Getting Data Back

Repatriation of Augmented Information to an Institutional Repository

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 [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

!

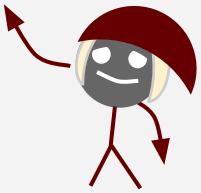
We've been collating and naming and standardising for a very long time.

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"



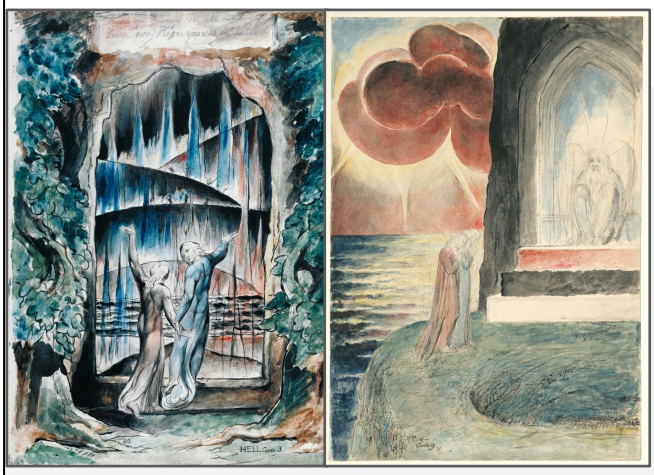
Artwork Reference: https://commons.wikimedia.org/wiki/File:Portrait_de_Dante.jpg

- 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 Aligheri our guide for the next 15 minutes or so. In

Artwork References:

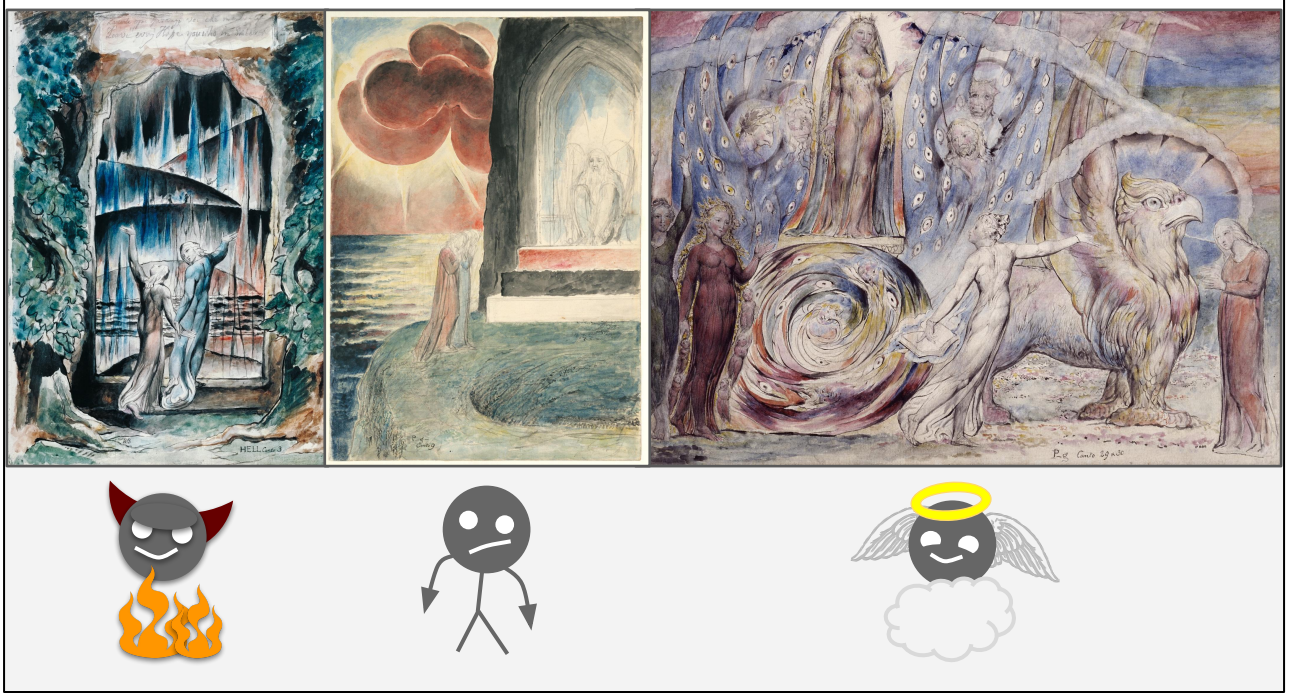
Hell:
 - N03352 **The Inscription over the Gate**
1824–7 [A00005-A00011; N03351-N03370; T01950-T01956; complete]
https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352



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...

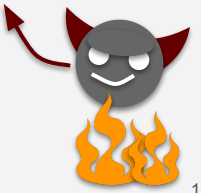
Hell:
 - N03352 **The Inscription over the Gate** 1824–7 [A00005-A00011; N03351-N03370; T01950-T01956; complete]
https://www.tate.org.uk/art/artworks/blake-the-inscription-over-the-gate-n03352

SETTING THE SCENE:

1980-90s

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 Domian: 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...]**

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.

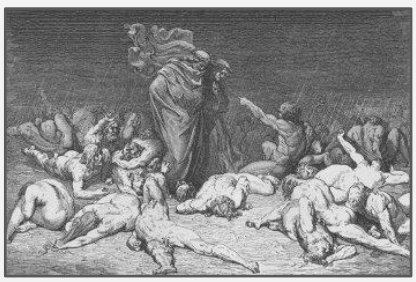
http://apps.kew.org/herbcat/gotoWhyPlantsLabeled.do
https://www.nybg.org/science-project/index-herbariorum-upgrade/

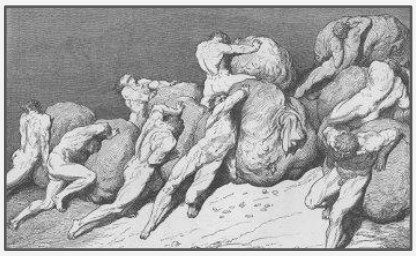
SETTING THE SCENE:

- 1980-90s
- **[...TDWG...]**
- Early 2000s

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 Domian: 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...]**

SETTING THE SCENE:

- 1980-90s
- **[...TDWG...]**
- Early 2000s
- **[...GBIF...]**
- Late 2000s
- **[...Darwin Core / ABCD / IPT...]**

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...]**

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

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

USE CASE:

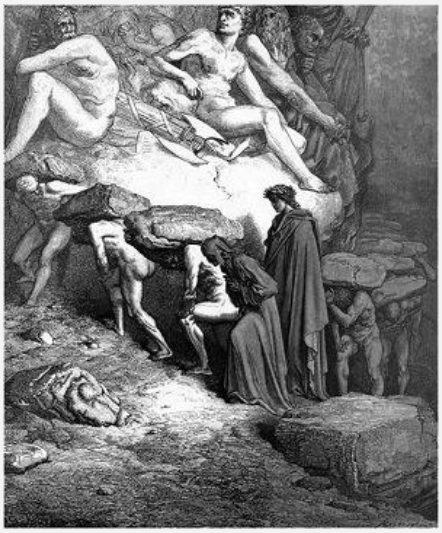
Purgatory: Humans can sin by using love towards improper or malicious ends (Wrath, Envy, Pride), or using it to proper ends but with love that is either not strong enough (Sloth) or love that is too strong (Lust, Gluttony, Greed). Below the seven purges of the soul is the Ante-Purgatory, containing the 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

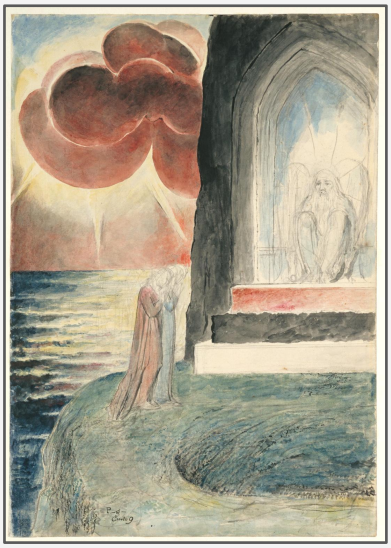
Staff at the museum always been active members of the community especially the vertebrate groups. At some time probably in 2006(ish) data was sent from our Fishes, Birds, Mammals and Herps Divisions to the Nets/ISs http://vertnet.org/about/classicnetworks.html to be georeferenced

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]



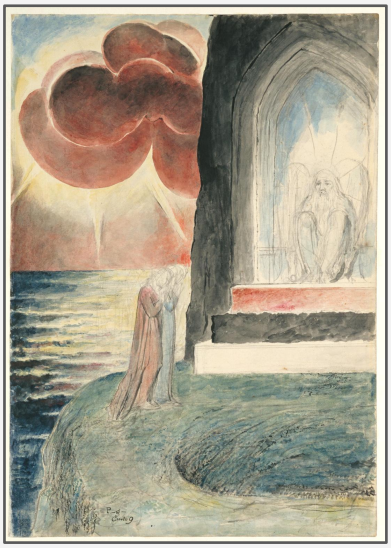
In 2005 each collection was using its own database using a variety of platforms maintained by a variety of people.

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?..]**

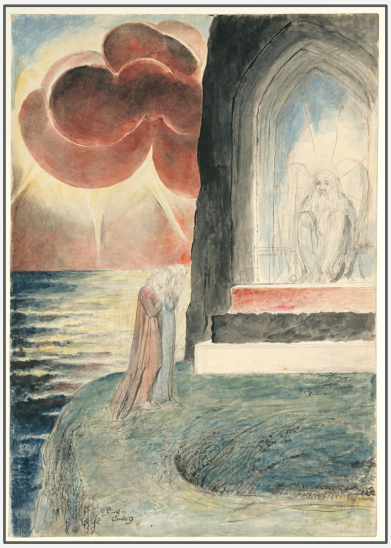


Variety is the spice of life…

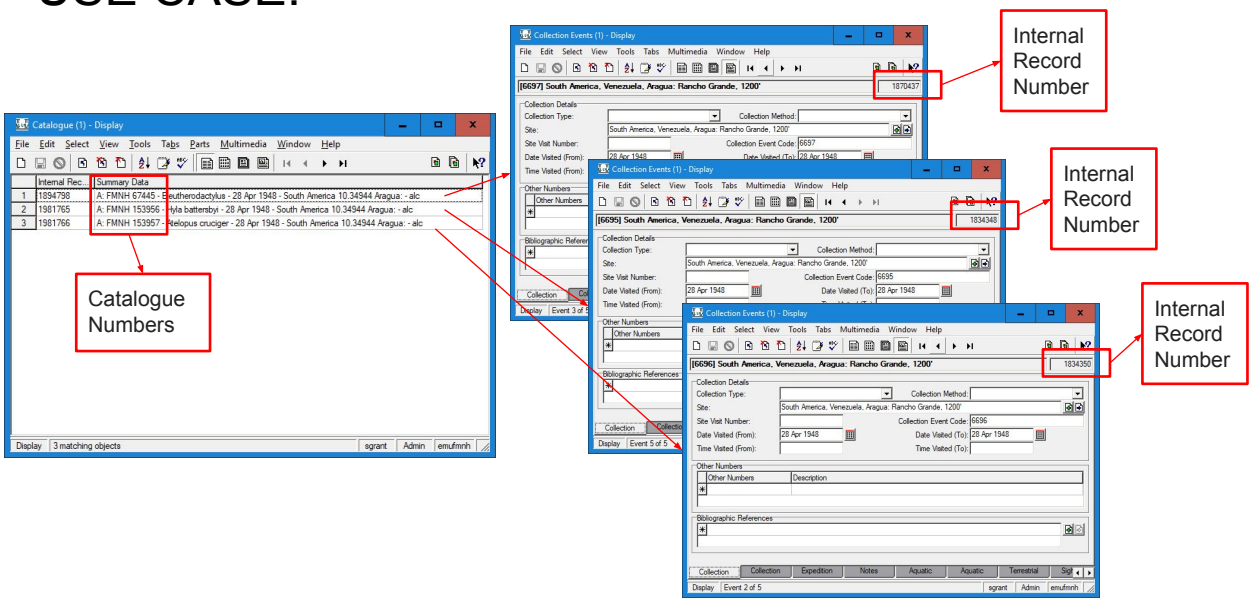
USE CASE:

USE CASE:

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



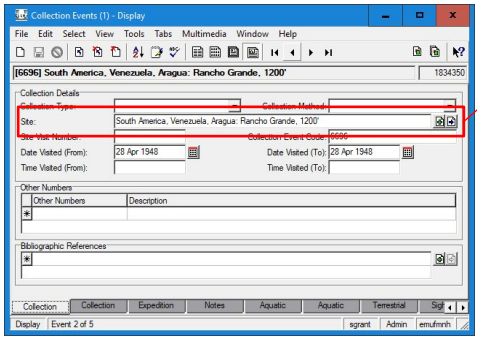
USE CASE:

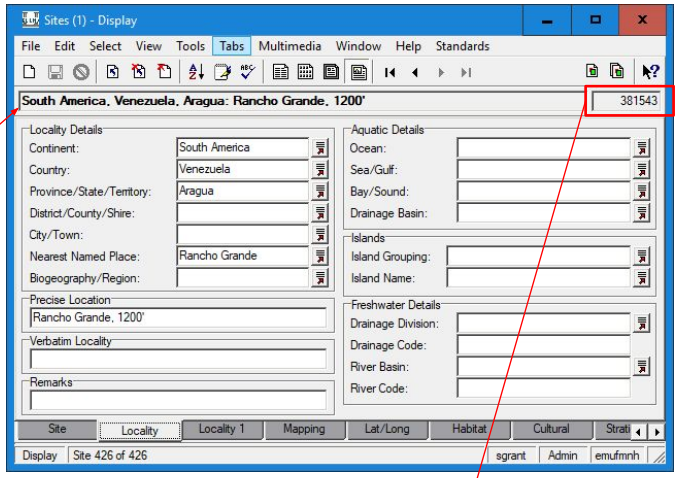


So data held at the Field Museum is no longer in the same form that was sent out.

Data is now held in relational structure.

USE CASE:

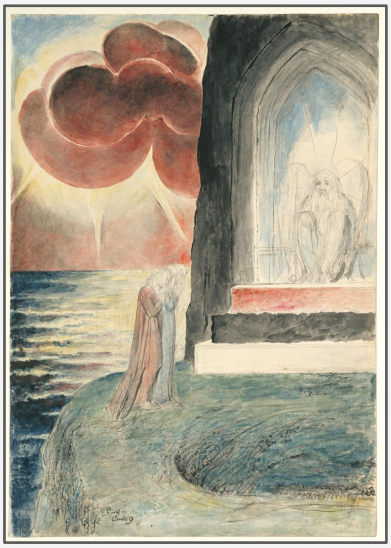




Internal Record Number

USE CASE:

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

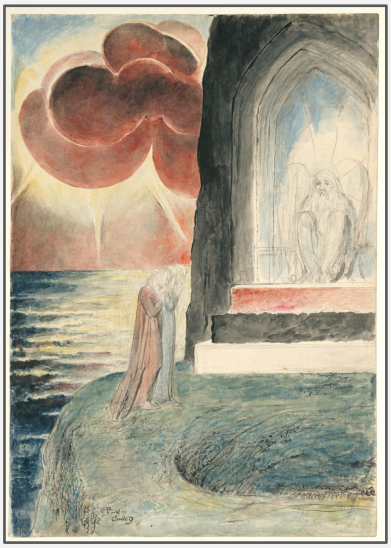


The Museum officially sanctions and provides ongoing support for a museum-wide collection management system.

- 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]



ORNIS Georeference Repatriation

GBIF/VertNet x GBIF/VertNet/ORNIS x

John Wieczorek <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://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."

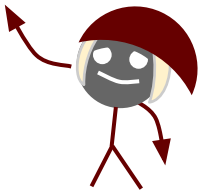
We would like to extend our heartfelt thanks for the diligent hard work and remarkable results from the georeferencers and those who managed them. Special thanks to Heather Constable who held the reins when John was in the field, and to Michelle Koo who .

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 Wieczorek <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://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.

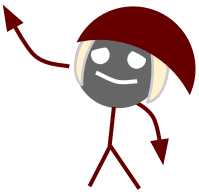
With excitement to see the fruition of all of this hard work,

John, Town, and Carla

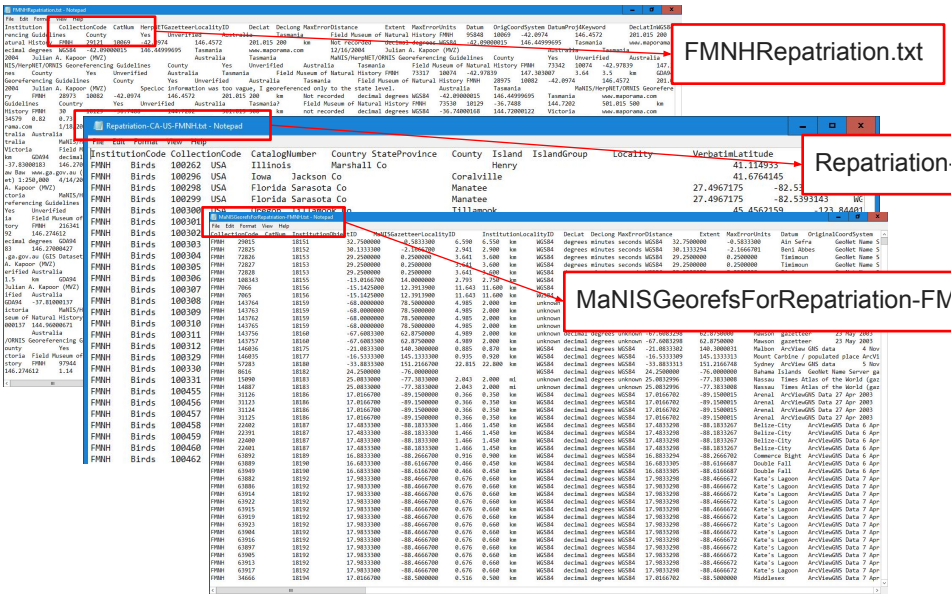
PS: If this message has reached you in error and you believe it should be sent to someone else, please reply and let John know to whom it should be addressed.

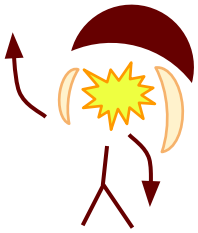
9th April or the 4th September 2010??

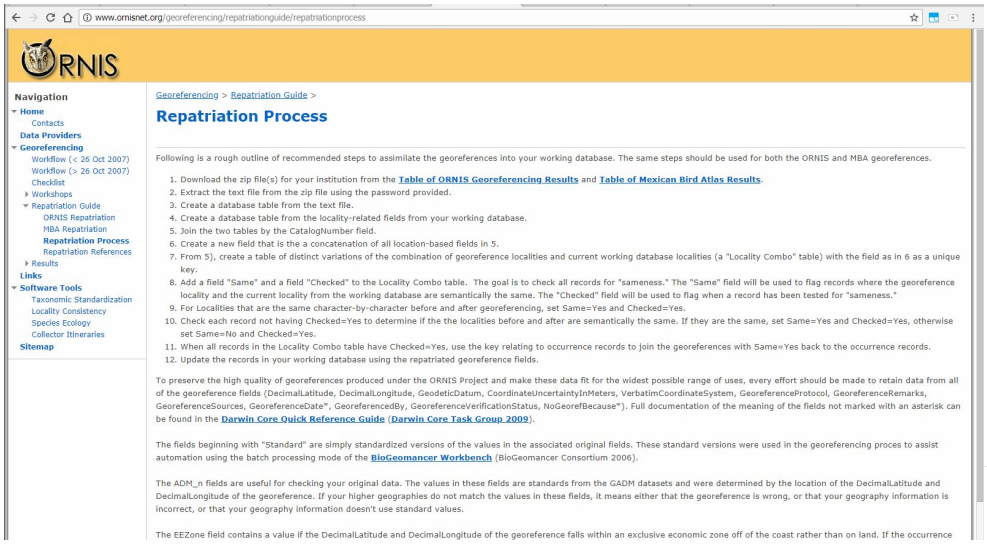
A noteworthy day has arrived. Georeferences from the ORNIS project are ready for repatriation. A document describing the results (http://olla.berkeley.edu/ornisnet/?q=node/31) and a guide to repatriation (http://olla.berkeley.edu/ornisnet/?q=node/32) are now available on the ORNIS web site.

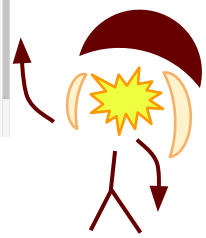


USE CASE:


FMNHRepatriation.txt
Repatriation-CA-US-FMNH.txt
MaNISGeorefsForRepatriation-FMNH.txt




www.ornisnet.org/georeferencing/repatriationguide/repatriationprocess
ORNIS
Navigation
Home
Contacts
Data Providers
Georeferencing
Workflow (< 26 Oct 2007)
Workflow (> 26 Oct 2007)
Checklist
Workshops
Repatriation Guide
ORNIS Repatriation
MBA Repatriation
Repatriation Process
Repatriation References
Results
Links
Software Tools
Taxonomic Standardization
Locality Consistency
Species Ecology
Collector Itineraries
Sitemap
Georeferencing > Repatriation Guide >
Repatriation 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 MBA 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 CatalogNumber 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*, GeoreferencedBy, 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



www.herpnet.org/Gazetteer/RepatriationFields.html

Summary of the HerpNET Georeferencing Process:

1. We constructed a gazetteer of unique localities in 2003 from the data you sent us for your collection. We standardized some of the country, state and county information for these data.
2. Other institutions downloaded regions from the HerpNET Gazetteer and georeferenced them following the MaNIS/HerpNET/ORNIS (MHO) guidelines and the Georeferencing Steps. Refer to the Georeferencing Checklist for a list of who georeferenced what.
3. The localities were then returned to MVZ, and we validated them for the correct data structure, whether they followed MHO guidelines, and whether they fell in the administrative boundaries (Country, State, County) of the original locality information. In cases where they did not fall in the correct administrative boundaries, we either corrected the georeference or made notes in the *LocalityAnnotations* and *LatLongRemarks* fields.
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.
5. MVZ then matched up the completed georeferenced localities to your catalog numbers and created two files, a **Repatriation File** and an **Exception File**. These files are explained below. Zipped copies of these tab-delimited text files were sent via e-mail to curators and collection managers of 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.
If you are opening these files in a database or spreadsheet, please note that several columns have 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, DeterminationRef* and *NoGeorefBecause*.
6. Institutions should evaluate the files they received and update 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 your data using the mapping feature.

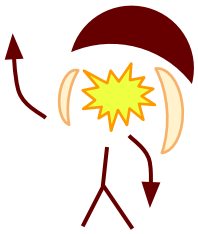
Fields found in Repatriation and Exception Files:

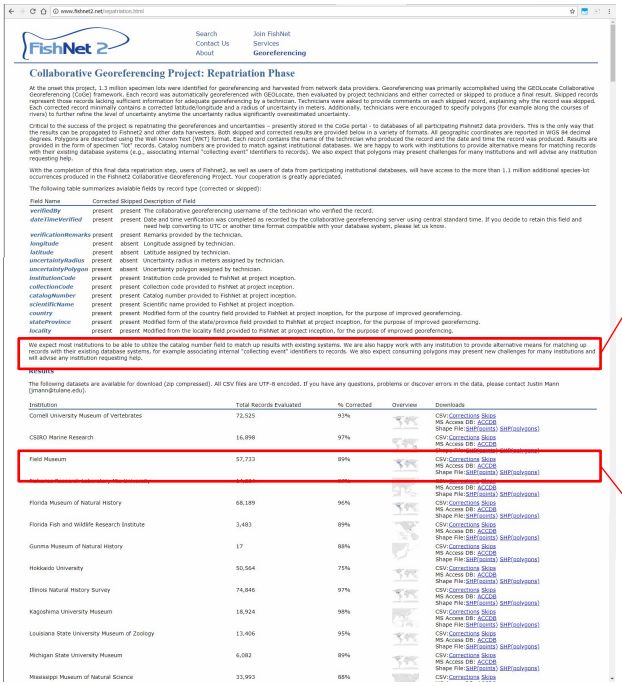
Unique Identifier Fields - These fields exist to help you match the georeferenced data to your collection catalog.

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.
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 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 MaNIS/ HerpNET/ 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.

DecLat - 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 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.





FishNet 2

Search | Join FishNet
Contact Us | Services
About | Georeferencing

Collaborative Georeferencing Project: Repatriation Phase

At the onset this project, 1.3 million specimen lots were identified for georeferencing and harvested from network data providers. Georeferencing was primarily accomplished using the GEOLocate Collaborative Georeferencing (CoGe) framework. Each record was automatically georeferenced with GEOLocate, 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 minimally contains a corrected latitude/longitude and a radius of uncertainty in meters. Additionally, technicians were encouraged to specify polygons (for example along the courses of rivers) to further refine the level of uncertainty anytime the uncertainty radius significantly overestimated uncertainty.

Critical to the success of the project is repatriating the georeferences and uncertainties – presently stored in the CoGe 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 results are provided below in a variety of formats. All geographic coordinates are reported in WGS 84 decimal degrees. Polygons are described using the Well Known Text (WKT) 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 specimen "lot" records. Catalog numbers are provided to match 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 polygons may present challenges for many institutions and will 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.
dateTimeVerified	present	present	Date and time verification was completed as recorded by the collaborative georeferencing server using central standard time. If you decide 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 from 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 polygons may present new challenges for many institutions and will 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 Mann (jmann@tulane.edu).

Institution	Total Records Evaluated	% Corrected	Overview	Downloads
Cornell University Museum of Vertebrates	72,525	93%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
CSIRO Marine Research	16,898	97%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Field Museum	57,733	89%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
[illegible]	[illegible]	[illegible]		[illegible]
Florida Museum of Natural History	68,189	96%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Florida Fish and Wildlife Research Institute	3,483	89%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Gunma Museum of Natural History	17	88%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Hokkaido University	50,564	75%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Illinois Natural History Survey	74,846	97%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Kagoshima University Museum	18,924	98%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Louisiana State University Museum of Zoology	13,406	95%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Michigan State University Museum	6,082	89%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Mississippi Museum of Natural Science	33,993	88%		CSV:Corrections Skips

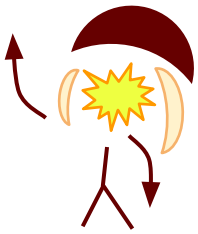
We expect most institutions to be able to utilize the **catalog number** field to match up results with existing systems.

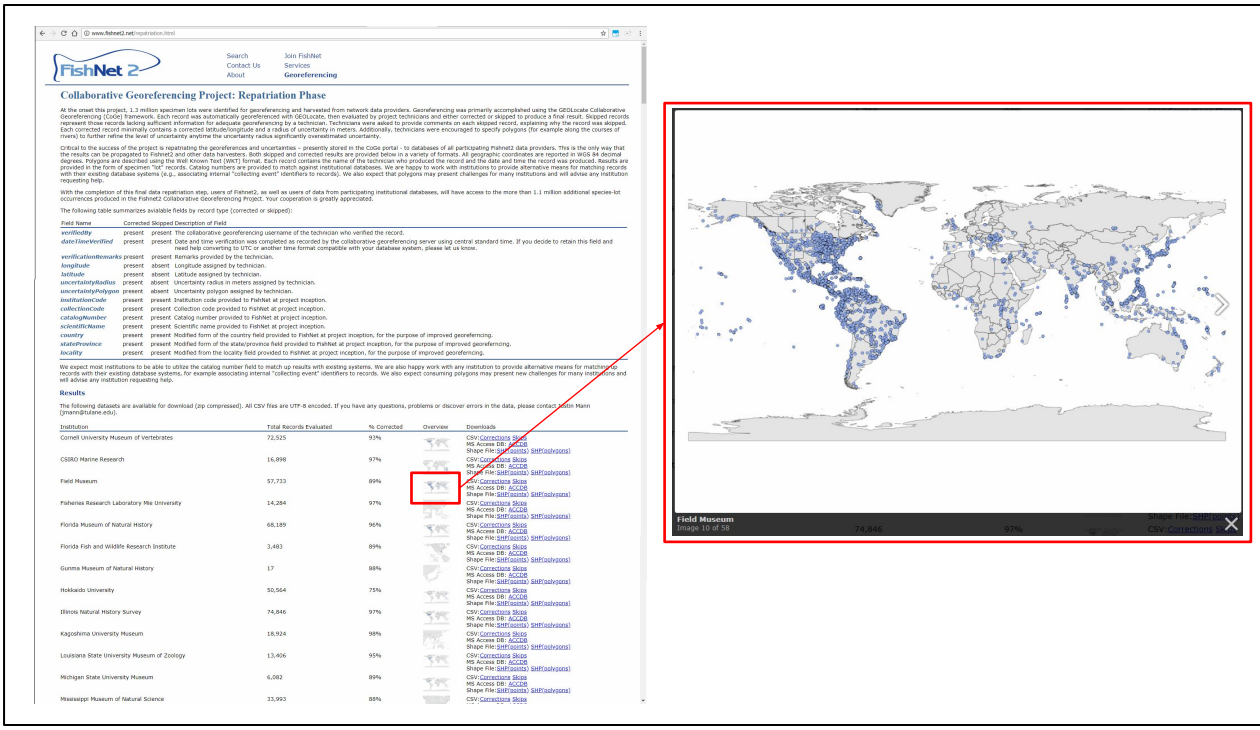
We are also happy work with any institution to provide alternative means for matching up records with their existing database systems, for example associating internal "collecting event" identifiers to records.

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)





FishNet 2

Search | Join FishNet
Contact Us | Services
About | **Georeferencing**

Collaborative Georeferencing Project: Repatriation Phase

At the onset this project, 1.3 million specimen lots were identified for georeferencing and harvested from network data providers. Georeferencing was primarily accomplished using the GEOLocate Collaborative Georeferencing (CoGe) framework. Each record was automatically georeferenced with GEOLocate, 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 minimally contains a corrected latitude/longitude and a radius of uncertainty in meters. Additionally, technicians were encouraged to specify polygons (for example along the courses of rivers) to further refine the level of uncertainty anytime the uncertainty radius significantly overestimated uncertainty.

Critical to the success of the project is repatriating the georeferences and uncertainties – presently stored in the CoGe 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 results are provided below in a variety of formats. All geographic coordinates are reported in WGS 84 decimal degrees. Polygons are described using the Well Known Text (WKT) 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 specimen "lot" records. Catalog numbers are provided to match 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 polygons may present challenges for many institutions and will 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.
dateTimeVerified	present	present	Date and time verification was completed as recorded by the collaborative georeferencing server using central standard time. If you decide 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 from 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 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.

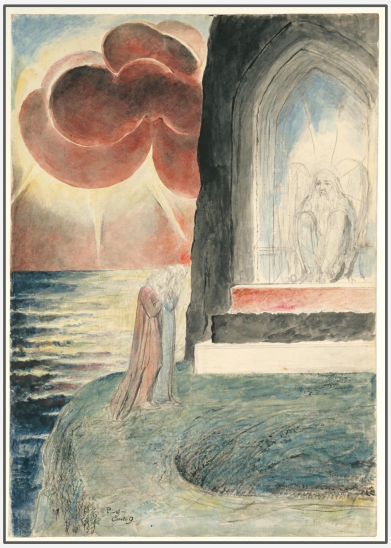
Results

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 Kristin Mann (jmann@tulane.edu).

Institution	Total Records Evaluated	% Corrected	Overview	Downloads
Cornell University Museum of Vertebrates	72,525	93%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
CSIRO Marine Research	16,898	97%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Field Museum	57,733	89%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Fisheries Research Laboratory Mie University	14,284	97%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Florida Museum of Natural History	68,189	96%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Florida Fish and Wildlife Research Institute	3,483	89%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Gunma Museum of Natural History	17	88%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Hokkaido University	50,564	75%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Illinois Natural History Survey	74,846	97%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Kagoshima University Museum	18,924	98%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Louisiana State University Museum of Zoology	13,406	95%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Michigan State University Museum	6,082	89%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Mississippi Museum of Natural Science	33,993	88%		CSV:Corrections Skips

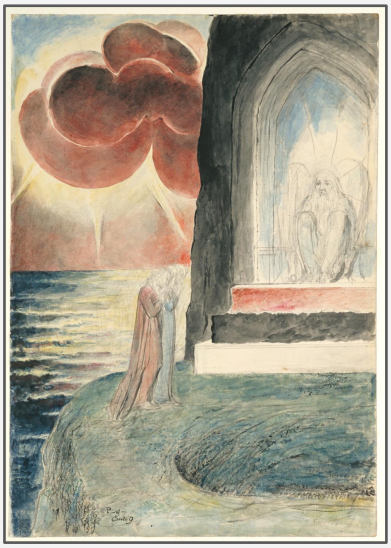
USE CASE:

- 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...]**



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...]**

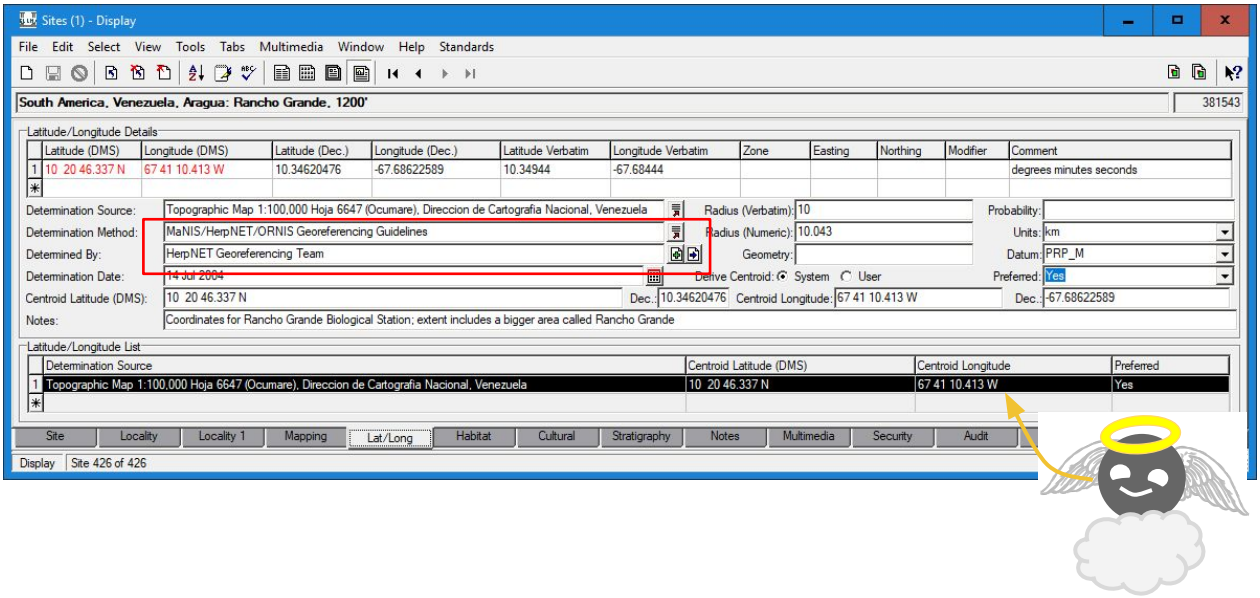


Three new collections managers

- Birds
- Fishes
- Mammals

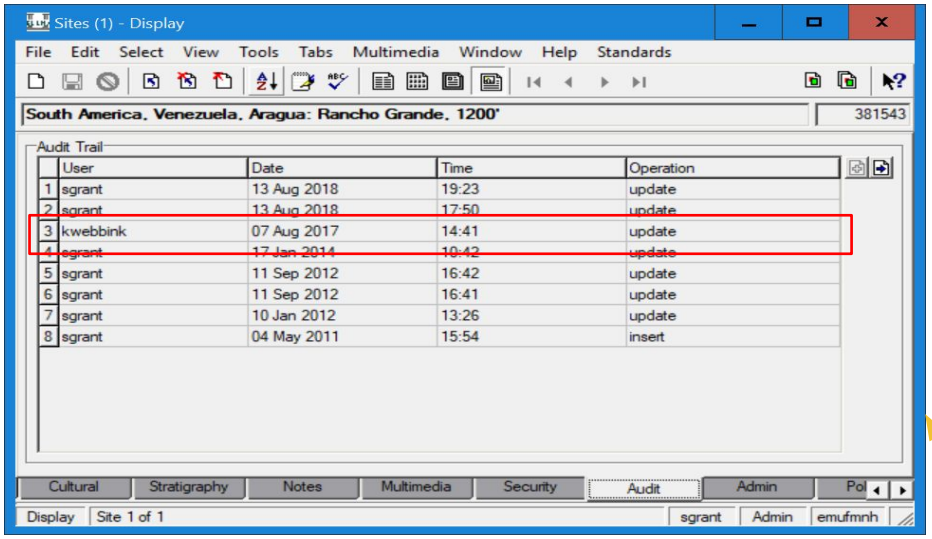
New trained data handlers on staff

USE CASE:



381543

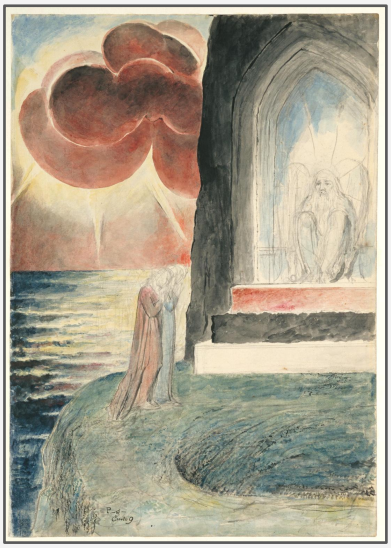
USE CASE:



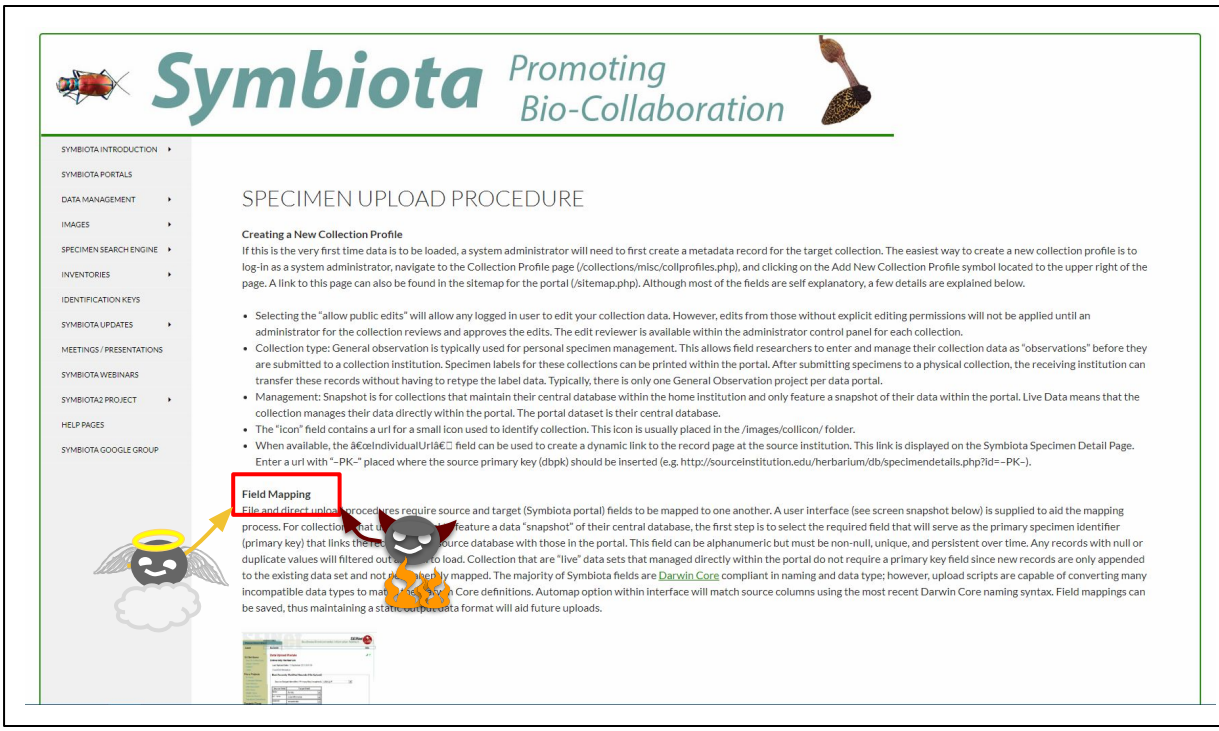
381543

USE CASE:

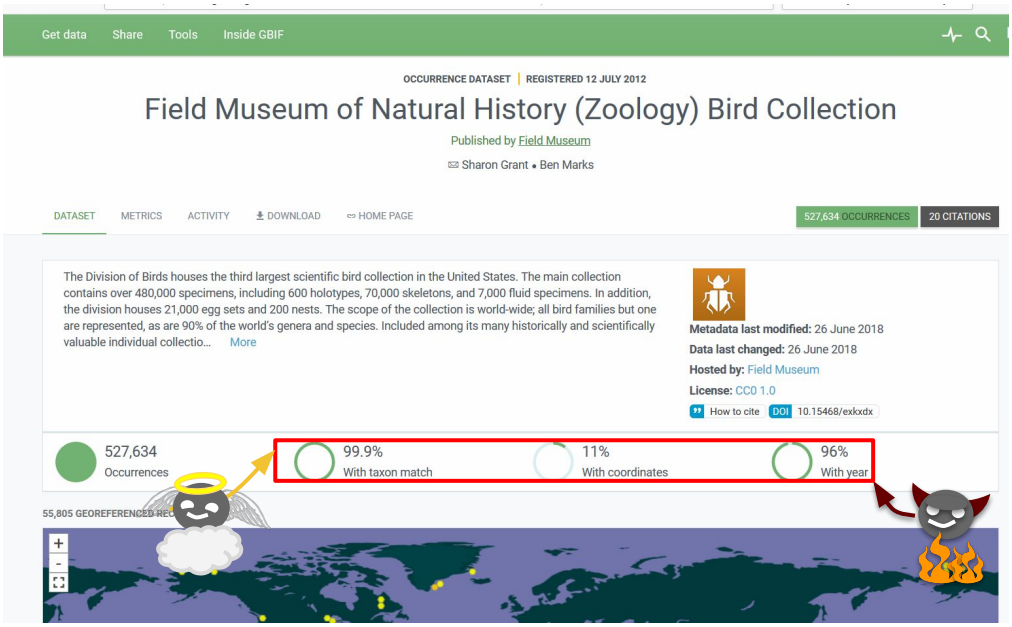
- Mid 2000-zeros [STUFF GOES OUT]
- **[...variety is the spice of life?..]**
- The 2000-tens [WE DID SOME STUFF]
- **[...things change...]**
- 2010 [OTHER PEOPLE DID STUFF]
- **[...prodigal returns...]**
- The 2000-teens [PEOPLE CHANGED]
- **[...more things change...]**
- The 2000-twenties? [WE KEPT DOING STUFF]



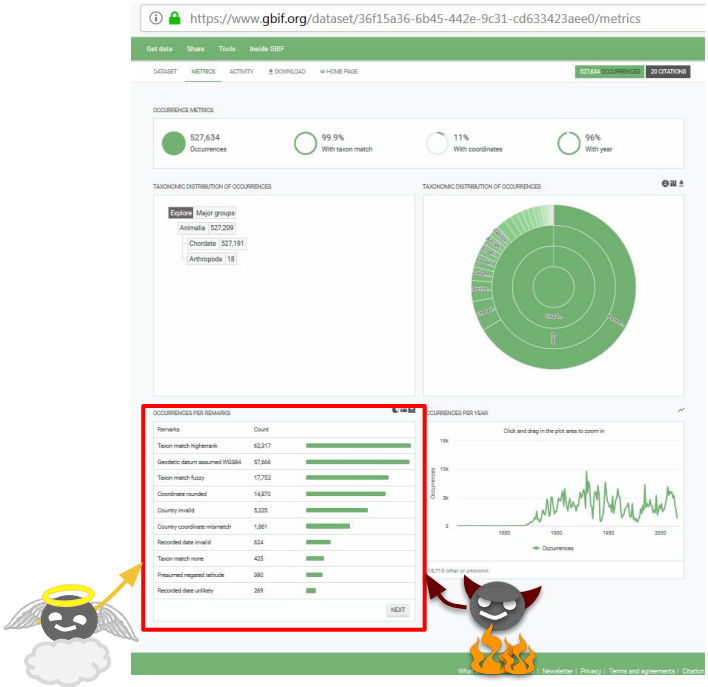

Secure | https://www.morphosource.org/Detail/SpecimenDetail/show/specimen_id/10810
MORPHO SOURCE
BY DUKE UNIVERSITY
ABOUT
BROWSE
DASHBOARD
LOGIN/REGISTER
Specimen: fmnh-amphibians and reptiles-1580, Adenomus kelaartii
VIEW SPECIMEN ON IDIGBIO
Project
Frog Diversity
Specimen Information
MorphoSource Identifier: S10810
Vouchered
Occurrence ID: d6a0693f-2423-47f7-ad2b-3141a9caeef8
Notes: imported from iDigBio. uuid:50c0b907-cd36-4bff-a212-b18b2db81ac4 Occurrence ID:d6a0693f-2423-47f7-ad2b-3141a9caeef8
Specimen Media
M14189, 3 files
CT scan of Adenomus kelaartii
(CT scan of skeleton and skull)
ABOUT | CONTACT | TERMS AND CONDITIONS | USER GUIDE
Commercial use of MorphoSource media is strictly prohibited.



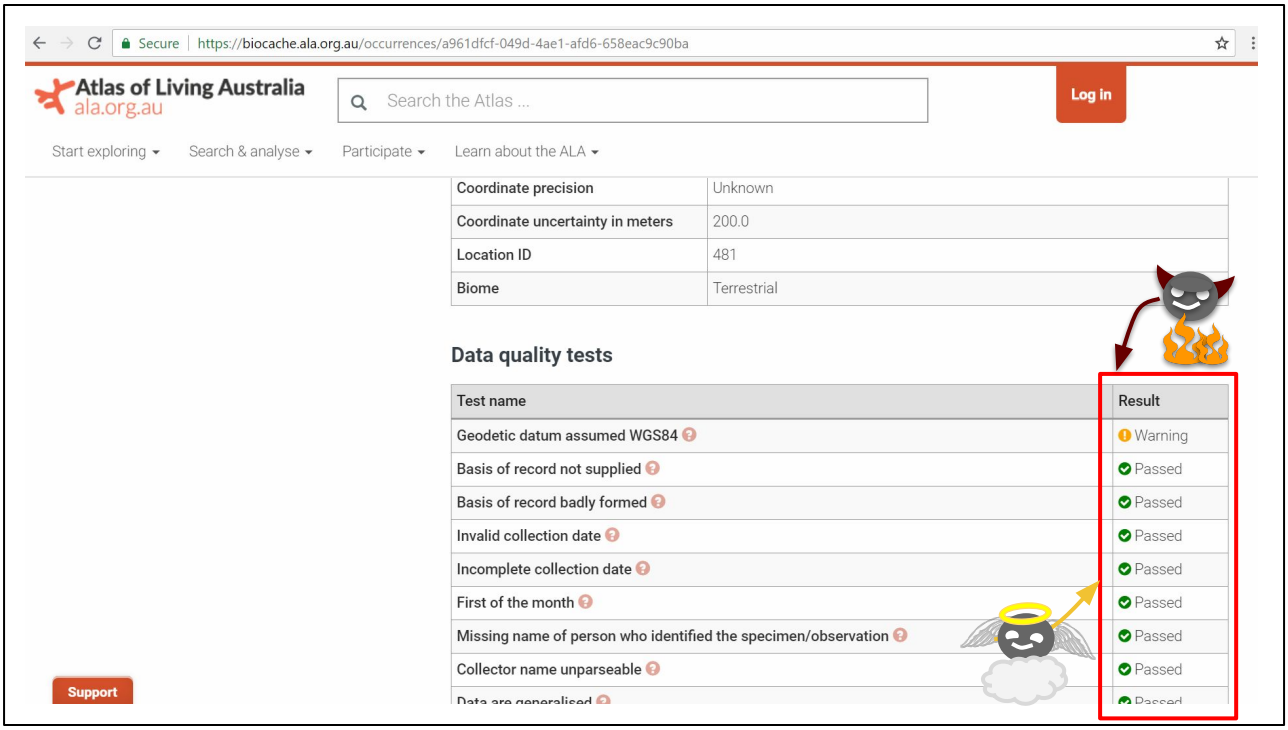
Symbiota portals

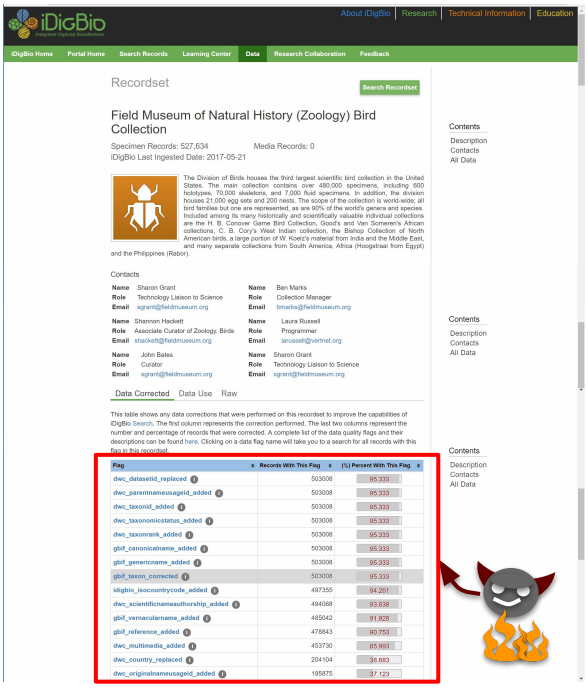


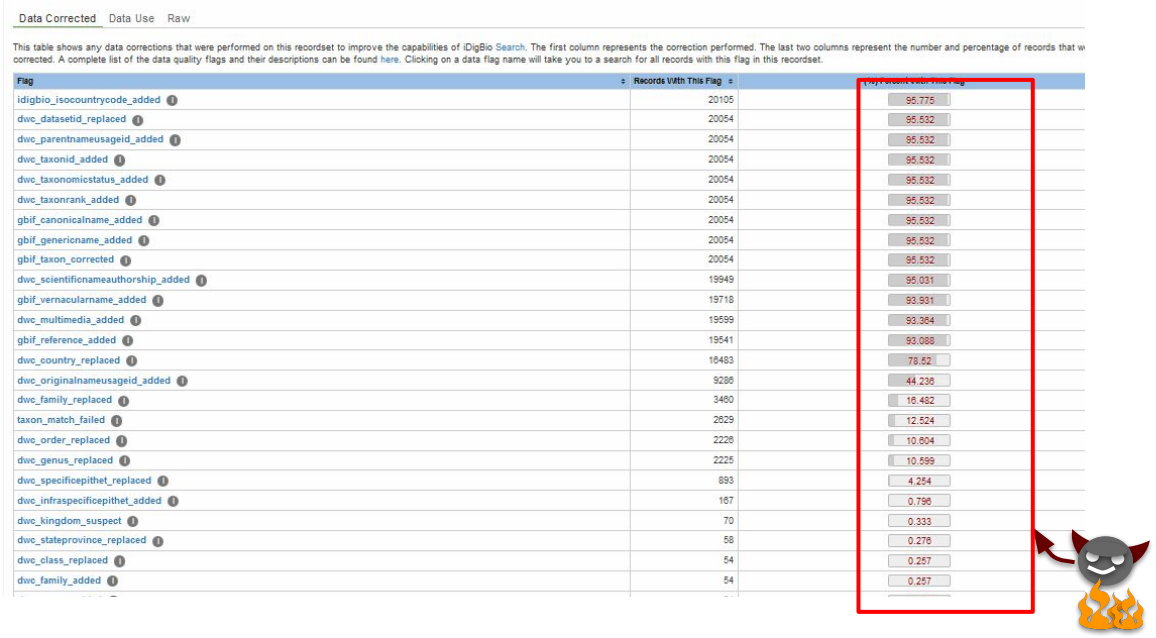
Field Museum Bird Collection
https://www.gbif.org/dataset/36f15a36-6b45-442e-9c31-cd633423aee0



Field Museum Birds collection GBIF metrics


Secure https://biocache.ala.org.au/occurrences/a961dfcf-049d-4ae1-afd6-658eac9c90ba
Atlas of Living Australia
ala.org.au
Search the Atlas ...
Log in
Start exploring
Search & analyse
Participate
Learn about the ALA
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
Support


iDigBio
About iDigBio | Research | Technical Information | Education
iDigBio Home
Portal Home
Search Records
Learning Center
Data
Research Collaboration
Feedback
Recordset
Search Recordset
Field Museum of Natural History (Zoology) Bird Collection
Specimen Records: 527,634
Media Records: 0
iDigBio Last Ingested Date: 2017-05-21
The 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 90% of the world's genera and species. Included among its many historically and scientifically valuable individual collections are the H. B. Conover Game Bird Collection, Good's and Van Someren's 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 Liaison to Science
Email sgrant@fieldmuseum.org
Name Ben Marks
Role Collection Manager
Email bmarks@fieldmuseum.org
Name Shannon Hackett
Role Associate Curator of Zoology, Birds
Email shackett@fieldmuseum.org
Name Laura Russell
Role Programmer
Email larussell@vertnet.org
Name John Bates
Role Curator
Email sgrant@fieldmuseum.org
Name Sharon Grant
Role Technology Liaison to Science
Email sgrant@fieldmuseum.org
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
dwc_datasetid_replaced 503008 95.333
dwc_parentnameusageid_added 503008 95.333
dwc_taxonid_added 503008 95.333
dwc_taxonomicstatus_added 503008 95.333
dwc_taxonrank_added 503008 95.333
gbif_canonicalname_added 503008 95.333
gbif_genericname_added 503008 95.333
gbif_taxon_corrected 503008 95.333
idigbio_isocountrycode_added 497355 94.261
dwc_scientificnameauthorship_added 494088 93.638
gbif_vernacularname_added 485042 91.928
gbif_reference_added 478843 90.753
dwc_multimedia_added 453730 85.993
dwc_country_replaced 204104 38.683
dwc_originalnameusageid_added 195875 37.123
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 w
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	[illegible]
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_genericname_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	93.364
gbif_reference_added	19541	93.088
dwc_country_replaced	16483	78.52
dwc_originalnameusageid_added	9286	44.236
dwc_family_replaced	3460	16.482
taxon_match_failed	2629	12.524
dwc_order_replaced	2226	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

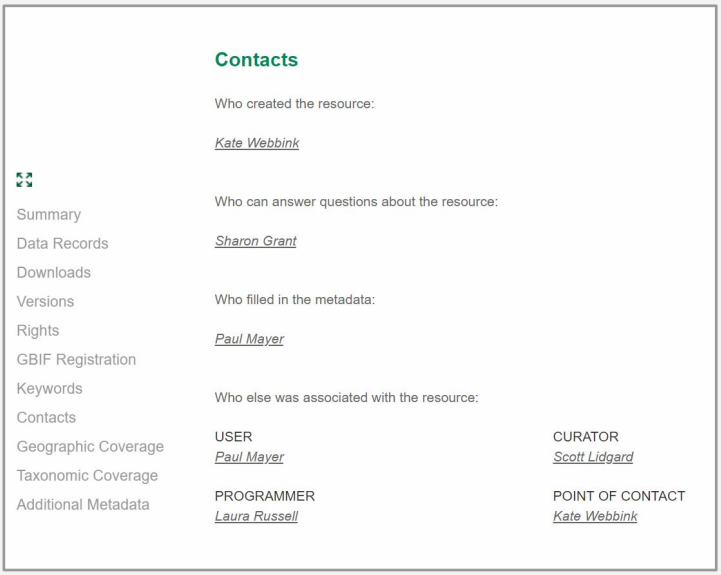
WHAT COULD HELP?:

WHAT COULD HELP?

- Requesting
- **[...right people]**

Let’s talk about the right people...

WHAT COULD HELP?



Right people, who to contact

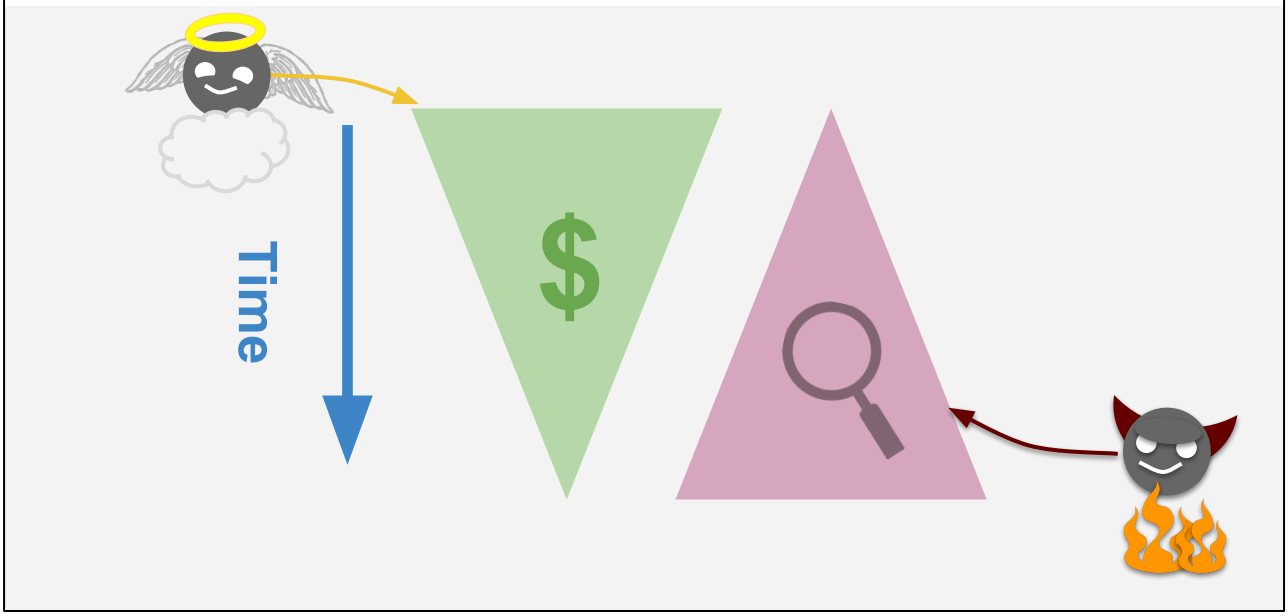
Collections managers AND Data managers. Take the time to work out the data landscape surrounding the dataset you are requesting.

WHAT COULD HELP?

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

And timing...

WHAT COULD HELP?



When to ask for it

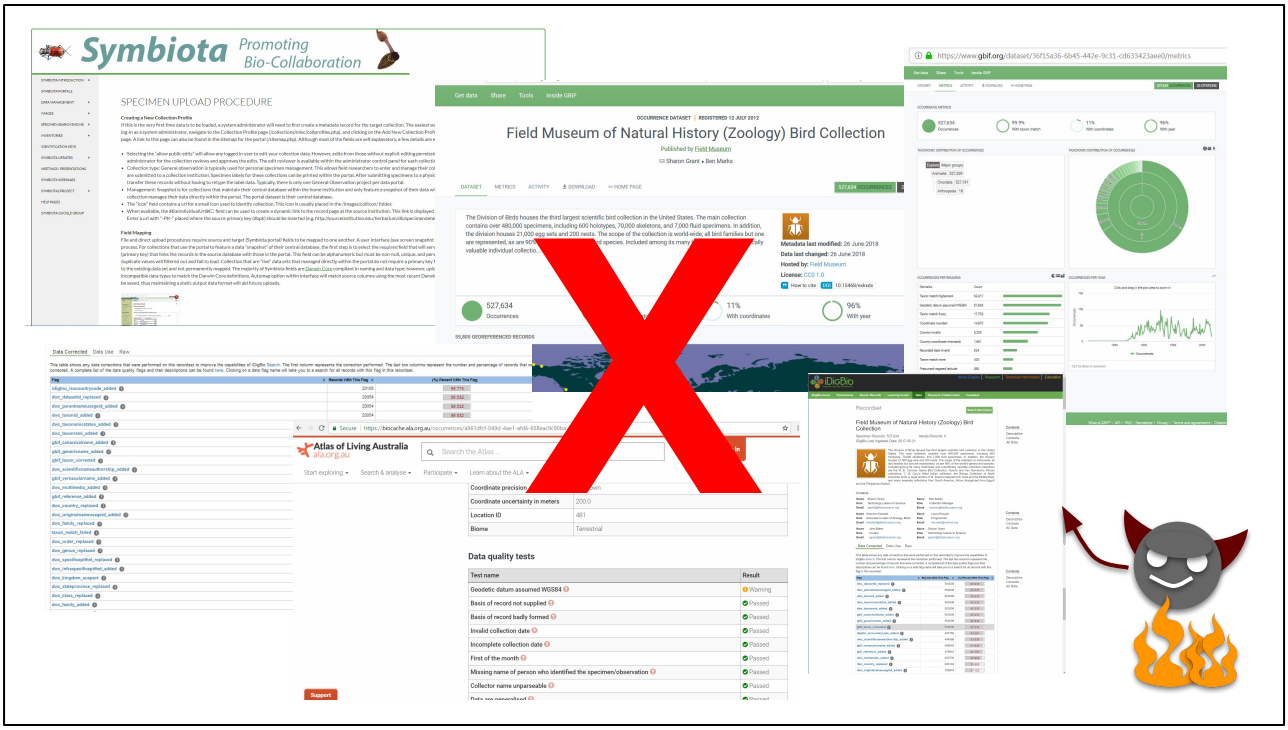
Depends on what you want to do. Tell us first. Include us in the project proposal.

Funding decreases as project progresses
Scope expands as project progresses

= EVIL

WHAT COULD HELP?

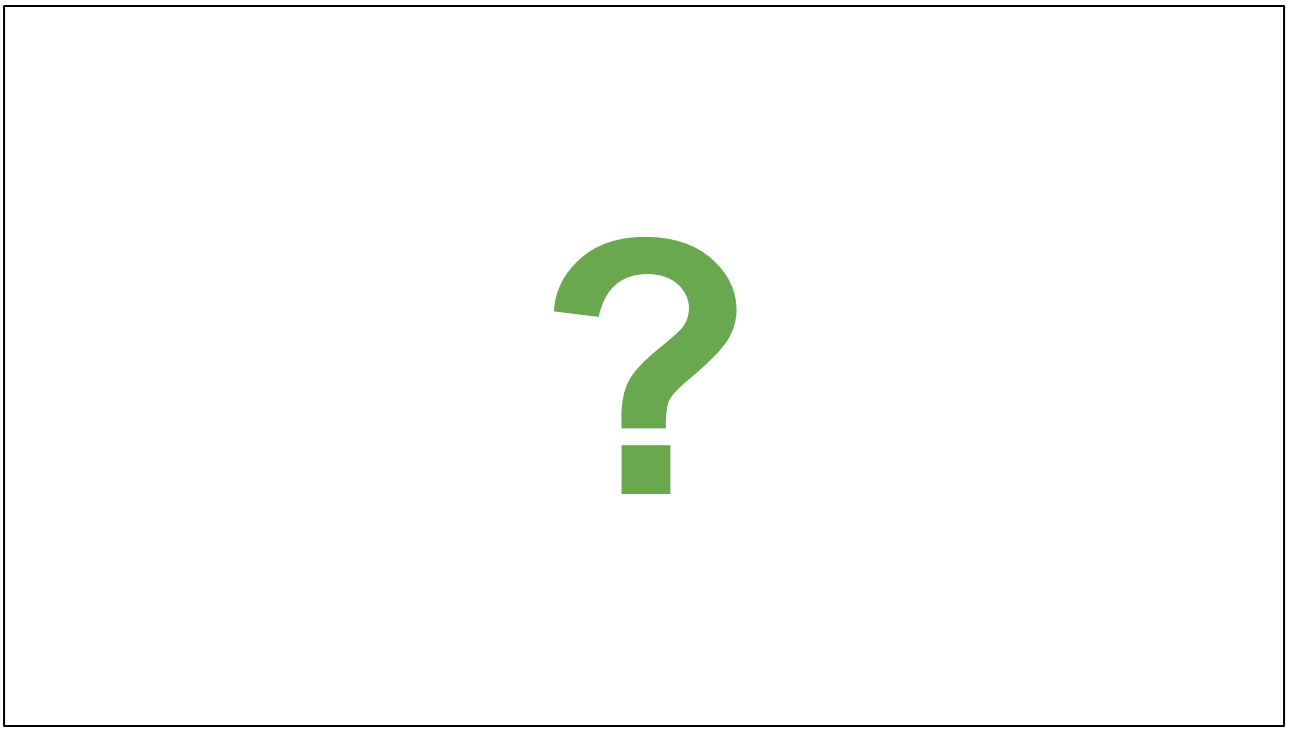
- Requesting
- **[...right people, right time...]**
- Reporting



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...]**



WHAT COULD HELP?:

- Requesting
- **[...right people, right time...]**
- Reporting
- **[...one place to find them all...]**
- Repatriating
- **[...standard tools...]**

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: 1830-01-01                     # No dates before 1830
  empty: True                             # Empty values are allowed
```

Documentation

- Syntax

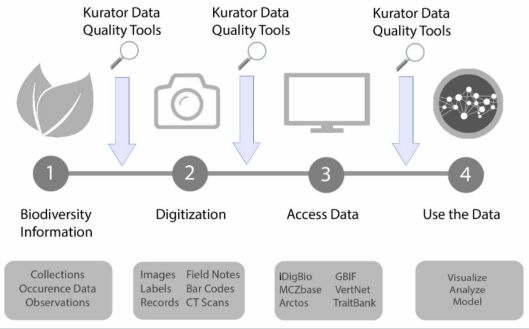
Implementations

You can test whip specifications with pywhip.

Contributors

Kurator Web Home About Kurator Help Wiki

How Kurator fits in to the biodiversity data workflow



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://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 Dudu, Li Tiezi, and Zhang An, 2006, oil on canvas:" http://cliptank.com/PeopleofInfluencePainting.html

Artwork References:

Chinese Artists Dai Dudu, 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

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 [accessed Jun 22 2018].

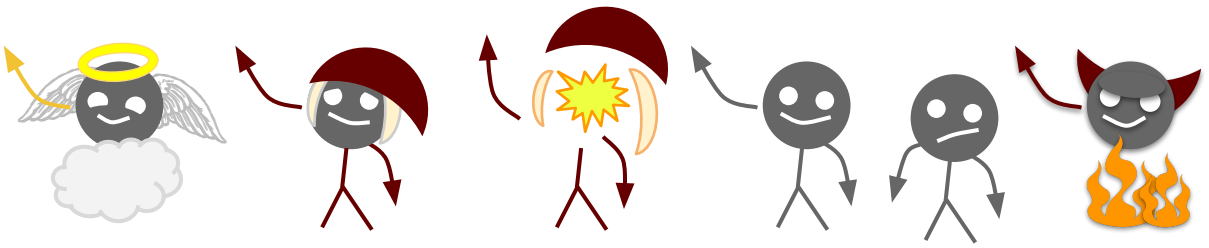
STEP 1 - You ask us for data and/or images (sometimes, ‘cause we make it all public so you can just come get it yourself. So I guess we like that and hate that.) but we really don’t know what the heck you are going to do with it. The project description is often vague and/or over ambitious. Possibly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