Issues in Data Preservation for Biodiversity Collections

Cody Thompson, University of Michigan
Laura Abraczinskas, Michigan State University
Diego Barroso, University of Michigan

Overview

Data issues from three perspectives:

- Issues in a single collection (Cody)
- Issues across vertebrate collections (Laura)
- Issues across biological collections (Diego)

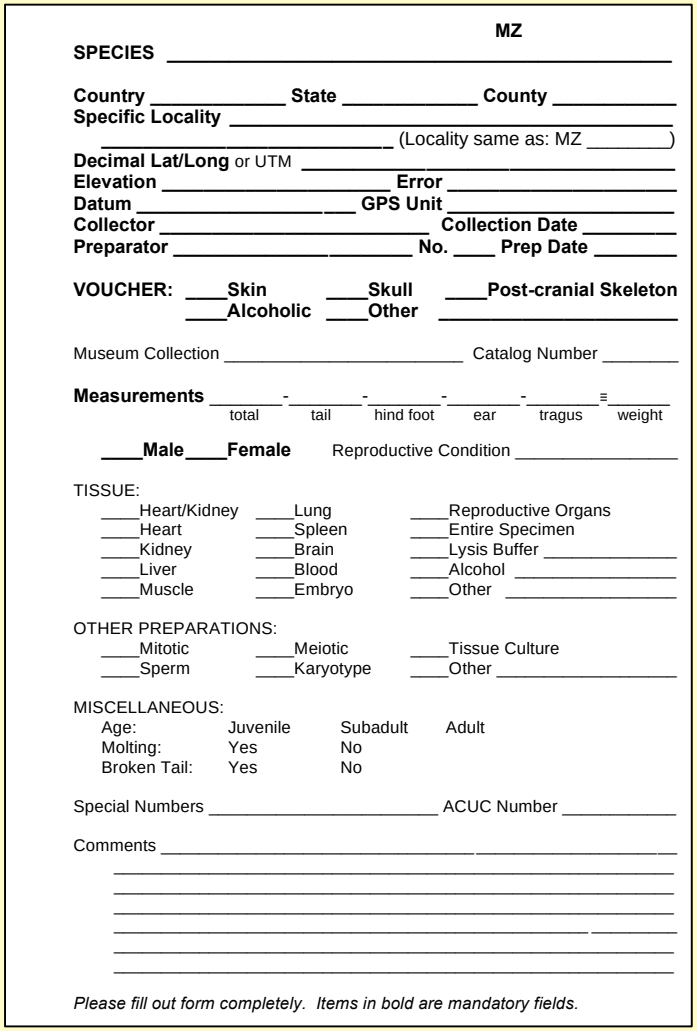
Group Discussion

What are the data issues?

- **Inconsistent data collection in the field**
  - Lack of upfront communication
  - Delay in receiving field notes
- **No recent OP or pick lists for entering data in database**
  - Duplicate numbers, data redundancy, etc.
- **Digitization limited to core data**
  - Very little “extra” data (e.g., measurements, field notes, images, etc.)
- **Prior to Specify migration, data were not relational**
  - No connection between vouchers and associated tissues
- **Issues with specimen tracking**

What are the solutions?

- **Repurposed existing field series system**
  - Standardized data fields
  - Included bar code voucher tags and tissue labels
  - Increased data availability for immediate collection use
- **Updating standard OPs and creating pick lists**
  - The most recent mammal collection OP was from 1973
  - Using Wilson & Reeder as taxonomic standard
  - Other pick lists generated on “cleaned” data



MZ

SPECIES _____

Country _____ **State** _____ **County** _____
Specific Locality _____
_____ (Locality same as: MZ _____)
Decimal Lat/Long or UTM _____
Elevation _____ **Error** _____
Datum _____ **GPS Unit** _____
Collector _____ **Collection Date** _____
Preparator _____ **No.** ____ **Prep Date** _____

VOUCHER: **____Skin** **____Skull** **____Post-cranial Skeleton**
____Alcoholic **____Other** _____

Museum Collection _____ Catalog Number _____

Measurements _____-_____-_____-_____-_____≡_____
total tail hind foot ear tragus weight

____Male____Female Reproductive Condition _____

TISSUE:

____Heart/Kidney	____Lung	____Reproductive Organs
____Heart	____Spleen	____Entire Specimen
____Kidney	____Brain	____Lysis Buffer _____
____Liver	____Blood	____Alcohol _____
____Muscle	____Embryo	____Other _____

OTHER PREPARATIONS:

____Mitotic	____Meiotic	____Tissue Culture
____Sperm	____Karyotype	____Other _____

MISCELLANEOUS:

Age:	Juvenile	Subadult	Adult
Molting:	Yes	No	
Broken Tail:	Yes	No	

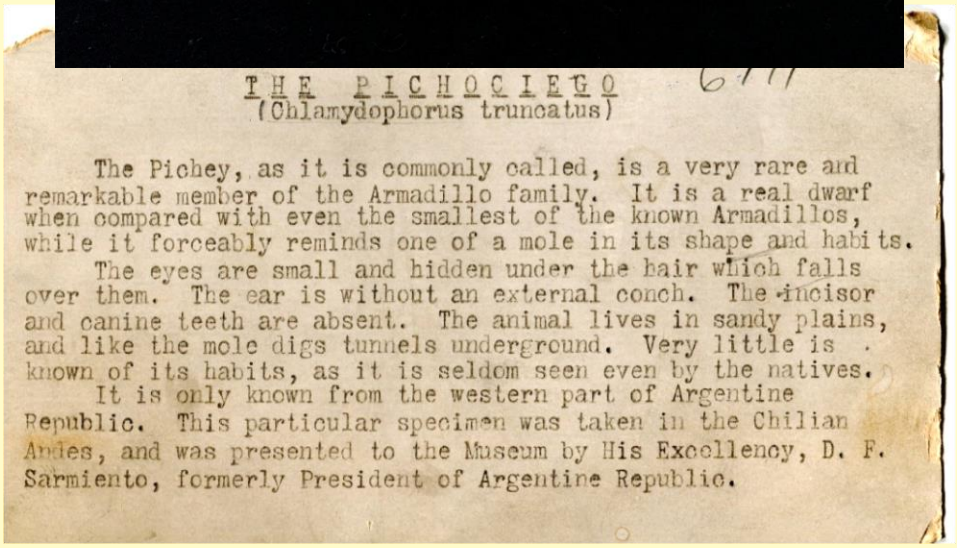
Special Numbers _____ ACUC Number _____

Comments _____
_____
_____
_____
_____
_____
_____

Please fill out form completely. Items in bold are mandatory fields.

What are the solutions?

- **Digitize “any and all” data**
  - May not seem important now but likely helpful later
  - Opportunistic data entry when specimen in hand
- **Use unique key to relate tissue collections (e.g., MZ series)**
  - MZ numbers assigned *post hoc* to previously archived tissue
  - Specify, however, does not seem to update these “related” collections
- **Scan and enter all transactional records**
  - Priority given to current loans and accession records



THE PICHOCIEGO
(Chlamydophorus truncatus)

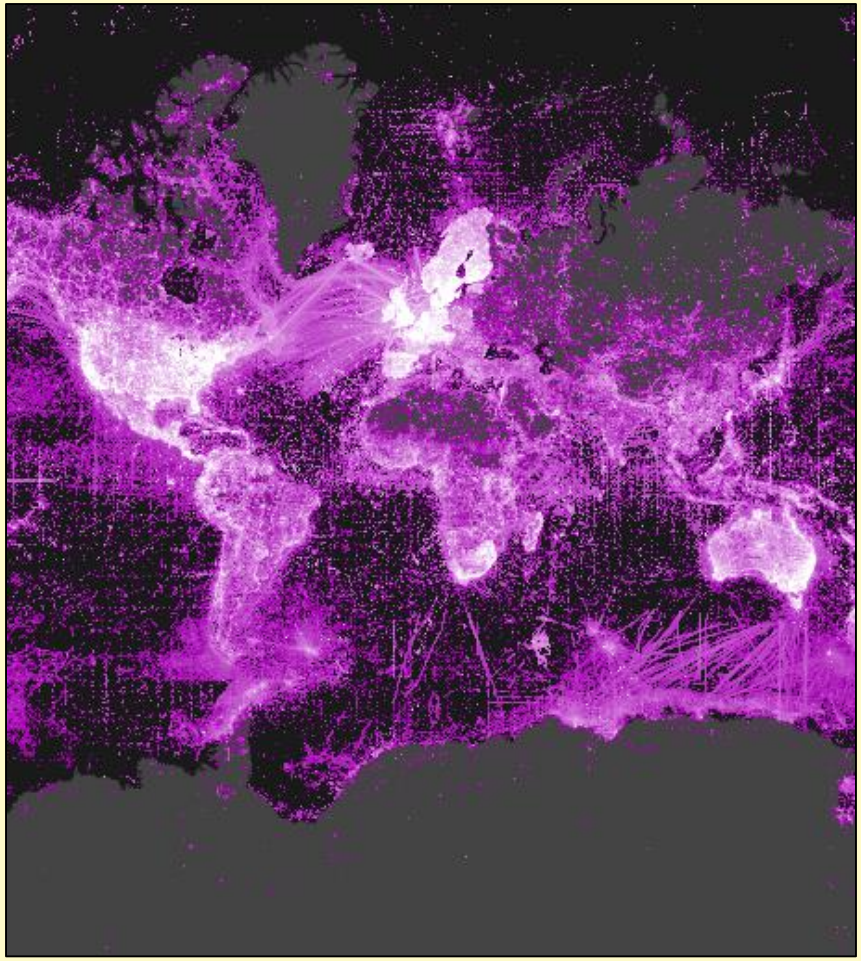
The Pichey, as it is commonly called, is a very rare and remarkable member of the Armadillo family. It is a real dwarf when compared with even the smallest of the known Armadillos, while it forceably reminds one of a mole in its shape and habits.

The eyes are small and hidden under the hair which falls over them. The ear is without an external conch. The incisor and canine teeth are absent. The animal lives in sandy plains, and like the mole digs tunnels underground. Very little is known of its habits, as it is seldom seen even by the natives.

It is only known from the western part of Argentine Republic. This particular specimen was taken in the Chilian Andes, and was presented to the Museum by His Excellency, D. F. Sarmiento, formerly President of Argentine Republic.

What are the future data issues?

- **Unfortunately, we do not live in a vacuum…**
  - Big data is here and it is here to stay!
- **When is enough data, enough?**
  - Probably not possible, but we are biologists not computer scientists
- **How do I connect the UMMZ mammal collection to the world without losing its integrity?**
  - Maintain visitor traffic
  - Attribution in citations
- **Will push to digitize eliminate the gold standard?**
  - Vouchers, paper copies, etc.
  - Keep doing mammalogy!

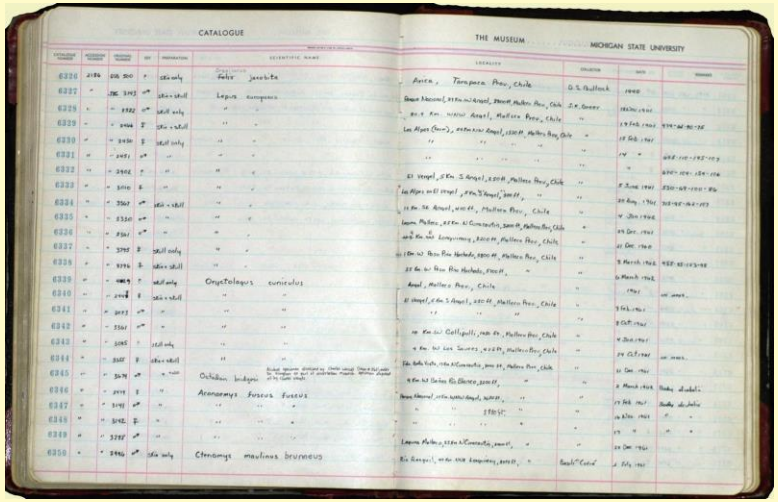


Source: GBIF

Issues Across Vertebrate Collections

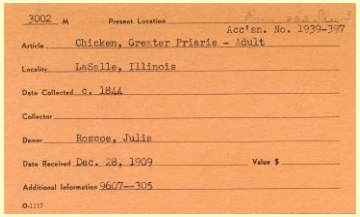
MICHIGAN STATE UNIVERSITY MUSEUM
the science and culture museum at Michigan State University

- Founded in 1857 – one of the oldest museums in the U.S.
- Houses over **111,000 vertebrate specimens**
- Earliest specimens date from 1844
- Began digitizing vertebrate collections in 1993

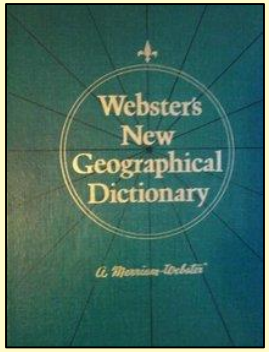


Data converted to *Specify* in 2004

DwC Database modified to Darwin Core



Strategies and Impacts



1) Entered data directly from catalog ledgers & cards

2) Wrote digitization protocols

3) Same data fields across extant vertebrate disciplines

4) Different data fields for vertebrate paleontology

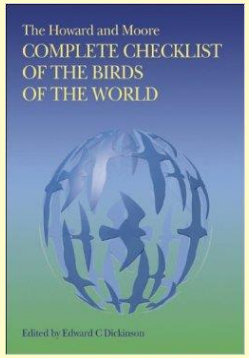
5) Utilized standard resources for taxonomy & geography

6) Standardized names of collectors

when certain (e.g. affiliated with our collections or institution)

Example: J. Alan Holman

instead of J.A. Holman, Al Holman, Holman



Strategies and Impacts

7) Utilized standard terminology for **specimen preparation types** across all vertebrate groups

From ***Element Names and Modifiers***, J. Howard Hutchison, Appendix 2, pp119-124 In: *Guidelines and Standards for Fossil Vertebrate Databases*. Blum, Stanley, D (ed) 1991. SVP.

• Employed standard "**Collective**" terms (skull, skeleton) and expanded the list to include all preparations of our specimens

• Employed standard "**Elements**" terms from comparative anatomy (over vernacular or medical terminology)

• Combined **collective** and **element**s terms into a preparation types pick list in our database

*A specimen in our collection (and database) may have **multiple preparation types**: skull, axis, atlas, baculum, skin*

Data Issues

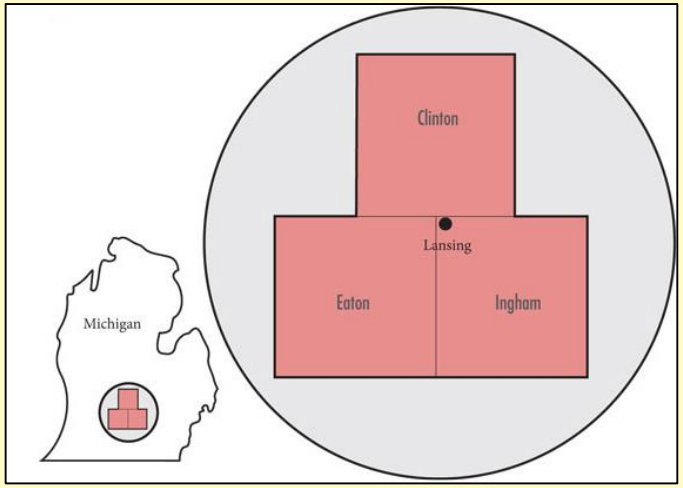
- Over its 159-year history, the Museum employed multiple cataloging and numbering systems for the vertebrate collections
- We have two **institution codes** for our vertebrate collections

  MSU for mammalogy, ornithology, & vertebrate paleontology
  MSUM for herpetology & ichthyology

- Duplicate field number series inadvertently assigned by prolific collector
- Local county confusion

  City of Lansing in Michigan is at the intersection of 3 counties

  City of East Lansing in Michigan is in parts of 2 counties



Data Issues

- Specimens that are hybrid crosses

  Numerous hybrid crosses are indicated in our ichthyology, herpetology, ornithology , and mammalogy records

- Specimen measurement and weight data are not currently online

Researcher Needs

•Our two acronyms may be causing confusion to users of online data (incorrect searches or misinterpreted returned data records)

•Increase in **requests for specimen color morph data** and/or images from researchers (herpetologists, ornithologists, and mammalogists)

•Increase in **requests for specimen measurements & weights** from researchers (mammalogists in particular)

GREAT LAKES
INVASIVES NETWORK

Data Across Collections

Herbarium

UMMZ

MICH

Museum of Zoology

UMMZ

Future “repatriation”

Symbiota Promoting Bio-Collaboration

Herbarium DB

MICH

(B. Walters)

snapshots

migrating to

(J. Torgersen)

DwC
CSV

Tri-Trophic Thematic Network

New York Botanical Garden

completed
(E. Gilbert)
(B. Brandt)

Consortium of
Midwest
Herbaria

IIT

(R. Anglin)

GREAT LAKES
INVASIVES NETWORK

In-house DB ≠ DwC ≠ Symbiota

GREAT LAKES
INVASIVES NETWORK

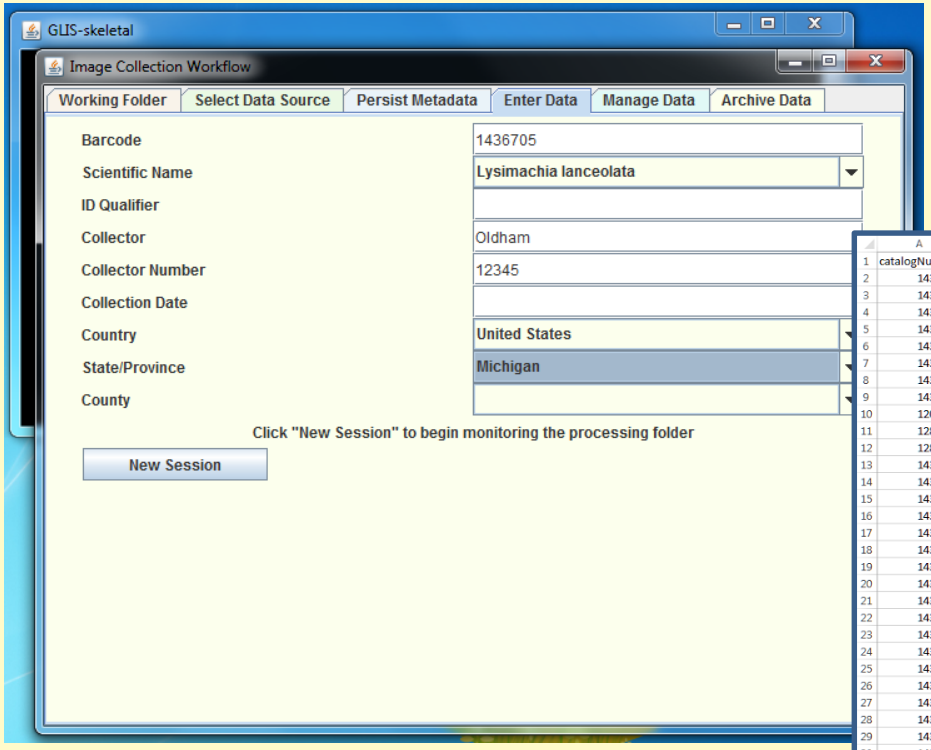
Herbarium MICH

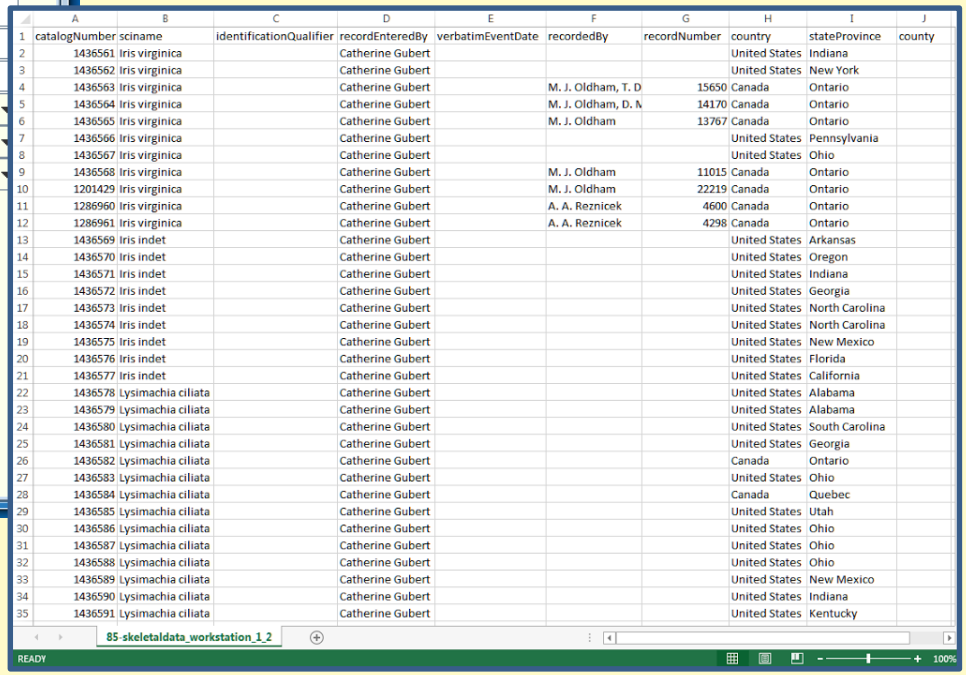
- MICH-V- prefix for Vascular Plants (not in barcodes!)
- Collectors vs. (Collector+Associated Collectors); initials et al.
- Date format: 04 Apr 2016 vs. 2016-04-04
- (Elevation + Units) → Elevation in Meters
- Verbatim Lat/Long ; Decimal Lat/Long ; UTM , TRS
- previousIdentifications field

GREAT LAKES
INVASIVES NETWORK

Herbarium MICH

- Michigan records are complete
- Others (North America), Java App for CSV skeletal records





	A	B	C	D	E	F	G	H	I	J
1	catalogNumber	sciname	identificationQualifier	recordEnteredBy	verbatimEventDate	recordedBy	recordNumber	country	stateProvince	county
2	1436561	Iris virginica		Catherine Gubert				United States	Indiana	
3	1436562	Iris virginica		Catherine Gubert				United States	New York	
4	1436563	Iris virginica		Catherine Gubert		M. J. Oldham, T. D	15650	Canada	Ontario	
5	1436564	Iris virginica		Catherine Gubert		M. J. Oldham, D. N	14170	Canada	Ontario	
6	1436565	Iris virginica		Catherine Gubert		M. J. Oldham	13767	Canada	Ontario	
7	1436566	Iris virginica		Catherine Gubert				United States	Pennsylvania	
8	1436567	Iris virginica		Catherine Gubert				United States	Ohio	
9	1436568	Iris virginica		Catherine Gubert		M. J. Oldham	11015	Canada	Ontario	
10	1201429	Iris virginica		Catherine Gubert		M. J. Oldham	22219	Canada	Ontario	
11	1286960	Iris virginica		Catherine Gubert		A. A. Reznicek	4600	Canada	Ontario	
12	1286961	Iris virginica		Catherine Gubert		A. A. Reznicek	4298	Canada	Ontario	
13	1436569	Iris indet		Catherine Gubert				United States	Arkansas	
14	1436570	Iris indet		Catherine Gubert				United States	Oregon	
15	1436571	Iris indet		Catherine Gubert				United States	Indiana	
16	1436572	Iris indet		Catherine Gubert				United States	Georgia	
17	1436573	Iris indet		Catherine Gubert				United States	North Carolina	
18	1436574	Iris indet		Catherine Gubert				United States	North Carolina	
19	1436575	Iris indet		Catherine Gubert				United States	New Mexico	
20	1436576	Iris indet		Catherine Gubert				United States	Florida	
21	1436577	Iris indet		Catherine Gubert				United States	California	
22	1436578	Lysimachia ciliata		Catherine Gubert				United States	Alabama	
23	1436579	Lysimachia ciliata		Catherine Gubert				United States	Alabama	
24	1436580	Lysimachia ciliata		Catherine Gubert				United States	South Carolina	
25	1436581	Lysimachia ciliata		Catherine Gubert				United States	Georgia	
26	1436582	Lysimachia ciliata		Catherine Gubert				Canada	Ontario	
27	1436583	Lysimachia ciliata		Catherine Gubert				United States	Ohio	
28	1436584	Lysimachia ciliata		Catherine Gubert				Canada	Quebec	
29	1436585	Lysimachia ciliata		Catherine Gubert				United States	Utah	
30	1436586	Lysimachia ciliata		Catherine Gubert				United States	Ohio	
31	1436587	Lysimachia ciliata		Catherine Gubert				United States	Ohio	
32	1436588	Lysimachia ciliata		Catherine Gubert				United States	Ohio	
33	1436589	Lysimachia ciliata		Catherine Gubert				United States	New Mexico	
34	1436590	Lysimachia ciliata		Catherine Gubert				United States	Indiana	
35	1436591	Lysimachia ciliata		Catherine Gubert				United States	Kentucky	

85-skeletaldata_workstation_1_2

READY

(R. Anglin)

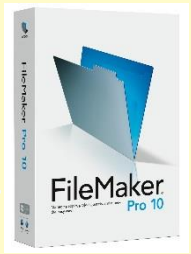
Fish DB

(D. Nelson)

MUSE

1990's , limited fields

snapshots;



server

migrated to

(J. Torgersen)

DwC
CSV

IIT

georeferencing done

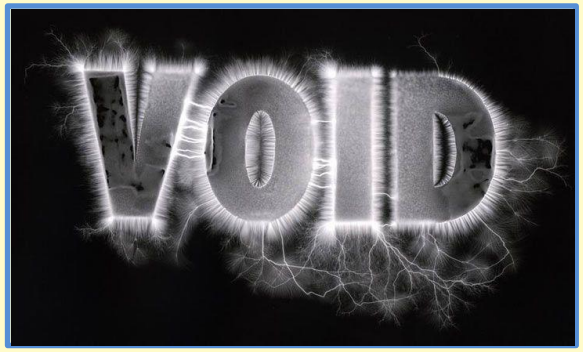
GREAT LAKES
INVASIVES NETWORK

- UMMZ-F- prefix
- Original taxonomy
- Mapping remarks to original fields?
- Leveraging georef. & existing images

In-house DB ≠ DwC ≠ Symbiota

Mollusk DB

(T. Lee)



Invert eBase

label transcription from scratch
100% done by army of students

DwC
CSV

IIT

- UMMZ-MOL- prefix; renaming images
- Skeletal records only
- Leveraged transcription; contributing images

Group Discussion

http://www.spokesman.com/stories/2013/jul/14/the-slice-last-one-in-the-pool-is-a-sweltering/