

# What to Transcribe & Database? Fully Populated vs Skeletal Records



Carla Cicero  
Carol Spencer  
Chris Conroy



iDigBio Workshop  
Berkeley, California  
4-6 April 2016

# Why Database Collections?

## Three Primary Motivations:

- Assist with Collection Management
- Facilitate Specimen-based Research
- Promote Institutional Support & Funding

# Legacy Data: Primary Sources



\* GPS coordinates on back,  
data entered into Syb. 1 Aug 6, 2001 M. Allie

DEPARTMENT CATALOGUE—BIRDS UNIVERSITY OF CALIFORNIA  
MUSEUM OF VERTEBRATE ZOOLOGY

ACC'N No. 13483 DATE OF ENTRY 11 September 2000 ENTERED BY J.LOR 1/9

GENERAL LOCALITY Oregon: Harney Co., Umatilla Co., Morrow Co.

DEPT. No.	ORIG. No.	NAME	DATE	COLLECTOR	EXACT LOCALITY
+ tissue 179481	2034	Himantopus mexicanus	Ad. ♂ 18 June 2000	C. Cicero	① Hwy. 205, 4600ft., 2 mi E + 7mi S Burns
+ tissue 179482	2039	Picoides arcticus	Ad. ♀ 22 June 2000	"	② USFS Rd 5320, 4800ft. 2 1/2 mi S + 10 mi W Ukiah
+ tissue 179483	2046	Perisoreus canadensis	Ad. sk. ♀ 25 June 2000	"	③ Little Potamus Cr., 4700ft., 3 1/2 mi S + 17 mi W Ukiah
+ tissue 179484	2045	" "	Ad. sk. ♂ 21 June 2000	"	④ USFS Rd 2105, 4700ft., 4 mi S + 15 mi W Ukiah
+ tissue 179485	2036	" "	Juv. ♂ 20 June 2000	"	⑤ Fremont Cr., 6000ft., 1 mi W Ukiah
+ tissue 179486	2040	" "	Ad. sk. ♀ 22 June 2000	"	⑥ USFS Rd 5320, 4800ft. 2 1/2 mi S + 10 mi W Ukiah
+ tissue 179487	2041	" "	Ad. sk. ♂ "	"	"
+ tissue 179488	2042	" "	Ad. sk. ♀ "	"	"
+ tissue 179489	2043	Cyanocitta stelleri	Juv. ♀ 23 June 2000	"	⑦ Potamus Cr., 4500ft., 3 mi S + 14 mi W Ukiah
+ tissue 179490	7076	" "	Ad. sk. ♂ 25 June 2000	A. K. Johnson	⑧ Little Potamus Cr., 4700ft., 3 1/2 mi S + 17 mi W Ukiah

NAME	SEX	LOCALITY	DATE COLLECTED	EMBEDDED FROM	COLLECTED BY	WEEK ENTERED	NO. OF EGGS	REMARKS
scitophorus e. californicus		Winst. Shofield, Calif. Co.	May 1, 1932		Dr. C.E. Burt	132		
knideactylus		Arroyo Viejo, Cal.	1931		Dr. C.E. Burt	1		
typhlus celleris		4 mi. NW Shofield, Calif. Co.	May 17, 1932		Dr. C.E. Burt		1	
lyda serpentina		Island Park	May 4, 1932		"		1	
capra ornata	♂	3 mi. E	May 26, 1932		"		1	
	♀	3 mi. NE Woodlark	May 6, 1932		"		1	
mosca fasciatus		Arroyo Viejo, Calif.	June 10, 1932		Dr. John F. Doherty	June 30	1	
ple. shabete divisa		Hill Timbers	June 25, 1932		J. E. Benedict, Jr.		1	
capra bonni		Big Pine Key, Florida	June 30, 1932		Francis K. Bond		1	Shed Big woods near shed of island
							1	Shed
lyda osceola		St. Ignace, Mich., 13 mi. from	July 7, 1932				1	Shed & shell

June 19

- 6887 *Pipodonyx o. columbianus* ♂ 35.9g. 220x15x3
- 6888 *Peromyscus m. sonoriensis* ♂ 19.2g. 163x75x1
- 6889 *Micropodops* ♀ 12.3g. 158x80x25
- 6890 " ♀ 12.4g. 160x84x24
- 6891 *Sage Thrasher* ♂ testes large. 37.2g.  
full song near Mockingbird's nest!
- 6892 *Mockingbird* ♀ 47.9g. Taken with  
4 hatching eggs, as follows:
- 6893 *Mockingbird* 4 young<sup>in formalin</sup>, of which one had  
in the nest naturally, the other three taken  
shells, by cutting around larger end. Shell
- 6894 *Scaphiopus* (?) Picked up in dry  
where we camped, about 6 p.m.  
visible water short of 2 miles.

June 22

On "picnic" to lower Viola Road,  
of Mineral. Saw two *Blue-tailed Skink*  
on ground on floor of open woods among  
yellow pines and incense cedars; in-  
lizard found successful retreat in  
logs. A family of *Hermit Thrush*  
conveyed scattering by at least  
the darkly shaded forest floor;  
males (2) in full song in  
sight thru the heat of the

In evening, fishing do  
found by Mrs. J.: *June*  
eggs, ♀ flushed from it;

Ned K. Johnson  
1996

Catalogue

2.

2 mi. E + 1/2 mi. N Cassel, 3300 ft., Shasta Co., Calif.

May 22, 1996 (cont.)

- 6520 ♂ *Pipilo maculatus* Testis 13x8 mm. Ad. Sk. No fat. Large cloacal protuberance. gms. Wt. 40.2
- Tissue saved. Song not recorded on tape. Sang only trills. Irides orange.

May 23 1996

- 6521 ♂ *Pipilo maculatus* Testis 11x6 mm. Ad. Sk. No fat. Large cloacal protuberance. gms. Wt. 39.0

Tissue saved. "Spotted Towhee #1" on tape - 40 cues. Irides deep red.

- 6522 ♂ *Pipilo maculatus* Testis 12x7 mm. Ad. Sk. No fat. Large cloacal protuberance. Wt. 40.8

Tissue saved. "Spotted Towhee #2" on tape - 42 cues. Irides orange.

- Pipilo maculatus* Testis 11x6 mm. Ad. Sk. No fat. Large cloacal protuberance. Wt. 40.7

+ 1 Towhee #3 on tape - 55 cues. Irides orange.

- Pipilo maculatus* Testis 12x7 mm. Ad. Sk. No fat. Large cloacal protuberance. Wt. 43.2

Ad. Sk. No fat. Irides orange-red.

- Pipilo maculatus* Testis 12x7 mm. Ad. Sk. No fat. Large cloacal protuberance. Wt. 40.0

YR:	2012	NAME:	Carla Cicero
#	3909	Locality:	Spirit Lake, Utah
<input checked="" type="checkbox"/> Skin		GPS Error:	19 ft
<input type="checkbox"/> Skel.		Date collected:	10 Sep 2012
<input type="checkbox"/> ETOH		Species:	<i>Perisoreus canadensis</i>
<input type="checkbox"/> Other		Tiss:	<input checked="" type="checkbox"/> Froz <input type="checkbox"/> Other
<input type="checkbox"/> Audio		Repro.:	<input checked="" type="checkbox"/> M <input type="checkbox"/> F <input type="checkbox"/> sex?
<input type="checkbox"/> Photo		Molt:	Testis 2x1 mm
FN		Colors:	Slight body molt.
<input checked="" type="checkbox"/> Shot		Habitat:	Spruce forest.
<input type="checkbox"/> Net		Notes:	Collected with cc # 3910-3911.
<input type="checkbox"/> Salvage		Stomach:	
#	3910	Locality:	Same as cc # 3910-3911.
<input checked="" type="checkbox"/> Skin		GPS Error:	
<input type="checkbox"/> Skel.		Date collected:	10 Sep 2012
<input type="checkbox"/> ETOH		Species:	<i>Perisoreus canadensis</i>
<input type="checkbox"/> Other		Tiss:	<input checked="" type="checkbox"/> Froz <input type="checkbox"/> Other
<input type="checkbox"/> Audio		Repro.:	<input type="checkbox"/> M <input checked="" type="checkbox"/> F <input type="checkbox"/> sex?
<input type="checkbox"/> Photo		Molt:	Ovary granular
FN		Colors:	Trace body
<input checked="" type="checkbox"/> Shot		Habitat:	
<input type="checkbox"/> Net		Notes:	

SPECIMEN CATALOG

PAGE #: 16

Date prepared: 10 Sep 2012

Datum: WGS 84

Barcode #: 199308

Wt: 82.5 gm

Fat: No fat

41.2

Wt. 39.7

Wt. 40.1

red.

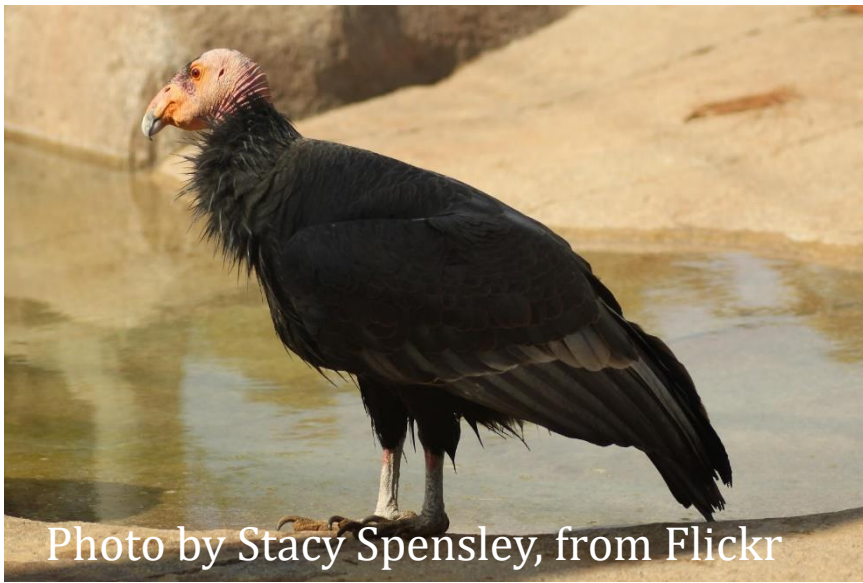


Photo by Stacy Spensley, from Flickr

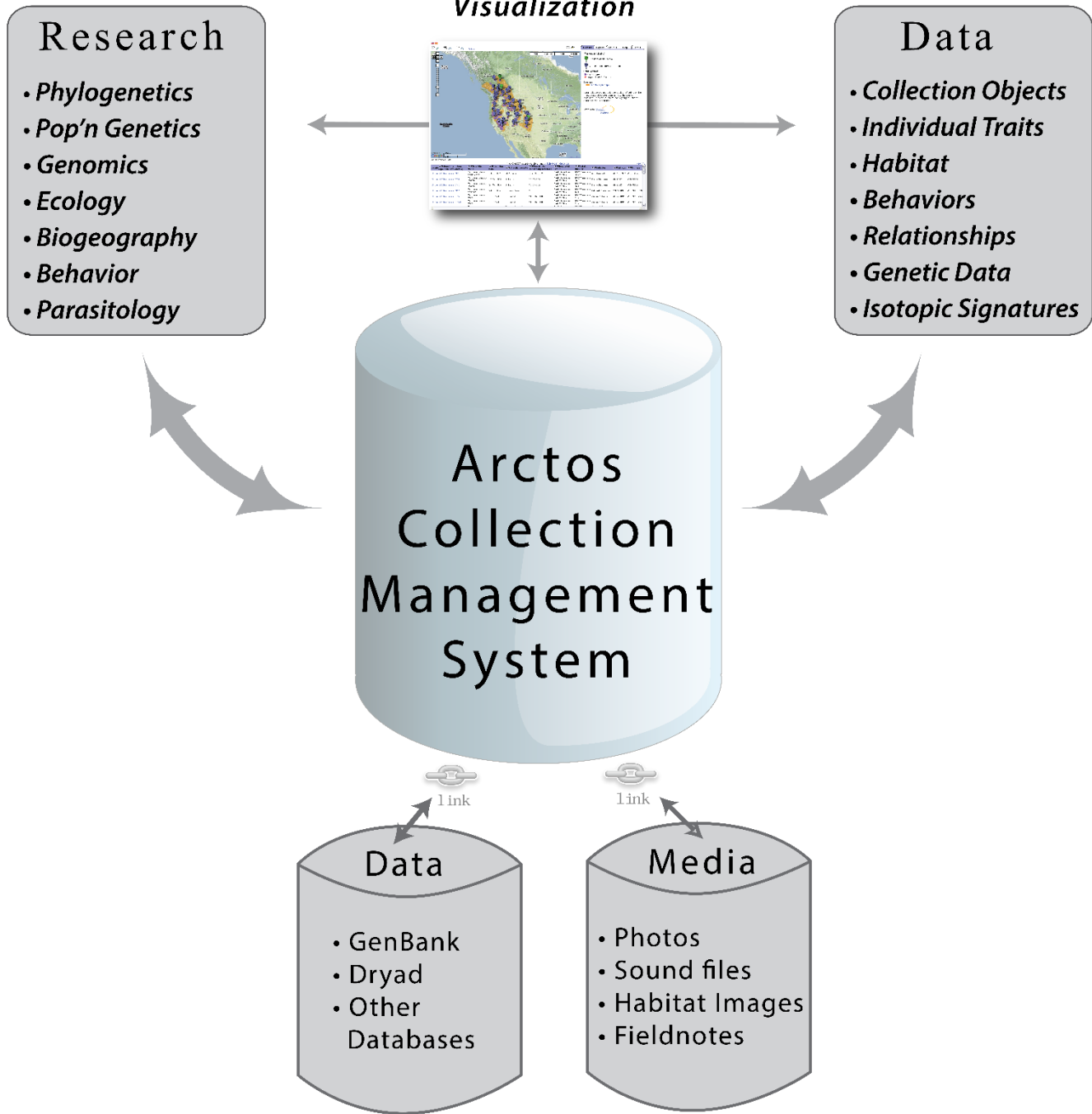


UNIVERSITY OF CALIFORNIA  
MUSEUM OF VERTEBRATE ZOOLOGY  
No. 13670 Orig. No.  
*Gymnogyps californianus*  
Eaton Canyon, 5 1/2 mi. NE Pasadena  
Los Angeles Co., California  
10 Feb 1907 J. Grinnell

MVZ 13670 Egg of  
California Condor (*Gymnogyps californianus*)  
Collected by J. Grinnell (witness Willard S. Wood), fresh,  
February 10, 1907, in Eaton Canyon, 5 1/2 miles  
northeast of Pasadena, in Los Angeles County,  
California. Egg laid on naturally accumulated  
layer of leaves and twigs of golden-oak on floor of  
cavity beneath slanting slab of rock leaning against  
vertical wall of rock; the site on steep canyon side.  
See description of site and photos by Finley, whom I  
conducted to the place in 1906. See also notebook of  
J. Grinnell for 1906, 1907 & 1908.  
J. Grinnell MVZ 13670

# Current Challenges

- **Complexity in parts preserved has increased**
  - both the number (e.g., # vials) and types of parts
- **More specimen attributes being recorded in field and also captured in databases**
  - habitat, behaviors, traits (age, sex, reproductive, molt, measurements, colors), relationships (host/parasite, predator/prey, parent/child), genetic or isotopic data, etc.
- **Media are now a standard type of data**
  - field photos, specimen images, audio and video, etc.
- **Related specimens (or their parts) can be distributed across different institutions**
  - tissues/vouchers; parasites/hosts



# Databasing at the MVZ

One of the first natural history museums to  
database its specimen collections (1978-1984)

➤ online in 1997





- Stopped using cards for cataloging in 2003 (herps, mammals) and 2007 (birds)
  - specimens cataloged directly into the database



## Arctos Migration 2005

- richer data capture
- easy relationships to other data (specimens, sequences, media, etc.)

Data conform to the DarwinCore standard for sharing data

# Who, What, Where, When

- Standard data captured for all specimens:
  - **WHO:** Collector/Preparator and their individual #
  - **WHAT:** Identification (genus, species, subspecies for birds and mammals)
  - **WHERE:** Higher geography, location, elevation, coordinates, datum, max error – georeferencing with GPS, GEOLocate, Google, etc.; retrospectively through MaNIS, HerpNET, ORNIS
  - **WHEN:** Verbatim date (text), begin and end dates

# Databasing – Mammals

## 232,255 specimens

Data entered from catalog cards until 2003

DEPARTMENT CATALOG – MAMMALS		UNIVERSITY OF CALIFORNIA MUSEUM OF VERTEBRATE ZOOLOGY			
ACC'N No. 891... (PT)		DATE OF ENTRY 7 Jan. 1916		ENTERED BY WPT	
GENERAL LOCALITY Yosemite Section, California, Mariposa & Tuolumne Co.'s					
DEPT. No.	ORIG. No.	NAME	DATE	COLLECTOR	EXACT LOCALITY
+ skeleton	22104	<i>Carnia latrans lestes</i>	June 25, 15	C. L. Camp	Yosemite Val.
+ skull	22105	<i>Urocyon cinereogentens californicus</i>	May 21	Ferris Holliger	Pleasant Val.
+ skull	22106	..	.. 25 ..	Ferris & Grinnell	..
22107	342	..	.. ..	C.D. Holliger	..
+ skull	22108	<i>Bassariscus astutus raptor</i>	.. 27 ..	G.F. Ferris	..
22109	523	<i>Martes americana sierrae</i>	July 19	..	Head of Squell Canyon
+ skull	22110	..	.. ..	Storer Ferris	..
Head + body in alc.	22111	..	.. 20 ..	Camp & Ferris	.. 9800'
+ skeleton	22112	<i>MARTES CAONIA SIERRAE</i>	.. 24 ..	C.D. Holliger	.. 9800'
+ skeleton	22113	..	.. 25 ..	..	.. 9800'

- ~90% of MVZ mammals in Arctos missing trait data from labels
- sex recorded for all
  - age, weight, measurements, and reproductive details only recorded in last 13 years

# Mammals – Since 2003

- Data captured from labels and field catalogs
- Traits standardly recorded, <1% of specimens imaged (primarily holotypes)



# Databasing – Herps

## 268,688 specimens

Data entered from catalog cards until 2003

ACQ'N No. 11011		DATE OF ENTRY Oct 11/77		ENTERED BY R. Alberch	
GENERAL LOCALITY Guatemala San Marcos <sup>①</sup>					
DEPT. No.	ORIG. No.	NAME	DATE	COLLECTOR	EXACT LOCALITY
144470	TP 14066	Ninia sebae	22 Aug 1976	T. Papenfuss	① Finca Santa Julia 1 1/2 Km. E. 3/4 Km. S. San Rafael Pie de la Cuesta. elev. 1100
144471	TP 14067	" "	"	"	"
144472	TP 14068	" "	"	"	"
144473	TP 14069	" "	"	"	"
144474	TP 14070	" "	"	"	"
144475	TP 14071	" "	"	"	"
144476	TP 14072	Rhadinea hannsteini	"	"	"
144477	TP 14073	" "	"	"	"
144478	TP 14074	" "	"	"	"

Limited information entered from field catalogs 2003-2008

~93% of MVZ herp records in Arctos missing trait data



Many specimens lack data tags

# Herps - Since 2008

Carol Spencer hired as Staff Curator in Herpetology  
– dedicated position for the first time



More complete data captured  
if recorded in field catalogs -  
age, sex, weight, SVL, colors,  
ecology, etc.



Color photos taken in field

Age/sex determined from  
specimens if not provided

# Databasing - Birds

## 186,246 specimens

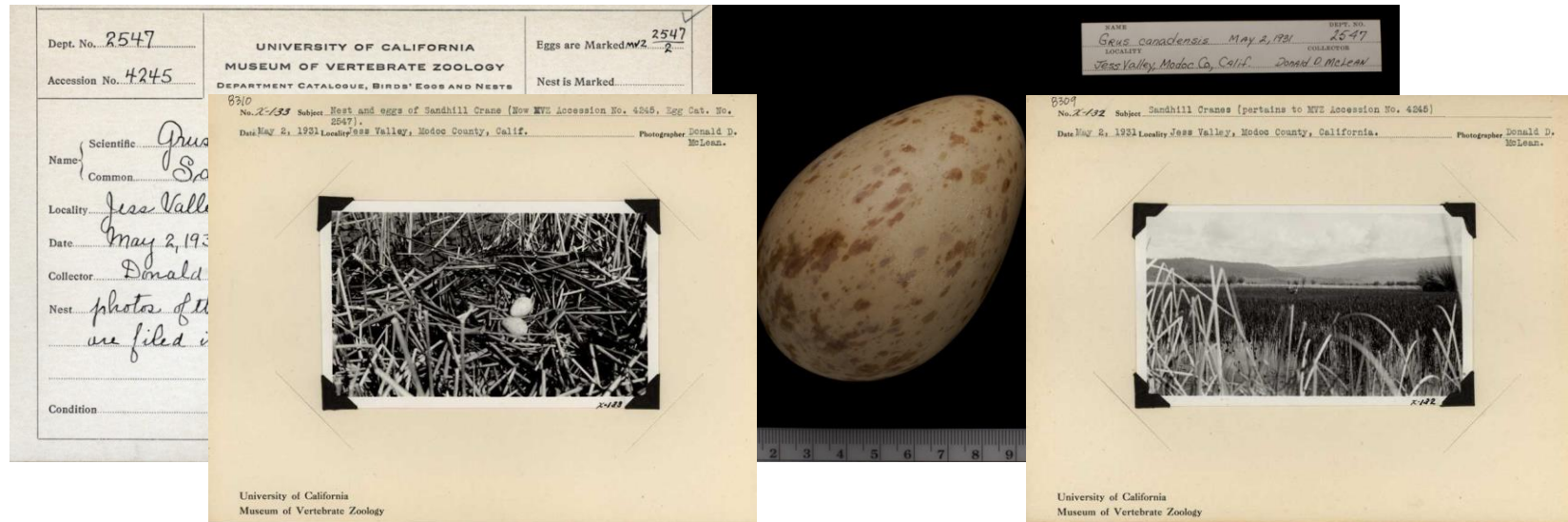
- Pre-1980s data entered from specimen labels (skins, skels) or ledger (fluids)
- Trait information captured if provided on tags
  - age, sex, weight, reproductive status, molt, fat, etc.
  - some traits determined for skins at time of data entry if not provided



# Databasing – Eggs/Nests

## 14,532 specimens

- Data entered from specimen labels
- NSF grant 2007-2012: Data slips scanned, eggs & nests photographed





# Birds – Current Practices

- Since mid-1980s, data captured from specimen labels and field catalogs
- Growth over the past 10+ years in complexity of data captured
- Arctos bulkloader for MVZ birds: ~100 fields
- Records fully populated at time of cataloging – eggs photographed, data slips scanned

# Trade-offs

- Fully populated records take more time to input, so may not be feasible or desirable
  - collection size, staff, management system, resources
  - test what is feasible – estimated time to complete
  - more time per specimen = slower collection growth?
- Quality versus quantity of data
  - goal = high quality and fully populated data
- Database should promote easy input of high quality data
  - standardized fields, authority tables
  - avoid free text typing as much as possible

# Large Collections



Impractical to digitize retrospectively from primary sources

Capture as much data as possible, as quickly and easily as possible (ledgers, cards)

If data are recorded, do not ignore them!

# Small Collections



Feasible to  
fully populate  
records

Database by  
Image Capture

# New Incoming Material...

Capture as much...



...as you can now!



# Facilitate Data Entry



YR: 2012	NAME: Carla Cicero	SPECIMEN CATALOG	PAGE #: 16	
<input checked="" type="checkbox"/> Skin <input type="checkbox"/> Skel. <input type="checkbox"/> ETOH <input type="checkbox"/> Other  <input type="checkbox"/> Audio <input type="checkbox"/> Photo FN:	# 3909	Locality: Spirit Lake, 10 200 ft, Uintah Mts., Daggett Co., Utah		
	<input checked="" type="checkbox"/> M <input type="checkbox"/> F <input type="checkbox"/> sex?		Skull: 100% ossified	
	Repro.: Testis 2x1 mm		Molt: Slight body molt.	
	Colors:		Stomach:	
	Habitat: Spruce forest.		Notes: Collected with CC # 3910-3911.	
	Lat: 40.84353°N Long: 109.99489°W GPS Error: 19 ft Extent: Datum: WGS 84 Date collected: 10 Sep 2012 Date prepared: 10 Sep 2012 Species: <i>Perisoreus canadensis</i> Wt: 82.5 gm Tiss: <input checked="" type="checkbox"/> Froz <input type="checkbox"/> Other Types: H, L, M, K Barcode #: 199308			
	<input type="checkbox"/> Net <input type="checkbox"/> Salvage			
	<input checked="" type="checkbox"/> Skin <input type="checkbox"/> Skel. <input type="checkbox"/> ETOH <input type="checkbox"/> Other  <input type="checkbox"/> Audio <input type="checkbox"/> Photo FN:	# 3910	Locality: Same as CC # 3909.	
		<input checked="" type="checkbox"/> M <input checked="" type="checkbox"/> F <input type="checkbox"/> sex?		Skull: 100% ossified
		Repro.: Ovary granular, ova minute		Molt: Trace body molt.
Colors:		Stomach:		
Habitat: Spruce forest.		Notes: Collected with CC # 3909 + 3911.		
Lat: Long: GPS Error: Extent: Datum: Date collected: 10 Sep 2012 Date prepared: 10 Sep 2012 Species: <i>Perisoreus canadensis</i> Wt: 69.5 gm Tiss: <input checked="" type="checkbox"/> Froz <input type="checkbox"/> Other Types: H, L, M, K Barcode #: 199309				
<input type="checkbox"/> Net <input type="checkbox"/> Salvage				
<input checked="" type="checkbox"/> Skin <input type="checkbox"/> Skel. <input type="checkbox"/> ETOH <input type="checkbox"/> Other  <input type="checkbox"/> Audio <input type="checkbox"/> Photo FN:		# 3911	Locality: Same as CC # 3909	
		<input checked="" type="checkbox"/> M <input type="checkbox"/> F <input type="checkbox"/> sex?		Skull: 100% ossified
		Repro.: Ovary granular, ova minute		Molt: Trace body molt.
	Colors:		Stomach:	
	Habitat: Spruce forest.		Notes: Collected with CC # 3909 + 3911.	
	Lat: Long: GPS Error: Extent: Datum: Date collected: 10 Sep 2012 Date prepared: 10 Sep 2012 Species: <i>Perisoreus canadensis</i> Wt: 73.5 gm Tiss: <input checked="" type="checkbox"/> Froz <input type="checkbox"/> Other Types: H, L, M, K Barcode #: 199310			
	<input type="checkbox"/> Net <input type="checkbox"/> Salvage			

# Undergraduates Can Help!



# Why Record It All?

- Can't predict how data will be used in the future
  - e.g., VertLife – trait databases
- Much easier to record full data at the time of capture – unlikely that you will capture it retrospectively unless there is a specific reason
  - e.g., research project that requires the data
- Increase in time is worth the value-added benefit for long-term research use



*Ward Russell (1992 interview) describing Joseph Grinnell's philosophy*





Higher research value for your collection = Higher institutional and funding support!



# Questions?

