Marcia A. Revelez
Angelo State Natural History Collections
Angelo State University, San Angelo, TX

Digitization at ASNHC:

A SMALL COLLECTIONS APPROACH

Overview of Presentation

1. Overview of Small Collections
2. ASNHC Workflow
3. Helpful Tools & Resources

1 **Overview** of Small Collections

History of Collection Surveys by ASM

1923 by A. Brazier Howell

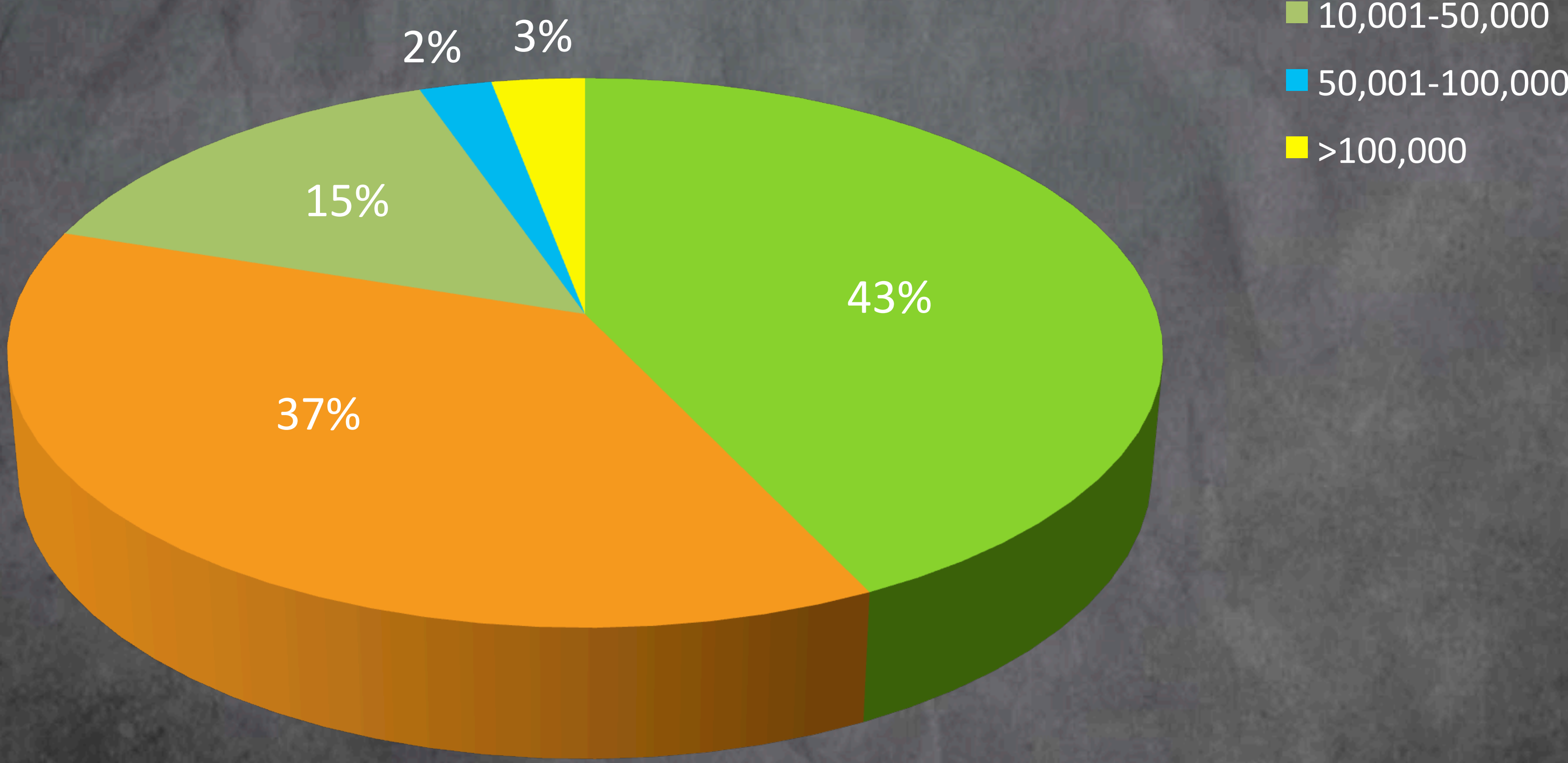
- 37 public and 40 private collections in North America (U.S. & Canada only, sorry Mexico)
- ½ of all specimens at Smithsonian

> "The larger museums are relatively sufficient unto themselves; but the smaller ones, as well as unattached individuals, are often handicapped…it is hoped the appearance of this list will have is the stimulation of the interest of the small collector, the beginner, and the amateur."

Year	No. Collections	Growth of Specimens Added Between Surveys	% Growth	Comment
1945	297			
1961	307			
1973	388	1963-1973: 900,000 specimens = 90,000/yr	57%	
1983	474	1973-1983: 600,000 specimens = 60,000/yr 3,136,557 specimens	24%	First mention of digitization
1996	329	1983-1995: 670,000 specimens = 56,000/yr 3,806,557 specimens	21%	Increase in number of collections with 10,000 or less
2007	302	1995-2007: 317,508 specimens = 26,500/yr 4,124,065 specimens	8%	

Size of Mammal Collections in United States

307 Collections
3,886,071 specimens

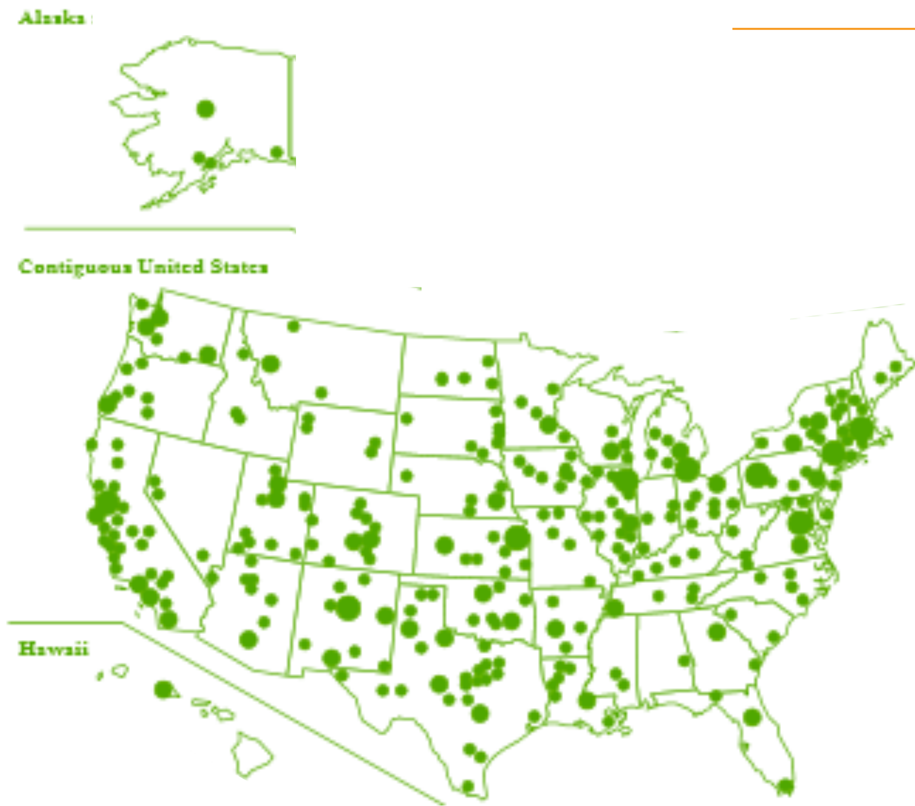


ASM Systematics Committee, 2007

Location of Collections in United States

ASM Systematics Committee, 2007

307 Collections
3,886,071 specimens



Map from
ASM, 1997

104 databased **34%**

41 accessible to public **13%**

Collections 2015
Comparison of Portals

iDigBio

- 79 collections
- 3,341,640 specimens

ASM Systematics Committee Survey, 2007

- 307 Collections
- 3,886,071 specimens

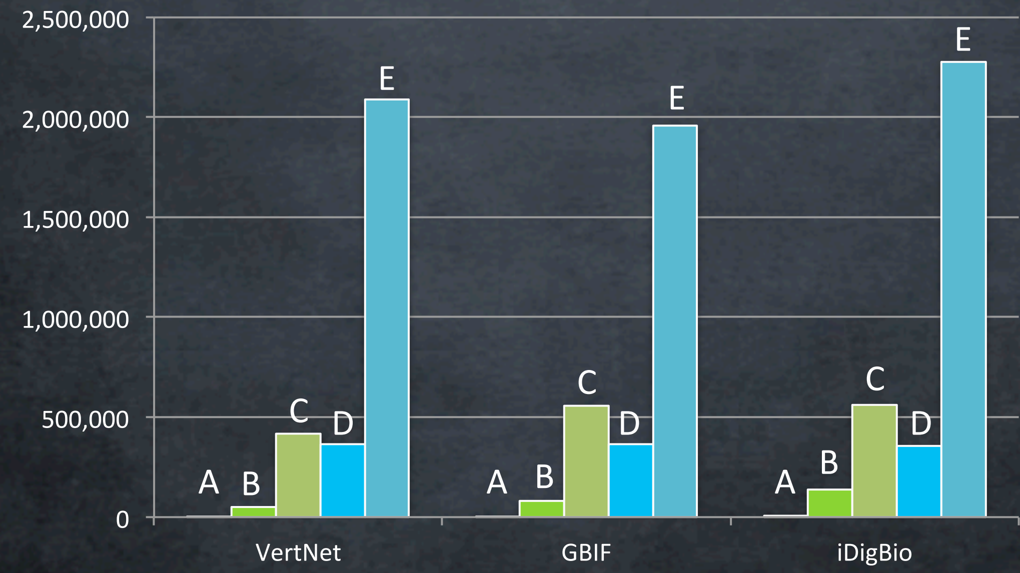
Over half million specimens left
Mostly from small collections
LIKELY MORE

225+ Collections

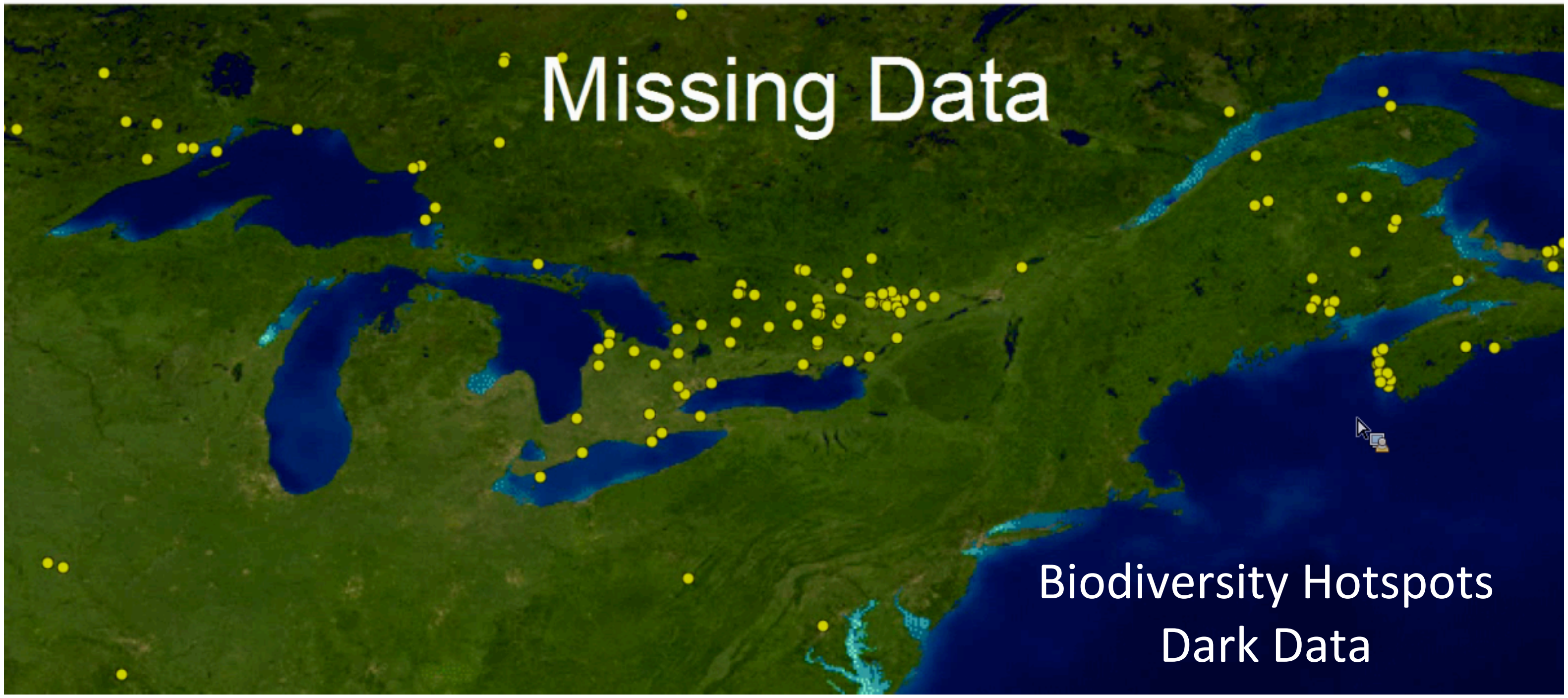
Number of Collections in Each Portal by Size

	A	B	C	D	E	
Portal	**<1,000**	**1,001-10,000**	**10,001-50,000**	**50,001-100,000**	**>100,000**	**Total**
VertNet	1	15	18	5	9	48
GBIF	3	16	22	5	8	54
iDigBio	10	30	23	6	10	79
ASM (2007)	169	154	57	9	10	307

Portal Breakdown: Who has the most specimens?



Why Digitize Small Collections



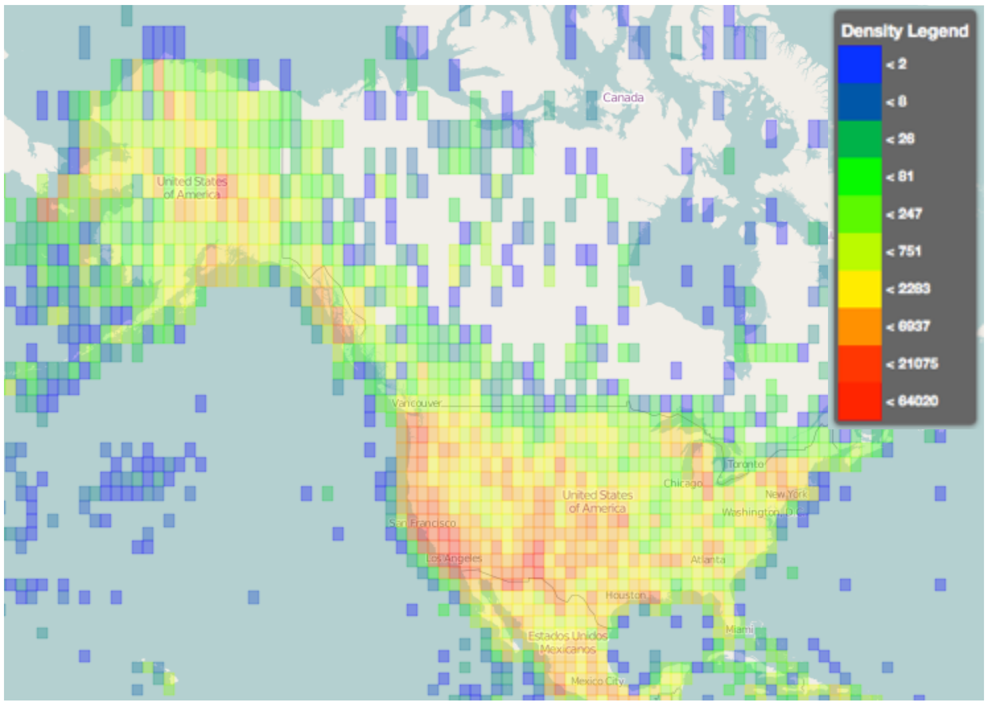
There are many large natural history collections, but it's really the small collections that provide detailed knowledge of the flora and fauna of local and regional communities.

- Hilary Moore, Importance of Small Museum Collections

DIGITIZATION INCREASES VALUE OF YOUR COLLECTION

Advocating your collection = Advocating your job

Concentration of Records vs Location of Collections



iDigBio, 2015

ASM, 1997

Small Collections
Big Benefits

CHALLENGES

- Limited Budget
- Limited Time
- Limited Staff
- Lack of IT Support

Successfully Curating Smaller Herbaria and Natural History Collections in Academic Settings, Neil Snow, *BioScience* 2005

2

ASNHC Workflow

Angelo State Natural History Collections

- 5 Collections
  - ❖ Mammalogy
  - ❖ Ornithology
  - ❖ Herpetology
  - ❖ Herbarium
  - ❖ Frozen Tissues

~ 150,000 objects

- 5 Curators – all full-time faculty with no formal time allocated toward collections
- 1 Full-time Collections Manager – hired 2013

Angelo State University

ASNHC
Angelo State Natural History Collections

- Began in 1968 (3 curators)
- 4^{th} largest in Texas
- Nearly 50% of collection is from Texas. Also has concentrations from Mexico, Guyana, and Ecuador

- Mostly databased
- Specify
- Approximately 18,000

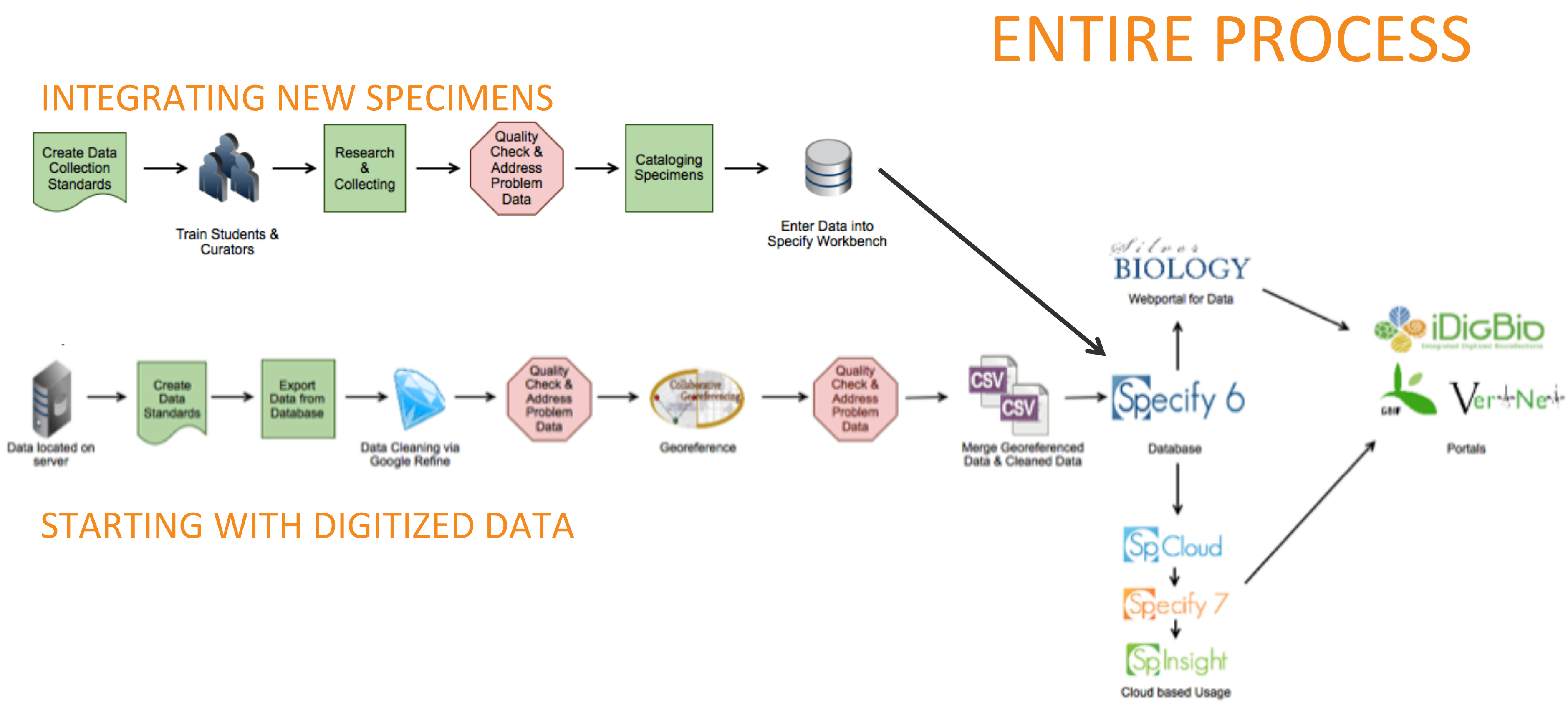
Problems
Nothing Georeferenced
Standardization

Mammalogy Collection

CURATOR

Dr. Robert C. Dowler

Digitization Workflow for Mammalogy Collection at ASNHC



3 Helpful Tools & Resources

digitization at ASNHC: a small collections approach

Google refine

Now Called Open Refine

- Power tool for working with messy data, cleaning it up, transforming it from one format into another, extending it with web services, and linking it databases like Freebase
- Not a web service, but a desktop application, interact with it on web browser
  Data can be in spreadsheet or other format: TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, Wiki markup, and Google Data documents

OpenRefine Core

Google Refine 2.5 - Stable version

- **Windows kit**, Download, unzip, and double-click on *google-refine.exe*. If you're having issues with the above, try double-clicking on *refine.bat* instead.
- **Mac kit**, Download, open, drag icon into the Applications folder and double click on it. **NOTE:** If you have issues installing Refine on Mac, please refer to issue 590 - OpenRefine 2.5 for mac support java 6 and 7 only
- **Linux kit**, Download, extract, then type *./refine* to start. **NOTE:** OpenRefine 2.5 for linux support java 6 and 7 only

http://openrefine.org/download.html

Data Cleaning Tools

Configuring the File

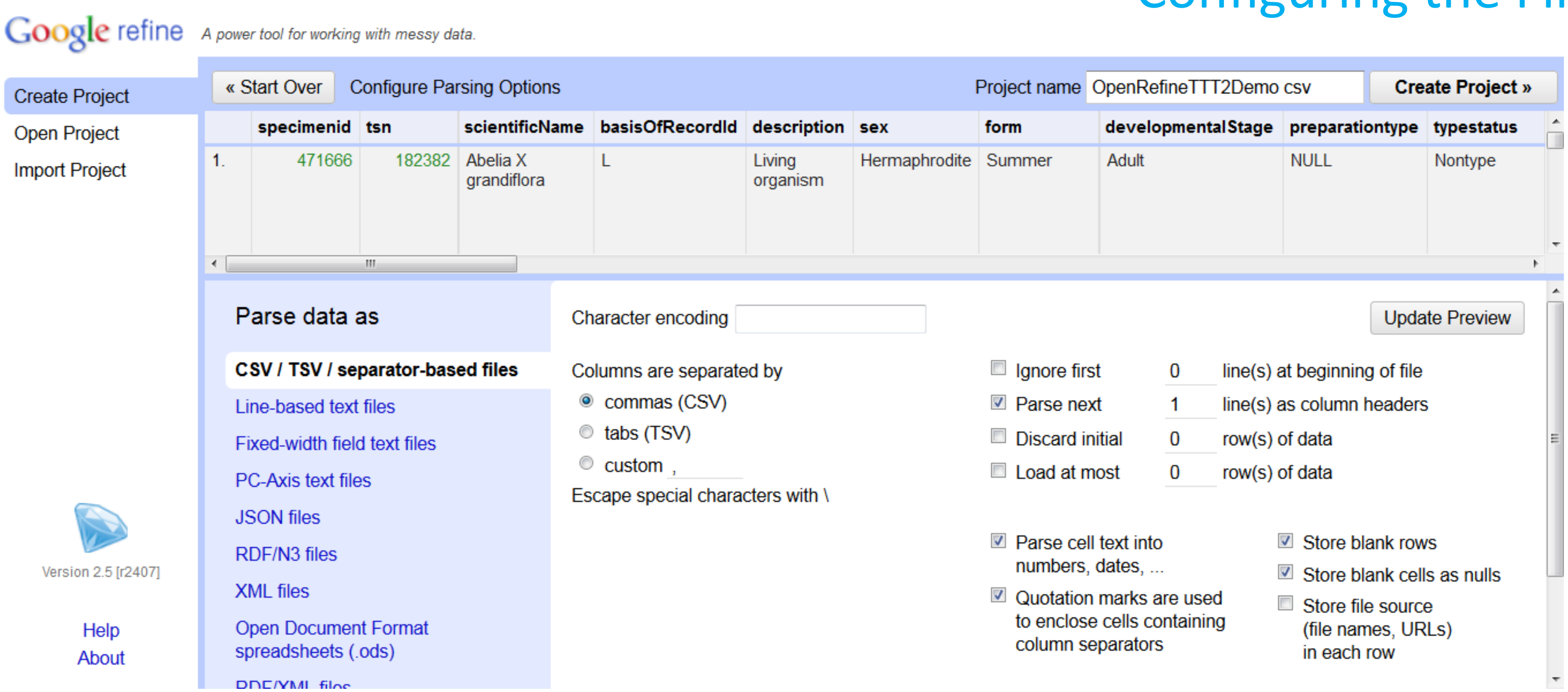
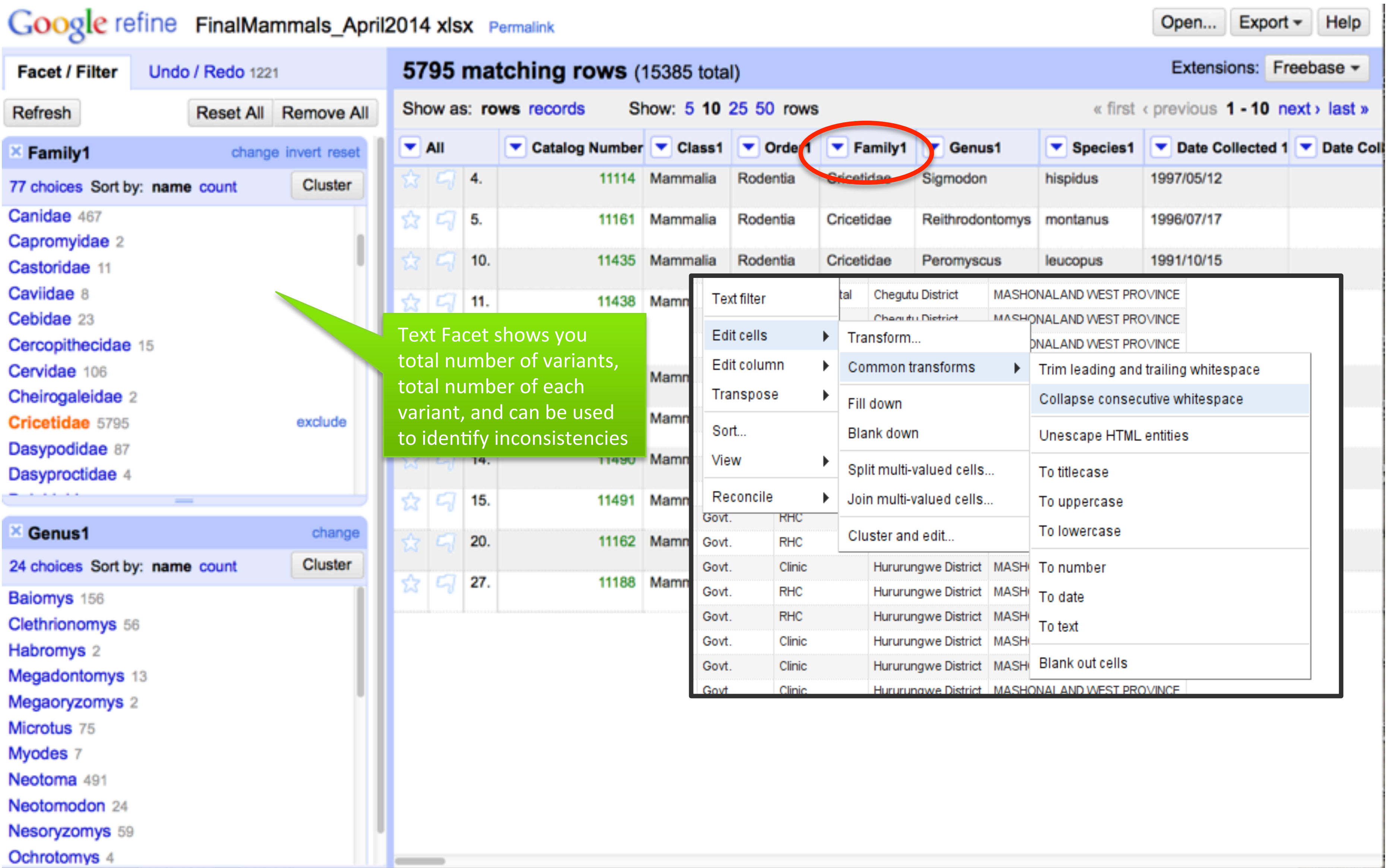


Image by Deb Paul

- Special Characters (ä, ü) - choose “UTF-8” encoding
- Custom radio button if your file is tab-delimited, or has a delimiter different than comma
- Discard rows, ignore rows, parse data, in this initial step to create a project.
- NOTICE: “Parse cell text into numbers, dates, …” Untick this box unless you want this program to convert your data into numeric or date format. If you want your dates verbatim, and the significant digits of your numbers unchanged, untick this box.

Data Cleaning Tools



Data Cleaning Tools

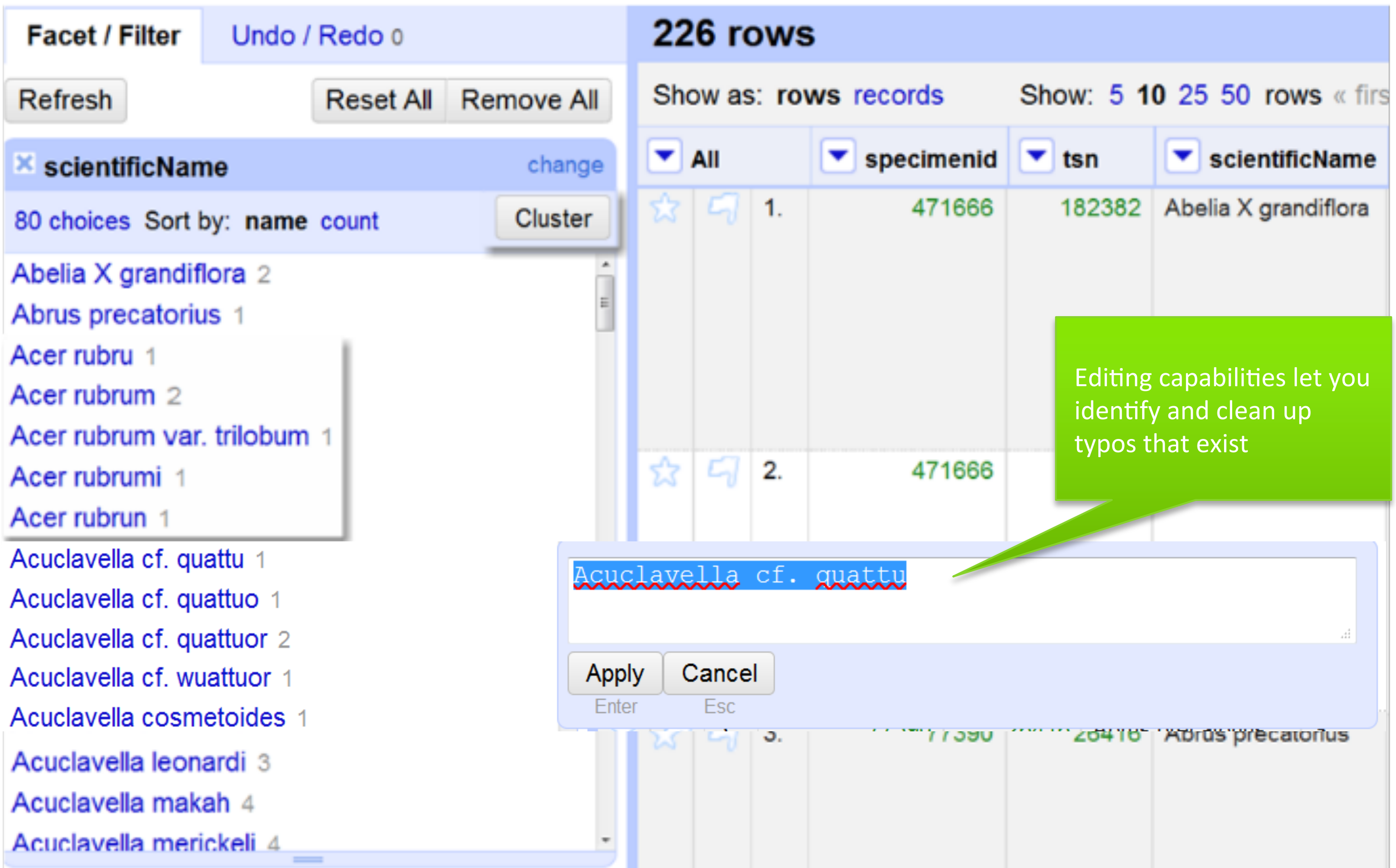
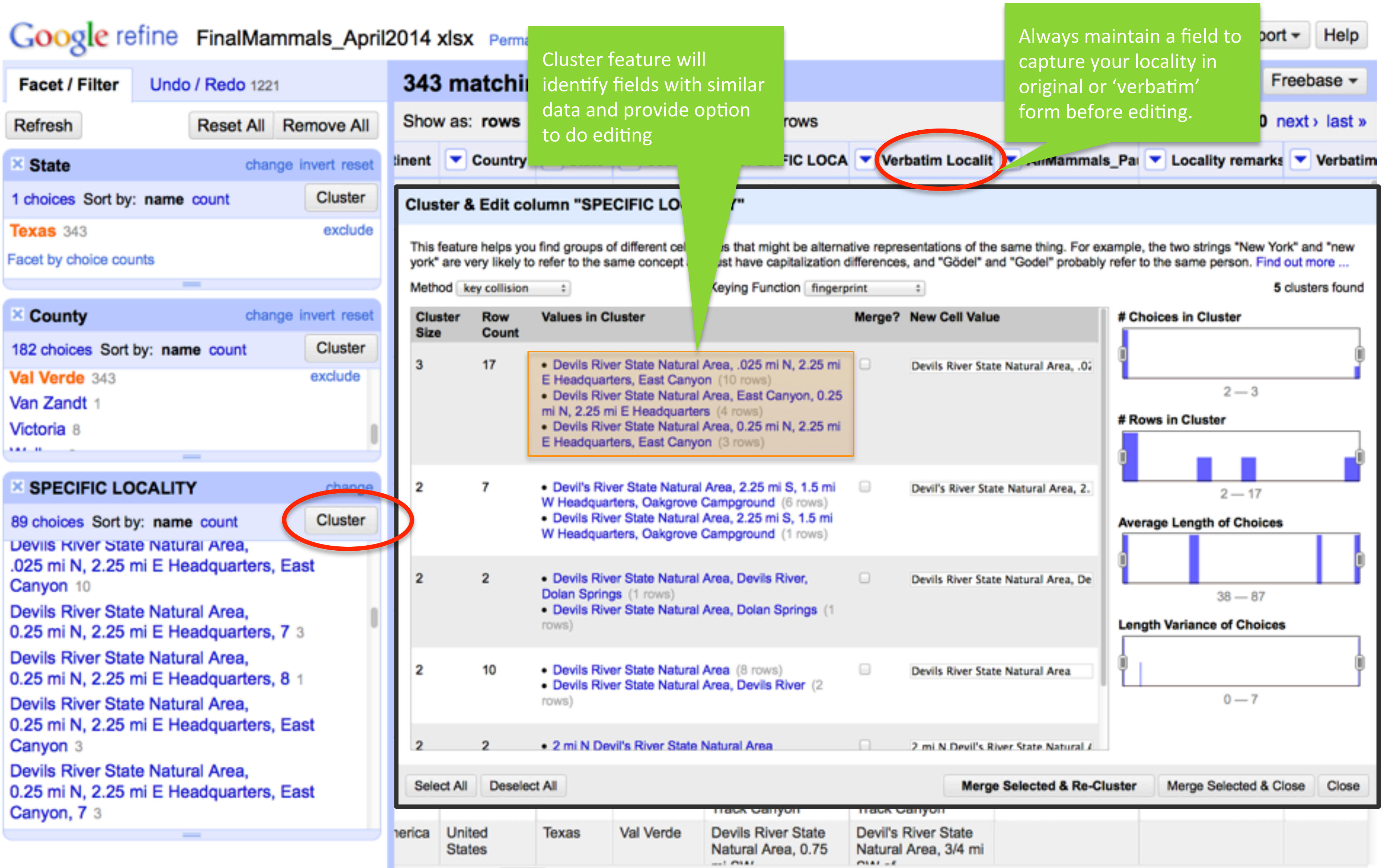
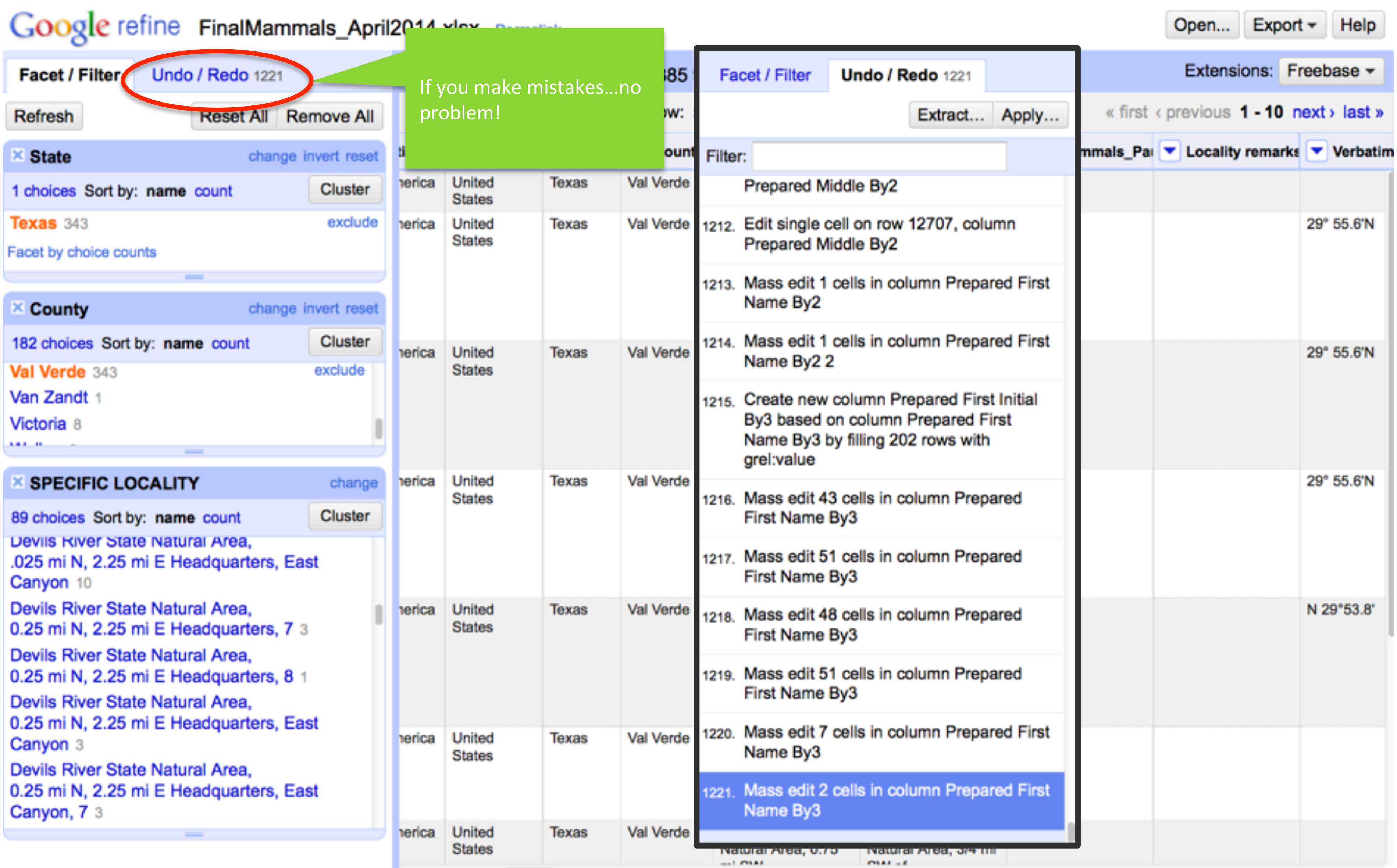


Image by Deb Paul

Data Cleaning Tools

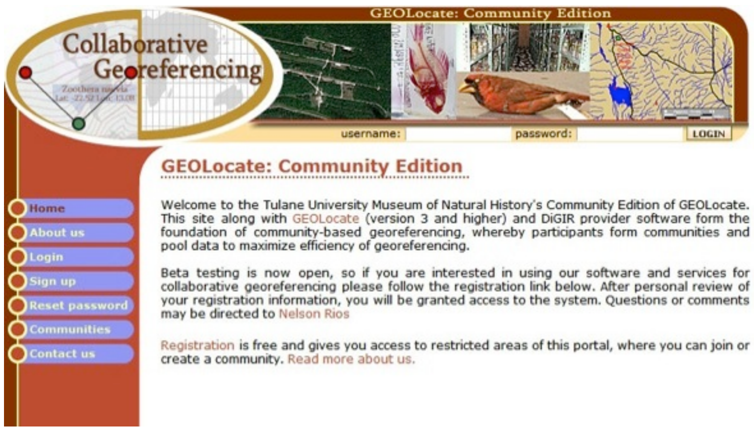


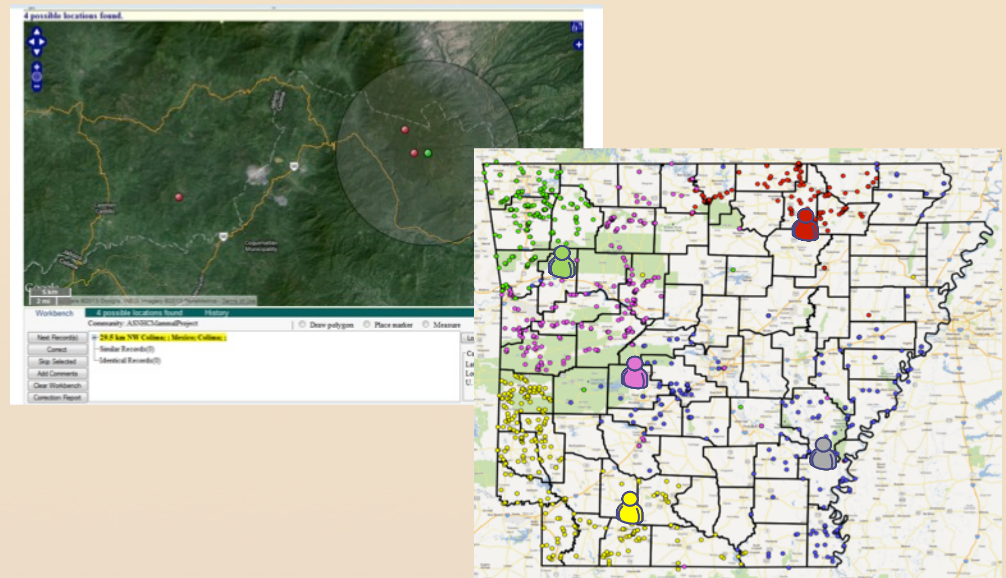
Data Cleaning Tools



Georeferencing Tools

GEOLocate





Collaborative Georeferencing

GEOREFERENCING
QUICK REFERENCE GUIDE
Version: 2012-10-02
John Wieczorek, David Bloom, Heather Constable, Janet Fang, Michelle Koo, Carol Spencer, Kristina Yamamoto

This is a practical guide for georeferencing using the point-radius method [1, 2, 3] using the Georeferencing Calculator [4, 5], maps, gazetteers, and other resources from which coordinates and spatial boundaries for places can be found. This guide is an update of "Georeferencing for Dummies" [6], and explains the recommended calculation procedure for localities encountered in the georeferencing process.

Georeferences using the methods in this guide will be maximally useful if as much information as possible is captured about and during the georeferencing process in the following fields defined in the Darwin Core standard [7]. For additional community discussion and recommendations, see the Darwin Core Project wiki [8].

Guidelines

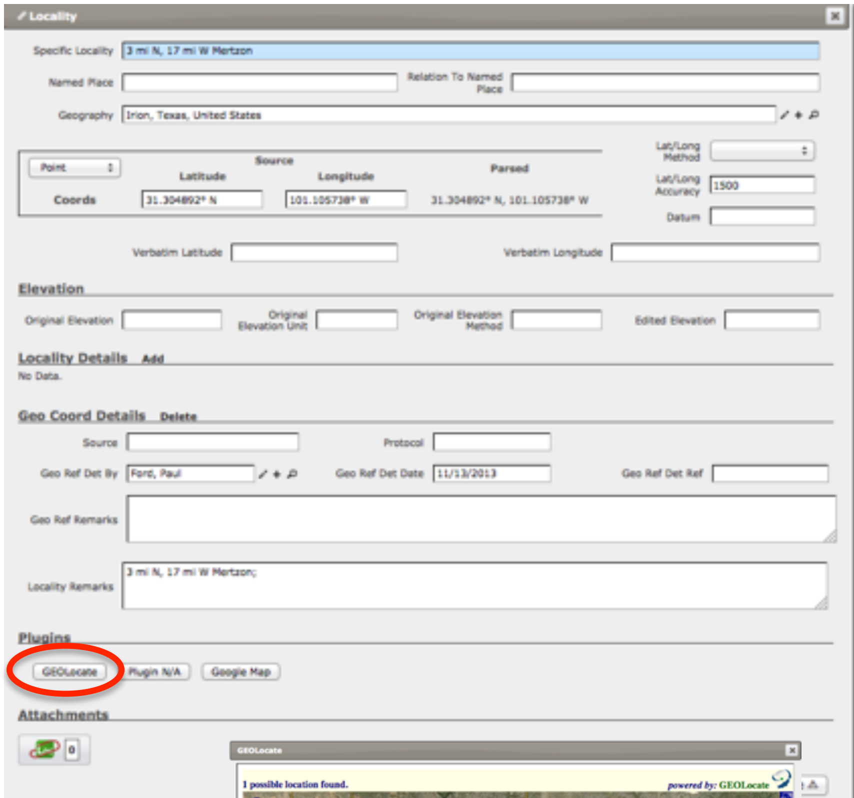
iDigBio
Integrated Digitized Biocollections

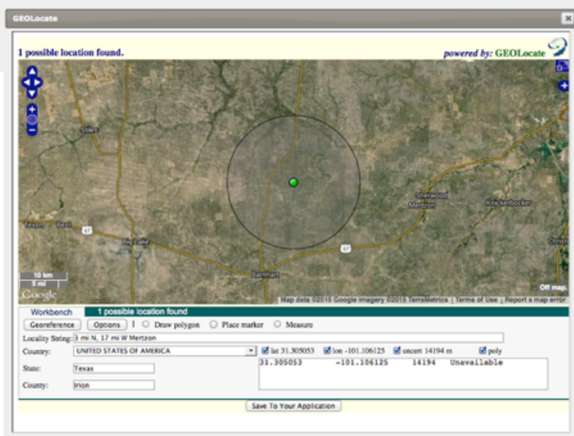
- Train the Trainers
- Georeferencing Working Group

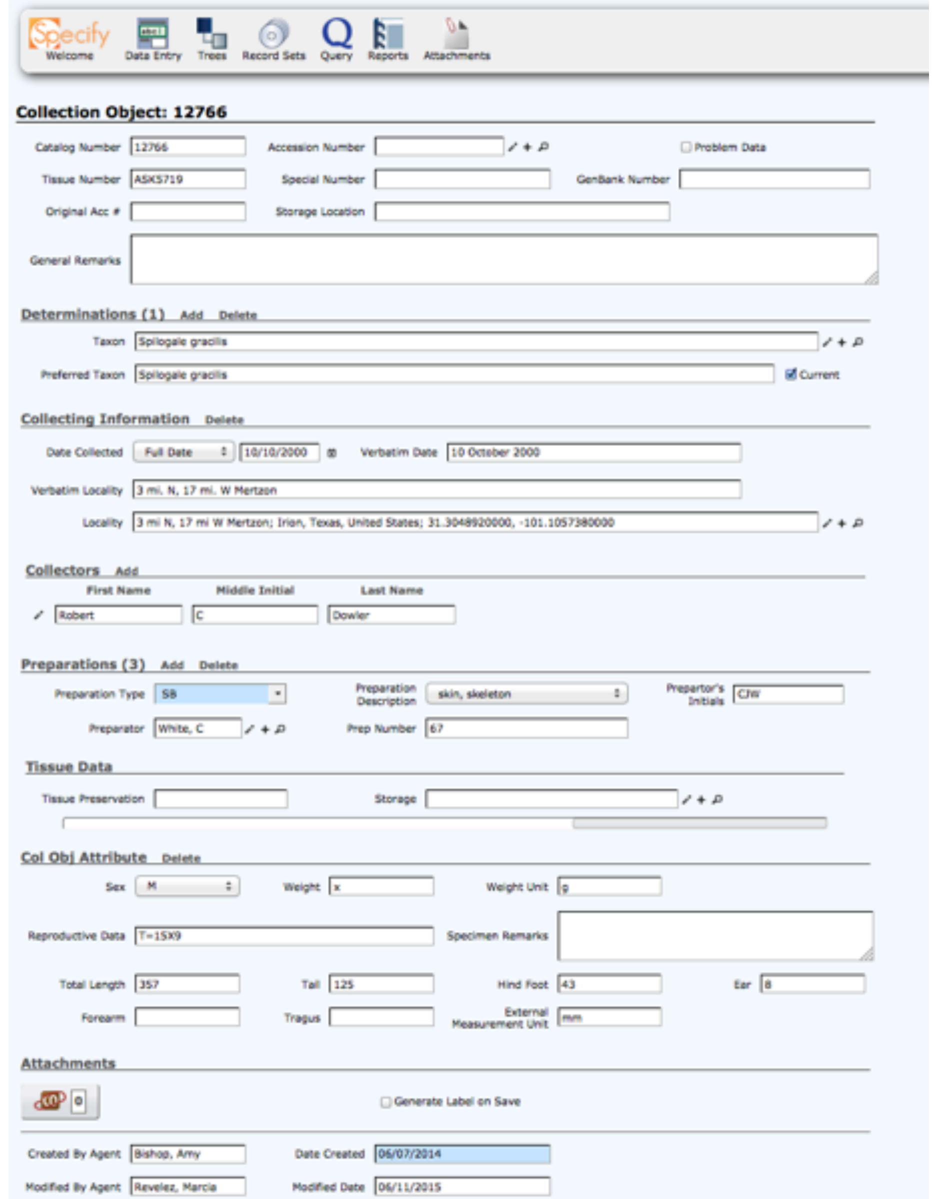
Database

Specify 6 → Sp Cloud → Specify 7

Database


Locality
Specific Locality 3 mi N, 17 mi W Mertzon
Named Place
Relation To Named Place
Geography Irion, Texas, United States
Source
Latitude
Longitude
Parsed
Coords 31.304892° N 101.105738° W 31.304892° N, 101.105738° W
Lat/Long Method
Lat/Long Accuracy 1500
Datum
Verbatim Latitude
Verbatim Longitude
Elevation
Original Elevation
Original Elevation Unit
Original Elevation Method
Edited Elevation
Locality Details Add
No Data.
Geo Coord Details Delete
Source
Protocol
Geo Ref Det By Ford, Paul
Geo Ref Det Date 11/13/2013
Geo Ref Det Ref
Geo Ref Remarks
Locality Remarks 3 mi N, 17 mi W Mertzon;
Plugins
GEOLocate
Plugin N/A
Google Map
Attachments

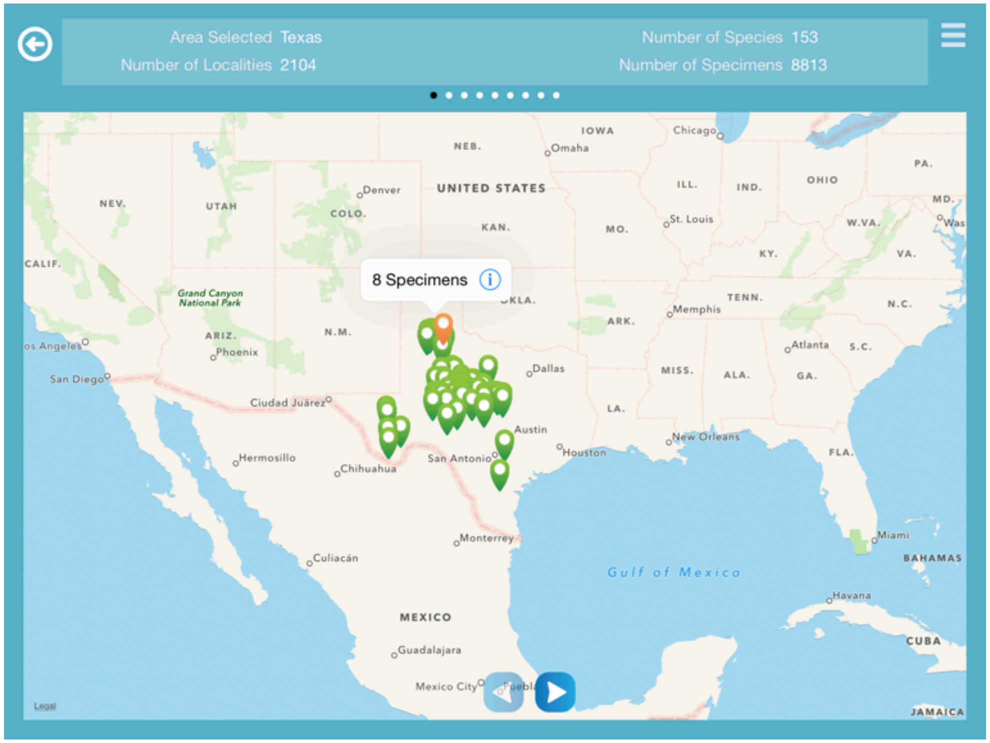

GEOLocate
1 possible location found.
powered by: GEOLocate
Save To Your Application

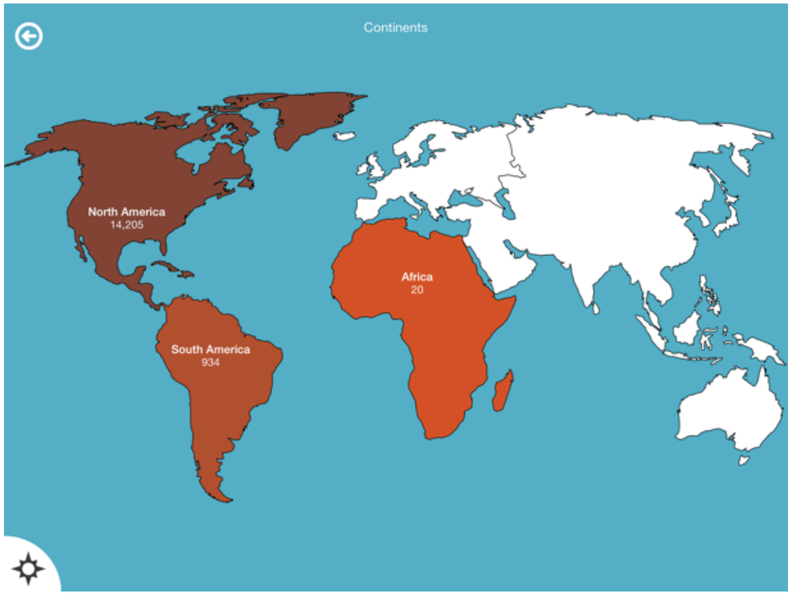

Specify
Welcome
Data Entry
Trees
Record Sets
Query
Reports
Attachments
Collection Object: 12766
Catalog Number 12766
Accession Number
Problem Data
Tissue Number ASK5719
Special Number
GenBank Number
Original Acc #
Storage Location
General Remarks
Determinations (1) Add Delete
Taxon Spilogale gracilis
Preferred Taxon Spilogale gracilis
Current
Collecting Information Delete
Date Collected Full Date 10/10/2000
Verbatim Date 10 October 2000
Verbatim Locality 3 mi. N, 17 mi. W Mertzon
Locality 3 mi N, 17 mi W Mertzon; Irion, Texas, United States; 31.3048920000, -101.1057380000
Collectors Add
First Name
Middle Initial
Last Name
Robert
C
Dowler
Preparations (3) Add Delete
Preparation Type SB
Preparation Description skin, skeleton
Preparator's Initials CJW
Preparator White, C
Prep Number 67
Tissue Data
Tissue Preservation
Storage
Col Obj Attribute Delete
Sex M
Weight x
Weight Unit g
Reproductive Data T=15X9
Specimen Remarks
Total Length 357
Tail 125
Hind Foot 43
Ear 8
Forearm
Tragus
External Measurement Unit mm
Attachments
Generate Label on Save
Created By Agent Bishop, Amy
Date Created 06/07/2014
Modified By Agent Revelez, Marcia
Modified Date 06/11/2015

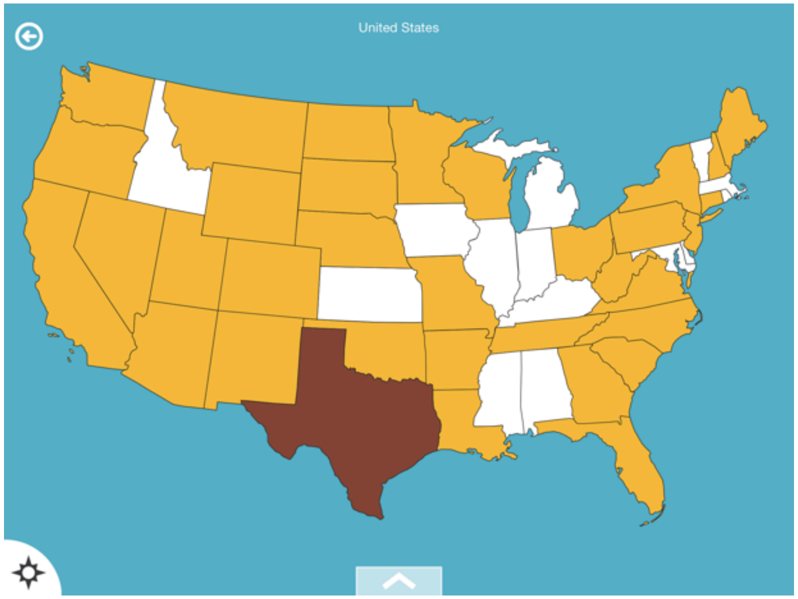
Database Tool

Database Tool

GEOGRAPHY





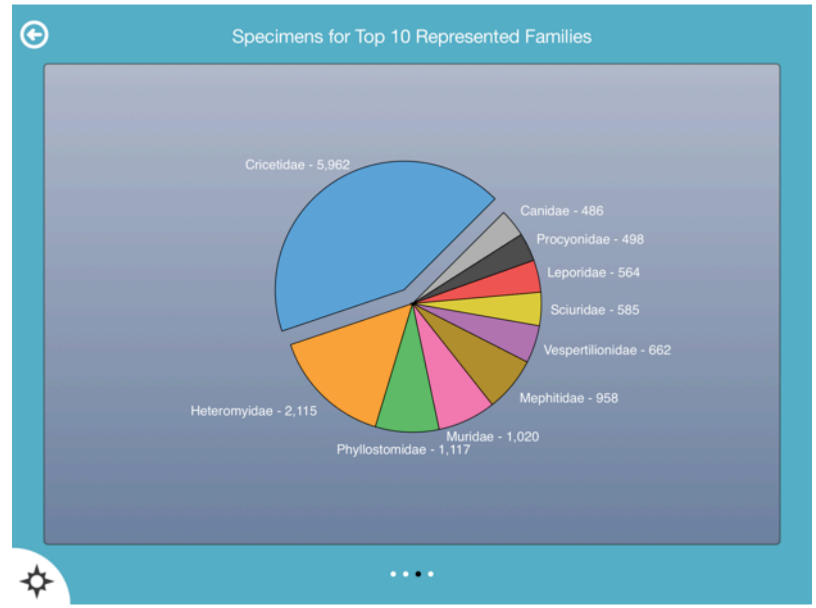


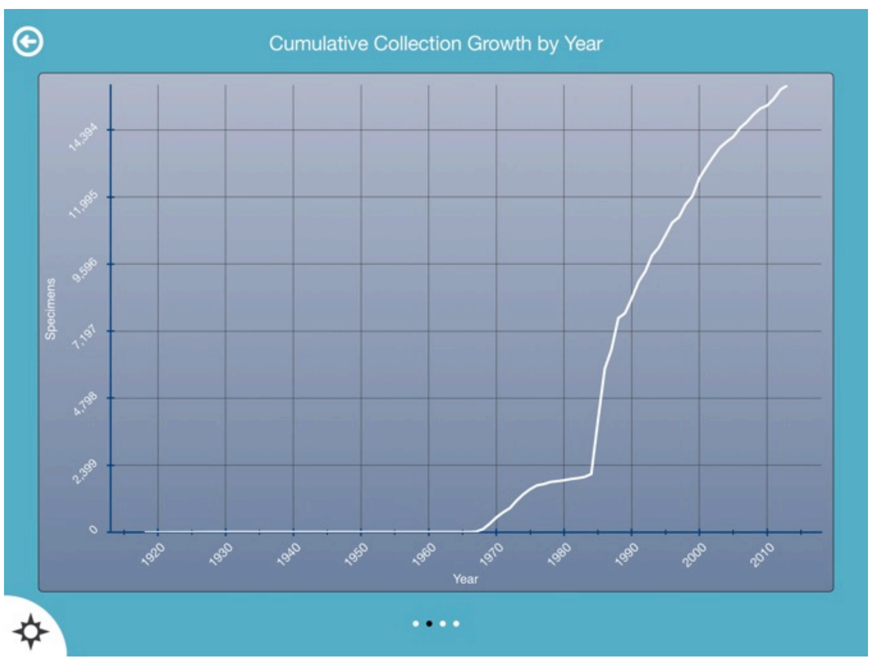
Database Tool

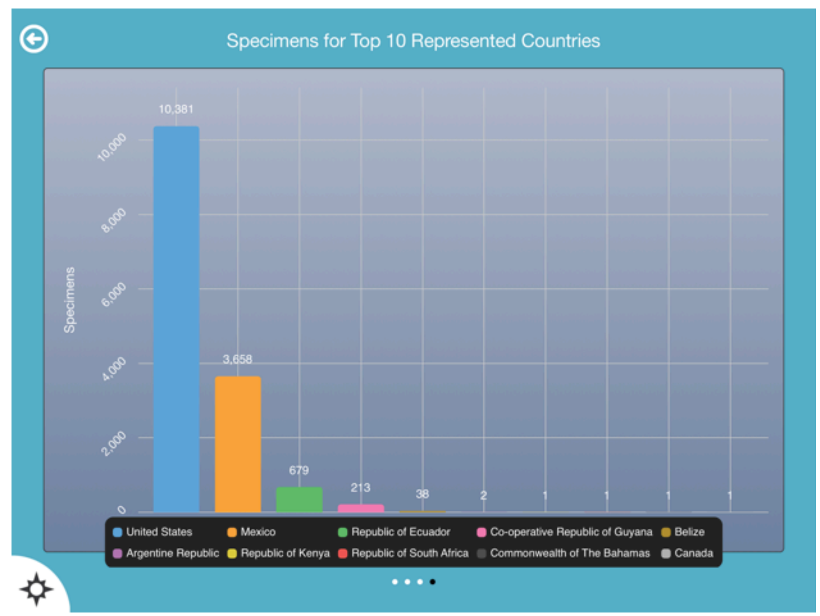
SpInsight

SpInsight

STATISTICS





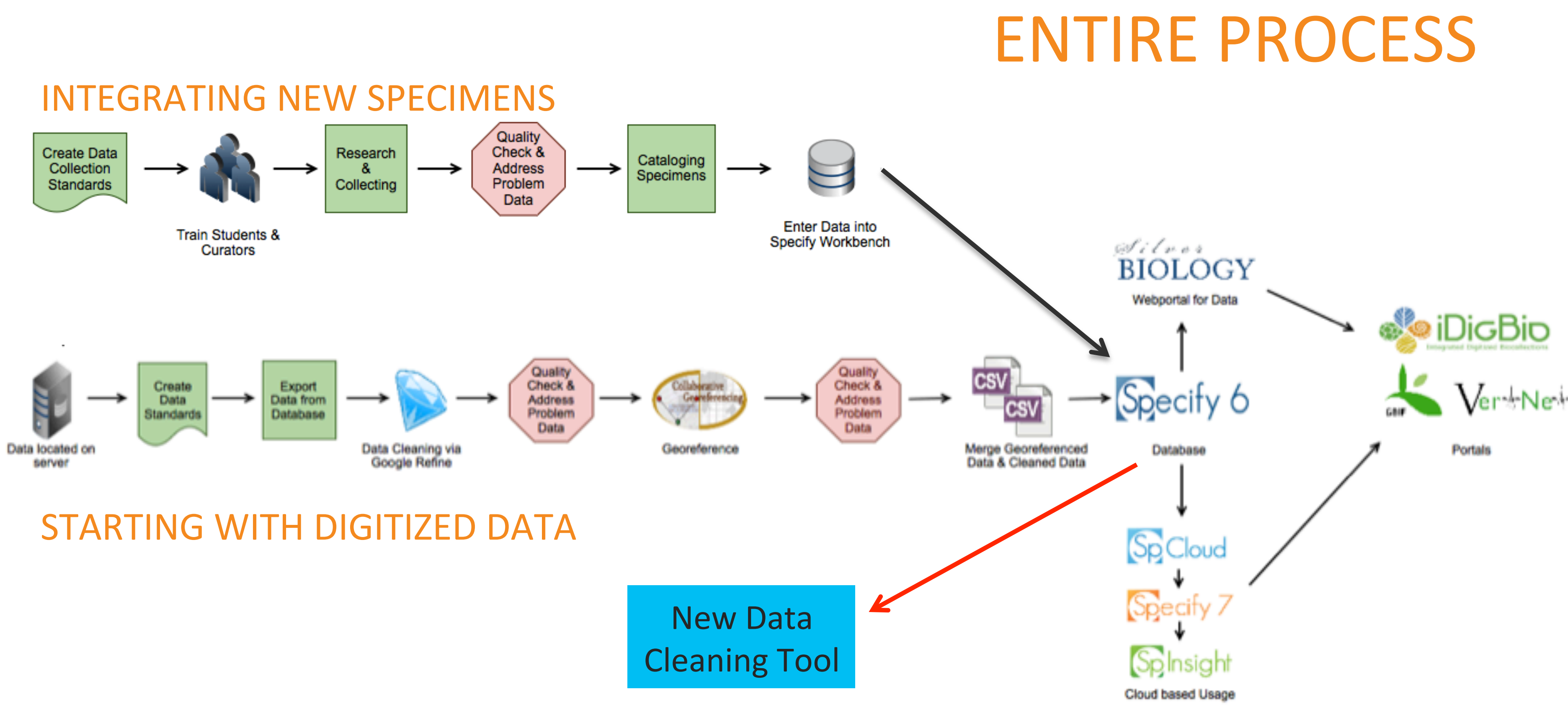


Inventories

Pocket

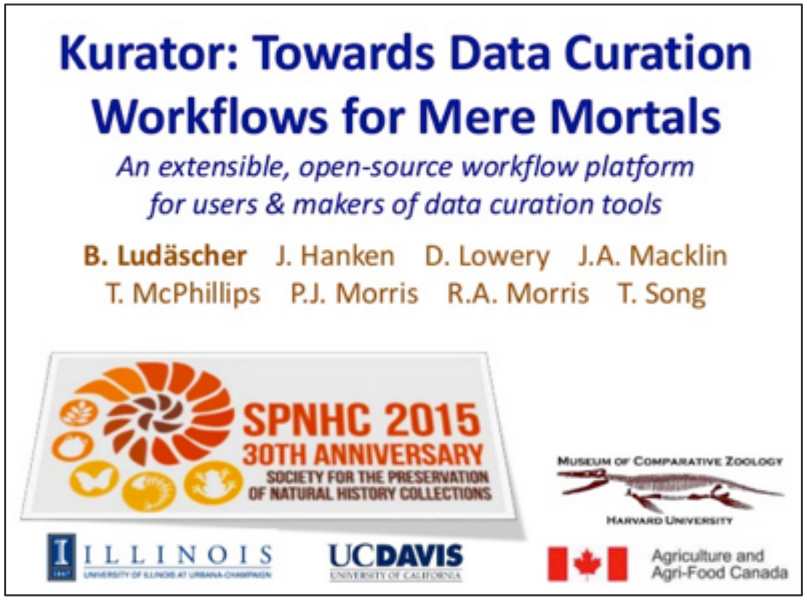
- Utilize Database from smartphone
- Be in the cases, not anchored to computer
- Create Inventories, Loans, Reports, etc.

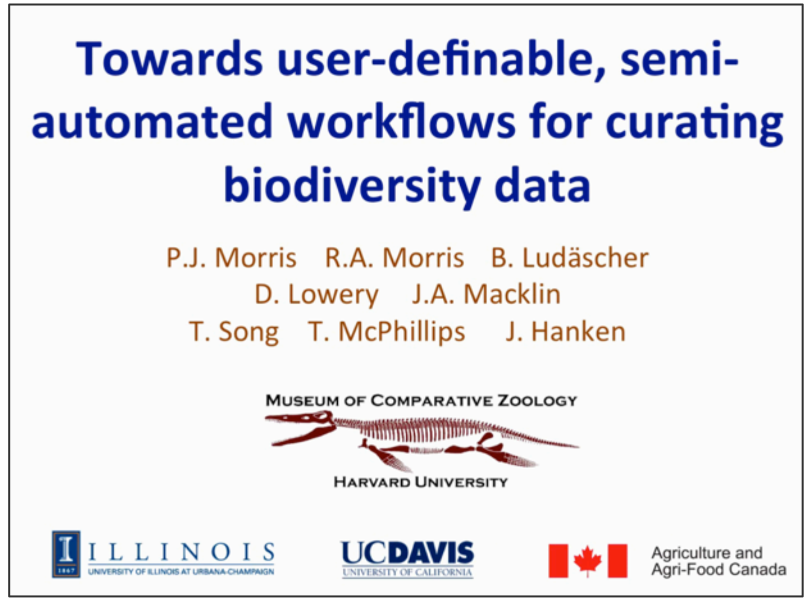
Digitization Workflow for Mammalogy Collection at ASNHC



Data Cleaning Tools

Kurator





Webinar hosted by

iDigBio
Integrated Digitized Biocollections

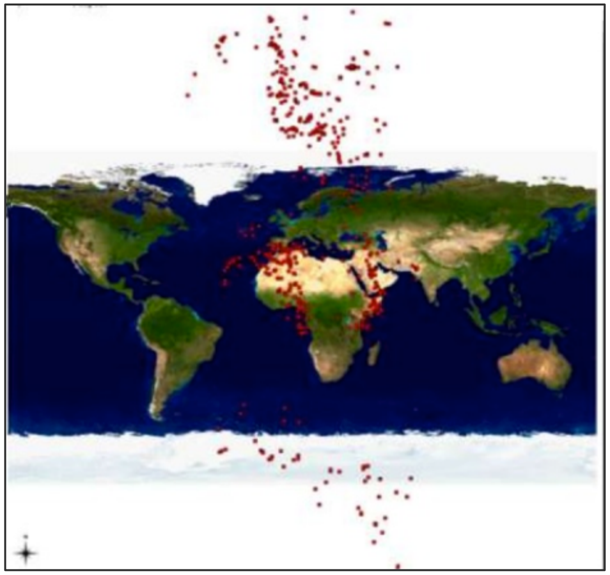
Kurator: Kepler Data Curation Package

- Software in development – mechanism for quality control of biodiversity data

- Problems with Data and Metadata Quality
  - Lat/Long transposition and other georeferencing issues
  - Scientific Names (spelling errors, out-of-date nomenclature, etc.)
  - Data entry/creation can be full of errors associated with data conversions and transformations, schema mappings, or human error

- Problems Kurator is Attempting to Solve
  - Detect and flag data quality issues
  - Provide opportunity for repair
  - Employ workflow (semi-)automation

GOALS: 1) get data quality control tools out into the hands of people who are vested in the quality of their own data, and 2) get feedback from them to improve the capabilities of our tools

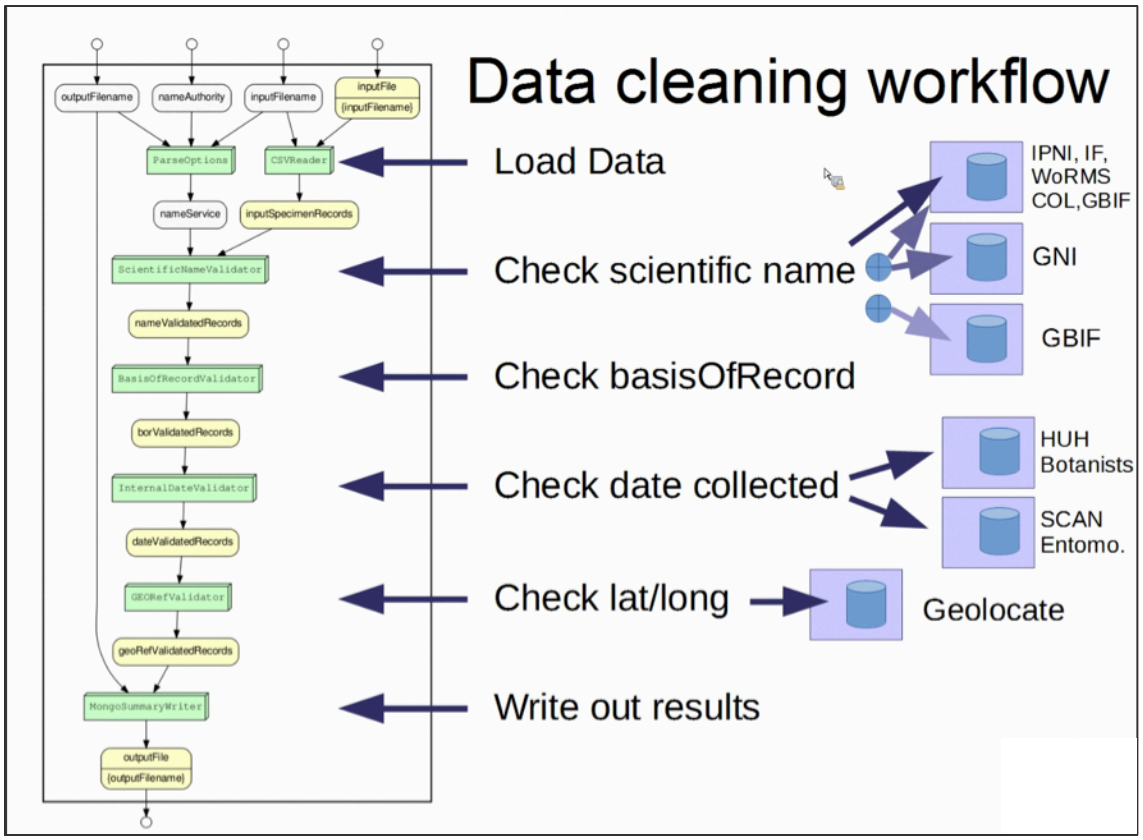


Data Cleaning Tools

Kurator

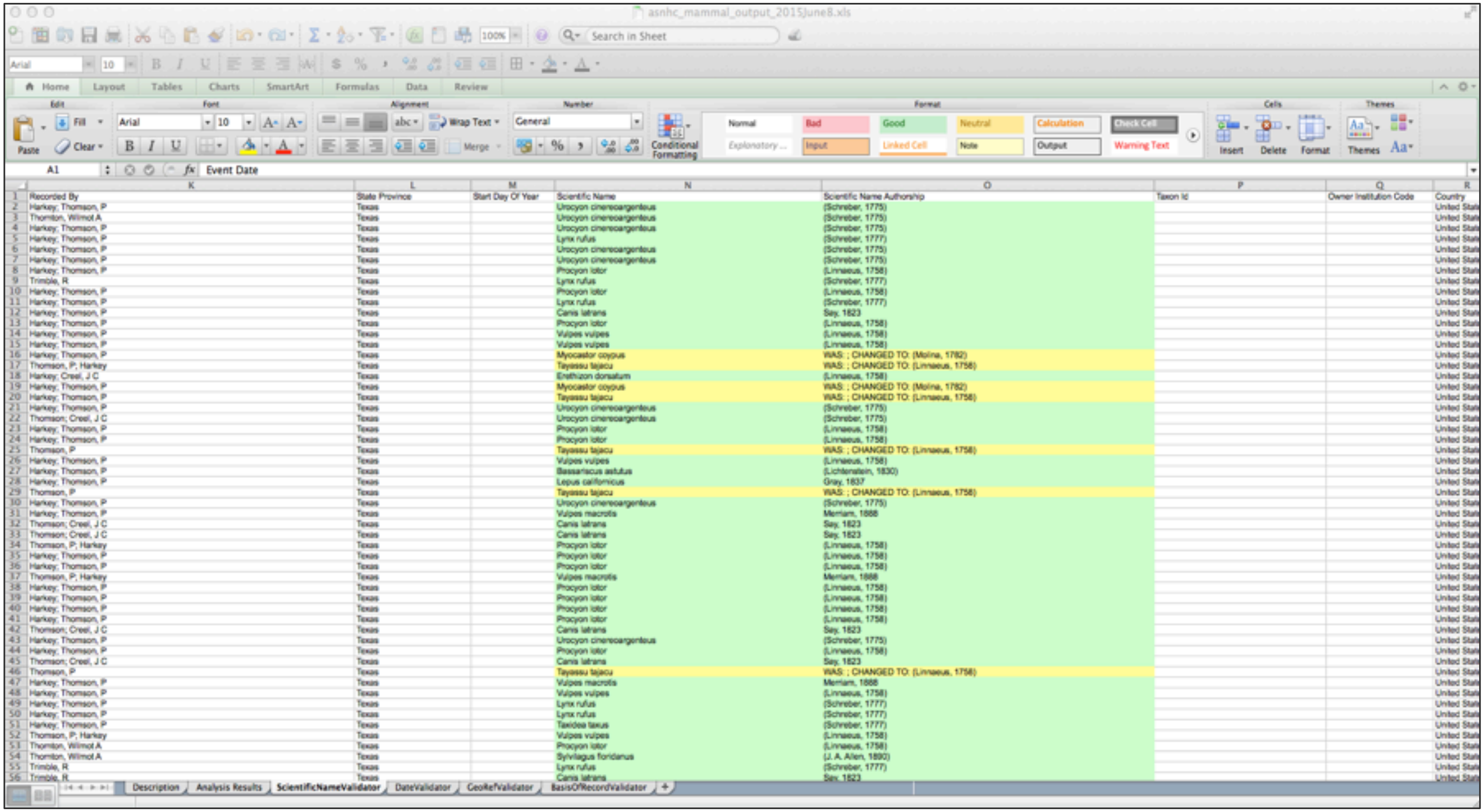
3 Key Questions to Test Quality of Data

- What? – scientific name validation
- When? – collecting event date validation checks a collector's life span vs. date collected
- Where? – georeference validator



Data Cleaning Tools

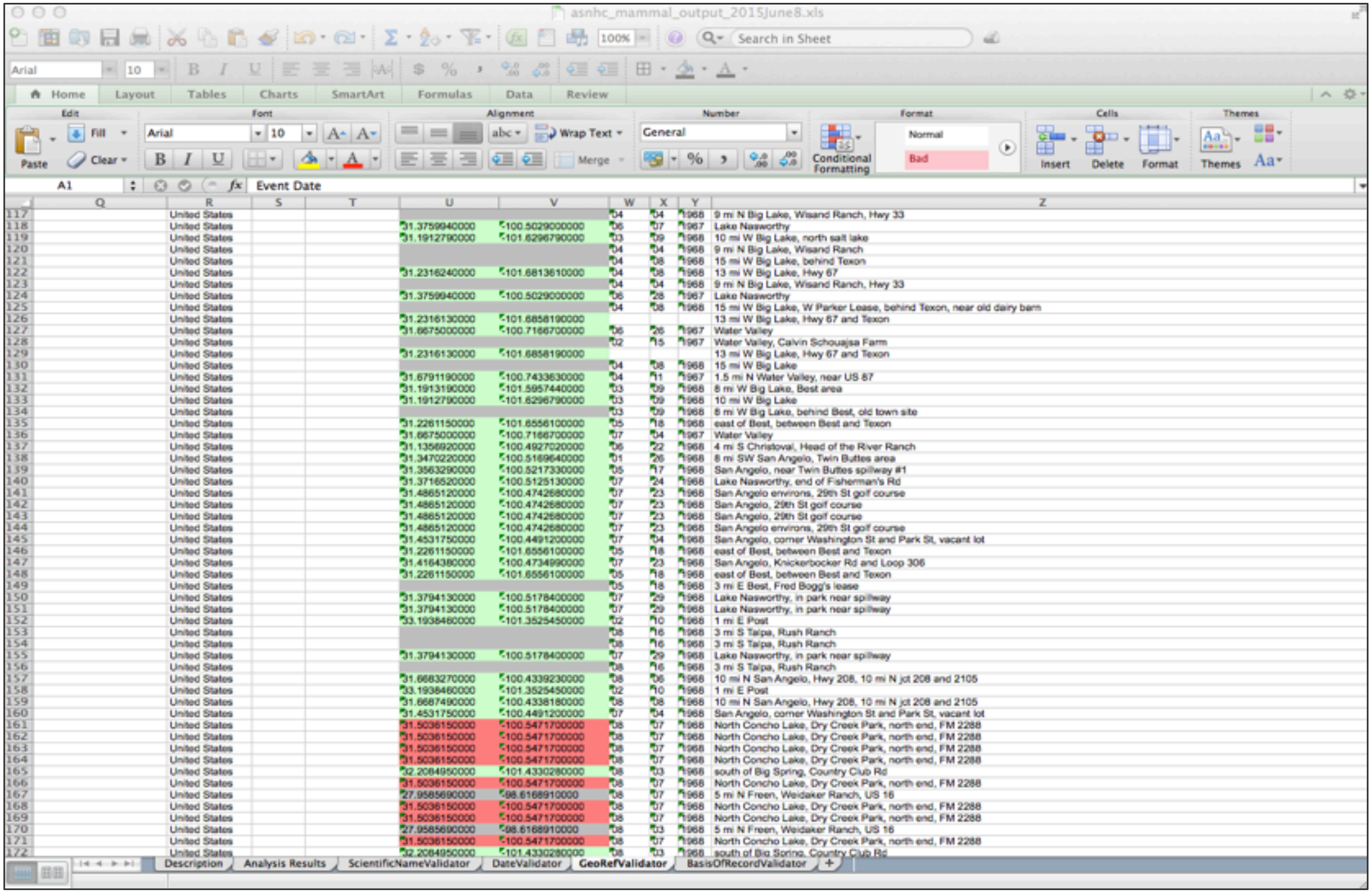
Kurator



Taxonomic Validation

Data Cleaning Tools

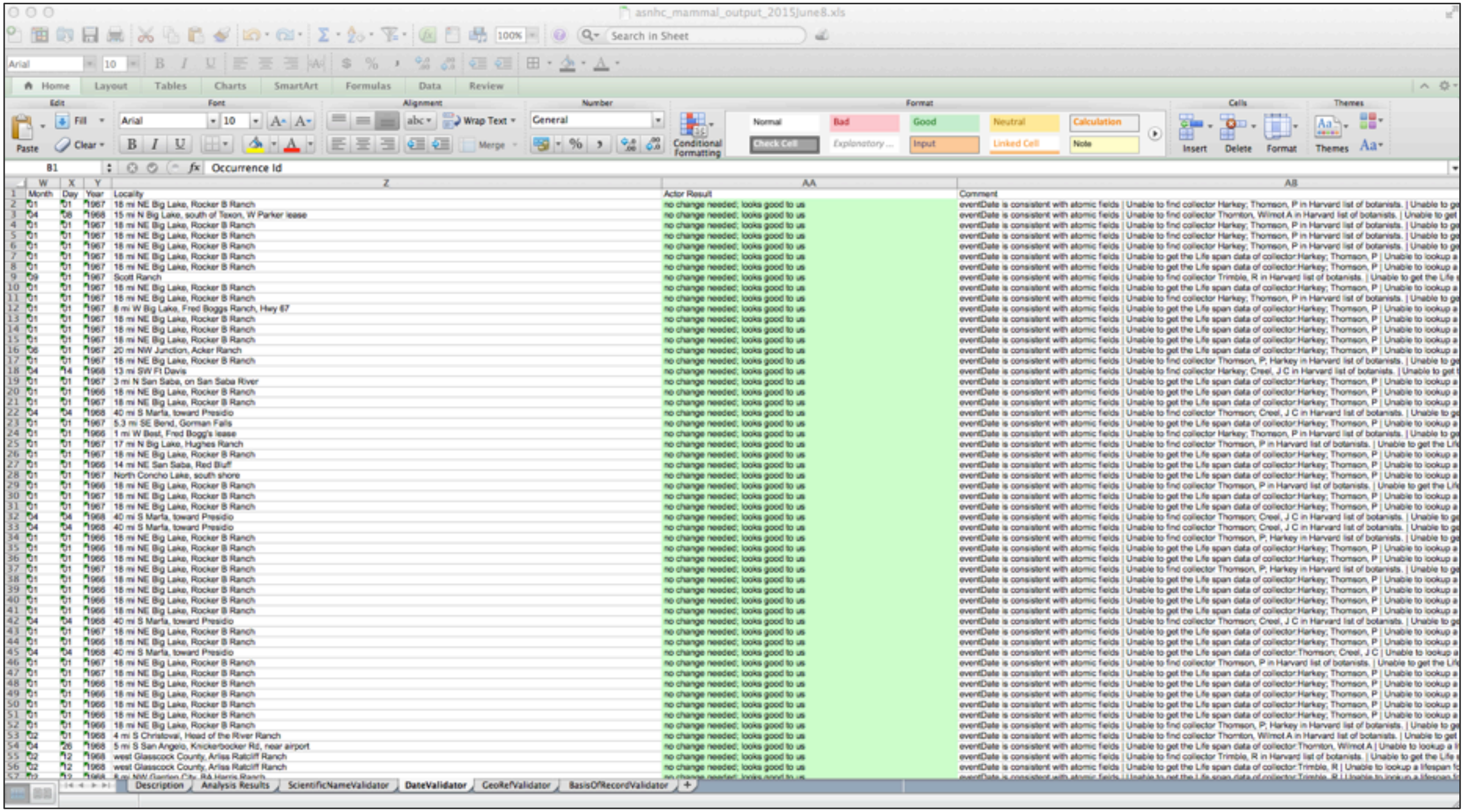
Kurator



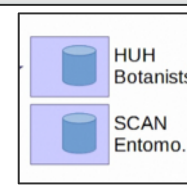
Georeference Validation

Data Cleaning Tools

Kurator



Collecting Event Date Validation



Resources

SPNHC
ADVANCING COLLECTIONS CARE

iDigBio
Integrated Digitized Biocollections

SC net
Small Collections Network
Serving, Supporting, Connecting Small Natural History Collections

Specify
Software Project

GEOLocate

Websites, Listservs, Meetings, Symposia, Workshops, YouTube videos, documents, powerpoint presentations, recordings, and so much more!

Resources

Students

Get Involved

EDUCATE AT EVERY OPPORTUNITY

Training the next generation of stewards

Benefits of Student Involvement

Become advocates for natural history collections

GIVE

- ✓ Critical to the successful execution of the collections goals
- ✓ Carry out the high quality of research conducted by faculty and staff
- ✓ Contribute to the university's mission

TAKE

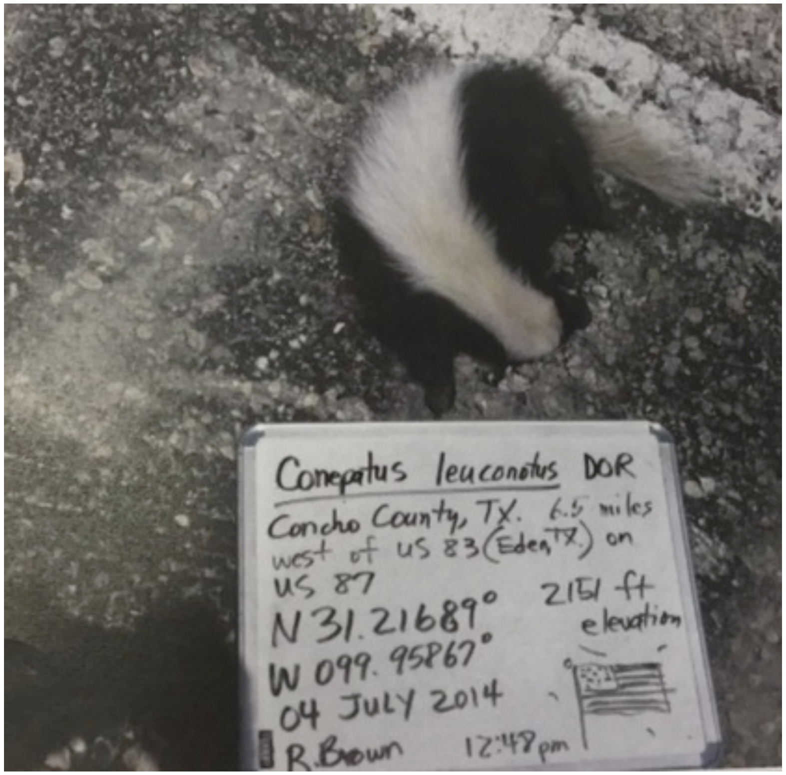
- ✓ Gain a better understanding and appreciation for best practices, natural history museums, and the greater scientific community
- ✓ Gain an appreciation for the area flora & fauna of the Concho Valley and beyond – understanding in biodiversity, conservation, climate change, etc.
- ✓ Obtain skills in research, work ethic, time management, accuracy, responsibility, and professional courtesy
- ✓ Inspiration in future career options

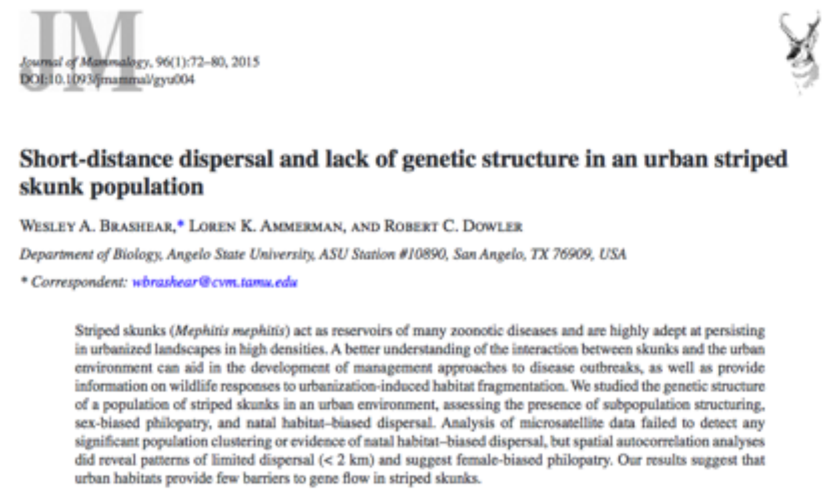
I CHOSE ANGELO STATE

But Wait

There's More…

Advance Digitization

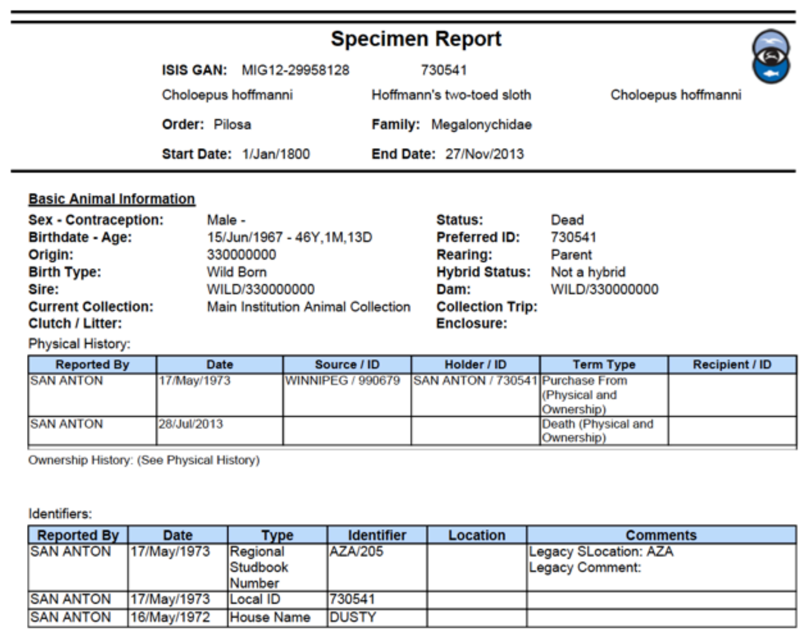



Journal of Mammalogy, 96(1):72–80, 2015
DOI:10.1093/jmammal/gyu004

Short-distance dispersal and lack of genetic structure in an urban striped skunk population

WESLEY A. BRASHEAR,* LOREN K. AMMERMAN, AND ROBERT C. DOWLER
Department of Biology, Angelo State University, ASU Station #10890, San Angelo, TX 76909, USA
* *Correspondent: wbrashear@cvm.tamu.edu*

Striped skunks (*Mephitis mephitis*) act as reservoirs of many zoonotic diseases and are highly adept at persisting in urbanized landscapes in high densities. A better understanding of the interaction between skunks and the urban environment can aid in the development of management approaches to disease outbreaks, as well as provide information on wildlife responses to urbanization-induced habitat fragmentation. We studied the genetic structure of a population of striped skunks in an urban environment, assessing the presence of subpopulation structuring, sex-biased philopatry, and natal habitat–biased dispersal. Analysis of microsatellite data failed to detect any significant population clustering or evidence of natal habitat–biased dispersal, but spatial autocorrelation analyses did reveal patterns of limited dispersal (< 2 km) and suggest female-biased philopatry. Our results suggest that urban habitats provide few barriers to gene flow in striped skunks.



Specimen Report

ISIS GAN: MIG12-29958128 730541

Choloepus hoffmanni Hoffmann's two-toed sloth Choloepus hoffmanni

Order: Pilosa **Family:** Megalonychidae

Start Date: 1/Jan/1800 **End Date:** 27/Nov/2013

Basic Animal Information

Sex - Contraception:	Male -	**Status:**	Dead
Birthdate - Age:	15/Jun/1967 - 46Y,1M,13D	**Preferred ID:**	730541
Origin:	330000000	**Rearing:**	Parent
Birth Type:	Wild Born	**Hybrid Status:**	Not a hybrid
Sire:	WILD/330000000	**Dam:**	WILD/330000000
Current Collection:	Main Institution Animal Collection	**Collection Trip:**	
Clutch / Litter:		**Enclosure:**	

Physical History:

Reported By	Date	Source / ID	Holder / ID	Term Type	Recipient / ID
SAN ANTON	17/May/1973	WINNIPEG / 990679	SAN ANTON / 730541	Purchase From (Physical and Ownership)	
SAN ANTON	28/Jul/2013			Death (Physical and Ownership)	

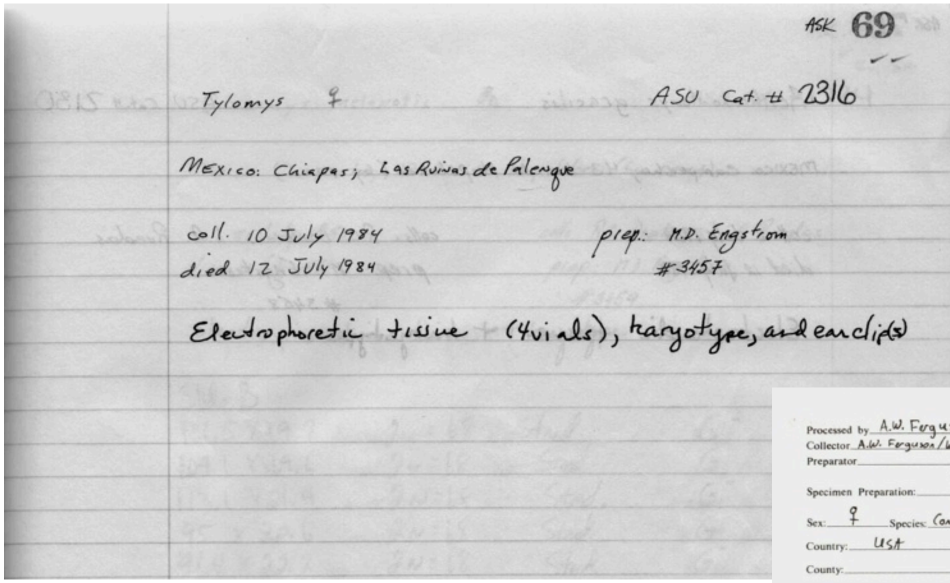
Ownership History: (See Physical History)

Identifiers:

Reported By	Date	Type	Identifier	Location	Comments
SAN ANTON	17/May/1973	Regional Studbook Number	AZA/205		Legacy SLocation: AZA Legacy Comment:
SAN ANTON	17/May/1973	Local ID	730541		
SAN ANTON	16/May/1972	House Name	DUSTY		

- Genbank
- Zoo Record Sheets
- Texas Department of Health Services
- Literature Citations
- Images
- Video

Advance Digitization


ASK 69

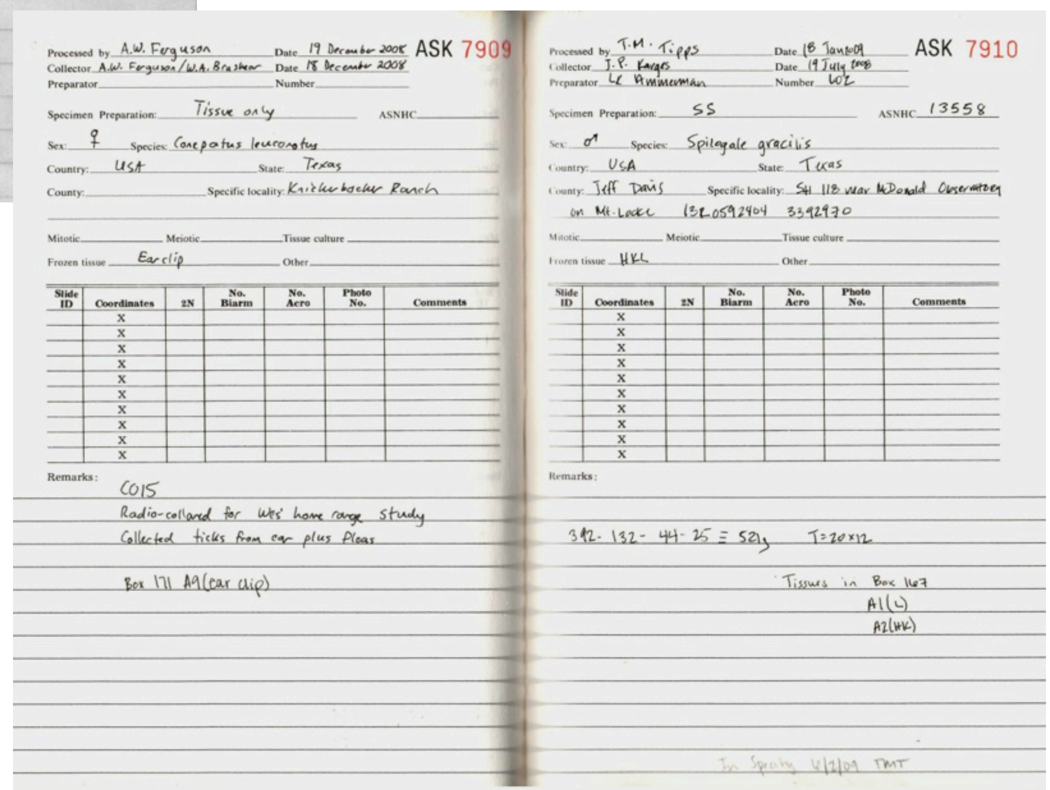
Tylomys ♀ ASU Cat. # 2316

Mexico: Chiapas; Las Ruinas de Palenque

coll. 10 July 1984 prep.: M.D. Engstrom

died 12 July 1984 #3457

Electrophoretic tissue (4 vials), karyotype, and earclip(s)



- Tissue Data pages (ASK #)
- Journals
- Catalogs
- Specimen (with tag data)

Acknowledgments

National Science Foundation's Collections in Support of Biological Research - *Modernization and Digitization of the Angelo State Natural History Collections*" (1203363)

Brian J. May, President, Angelo State University
Russell Wilke, Chair, Dept. of Biology, Angelo State University
Robert Dowler, Loren Ammerman, Bonnie Amos, Terry Maxwell, Ben Skipper, and Mike Dixon, Curators, ASNHC

Deb Paul, Gil Nelson, Kevin Love, Joanna McCaffrey

James Beach, Teresa Miller, Andy Bentley

Paul Morris, Bertram Ludascher, James Macklin, D. Lowery, R. A. Morris, T. Song, T. McPhilips, and J. Hanken

Cody Thompson, American Society of Mammalogists