elsewhere along the collections data pipeline?

We present a real life case study of repatriating an enhanced dataset to its institute of origin, including details on timelines, estimates of effort, and lessons learned. The best laid repatriation protocols might not prepare us for everything, but following them more closely might save us some sanity.

*Repatriation of Augmented Information to an Institutional Database.*



Available from:

[https://www.researchgate.net/publication/325745168\\_Repatriation\\_of\\_Augmented\\_Information\\_to\\_an\\_Institutional\\_Database](https://www.researchgate.net/publication/325745168_Repatriation_of_Augmented_Information_to_an_Institutional_Database) [accessed Jun 22 2018].

STEP 1 - You ask us for data and/or images (sometimes, 'cause we make it all public so you can just come get it yourself. So I guess we like that and hate that.) but we really don't know what the heck you are going to do with it. The project description is often vague and/or over ambitious. Possibly because they are loosely defined describing what you are going to do to my data after I send my data to you

- annotation?
- augmentation?
- value added?
- metadata?
- crowd-sourcing
- transcription?
- automated validation? (take lots of peoples opinions and come up with an average)
- 

STEP 2 - Now we can send data out in a standard format but it wasn't always the case that time was a) easy to do, b) easy to understand, c) commonly known. In 2010 Darwin Core was VERY new and we still had to use DiGir which was only great if you had a degree in computer science.

STEP 3 - No standard way of documenting the augmentation

STEP 4 - No standard way to physically get the data back



## ***Carl Linnaeus***

- Carl Linneaus
- Carl von Linné
- Carolus Linnæus
- Carolus a Linné
- L.
- Carolus Linneaus



CC-0 Public Domain: ["The 32-year-old Linnaeus in his wedding finery. Oil painting by J. H. Scheffel. 1739."](#)

We've been collating and naming and standardising for a very long time.

Linneaus - Mr Standards! Irony of ironies...

Artwork Reference:

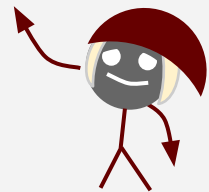
<https://commons.wikimedia.org/wiki/File:LinnaeusWeddingPortrait.jpg>



## ***Dante Alighieri***



CC-0 Public Domain: ["Portrait of Dante Alighieri, Sandro Botticelli"](https://commons.wikimedia.org/wiki/File:Portrait_of_Dante_Alighieri_Sandro_Botticelli)



Artwork Reference: [https://commons.wikimedia.org/wiki/File:Portrait\\_de\\_Dante.jpg](https://commons.wikimedia.org/wiki/File:Portrait_de_Dante.jpg)

- The divine Comedy: A **narrative poem** by **Dante Alighieri**, begun c. 1308 and completed in 1320, a year before his death in 1321.
- The narrative describes Dante's travels through the 9 levels each of Hell, Purgatory, and Paradise or Heaven.

Let me introduce Dante Alighieri our guide for the next 15 minutes or so. In





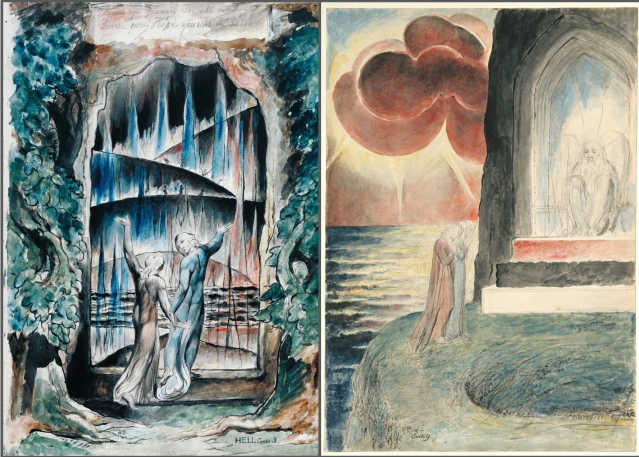
#### Artwork References:

Hell:

[CC-BY-NC-ND \(3.0 Unported\)](#) - [N03352 The Inscription over the Gate 1824–7 \[A00005-A00011; N03351-N03370; T01950-T01956; complete\]](#)

<https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>





## Artwork References:

### Hell:

[CC-BY-NC-ND \(3.0 Unported\)](#) - [N03352 The Inscription over the Gate 1824–7 \[A00005-A00011; N03351-N03370; T01950-T01956; complete\]](#)  
<https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>

### Purgatory:

[CC-BY-NC-ND \(3.0 Unported\)](#) - [N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory 1824–7](#)  
<https://www.tate.org.uk/art/artworks/blake-dante-and-virgil-approaching-the-angel-who-guards-the-entrance-of-purgatory-n03367>





## Artwork References:

### Hell:

[CC-BY-NC-ND \(3.0 Unported\)](#) - [N03352 The Inscription over the Gate 1824–7 \[A00005-A00011; N03351-N03370; T01950-T01956; complete\]](#)  
<https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>

### Purgatory:

[CC-BY-NC-ND \(3.0 Unported\)](#) - [N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory 1824–7](#)  
<https://www.tate.org.uk/art/artworks/blake-dante-and-virgil-approaching-the-angel-who-guards-the-entrance-of-purgatory-n03367>

### Heaven:

[CC-BY-NC-ND \(3.0 Unported\)](#) - [N03369 Beatrice Addressing Dante from the Car 1824–7](#)  
<https://www.tate.org.uk/art/artworks/blake-beatrice-addressing-dante-from-the-car-n03369>





## SETTING THE SCENE:



So let's start our journey through data repatriation as Dante did, on the edge of hell...

Hell:

[CC-BY-NC-ND \(3.0 Unported\)](#) - [N03352 The Inscription over the Gate 1824–7 \[A00005-A00011; N03351-N03370; T01950-T01956; complete\]](#)  
<https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>



# SETTING THE SCENE:

1980-90s



CC-BY-NC-ND (3.0 Unported) William Blake: <https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>

Prior to the 1980s and 1990 what we now call biodiversity information was collated in very individualistic, function specific ways. Systems were connected institutionally by word of mouth at best and externally not at all. Everyone was king of their own data.



## 1st Level of Hell - Limbo



1  
CC-0 Public Domain: Gustave Doré

I like to think of this period as “Limbo”, the first Circle of Hell, which contains the virtuous pagans who were not sinful but were ignorant of Christ.

It wasn't that folks were wilfully doing anything wrong they were simply ignorant of the fact that connectedness was useful, beneficial and ultimately the only way forward.



# SETTING THE SCENE:

1980-90s

[...TDWG...]



CC-BY-NC-ND (3.0 Unported) William Blake: <https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>

As we move through the 80s and 90s the community starts to connect digitally. Along comes the Taxonomic Databases Working Group (mostly Botanists). Disciplines that were only connected by word of mouth and expert personal (dare I say) old boys networks start to work in a coordinated digital way.

Eg Barcodes : Index Herbariorum goes back to the 1930s.

<http://apps.kew.org/herbcat/gotoWhyPlantsLabeled.do>

<https://www.nybg.org/science-project/index-herbariorum-upgrade/>



# SETTING THE SCENE:

1980-90s

[...TDWG...]

Early 2000s

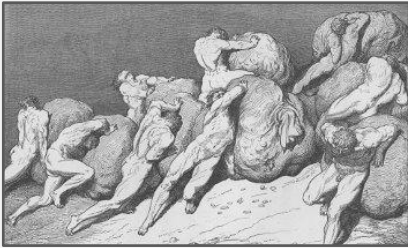


CC-BY-NC-ND (3.0 Unported) William Blake: <https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>

Things really start to crank up in the early 2000s databases galore, standards and systems to deploy data



## 2nd to the 5th Levels of Hell - The Sins of Indulgence



1  
CC-0 Public Domain: Gustave Doré

At this point we enter what I call the era of indulgence. Levels 2 - 5 of hell. I suspect most of the acronyms and abbreviations that we throw around today are generated in this period...



# SETTING THE SCENE:

1980-90s

**[...TDWG...]**

Early 2000s

**[...GBIF...]**



CC-BY-NC-ND (3.0 Unported) William Blake: <https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>



# SETTING THE SCENE:

1980-90s

**[...TDWG...]**

Early 2000s

**[...GBIF...]**

Late 2000s

**[...Darwin Core / ABCD / IPT...]**



CC-BY-NC-ND (3.0 Unported) William Blake: <https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>

History: <http://community.canadensys.net/2011/new-terms-in-darwin-core>



# SETTING THE SCENE:

1980-90s

**[...TDWG...]**

Early 2000s

**[...GBIF...]**

Late 2000s

**[...Darwin Core / IPT...]**

2000s

**[...MORE... of EVERYTHING...]**



CC-BY-NC-ND (3.0 Unported) William Blake: <https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>

## References:

<http://historylists.org/art/9-circles-of-hell-dantes-inferno.html>

Image: CC-BY-NC-ND (3.0 Unported) William Blake:

<https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>

History: <http://community.canadensys.net/2011/new-terms-in-darwin-core>



## 7th to the 9th Levels of Hell - The Sins of Violence & Malice



1  
CC-0 Public Domain: Gustave Doré

Circle 7 for the sins of violence and Circles 8 and 9 for the sins of malice (fraud and treachery)





USE CASE:



Purgatory: Humans can sin by using love towards improper or malicious ends ([Wrath](#), [Envy](#), [Pride](#)), or using it to proper ends but with love that is either not strong enough ([Sloth](#)) or love that is too strong ([Lust](#), [Gluttony](#), [Greed](#)). Below the seven purges of the soul is the Ante-Purgatory, containing the Excommunicated from the church and the Late repentant who died, often violently, before receiving rites.

Artwork References:

Hell:

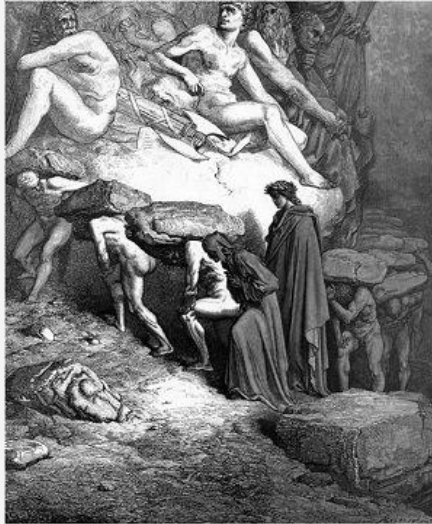
[CC-BY-NC-ND \(3.0 Unported\)](#) - [N03352 The Inscription over the Gate 1824–7 \[A00005-A00011; N03351-N03370; T01950-T01956; complete\]](#)  
<https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352>

Purgatory:

[CC-BY-NC-ND \(3.0 Unported\)](#) - [N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory 1824–7](#)  
<https://www.tate.org.uk/art/artworks/blake-dante-and-virgil-approaching-the-angel-who-guards-the-entrance-of-purgatory-n03367>



## 1st Level Purgatory - Pride



1

Staff at the museum always been active members of the community especially the vertebrate groups. At some time probably in 2006(ish) data was sent from our Fishes, Birds, Mammals and Herps Divisions to the Nets/ISs

<http://vertnet.org/about/classicnetworks.html> to be georeferenced

<http://www.ornisnet.org/georeferencing/georefworkflow>

Humans can sin by using love towards improper or malicious ends ([Wrath](#), [Envy](#), [Pride](#))



# USE CASE

Mid 2000-zeros [STUFF GOES OUT]



[CC-BY-NC-ND \(3.0 Unported\)](#) - N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory

In 2005 each collection was using its own database using a variety of platforms maintained by a variety of people.

Birds - CBASE

Herps - Access/SQL Server

Mammals - ???

Fishes - ???

All flat datasets.



# USE CASE

- Mid 2000-zeros [STUFF GOES OUT]

- [...variety is the spice of life?..]



[CC-BY-NC-ND \(3.0 Unported\)](#) - N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory

Variety is the spice of life...



## USE CASE:



```

,,,,,,,,,,,,,1,,,,,,,,,FMNH,,,,,,,,,163,"Animalia",38.78302,,,,,"adult unsexed",,,,,-120.508168,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropo
,,,,,,,,,"Dr. Jiri Hava : National Museum in Prague",,,,,1,,,,,,,,,FMNH,,,,,,,,,106,"Animalia",37.232484,,,,,"adult unsexed",,,,,-121.682546,,,
Oygur",,,,,1,,,,,,,,,FMNH,,,,,,,,,159,"Animalia",38.78302,,,,,"adult unsexed",,,,,-120.508168,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,FMNH,,,,,,,,,344,"Animalia",,,,,,"adult unsexed",,"Lake Lucy",,,,,,,,,,12,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",,,,,,"No",,,,,,"No",,,,,,"No",
edius (Fall, 1995),,,,,,"intermedius",,"intermedius",,,,,,,,,,Liocellus",,,,,,"No",,,,,,"No",,,,,,"No",,,,,,"No",,,,,,"No",,,,,,"No",,,,,,"No",
Jiri Hava",,,,,1,,,,,,,,,FMNH,,,,,,,,,153,"Animalia",40.717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,FMNH,,,,,,,,,adult male",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",,,,,,"Grainger Voice",,,,,,"Nicrophorus carolinus Linnaeus, 1771",,"Linnaeus, 1771",,,,,,"adult uns
optera",,,,,,"Grainger Voice",,,,,,"Nicrophorus orbicollis Say, 1825",,"Say, 1825",,,,,,"adult unsexed",,,,,,9,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",,,,,,"Insecta",,,,,,"Haliplus leechi Wallis
malia",,,,,,"adult unsexed",,"Salmon Arm",,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",,,,,,717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,1,,,,,,,,,FMNH,,,,,,,,,153,"Animalia",40.717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
Jiri Hava",,,,,1,,,,,,,,,FMNH,,,,,,,,,153,"Animalia",40.717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
e",,,,,1,,,,,,,,,FMNH,,,,,,,,,153,"Animalia",40.717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,1,,,,,,,,,FMNH,,,,,,,,,153,"Animalia",40.717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
ae",,,,,1,,,,,,,,,FMNH,,,,,,,,,153,"Animalia",40.717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,8899,,,,,"adult unsexed",,,,,-118.0009,,,,,,,,,8,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",,,,,,Animalia",37.925586,,,,,"adult unsexed",,,,,-121.984171,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,ate University",,,,,1,,,,,,,,,FMNH,,,,,,,,,154,"Animalia",37.925586,,,,,"adult unsexed",,,,,-121.984171,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,1,,,,,,,,,FMNH,,,,,,,,,154,"Animalia",37.925586,,,,,"adult unsexed",,,,,-121.984171,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
Jiri Hava",,,,,1,,,,,,,,,FMNH,,,,,,,,,153,"Animalia",40.717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
1,,,,,,,,,FMNH,,,,,,,,,213,"Animalia",38.697559,,,,,"adult unsexed",,,,,-121.89838,,,,,,,,,9,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
Jiri Hava : National Museum in Prague",,,,,1,,,,,,,,,FMNH,,,,,,,,,153,"Animalia",40.717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,FMNH,,,,,,,,,325,"Animalia",,,,,,"adult unsexed",,"Lake Lucy",,,,,,,,,,11,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",,,,,,"A. Smetana : Agriculture Canada",,,,,1,,,,,,,,,FMNH,,,,,,,,,149,"Animalia",41.728546,,,,,"adult unsexed",,,,,-123.933082,,,,,,,,,5,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
t unsexed",,"Thunder Bay",,,,,,,,,,8,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",,,,,,"Helocombus bifidus (J. L. LeConte",,,,,,"Helocombus bifidus (J. L. LeConte",
,,,,,,,,,1,,,,,,,,,FMNH,,,,,,,,,201,"Animalia",37.925586,,,,,"adult unsexed",,,,,-121.984171,,,,,,,,,7,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,FMNH,,,,,,,,,170,"Animalia",39.348113,,,,,"adult unsexed",,,,,-120.708327,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
Jiri Hava : National Museum in Prague",,,,,1,,,,,,,,,FMNH,,,,,,,,,153,"Animalia",40.717985,,,,,"adult unsexed",,,,,-122.183344,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
,,,,,,,,,212,"Animalia",,,,,,"adult unsexed",,"Val-Morin",,,,,,,,,,7,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",,,,,,"Cercyon (Cer
,,,,,,,,,1,,,,,,,,,FMNH,,,,,,,,,158,"Animalia",,,,,,"adult unsexed",,"Edmonton",,,,,,,,,,6,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",
ersity",,,,,1,,,,,,,,,FMNH,,,,,,,,,262,"Animalia",,,,,,"adult unsexed",,"Lumby",,,,,,,,,,9,,,,,,,,,Coleoptera,,,,,,,,,Arthropoda",

```



# USE CASE:

- Mid 2000-zeros [STUFF GOES OUT]
- **[...variety is the spice of life?..]**
- The 2000-tens [WE DID SOME STUFF]



[CC-BY-NC-ND \(3.0 Unported\)](#) - N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory



## USE CASE:

The image displays a museum database interface with three overlapping windows. The leftmost window, titled 'Catalogue (1) - Display', shows a table of records. A red box highlights the 'Internal Rec.' column, and another red box highlights the 'Summary Date' column. Red arrows point from these boxes to the other two windows. The top-right window, titled 'Collection Events (1) - Display', shows details for event [6697] with an internal record number of 1819407. The bottom-right window, titled 'Collection Events (1) - Display', shows details for event [6696] with an internal record number of 1834350. Both of these windows also have red boxes highlighting their internal record numbers, with red arrows pointing to labels 'Internal Record Number'.

Internal Record Number

Internal Record Number

Internal Record Number

Catalogue Numbers

So data held at the Field Museum is no longer in the same form that was sent out.

Data is now held in relational structure.



# USE CASE:

Collection Events (1) - Display

File Edit Select View Tools Tabs Multimedia Window Help Standards

[6696] South America, Venezuela, Aragua: Rancho Grande, 1200'

1834350

Collection Details

Collection Type:

Site:

Date Visited (From):  Date Visited (To):

Time Visited (From):  Time Visited (To):

Other Numbers

Other Numbers	Description

Bibliographic References

Collection:  Expedition:  Notes:  Aquatic:  Terrestrial:  Sig:

Display | Event 2 of 5

Sites (1) - Display

File Edit Select View Tools Tabs Multimedia Window Help Standards

South America, Venezuela, Aragua: Rancho Grande, 1200'

381543

Locality Details

Continent:

Country:

Province/State/Territory:

District/County/Shire:

City/Town:

Nearest Named Place:

Biogeography/Region:

Precise Location

Verbatim Locality

Remarks

Aquatic Details

Ocean:

Sea/Gulf:

Bay/Sound:

Drainage Basin:

Islands

Island Grouping:

Island Name:

Freshwater Details

Drainage Division:

Drainage Code:

River Basin:

River Code:

Site Locality Locality 1 Mapping Lat/Long Habitat Cultural Strat

Display | Site 426 of 426

sgrant Admin emufmnh

Internal Record Number



# USE CASE:

- Mid 2000-zeros [STUFF GOES OUT]
- [...variety is the spice of life?..]
- The 2000-tens [WE DID SOME STUFF]
- [...things change...]



[CC-BY-NC-ND \(3.0 Unported\)](#) - N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory

The Museum officially sanctions and provides ongoing support for a museum-wide collection management system.

- Fishes goes into it in 2007,
- Mammals in 2009,
- Birds in 2010,
- Amphibians and Reptiles in 2011.



# USE CASE:

- Mid 2000-zeros [STUFF GOES OUT]
- [...**variety is the spice of life?..**]
- The 2000-tens [WE DID SOME STUFF]
- [...**things change...**]
- 2010 [OTHER PEOPLE DID STUFF]



[CC-BY-NC-ND \(3.0 Unported\)](#) - N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory



## ORNIS Georeference Repatriation

GBIF/VertNet x GBIF/VertNet/ORNIS x



John Wiczorek <tuco@berkeley.edu>

09/04/2010 ☆

Reply

to tuco

Dear all,

A noteworthy day has arrived. Georeferences from the ORNIS project are ready for repatriation. A document describing the results (<http://olla.berkeley.edu/ornisnet/?q=node/31>) and a guide to repatriation (<http://olla.berkeley.edu/ornisnet/?q=node/32>) are now available on the ORNIS web site.

We are fortunate to have the support of VertNet (<http://www.vertnet.org>), and more specifically, Laura Russell (cc'd), to assist with repatriation for any institutions that need it. In order to know who needs help and who does not, and to schedule this monumental task, please reply with the following information:

Database management system:

Operating system:

Web server:

Technical contact for repatriation:

When:

HerpNet:

The first three of these are not only to complete our records, but also to help in grouping institutions for similar repatriation methods. The "Technical contact" should be the person with whom we will need to work to get the data into the working databases. "When" should be a range of dates that would be reasonable to do the repatriation work, given any constraints, windows of opportunity and the like. If your institution has a pending HerpNet georeference repatriation which could be done in conjunction with that for ORNIS, please indicate this next to "HerpNet."

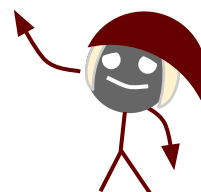
We would like to extend our heartfelt thanks for the diligent hard work and remarkable results from the georeferencers and those who managed them. Special thanks to Heather Constable who held the reins when John was in the field, and to Michelle Koo who.

We will be in further contact case by case as we accumulate these responses and take the next steps.

With excitement to see the fruition of all of this hard work.

John, Town, and Carla

PS: If this message has reached you in error and you believe it should be sent to someone else, please reply and let John know to whom it should be addressed.





## ORNIS Georeference Repatriation

GBIF/VertNet x GBIF/VertNet/ORNIS x



John Wiczorek <tuco@berkeley.edu>

09/04/2010

Reply

to tuco

Dear all,

A noteworthy day has arrived. Georeferences from the ORNIS project are ready for repatriation. A document describing the results (<http://olla.berkeley.edu/ornisnet/?q=node/31>) and a guide to repatriation (<http://olla.berkeley.edu/ornisnet/?q=node/32>) are now available on the ORNIS web site.

We are fortunate to have the support of VertNet (<http://www.vertnet.org>), and more specifically, Laura Russell (cc'd), to assist with repatriation for any institutions that need it. In order to know who needs help and who does not, and to schedule this monumental task, please reply with the following information:

Database management system:  
Operating system:  
Web server:  
Technical contact for repatriation:  
When:  
HerpNet:

The first three of these are not only to complete our records, but also to help in grouping institutions for similar repatriation methods. The "Technical contact" should be the person with whom we will need to work to get the data into the working databases. "When" should be a range of dates that would be reasonable to do the repatriation work, given any constraints, windows of opportunity and the like. If your institution has a pending HerpNet georeference repatriation which could be done in conjunction with that for ORNIS, please indicate this next to "HerpNet."

We would like to extend our heartfelt thanks for the diligent hard work and remarkable results from the georeferencers and those who managed them. Special thanks to Heather Constable who held the reins when John was in the field, and to Michelle Koo who.

We will be in further contact case by case as we accumulate these responses and take the next steps.

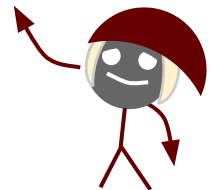
With excitement to see the fruition of all of this hard work,

John, Town, and Carla

PS: If this message has reached you in error and you believe it should be sent to someone else, please reply and let John know to whom it should be addressed.

9th April or the 4th September 2010??

A noteworthy day has arrived. Georeferences from the ORNIS project are ready for repatriation. A document describing the results (<http://olla.berkeley.edu/ornisnet/?q=node/31>) and a guide to repatriation (<http://olla.berkeley.edu/ornisnet/?q=node/32>) are now available on the ORNIS web site.





# USE CASE:

Institution	CollectionCode	Country	StateProvince	County	Island	IslandGroup	Locality	VerbatimLatitude	VerbatimLongitude
Field Museum of Natural History	100296	USA	Iowa	Marshall Co			Coraville	41.6764145	-82.51
Field Museum of Natural History	100298	USA	Florida	Sarasota Co			Manatee	27.4967175	-82.51
Field Museum of Natural History	100300	USA	Florida	Sarasota Co			Manatee	27.4967175	-82.51

FMNHRepatriation.txt

Repatriation-CA-US-FMNH.txt

MaNISGeorefsForRepatriation-FMNH.txt





ORNIS

Navigation

Home

Contacts

Data Providers

Georeferencing

Workflow (< 26 Oct 2007)

Workshops

Repatration Guide

ORNIS Repatriation

HBA Repatriation

Repatration Process

Repatration References

Results

Links

Software Tools

Taxonomic Standardization

Locality Consistency

Species Ecology

Collector Itineraries

Sitemap

Georeferencing > Repatriation Guide >

Repatration Process

Following is a rough outline of recommended steps to assimilate the georeferences into your working database. The same steps should be used for both the ORNIS and HBA georeferences.

1. Download the zip file(s) for your institution from the [Table of ORNIS Georeferencing Results](#) and [Table of Mexican Bird Atlas Results](#).
2. Extract the text file from the zip file using the password provided.
3. Create a database table from the text file.
4. Create a database table from the locality-related fields from your working database.
5. Join the two tables by the Cataloglumber field.
6. Create a new field that is the a concatenation of all location-based fields in 5.
7. From 5), create a table of distinct variations of the combination of georeference localities and current working database localities (a "Locality Combo" table) with the field as in 6 as a unique key.
8. Add a field "Same" and a field "Checked" to the Locality Combo table. The goal is to check all records for "sameness." The "Same" field will be used to flag records where the georeference locality and the current locality from the working database are semantically the same. The "Checked" field will be used to flag when a record has been tested for "sameness."
9. For Localities that are the same character-by-character before and after georeferencing, set Same=Yes and Checked=Yes.
10. Check each record not having Checked=Yes to determine if the the localities before and after are semantically the same. If they are the same, set Same=Yes and Checked=Yes, otherwise set Same=No and Checked=Yes.
11. When all records in the Locality Combo table have Checked=Yes, use the key relating to occurrence records to join the georeferences with Same=Yes back to the occurrence records.
12. Update the records in your working database using the repatriated georeference fields.

To preserve the high quality of georeferences produced under the ORNIS Project and make these data fit for the widest possible range of uses, every effort should be made to retain data from all of the georeference fields (DecimalLatitude, DecimalLongitude, GeodeticDatum, CoordinateUncertaintyInMeters, VerbatimCoordinateSystem, GeoreferenceProtocol, GeoreferenceRemarks, GeoreferenceSources, GeoreferenceDate\*, GeoreferenceBy, GeoreferenceVerificationStatus, NoGeorefBecause\*). Full documentation of the meaning of the fields not marked with an asterisk can be found in the [Darwin Core Quick Reference Guide](#) ([Darwin Core Task Group 2009](#)).

The fields beginning with "Standard" are simply standardized versions of the values in the associated original fields. These standard versions were used in the georeferencing proces to assist automation using the batch processing mode of the [BioGeomancer Workbench](#) (BioGeomancer Consortium 2006).

The ADM\_n fields are useful for checking your original data. The values in these fields are standards from the GADM datasets and were determined by the location of the DecimalLatitude and DecimalLongitude of the georeference. If your higher geographies do not match the values in these fields, it means either that the georeference is wrong, or that your geography information is incorrect, or that your geography information doesn't use standard values.

The EEZone field contains a value if the DecimalLatitude and DecimalLongitude of the georeference falls within an exclusive economic zone off of the coast rather than on land. If the occurrence





1. We constructed a [gazetteer](#) of unique localities in 2003/2016 the data you sent us for your collection. We standardized some of the country, state and county information for these. Refer to the [Georeferencing Checklist](#).  
2. Other institutions downloaded records from the HerpetNet Gazetteer, and georeferenced them following the [HABITAT/HERPNET/DONORS/HMOO](#) guidelines and the [Georeferencing Steps](#). Data to the [Georeferencing Checklist](#) for a list of who georeferenced what.  
3. The localities were then returned to HNPV, and we validated them for the correct data structure, whether they followed HMOO guidelines, and whether they fell in the administrative boundaries (Country, State, County) of the original gazetteer. If they did not, we corrected the values and re-uploaded them to the database.  
4. In 2008 most institutions sent a new copy of their database. If you sent this copy, we matched up the 2003 and the 2008 databases and marked which localities were no longer the same. These are listed in the [HerpetNet Gazetteer](#) under "Discontinued".  
5. We also received a number of handwritten notes from our contributors. We scanned these and uploaded them as PDF files. Please email us if you have additional scanned copies of these tab-delimited PDF files we sent via e-mail to curators and collection managers at participating institutions. **Please keep a copy of these files somewhere safe as a future reference for georeferencing questions and please read these descriptions below carefully.** Please direct any questions about these fields to the HerpNet coordinator.  
6. Institutions should update their data to include more than 250 characters and you will have to adjust for this in your database first so that you do not cut off these fields when importing the data. These fields are LatLongRemarks, LocalityAnnotations, DeterminationDate and NOGEOGRAPHICUSE.  
7. Institutions should evaluate the files they received and upload their database and the copy of their database connected to their HerpNet server as soon as possible. We recommend putting these data up right away, as you can use the HerpNet portal to check upon the data using the mapping feature.

**Fields found in Repatriation and Exception Files:**

**Unique Identifier Fields** - These fields exist to help you match the georeferenced data to your collection catalog.

Institution - Name of originating institution.

**CollectionCode** - An alphabetic code used during the HerpNet project indicating your collection and collection type (for example, if MVZ had separate Amphibian and Reptile collections, collection code would be MVZ-A and MVZ-R).

**CatNum** - Catalog Number or number indicated by your institution as unique identifier for the specimen. This may be a field number or locality number for some collections  
**HermNETCatalogueLocalityID** - A unique identifier used by HermNET to condense identical verbatim locality descriptions into one record per institution per unique locality

HerpNETGazetteerLocalityID - A unique identifier used by HerpNET to condense identical verbatim locality descriptions into one record per institution per unique locality.

**Georeferencing Fields** - These 13 fields are required for georeferenced data to comply with the MaNIS/ HerpNet/ ORNIS Georeferencing Guidelines. Please include all of these fields in your database or exclude them from the submission file. Without all of these fields, the georeferencing is incomplete. There are

Please include **all** of these fields in your database and update them from the repatriation file. Without **all** of these fields, the georeference is incomplete. These are necessary information during the georeferencing process, and without them, the georeference will not be certified with the [Halis/ HergNE7/ ORNIS Georeferencing Guidelines](#). By keeping these fields, you are allowing users of your data the ability to judge the quality of the georeference and ensuring the repeatability of the georeferencing process.

DeclLat - Latitude in decimal degrees.

**DecLong** - Longitude in decimal degrees.

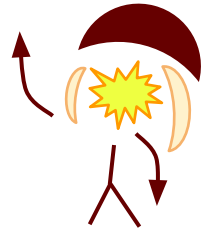
**MaxErrorDistance** - Uncertainty is a distance calculated according to MHO Guidelines.

Extent - Distance from center of named place to outer most edge according to maps or gazetteers.  
MaxErrorUnits - Length units for MaxErrorDistance and Extent

**Datum** - This is the three-dimensional projection of the globe

**Datum** - This is the three-dimensional projection of the globe associated with the coordinates. This information must be included with the Declat and Declong field, or coordinates will not plot correctly.

**DeterminationRef** - Reference map, website or gazetteer used in the georeferencing process to find coordinates. This field may contain more than 250 characters.





Search San FishNet Services  
Contact Us About Georeferencing

### Collaborative Georeferencing Project: Repatriation Phase

At the onset of the project, 1.3 million specimen lots were identified for georeferencing and harvested from network data providers. Georeferencing was primarily accomplished using the Globalstar Collaborative Georeferencing (CCG) framework. Each record was automatically georeferenced with GEOCODE, then evaluated by project technicians and either corrected or skipped to produce a final result. Skipped records represent those records lacking sufficient information for adequate georeferencing by a technician. Technicians were asked to provide comments on each skipped record, explaining why the record was skipped. Each corrected record internally contains a corrected latitude/longitude and a value of uncertainty in meters. Additionally, technicians were encouraged to specify copyright (for example along the course of rivers) to further refine the level of uncertainty anytime the uncertainty values significantly overestimated uncertainty.

Critical to the success of the project is separating the georeferencing and uncertainty – presently stored in the CCG portal – to databases of all participating FishNet2 data providers. This is the only way that the results can be propagated to FishNet2 and other data harvesters. Both skipped and corrected records are provided below in a variety of formats. All geographic coordinates are reported in WGS 84 decimal degrees. Primitives defined using the final export "Net DWT" format. Each record contains the name of the technician who produced the record and the date and time the record was produced. Results are provided in the form of a decision "for" records. Catalog numbers are provided to assist against institutional databases. We are happy to work with institutions to provide alternative means for matching records with their existing database systems (e.g., associating internal "collecting event" identifiers to records). We also expect that programs may present challenges for many institutions and we advise any institution requesting help.

With the completion of this final data repatriation step, users of FishNet2, as well as users of data from participating institutional databases, will have access to the more than 1.1 million additional species lot occurrences produced in the FishNet2 Collaborative Georeferencing Project. Your cooperation is greatly appreciated.

The following table summarizes available fields by record type (corrected or skipped):

Field Name	Corrected	Skipped	Description of Field
verifiedBy	present	present	The collaborative georeferencing username of the technician who verified the record.
dateVerified	present	present	Date and time verification was completed as recorded by the collaborative georeferencing server using central standard time. If you desire to retain this field and need help converting to UTC or another time format compatible with your database system, please let us know.
verificationRemarks	present	present	Remarks provided by the technician.
longitude	present	absent	Longitude assigned by technician.
latitude	present	absent	Latitude assigned by technician.
uncertaintyRadius	present	absent	Uncertainty radius in meters assigned by technician.
uncertaintyPolygon	present	absent	Uncertainty polygon assigned by technician.
institutionCode	present	present	Institution code provided to FishNet at project inception.
collectionCode	present	present	Collection code provided to FishNet at project inception.
catalogNumber	present	present	Catalog number provided to FishNet at project inception.
scientificName	present	present	Scientific name provided to FishNet at project inception.
country	present	present	Modified form of the country field provided to FishNet at project inception, for the purpose of improved georeferencing.
stateProvince	present	present	Modified form of the state/province field provided to FishNet at project inception, for the purpose of improved georeferencing.
locality	present	present	Modified form of the locality field provided to FishNet at project inception, for the purpose of improved georeferencing.

We expect most institutions to be able to utilize the catalog number field to match up results with existing systems. We are also happy work with any institution to provide alternative means for matching up records with their existing database systems, for example associating internal "collecting event" identifiers to records. We also expect consuming programs may present new challenges for many institutions and we advise any institution requesting help.

#### RESULTS

The following datasets are available for download (zip compressed). All CSV files are UTF-8 encoded. If you have any questions, problems or discover errors in the data, please contact Justin Helm (jhelm@fishnet.us).

Institution	Total Records Evaluated	% Corrected	Overview	Downloads
Cornell University Museum of Vertebrates	72,525	93%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
CSIRO Marine Research	16,998	93%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a>
Field Museum	57,733	89%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
Florida Museum of Natural History	68,189	96%		MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
Florida Fish and Wildlife Research Institute	3,483	89%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
Gunnels Museum of Natural History	17	88%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
Haskins University	50,564	73%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
Illinois Natural History Survey	74,846	93%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
Kaplanian University Museum	18,024	98%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
Louisiana State University Museum of Zoology	13,406	95%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
Michigan State University Museum	6,082	89%		CSV: <a href="#">Corrections Skips</a> MS Access DB: <a href="#">ACCDB</a> Shape File: <a href="#">SHP(points)</a> <a href="#">SHP(polygons)</a>
Mississippi Museum of Natural Science	33,993	88%		CSV: <a href="#">Corrections Skips</a>

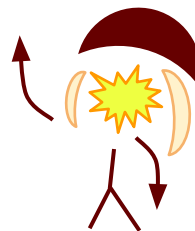
We expect most institutions to be able to utilize the **catalog number** field to match up results with existing systems.

We are also happy work with any institution to provide alternative means for matching up records with their existing database systems, for example associating internal "collecting event" identifiers to records.

We also expect consuming polygons may present new challenges for many institutions and will advise any institution requesting help.

Institution: Field Museum  
Total Records Evaluated: 57,733  
% Corrected: 89%

CSV: [Corrections Skips](#)  
MS Access DB: [ACCDB](#)  
Shape File: [SHP\(points\)](#) [SHP\(polygons\)](#)









# USE CASE:

Mid 2000-zeros [STUFF GOES OUT]

**[...variety is the spice of life?..]**

The 2000-tens [WE DID SOME STUFF]

**[...things change...]**

2010 [OTHER PEOPLE DID STUFF]

**[...prodigal returns...]**



[CC-BY-NC-ND \(3.0 Unported\)](#) - N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory



# USE CASE:

Mid 2000-zeros [STUFF GOES OUT]

**[...variety is the spice of life?..]**

The 2000-tens [WE DID SOME STUFF]

**[...things change...]**

2010 [OTHER PEOPLE DID STUFF]

**[...prodigal returns...]**

The 2000-teens [PEOPLE CHANGE]

**[...more things change...]**



[CC-BY-NC-ND \(3.0 Unported\)](#) - N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory

Three new collections managers

- Birds
- Fishes
- Mammals

New trained data handlers on staff



# USE CASE:

Sites (1) - Display

File Edit Select View Tools Tabs Multimedia Window Help Standards

South America, Venezuela, Aragua: Rancho Grande, 1200'

381543

Latitude/Longitude Details

	Latitude (DMS)	Longitude (DMS)	Latitude (Dec.)	Longitude (Dec.)	Latitude Verbatim	Longitude Verbatim	Zone	Easting	Northing	Modifier	Comment
1	10 20 46.337 N	67 41 10.413 W	10.34620476	-67.68622589	10.34944	-67.68444					degrees minutes seconds
*											

Determination Source: Topographic Map 1:100,000 Hoja 6647 (Ocumare), Direccion de Cartografia Nacional, Venezuela

Determination Method: MaNIS/HerpNet/ORNIS Georeferencing Guidelines

Determined By: HerpNet Georeferencing Team

Determination Date: 14-Jul-2004

Centroid Latitude (DMS): 10 20 46.337 N

Centroid Longitude (DMS): 67 41 10.413 W

Notes: Coordinates for Rancho Grande Biological Station; extent includes a bigger area called Rancho Grande

Latitude/Longitude List

	Determination Source	Centroid Latitude (DMS)	Centroid Longitude	Preferred
1	Topographic Map 1:100,000 Hoja 6647 (Ocumare), Direccion de Cartografia Nacional, Venezuela	10 20 46.337 N	67 41 10.413 W	Yes
*				

Site Locality Locality 1 Mapping Lat/Long Habitat Cultural Stratigraphy Notes Multimedia Security Audit

Display Site 426 of 426



381543



# USE CASE:

Sites (1) - Display

File Edit Select View Tools Tabs Multimedia Window Help Standards

South America, Venezuela, Aragua: Rancho Grande, 1200' 381543

Audit Trail

	User	Date	Time	Operation
1	sgrant	13 Aug 2018	19:23	update
2	sgrant	13 Aug 2018	17:50	update
3	kwebbink	07 Aug 2017	14:41	update
4	sgrant	17 Jan 2014	10:42	update
5	sgrant	11 Sep 2012	16:42	update
6	sgrant	11 Sep 2012	16:41	update
7	sgrant	10 Jan 2012	13:26	update
8	sgrant	04 May 2011	15:54	insert

Cultural Stratigraphy Notes Multimedia Security Audit Admin Poi

Display Site 1 of 1 sgrant Admin emufmnh



381543



# USE CASE:

Mid 2000-zeros [STUFF GOES OUT]

**[...variety is the spice of life?..]**

The 2000-tens [WE DID SOME STUFF]

**[...things change...]**

2010 [OTHER PEOPLE DID STUFF]

**[...prodigal returns...]**

The 2000-teens [PEOPLE CHANGED]

**[...more things change...]**

The 2000-twenties? [WE KEPT DOING STUFF]



[CC-BY-NC-ND \(3.0 Unported\)](#) - N03367 Dante and Virgil Approaching the Angel who Guards the Entrance of Purgatory









## SPECIMEN UPLOAD PROCEDURE

### Creating a New Collection Profile

If this is the very first time data is to be loaded, a system administrator will need to first create a metadata record for the target collection. The easiest way to create a new collection profile is to log-in as a system administrator, navigate to the Collection Profile page (/collections/misc/collprofiles.php), and clicking on the Add New Collection Profile symbol located to the upper right of the page. A link to this page can also be found in the sitemap for the portal (/sitemap.php). Although most of the fields are self explanatory, a few details are explained below.

- Selecting the "allow public edits" will allow any logged in user to edit your collection data. However, edits from those without explicit editing permissions will not be applied until an administrator for the collection reviews and approves the edits. The edit reviewer is available within the administrator control panel for each collection.
- Collection type: General observation is typically used for personal specimen management. This allows field researchers to enter and manage their collection data as "observations" before they are submitted to a collection institution. Specimen labels for these collections can be printed within the portal. After submitting specimens to a physical collection, the receiving institution can transfer these records without having to retype the label data. Typically, there is only one General Observation project per data portal.
- Management: Snapshot is for collections that maintain their central database within the home institution and only feature a snapshot of their data within the portal. Live Data means that the collection manages their data directly within the portal. The portal dataset is their central database.
- The "icon" field contains a url for a small icon used to identify collection. This icon is usually placed in the /images/collection/ folder.
- When available, the "IndividualUrl" field can be used to create a dynamic link to the record page at the source institution. This link is displayed on the Symbiota Specimen Detail Page. Enter a url with "-PK-" placed where the source primary key (dbpk) should be inserted (e.g. <http://sourceinstitution.edu/herbarium/db/specimendetails.php?id=-PK->).

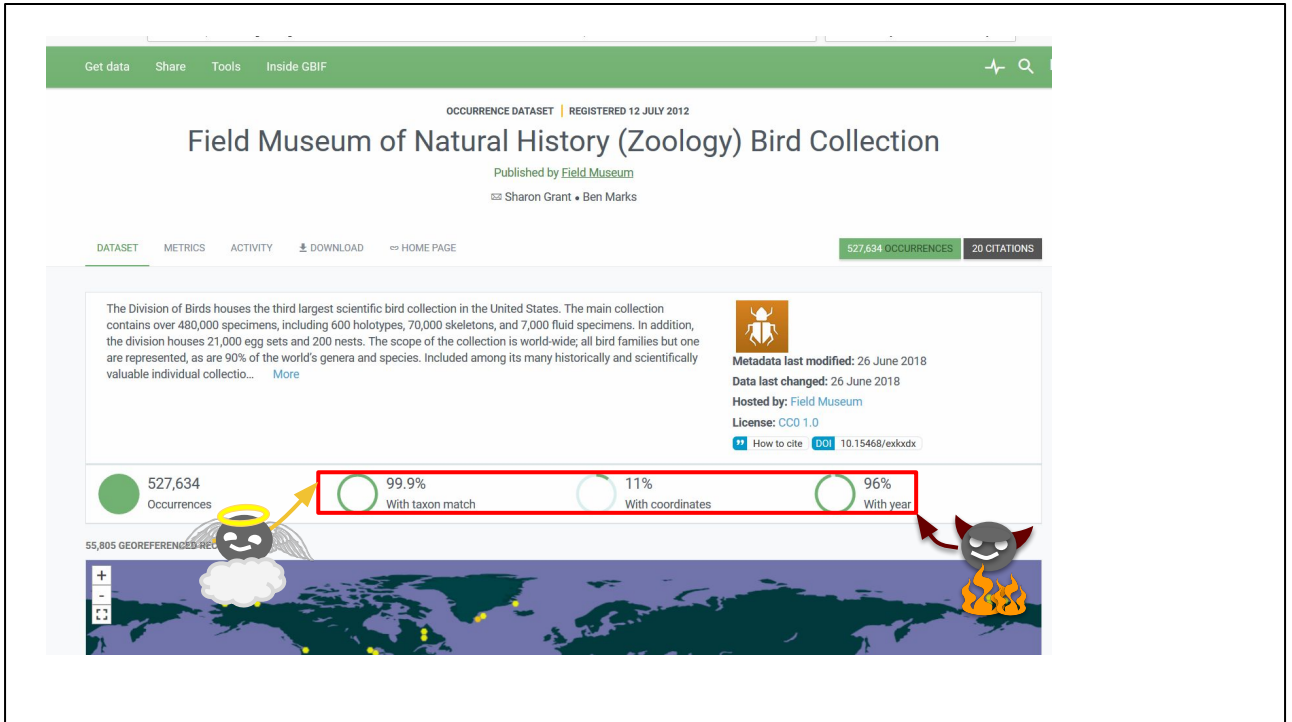
### Field Mapping

File and direct uploads procedures require source and target (Symbiota portal) fields to be mapped to one another. A user interface (see screen snapshot below) is supplied to aid the mapping process. For collections that use a "snapshot" of their central database, the first step is to select the required field that will serve as the primary specimen identifier (primary key) that links the records to the source database with those in the portal. This field can be alphanumeric but must be non-null, unique, and persistent over time. Any records with null or duplicate values will be filtered out and not loaded. Collection that are "live" data sets that managed directly within the portal do not require a primary key field since new records are only appended to the existing data set and not the previously mapped. The majority of Symbiota fields are [Darwin Core](#) compliant in naming and data type; however, upload scripts are capable of converting many incompatible data types to match the Darwin Core definitions. Automap option within interface will match source columns using the most recent Darwin Core naming syntax. Field mappings can be saved, thus maintaining a static upload data format will aid future uploads.



Symbiota portals





## Field Museum Bird Collection

<https://www.gbif.org/dataset/36f15a36-6b45-442e-9c31-cd633423aee0>





















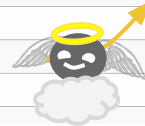




Coordinate precision	Unknown
Coordinate uncertainty in meters	200.0
Location ID	481
Biome	Terrestrial

### Data quality tests

Test name	Result
Geodetic datum assumed WGS84 	 Warning
Basis of record not supplied 	 Passed
Basis of record badly formed 	 Passed
Invalid collection date 	 Passed
Incomplete collection date 	 Passed
First of the month 	 Passed
Missing name of person who identified the specimen/observation 	 Passed
Collector name unparseable 	 Passed
Data are generalised 	 Passed





## Recordset


## Search Recordset

## Field Museum of Natural History (Zoology) Bird Collection

Specimen Records: 527,634

Media Records: C

iDigBio Last Ingested Date: 2017-05-21



Division of Birds houses the third largest scientific bird collection in the United States. The main collection contains over 480,000 specimens, including 600 holotypes, 70,000 skeletons, and 7,000 fluid specimens. In addition, the division houses 21,000 egg sets and 200 nests. The scope of the collection is world-wide; all bird families but one are represented, as are 95% of the world's genera and species. Included among its many historically and scientifically valuable individual collections are the H. B. Conrad, Garret Bird Collection, George and Vase Somers, African collections, C. B. Cory's West Indian collection, the Bishop Collection of North American birds, a large portion of W. Koelz's material from India and the Middle East, and many separate collections from South America, Africa (Hoogstraal from Egypt) and the Philippines (Rabor).

## Contacts

Name	Sharon Grant
Role	Technology Analyst

**Role** Technology Liaison to Scientists  
**Email** [scarr@fieldmuseum.org](mailto:scarr@fieldmuseum.org)

**Email** [agrant@fieldmuseum.org](mailto:agrant@fieldmuseum.org)

**Name** Shannon Hackett  
**Role** Associate Curator of Zoology Birds

**Role** Associate Curator of Zoology, Bird  
**Email** [shackett@fieldmuseum.org](mailto:shackett@fieldmuseum.org)

Name: \_\_\_\_\_ Job: \_\_\_\_\_ Rate: \_\_\_\_\_

Name	John Bates
Role	Curator

**Role** Curator  
**Email** [agrant@feldmuseum.org](mailto:agrant@feldmuseum.org)

Name	Ben Marks
Role	Collection 2

Role	Collection Manager
Email	<a href="mailto:bmacka@fieldmuseum.org">bmacka@fieldmuseum.org</a>

Email: [benmarko@gfz.de](mailto:benmarko@gfz.de)

Name	Laura Russell
Role	Programmer

Role	Programmer
Email	larussell@

Name: Sharon Grant

<b>Name</b>	Sharon Grant
<b>Role</b>	Technology Liaison to S

**Email** [sgrant@fieldmuseum.org](mailto:sgrant@fieldmuseum.org)

<u>Data Corrected</u>	Data Use	Raw
-----------------------	----------	-----

This table shows any data corrections that were performed on this recordset to improve the capabilities of DigBio 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.

File	Records with this Flag	% (of records with this Flag)
the_database	503008	95.3333
the_database_replaced	503008	95.3333
the_parentsnameaged_added	503008	95.3333
the_second_added	503008	95.3333
the_tamoneinstincts_added	503008	95.3333
the_tamounad	503008	95.3333
golf_coursename_added	503008	95.3333
golf_coursename_added	503008	95.3333
golf_teen_correct	503008	95.3333
sligbus_joystickcode_added	497355	94.2621
the_scientificnameotherorg_added	491698	91.698
golf_videocolumnname_added	485042	91.8208
golf_renamenumber_added	478843	89.7528
the_scientificname_added	458662	89.6662
the_country_added	424704	84.663
the_originatestname_added	195875	37.172

- Contents
- Description
- Contacts
- All Data

- [Contents](#)
- [Description](#)
- [Contacts](#)
- [All Data](#)

- Contents
- Description
- Contacts
- All Data





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.

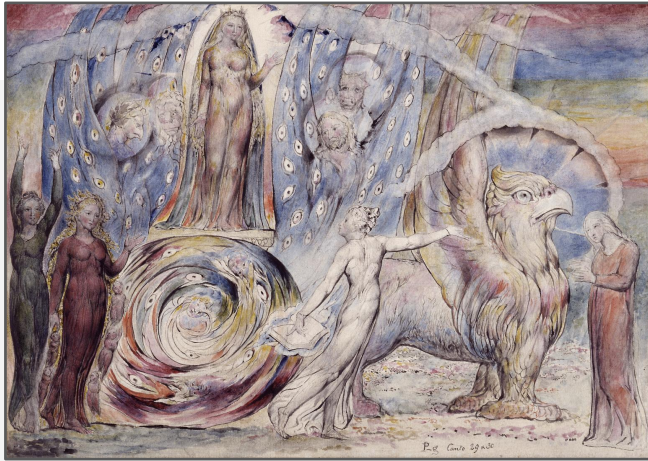
Flag	Records With This Flag	Count	Percentage
idigbio_isocountrycode_added	20105	95,775	
dwc_datasetid_replaced	20054	95,532	
dwc_parentnameusageid_added	20054	95,532	
dwc_taxonid_added	20054	95,532	
dwc_taxonomicstatus_added	20054	95,532	
dwc_taxonrank_added	20054	95,532	
gbif_canonicalname_added	20054	95,532	
gbif_genericoname_added	20054	95,532	
gbif_taxon_corrected	20054	95,532	
dwc_scientificnameauthorship_added	19949	95,031	
gbif_vernacularname_added	19718	93,931	
dwc_multimedia_added	19599	93,394	
gbif_reference_added	19541	93,068	
dwc_country_replaced	16483	78,52	
dwc_originalnameusageid_added	9286	44,236	
dwc_family_replaced	3480	16,482	
taxon_match_failed	2829	12,524	
dwc_order_replaced	2228	10,604	
dwc_genus_replaced	2225	10,599	
dwc_specificepithet_replaced	893	4,254	
dwc_infraspecificepithet_added	167	0,796	
dwc_kingdom_suspect	70	0,333	
dwc_stateprovince_replaced	58	0,276	
dwc_class_replaced	54	0,257	
dwc_family_added	54	0,257	



iDigBio Bird Egg collection

<https://www.idigbio.org/portal/recordsets/3ff3bf5c-7aba-40c3-80b2-1b00ea1abdd5>





WHAT COULD HELP?:





# WHAT COULD HELP?

Requesting

**[...right people ]**




CC-BY-NC-ND (3.0 Unported) - N03369 [Beatrice Addressing Dante from the Car 1824-7](#)

Let's talk about the right people...



# WHAT COULD HELP?

<b>Contacts</b>		
 Summary Data Records Downloads Versions Rights GBIF Registration Keywords Contacts Geographic Coverage Taxonomic Coverage Additional Metadata	Who created the resource:	<a href="#"><u>Kate Webbink</u></a>
	Who can answer questions about the resource:	<a href="#"><u>Sharon Grant</u></a>
	Who filled in the metadata:	<a href="#"><u>Paul Mayer</u></a>
	Who else was associated with the resource:	
	USER	CURATOR
	<a href="#"><u>Paul Mayer</u></a>	<a href="#"><u>Scott Lidgard</u></a>
	PROGRAMMER	POINT OF CONTACT
	<a href="#"><u>Laura Russell</u></a>	<a href="#"><u>Kate Webbink</u></a>



Right people, who to contact

Collections managers AND Data managers. Take the time to work out the data landscape surrounding the dataset you are requesting.



## WHAT COULD HELP?

## Requesting

**[...right people, right time...]**

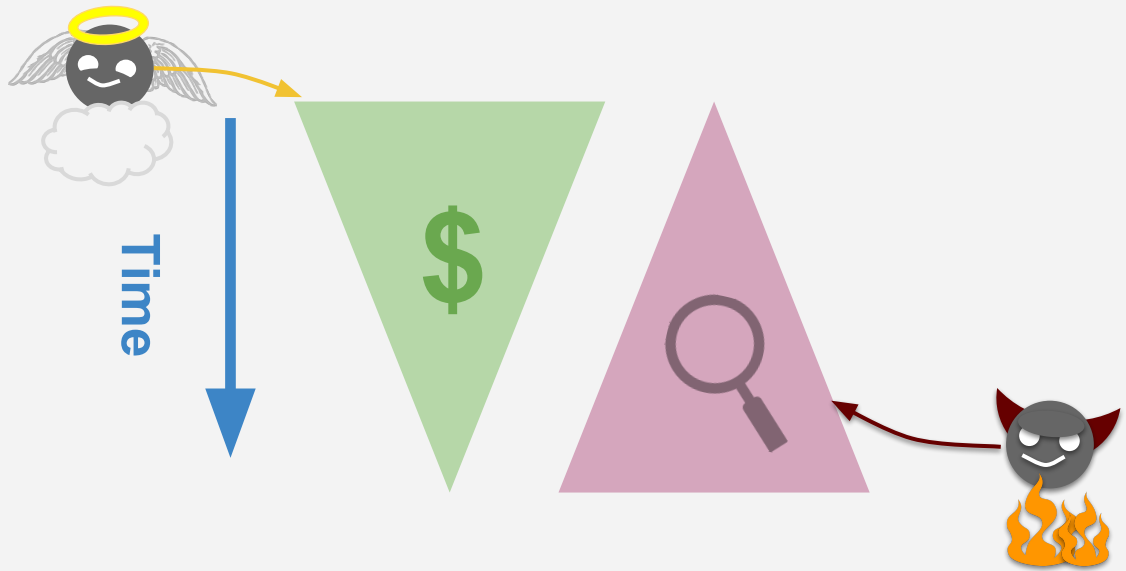


CC-BY-NC-ND (3.0 Unported) - [N03369 Beatrice Addressing Dante from the Car 1824-7](#)

And timing...



# WHAT COULD HELP?



When to ask for it

Depends on what you want to do. Tell us first. Include us in the project proposal.

Funding decreases as project progresses

Scope expands as project progresses

= EVIL



# WHAT COULD HELP?

Requesting

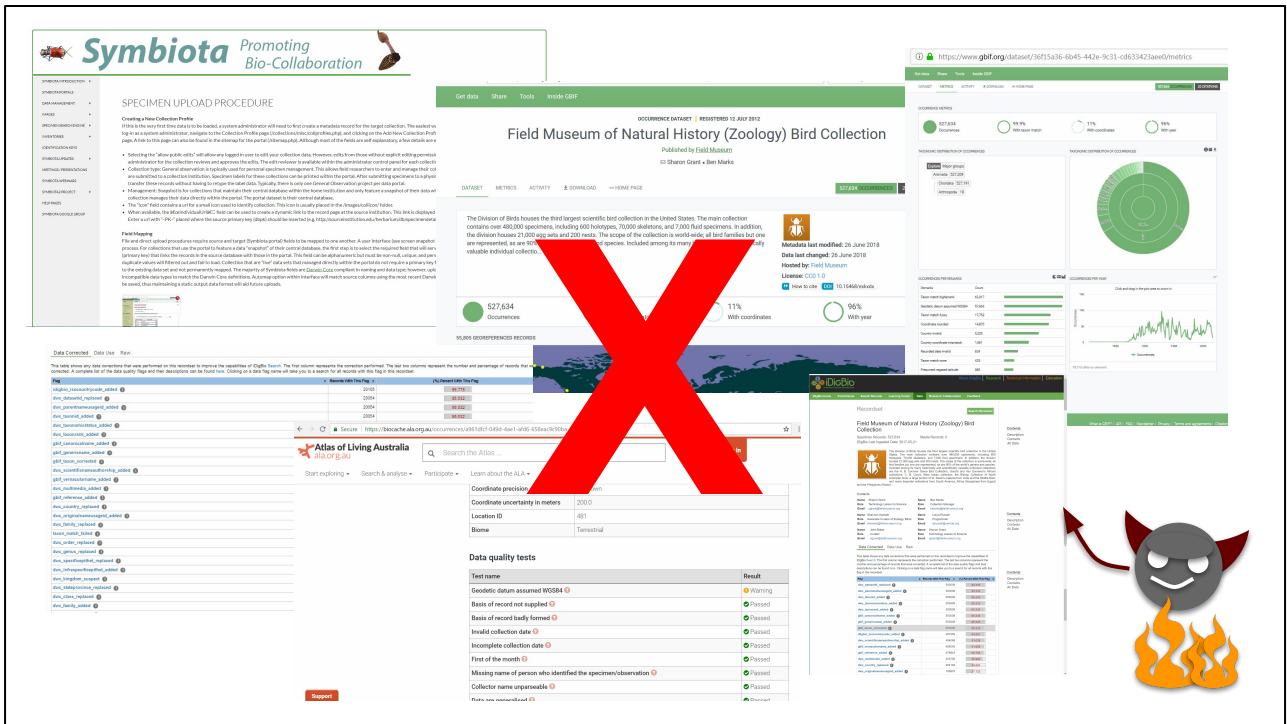
**[...right people, right time...]**

Reporting



CC-BY-NC-ND (3.0 Unported) - N03369 Beatrice Addressing Dante from the Car 1824-7





We really don't want anymore systems that validate our data after we've published it.



## WHAT COULD HELP?

Requesting

**[...right people, right time...]**

## Reporting

**[...one place to find them all...]**



CC-BY-NC-ND (3.0 Unported) - [N03369 Beatrice Addressing Dante from the Car 1824-7](#)







# WHAT COULD HELP?:

Requesting

**[...right people, right time...]**

Reporting

**[...one place to find them all...]**

Repatriating

**[...standard tools...]**



CC-BY-NC-ND (3.0 Unported) - N03369 [Beatrice Addressing Dante from the Car 1824-7](#)



README.md

## whip

Whip is a human and machine-readable syntax to express **specifications for data**. It can be used as a whip to test how well data meets certain specifications, be it a feather 🪶 or a chain whip 🪄.

Example:

```
my_date_field:
  dateFormat: ['%Y-%m-%d', '%Y-%m', '%Y'] # Needs to be ISO8601 format, but don't allow ranges
  minDate: 1838-01-01 # No dates before 1838
  empty: True # Empty values are allowed
```

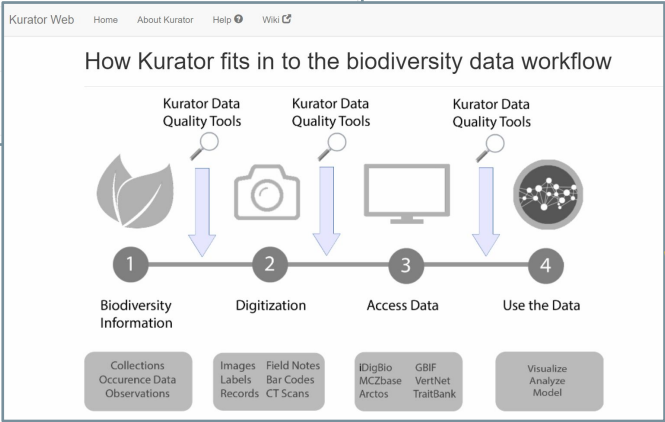
### Documentation

- [Syntax](#)

### Implementations

You can test whip specifications with [pywhip](#).

### Contributors







[CC-0 A depiction of a carrot from the Eastern Roman Juliana Anicia Codex.](#)



[CC-0 Young Boy with Whip.](#)

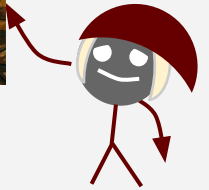
Artwork:

[https://en.wikipedia.org/wiki/Carrot#/media/File:Carrot,\\_Juliana\\_Anicia\\_Codex.jpg](https://en.wikipedia.org/wiki/Carrot#/media/File:Carrot,_Juliana_Anicia_Codex.jpg)

[https://commons.wikimedia.org/wiki/File:American\\_School,\\_Young\\_Boy\\_with\\_Whip,\\_ca.\\_1840.jpg](https://commons.wikimedia.org/wiki/File:American_School,_Young_Boy_with_Whip,_ca._1840.jpg)



## Famous People Painting - *Discussing the Divine Comedy with Dante*



Chinese Artists Dai Duda, Li Tiezi, and Zhang An, 2006, oil on canvas:” <http://cliptank.com/PeopleofInfluencePainting.html>

### Artwork References:

*Chinese Artists Dai Duda, Li Tiezi, and Zhang An, 2006, oil on canvas:”*

<http://cliptank.com/PeopleofInfluencePainting.html> -

<http://cliptank.com/PeopleofInfluencePainting.html>



# Questions, Comments, Thought?

## Repatriation of Augmented Information to an Institutional Repository



**Sharon Grant**

Janeen Jones

Kate Webbink

Pete Herbst

Rob Zschernitz

**SPNHC/TDWG**  
2018



On the 9th of April 2010 the Field Museum received a momentous email from the ORNIS (ORnithology Network Information System) team informing them that they could now access the products of a nationwide georeferencing project; its bird collection could be, quite literally, put on the map.

On the 7th of August 2017 those data (along with the sister datasets from FISHNet (FISH NETwork) and MaNIS (Mammal Network Information System) finally made their way into the Museum's collection management system.

It's easy to get data out, why is it so hard to get it back? To make it easier, what do we need to do in terms of coordination, staffing, and/or technological resources? How can tools like data quality flags better accommodate the needs of data-providers as well as data-users elsewhere along the collections data pipeline?

We present a real life case study of repatriating an enhanced dataset to its institute of origin, including details on timelines, estimates of effort, and lessons learned. The best laid repatriation protocols might not prepare us for everything, but following them more closely might save us some sanity.

*Repatriation of Augmented Information to an Institutional Database.*

Available from:

[https://www.researchgate.net/publication/325745168\\_Repatriation\\_of\\_Augmented\\_Information\\_to\\_an\\_Institutional\\_Database](https://www.researchgate.net/publication/325745168_Repatriation_of_Augmented_Information_to_an_Institutional_Database) [accessed Jun 22 2018].



STEP 1 - You ask us for data and/or images (sometimes, 'cause we make it all public so you can just come get it yourself. So I guess we like that and hate that.) but we really don't know what the heck you are going to do with it. The project description is often vague and/or over ambitious. Possibly because they are loosely defined describing what you are going to do to my data after I send my data to you

- annotation?
- augmentation?
- value added?
- metadata?
- crowd-sourcing
- transcription?
- automated validation? (take lots of peoples opinions and come up with an average)
- 

STEP 2 - Now we can send data out in a standard format but it wasn't always the case that time was a) easy to do, b) easy to understand, c) commonly known. In 2010 Darwin Core was VERY new and we still had to use DiGir which was only great if you had a degree in computer science.

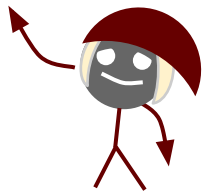
STEP 3 - NSW of documenting the augmentation

STEP 4 - No standard way to physically get the data back

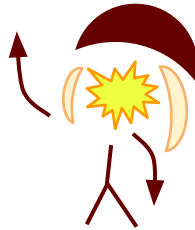




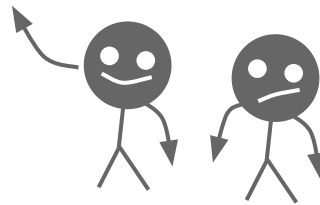
Angel thing



mere Dante  
(based on [this](#) / [this?](#))



[Exploding-head](#)  
Dante



mere mortal



"Devil doll"  
(based on [this one](#))