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.*
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
Dante Alighieri

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

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

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
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:
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**

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

Linnaeus - Mr Standards! Irony of ironies…

Artwork Reference:
https://commons.wikimedia.org/wiki/File:LinnaeusWeddingPortrait.jpg
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
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.
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.
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/
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...
Things really start to crank up in the early 2000s databases galore, standards and systems to deploy data.
SETTING THE SCENE:

1980-90s

[...TDWG...]

Early 2000s

[...GBIF...]
SETTING THE SCENE:

- 1980-90s
  - [...TDWG...]
- Early 2000s
  - [...GBIF...]
- Late 2000s
  - [...Darwin Core / 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...]
- 2000 teens
- [...]MORE… of EVERYTHING...]

References:

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


History: http://community.canadensys.net/2011/new-terms-in-darwin-core
Circle 6 contains the heretics who contradicted the doctrine and confused the spirit of Christ.
Circle 7 for the sins of violence and Circles 8 and 9 for the sins of malice (fraud and treachery)
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
Staff at the museum however had always been active members of the community especially the vertebrate groups. At some time probably in 2006(ish) data was sent from our 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)
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…
Workflow (< 26 Oct 2007)

This workflow is deprecated in favor of the new georeferencing workflow for projects downloaded and distributed after the release of BioGeomancer in which Projects could be created from projects, records could be deleted from projects, records could be re-georeferenced directly within the workbench, and localities could be edited within the workbench. This all occurred at 17:00 EDT 26 Oct 2007. If you wonder whether your project should use the steps on this page, look in the macros of your Access database and see if there is a macro called "Check Features". If there is, continue with the steps below.

1. WORK CENTER: Claim region to georeference via email to ORNIS.

2. JOHN: Download specimens from ORNIS using "ORNIS georeferencing result set" structure.

3. JOHN: Load file from step2 into a copy of the "ornisgeoreftemplate.mdb" database, named for the Georeferencing center and the contents (e.g., "UMMZ-MichiganGogebic.mdb"). This file contains queries, macros, import specifications, and export specifications for processing data for input into BioGeomancer.

4. JOHN: Compare localities in the file from step 2 to McKit, HerpNet, Mexican Bird Atlas, and Cattle's research georeferences. Set georeferences for those that do already exist from these other projects.

5. JOHN: Have BioGeomancer create a feature report from unique localities for a region, with source, coordinates, extent.

6. JOHN: Send Access database ready for processing in work center that claimed the region.

7. WORK CENTER: Run macros 'Check Features' and fill in the fields 'Correct?', 'Should Be' and 'Checked?' for every record. Records that have non-zero place names should have 'Correct?' and 'Should Be' left blank. Records that are misspelled or nonexistent (e.g., "Turner Township") should have 'Correct?' left blank and the correct form in 'Should Be'. Records that are correct should have 'Correct?' checked and 'Should Be' can be left blank. If you don't finish the 'Check Features' in one sitting you can reopen it. Only the unchecked features will show up in the list when it is reopened. When all features have been checked, go to step 8.

8. WORK CENTER: Run the macros 'Export Features for BioGeomancer', which will save a file 'FeaturesForBioGeomancer.txt' in the default database folder for Access (see footnotes about Access default database folders).

9. WORK CENTER: Log in to BioGeomancer and upload the file from step 8 as a new project with a name reflecting the region and the fact that this is a feature-only data set. Example: 'ORNIS Michigan Feature'.
USE CASE:
Workflow (> 26 Oct 2007)

This flow replaces the one used earlier for projects downloaded and distributed after the release of GeoGeomancer in which features projects could be created from projects, records could be deleted from projects, records could be re-georeferenced directly within the workflow, and localities could be edited within the workflow. This all occurred at 17:00 PDT 26 Oct 2007. If you wonder whether your project should use the steps below, look in the macros of your Access database and see if there is a macro called "Check Features". If there is, use the old georeferencing steps. Otherwise continue with the steps below.

Prerequisites: Learn the Longitude and Latitude macro.

1. **WORK CENTER**: Claim region to georeference via email to John and Heather.

2. **JOHN or HEATHER**: Download specimen from ORNIS using "ORNIS georeferencing result set" structure.

3. **JOHN or HEATHER**: Load file from step 2 into a copy of the "newxml2geoformat.platla.db" database, named for the georeferencing center and the contents (e.g., ORNIS-Michigan-Gigebic.mdb). This file contains queries, macros, import specifications, and export specifications for processing data for input into BioGeomancer.

4. **JOHN or HEATHER**: Compare localities in the file from step 3 to MNLIS, HerpNet, Mexican Bird Atlas, and Carla's research georeferences. Set georeferences for those that already exist from these other projects.

5. **JOHN or HEATHER**: Send Access database newly for processing to the work center that claimed the region.

6. **WORK CENTER**: Open the Access database sent by John or Heather in the previous step and run the 'Prepare Output for BioGeomancer' macro. This will export a file called 'OutputForBioGeomancer.txt' to the default Access data directory (see footprint about default database folders).

7. **WORK CENTER**: Log in to BioGeomancer and upload the file from step 6 as a new project with a name reflecting the contents. Example: "ORNIS Michigan Localities". BioGeomancer will begin georeferencing these records right away. You can do other things while BioGeomancer is georeferencing your project. You can even log out and it will continue. You should not, however, try to use BioGeomancer for more than one project at a time. You can safely use a second BioGeomancer window to do single record georeferencing without logging in.

8. **WORK CENTER**: When BioGeomancer is done uploading the file from step 7 (the progress bar will disappear and your new project will appear on the list of projects), edit the records until no record has more than one georeference and all of the georeferences are correct. NOTE: Records that already have coordinates in the descriptors are not to be georeferenced for ORNIS. These records can be used for reference, or they can be deleted (on by one) from the project. Projects created after 1 Feb 2008 will no longer include these records in the OutputForBioGeomancer.txt file.
USE CASE:

- Mid 2000-zeros [STUFF GOES OUT]
- [...variety is the spice of life?..]
- The 2000-tens [WE DID SOME STUFF]
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
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]
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](http://olla.berkeley.edu/ornisnet/?q=node/31)) and a guide to repatriation ([http://olla.berkeley.edu/ornisnet/?q=node/32](http://olla.berkeley.edu/ornisnet/?q=node/32)) are now available on the ORNIS web site.

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](http://olla.berkeley.edu/ornisnet/?q=node/31)) and a guide to repatriation ([http://olla.berkeley.edu/ornisnet/?q=node/32](http://olla.berkeley.edu/ornisnet/?q=node/32)) are now available on the ORNIS web site.
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...]
Referencing > Repatriation Guide >

Repatriation Process

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

1. Download the zip file(s) for your ingestion from the [Table of ORNIS Georeferencing Results](#) and [Table of Mexican Bird Atlas Results](#).
2. Extract the text file(s) 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 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 fields 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=True 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=False.
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, VerticalCoordinateSystem, GeoreferenceProtocol, Georeferencerelative, Georeferencesources, Georeferencequality, Georeferencespecificity, nonconformity, nonconformityreason). Full documentation of the meaning of the fields not marked with an asterisk can be found in the [Darwin Core Quick Reference Guide](http://www.dwc.org/quick-reference/). 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 process to assist automation using the batch processing mode of the [BioGeosmancer Workbench](http://www.biosystematics.org/). The ADM 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 "Decimallatitude" field contains a value if the Decimallatitude and Decimallongitude of the georeference falls within an exclusive economic zone off the coast rather than on land. If the occurrence
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
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 HerpNET Gazetteer Guidelines. Refer to the Georeferencing Checklist for a list of who georeferenced what.
3. The localities were then returned to HerpNet, and we validated them for the correct data structure, whether they followed MNO 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 location/notes and collection/localizations fields.
4. In 2006 most institutions sent a copy of their database. If you sent this copy, we matched up the 2003 and 2006 databases and marked unharmonized localities as no longer the same.
5. In 2007 the HerpNet matched up the completed georeferenced localities to your catalog numbers and created two files, a Repatriation Files and an Exception Files. 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.
6. 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 those fields when importing the data. These fields are Locality/Notes, Locality/Vernacular, Determination, and collector/vicecurator.
7. Institutions should evaluate the files they received and update their database and the copy of their database connected to the HerpNet server as soon as possible. We recommended putting these data up right away, as you can use the HerpNet portal to check your data using the mapping feature.

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

- Institution - Name of originating institution.
- CollectionCode - An alphanumerical code used during the HerpNet project indicating your collection and collection type (for example, if MZO had separate amphibian and reptile collections, collection code would be MZO-A and MZO-R).
- Catalog - 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.
- HerpNetGeoreferencedLocality - A unique identifier used by HerpNet to condense identical values in locality descriptions into one record per institution per unique locality.

Georeferencing Fields - These 13 fields are required for georeferenced data to comply with the MNO's HerpNet GEOREF Georeferencing Guidelines. Please include all of these fields in your database and update them from the replication files. 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 verified with the MNO's HerpNet GEOREF Georeferencing Guidelines. By keeping these fields, you are allowing users of your data the ability to judge the quality of the georeference and ensure the usability of the georeferencing process.

- DecimalLatitude - Latitude in decimal degrees.
- DecimalLongitude - Longitude in decimal degrees.
- Distance - Distance from center of named place to center most edge according to maps or gazetteers.
- ExactDate - Length unit for Distance and length.
- Format - This is the three-dimensional projection of the globe associated with the coordinates. This information must be included with the DecimalLatitude and DecimalLongitude, or coordinates will not plot correctly.
- Determination - Reference map, website or gazetteer used in the georeferencing process to find coordinates. This field may contain more than 250 characters.
We expect most institutions to be able to utilize the *catalog number* field to match up results with existing systems.

We are also happy to 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:

**Latitute-Longitude Details**

<table>
<thead>
<tr>
<th>Latitude (DMS)</th>
<th>Longitude (DMS)</th>
<th>Latitude (Dec.)</th>
<th>Longitude (Dec.)</th>
<th>Latitude (DMS)</th>
<th>Longitude (DMS)</th>
<th>Conversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.29.46.337 N</td>
<td>57.41.10.413 W</td>
<td>10.29.46.337 N</td>
<td>57.41.10.413 W</td>
<td>10.29.46.337 N</td>
<td>57.41.10.413 W</td>
<td>degrees minutes seconds</td>
</tr>
</tbody>
</table>

**Determination Source:**
Topographic Map 1:100,000 Hga 6647 (Zurose), Direccin de Cartografina Nacional, Venezuela.

**Determination Method:**

**Determined By:**

**Determination Data:**
14 July 2006

**Control Latitude (DMS):**
Dec: 2.3420474, Central Longitude: 67.41.10.413 W

**Notes:**
Coordinates for Rancho Grande Biological Station; extent includes a bigger area called Rancho Grande.
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]
Symbiota portals
Field Museum Bird Collection

https://www.gbif.org/dataset/36f15a36-6b45-442e-9c31-cd633423aee0
Field Museum Birds collection GBIF metrics
Coordinate precision: Unknown
Coordinate uncertainty in meters: 200.0
Location ID: 411
Biome: Terrestrial

Data quality tests

<table>
<thead>
<tr>
<th>Test name</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geodetic datum assumed WGS84</td>
<td>Warning</td>
</tr>
<tr>
<td>Basis of record not supplied</td>
<td>Passed</td>
</tr>
<tr>
<td>Basis of record badly formed</td>
<td>Passed</td>
</tr>
<tr>
<td>Invalid collection date</td>
<td>Passed</td>
</tr>
<tr>
<td>Incomplete collection date</td>
<td>Passed</td>
</tr>
<tr>
<td>First of the month</td>
<td>Passed</td>
</tr>
<tr>
<td>Missing name of person who identified the specimen/observation</td>
<td>Passed</td>
</tr>
<tr>
<td>Collector name unparseable</td>
<td>Passed</td>
</tr>
<tr>
<td>Data are generalised</td>
<td>Passed</td>
</tr>
</tbody>
</table>
iDigBio Bird Egg collection

https://www.idigbio.org/portal/recordsets/3ff3bf5c-7aba-40c3-80b2-1b00ea1abdd5
<table>
<thead>
<tr>
<th>Map</th>
<th>Records With This Tag</th>
<th>% Records With This Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>dan_persson Rutgers</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_santiago</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_jackson</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_university</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_university_university</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_university_santiago</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_university_rutgers</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_university_jackson</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_university</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_santiago_university</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_santiago_rutgers</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_santiago_jackson</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_rutgers_university</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_rutgers_santiago</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_rutgers_jackson</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_jackson_university</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_jackson_santiago</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt_jackson_rutgers</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>ptt</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_university_university</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_university_santiago</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_university_rutgers</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_university_jackson</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_santiago_university</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_santiago_rutgers</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_santiago_jackson</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_rutgers_university</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_rutgers_santiago</td>
<td>92/100</td>
<td>92%</td>
</tr>
<tr>
<td>dan_persson_rutgers_jackson</td>
<td>92/100</td>
<td>92%</td>
</tr>
</tbody>
</table>
What are we looking at?

We are looking through a microscope at a section of the underside of a plant. The rounded bulges, representing scale-like objects growing from either side are actually tiny modified leaves! Measuring the length and width of these modified leaves (microleaves) can tell us what kind of plant we have. Let's start by marking the length of the first microleaf.
Zooniverse - Rainforest flowers
https://www.zooniverse.org/projects/tomomi/rainforest-flowers
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...
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...]

- Repatriating
WHAT COULD HELP?

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

- Repatriating
  
  [...one place to find them all...]

CC BY-NC-ND 3.0 Unported - 5433203 Beatrix Addressing Dante from the Car: 1824–7
WHAT COULD HELP?:

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

- Repatriating
  - [...one place to find them all...]

- Reporting
  - [...even more standards...]
Famous People Painting - *Discussing the Divine Comedy with Dante*

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
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:
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
The goal of this project is to provide a mechanism whereby groups of users can form communities to collaboratively georeference and verify a shared dataset. This collaborative georeferencing framework consists of two end-user components:

1. The GEOLocate web-based collaborative client for reviewing and editing community records.
2. Web-based data management portal for creating and managing communities, their respective users and data sources.

Shared community datasets created via the portal may consist of multiple underlying data sources from either live DGSIR providers and/or uploaded text files. Support for TAPIR providers is currently under development. Data are stored using the full Darwin Core 1.2 specification, but subsets and/or alternatives schemas may be imported using the schema mapping interface. During import, data items are automatically normalized, georeferenced and related to one another via a similarity index. This index is used to identify all records that appear to describe the same collection locality regardless of syntax. During coordinate verification, users have the option to re-classify records that were incorrectly related to one another.

Verification and correction of the computer generated geographic coordinates is accomplished using the GEOLocate desktop application. GEOLocate allows users to login to their communities, retrieve and visualize results, make any necessary corrections, provide additional comments, define errors as polygons, and save the results back to the shared dataset. The verified results of georeferencing can then be downloaded via the portal data management interface for re-import to the parent database.

To examine the gains in efficiency over traditional georeferencing 2100 randomly selected collecting events from the TUMNH fish collection were imported and georeferenced using the collaborative georeferencing framework. The TUMNH fish collection was georeferenced by hand in the mid to late 90’s and therefore provides a useful test bed for assessing the efficiency and accuracy of automated methodologies. Of the 2100 records, 10% were identified as being similar to other records and an additional 33% were duplicates leaving a
Some history - part 1

I. No Standard way for requesting data
   1. Historically, no/wrong contact asked for data. (e.g., collections folks give different data/records/fields from what tech would give...e.g., +/- irm, site-vs catalog-records )
   2. Darwin Core was not always requested/specifed
   3. No agreement / document of what will be added/annotated/changed
   4. No project plan with dates (in theory we can have data out there now we don’t know about)

...Big mystery....

[added 10-aug-2018]
II. When we get data back, we reap the consequences...

1) Data came back to [who at FM?] with inappropriate structure/missing fields
2) In the 7 years data in our catalog had changed
   a) Who’s georeferencing do we use?
   b) Makes matching difficult w/o unique id
   c) In this case completely different DBP & structure as well
3) Data went out flat & had to be integrated into relationships

[added 10-aug-2018]
Some history - part 3

III. No / Incomplete documentation of annotated data decision process

1) Judgement calls had to be made by staff for validity-judgements because documentation wasn’t given
2) Georeferencing taxonomically was not noted therefore data ended up in Sites table, not Taxonomy table

[added 10-aug-2018]
[Heavenly/Less-Hellish?] things that might help

In case this ties into where you want to go:
- Standardizing how “annotations/augmentations/etc” are structured
  - [ALA + GBIF]
- & awareness of other potentially related fields outside biodiversity informatics & their diabolical data management & standards schemes
Levels are in order of [ Heaven | Purgatory | Hell ]

Hell: and Circles 8 and 9 for the sins of malice (fraud and treachery)
Purgatory: Humans can sin by using love that is too strong (Lust, Gluttony, Greed).
Levels are in order of [ Heaven | Purgatory | Hell ]

Hell: and Circles 8 and 9 for the sins of malice (fraud and treachery)
Purgatory: Humans can sin by using love that is too strong (Lust, Gluttony, Greed).
Levels are in order of [ Heaven | Purgatory | Hell ]

Hell: Circle 7 for the sins of violence
Purgatory: Humans can sin by using love that is too strong (Lust, Gluttony, Greed).
Levels are in order of [ Heaven | Purgatory | Hell ]

Hell: Circle 6 contains the heretics who contradicted the doctrine and confused the spirit of Christ.

Purgatory: Humans can sin by using it to proper ends but with love that is either not strong enough (Sloth)
Levels are in order of [ Heaven | Purgatory | Hell ]

for the four sins of indulgence (lust, gluttony, avarice, anger);
Purgatory: Humans can sin by using love towards improper or malicious ends (Wrath, Envy, Pride),
Levels are in order of [ Heaven | Purgatory | Hell ]

Hell: for the four sins of indulgence (lust, gluttony, avarice, anger);
Purgatory: Humans can sin by using love towards improper or malicious ends (Wrath, Envy, Pride),
3rd Level of [Hv/Pu/He] - [ Venus | Pride | Gluttony ]

Levels are in order of [ Heaven | Purgatory | Hell ]

Hell: for the four sins of indulgence (lust, gluttony, avarice, anger);
Purgatory: Humans can sin by using love towards improper or malicious ends (Wrath, Envy, Pride).
Levels are in order of [ Heaven | Purgatory | Hell ]

Hell: for the four sins of indulgence (lust, gluttony, avarice, anger);

Purgatory: 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.
Levels are in order of [ Heaven | Purgatory | Hell ]

**Hell:** Limbo, in Circle 1, contains the virtuous pagans who were not sinful but were ignorant of Christ,

**Purgatory:** Below the seven purges of the soul is the Ante-Purgatory, containing the Excommunicated from the church.
Angel thing
mere Dante (based on this / this?)
Exploding-head Dante
mere mortal
“Devil doll” (based on this one)