Smithsonian
National Museum of Natural History

Digitization: Past, Present, and Future

Paul Kimberly and Rebecca Snyder

Overview

- The Smithsonian Institution
- The National Museum of Natural History (NMNH)
- NMNH Collections
- NMNH Digitization

The Smithsonian Institution

- Established in 1846 "for the increase and diffusion of knowledge"
- 19 museums and the National Zoological Park
- 168 affiliated museums throughout the U.S., Panama, and Puerto Rico

The NMNH

NMNH Consists of:

- 2 buildings: NHB & MSC
- 7 scientific departments
- National Anthropological Archives & Human Studies Film Archives (NAA & HSFA)
- Education & Outreach (Q?rius)
- Onsite affiliate organizations
- Marine station in Ft. Pierce, FL
- Artic Studies Center in Anchorage, AK
- Laboratories of Analytical Biology
- Onsite libraries with ~20 special collections

Staffing

- 470 staff (350 science staff)
- 50 contractors (non-construction)
- 42 staff from Affiliated Agencies
- 750 Fellows and Academic Appointments
- 11,000 short-term visiting researchers (<30 days)
- 600+ volunteers and interns

Holdings

SD600* collections

- Collection records
- Descriptive records
- Digital Surrogates
- Finding aids
- Specimens/objects
- Taxonomic inventories
- Documentation such as field notebooks and ledger books
- Illustrations
- Archival materials

Non-SD600 materials

- Conservation records
- Illustrations
- Maps
- Publications and manuscripts
- Exhibit materials

* Smithsonian Directive 600 is the SI's official documentation for collections care and preservation.

How Much We Have

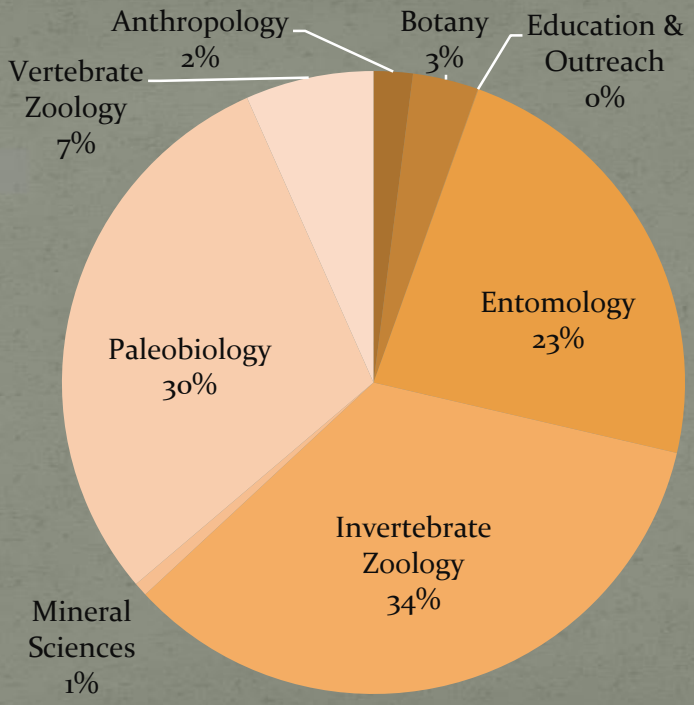
Our best guess is

144.6 million

Specimens, Genetic Samples & Objects

What Does 144 million look like?

Collections



What is Digitization?

Digitization is:

- Creation of digital descriptive records (databasing)
- Creation of digital surrogates (images representing collection items)
- Transcription of printed data, metadata, field books, and other supporting materials including collection item labels
- Transferring analog information to digital (35 mm slides, audio recordings, motion picture film, etc.)

What is Digitization?

Digitization is:

- Creation of digital descriptive records (databasing)
- Creation of digital surrogates (images representing collection items)
- Transcription of printed data, metadata, field books, and other supporting materials including collection item labels
- Transferring analog information to digital (35 mm slides, audio recordings, motion picture film, etc.)

Digitization is not:

- Simply taking digital imagery

How Much Is Digitized

6.8 Million Database Records

38.5 Million Objects Represented

~900,000 Surrogates (images of objects)

6.4 Million Web Accessible Records

THE PAST

Paleobiology Laboratory
(date unknown)

How We Got Here

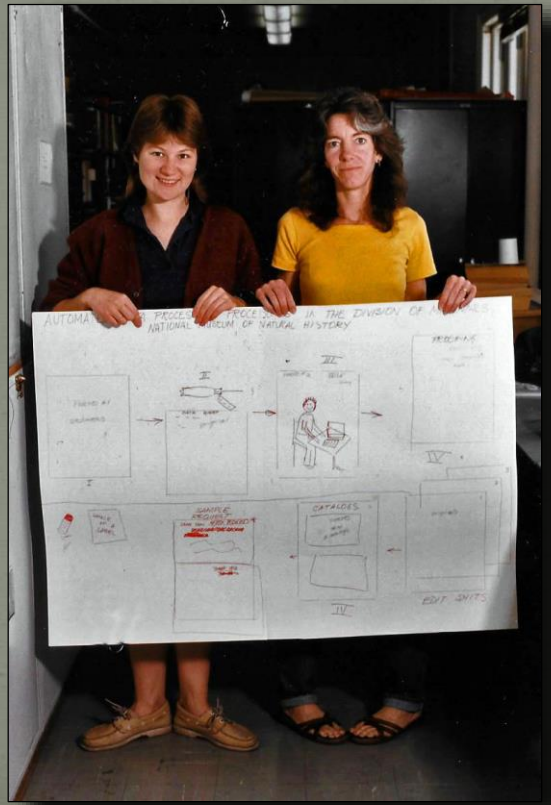
- Early 1960's computers begin to appear at NMNH
- 1963 - Creation of Automated Data Processing Committee (ADP)
- 1967 - SI purchases a series of mainframes, first database online, 10,000 specimen records created
- 1970 - NMNH creates custom database – SELGEM; Early scanning equipment is purchased
- *1979* - Begin the "Great Inventory"
- 1985 - Begin migration from SELGEM to INQUIRE
- *1992* - NMNH and parts of SI join the internet; Digital cameras begin to appear
- 2002 - Begin migration to EMu, ~1 million INQUIRE records; Hundreds of scanners, cameras and digitizing equipment onsite
- 2011 - Transaction Management integrated with EMu
- 2013 - SI Transcription Center goes live
- 2014 - Bumble Bee Rapid Capture Pilot Project

The Great Inventory 1979-1985

- Federal government mandated the Smithsonian to take an inventory of its holdings
- Approximately 20 technicians were hired temporarily for inventory
- The final specimen count was the foundation for our current totals
- It is accepted that that number is an approximation of what we held at that time
- I'm still searching for the final number that was generated
- Since then we have built on that base by tracking transactions in our system and then adding and subtracting these data each year

The Great Inventory 1979-1985

The Great Inventory 1979-1985



THE PRESENT

NMNH Digitization Steering Committee

(formed 2012)

Aka: DigiComm

DigiComm Looking Both Ways

The Castle

OCIO

DPO

DigiComm

Anthropology

Botany

Entomology

Invertebrates

Mineral Sci

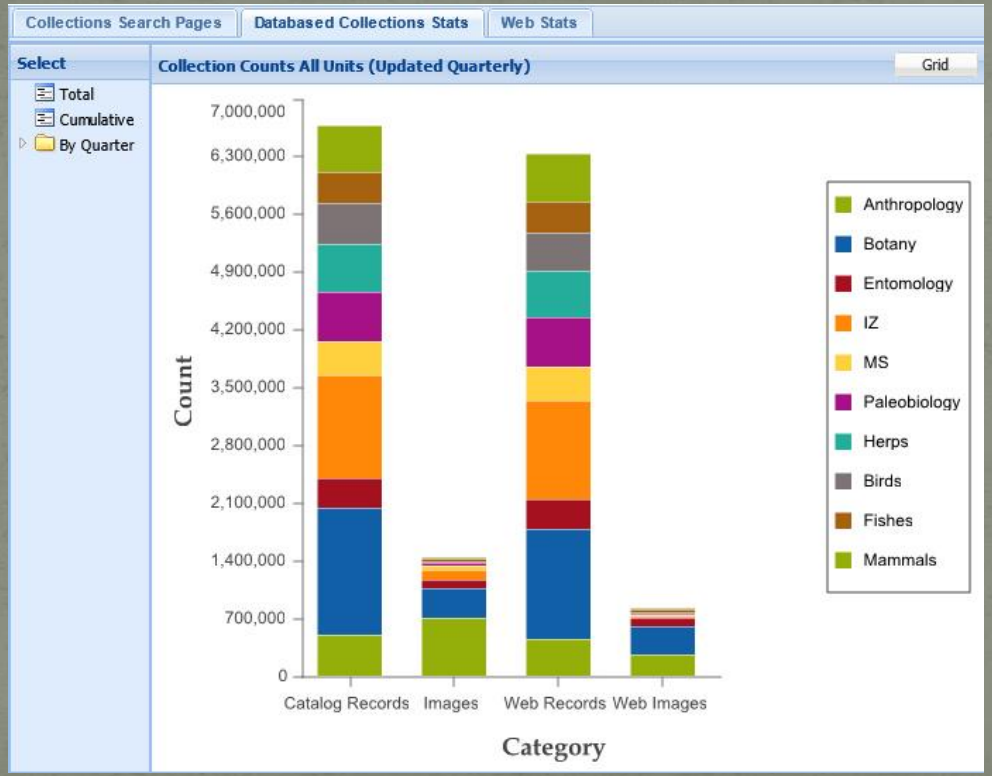
Paleobiology

Vertebrates

Education

March 2015 Statistics

6.8 million catalog records

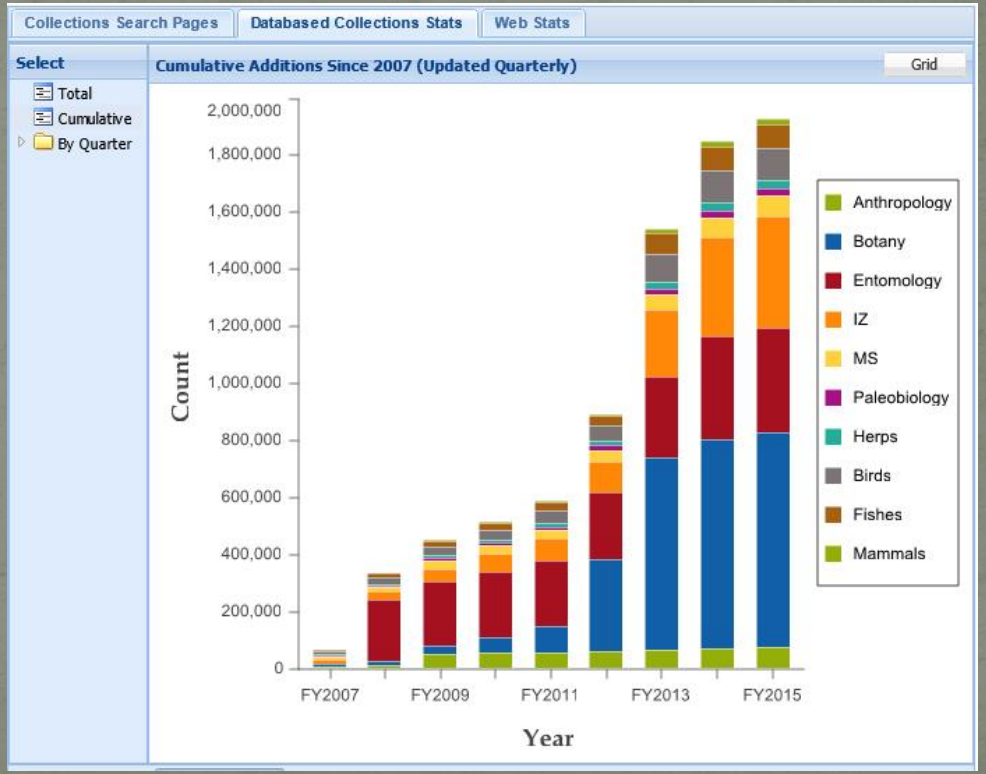


~900 K digital surrogates

1.2 million multimedia assets

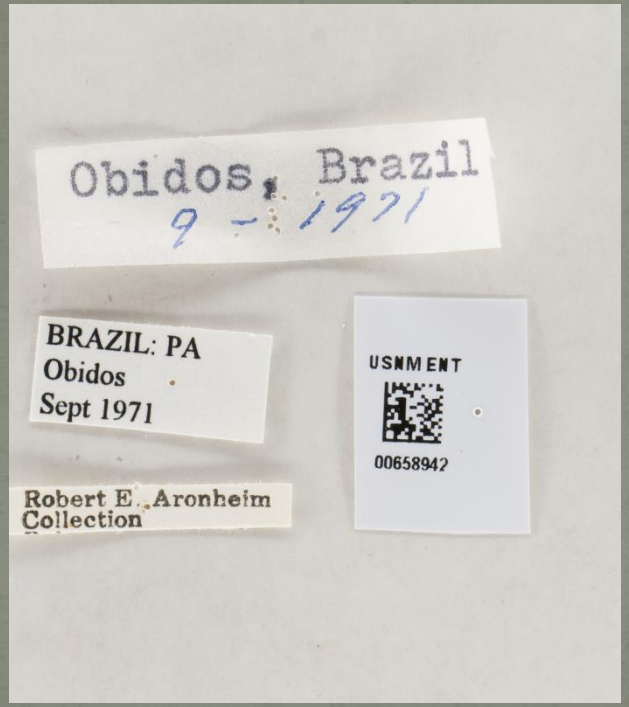
Cumulative Progress since 2007

- Large increase in 2013 is partly due to migration of 300 K records from disparate databases
- Hiring of 6 roaming technicians
- http://collections.nmnh.si.edu/search/

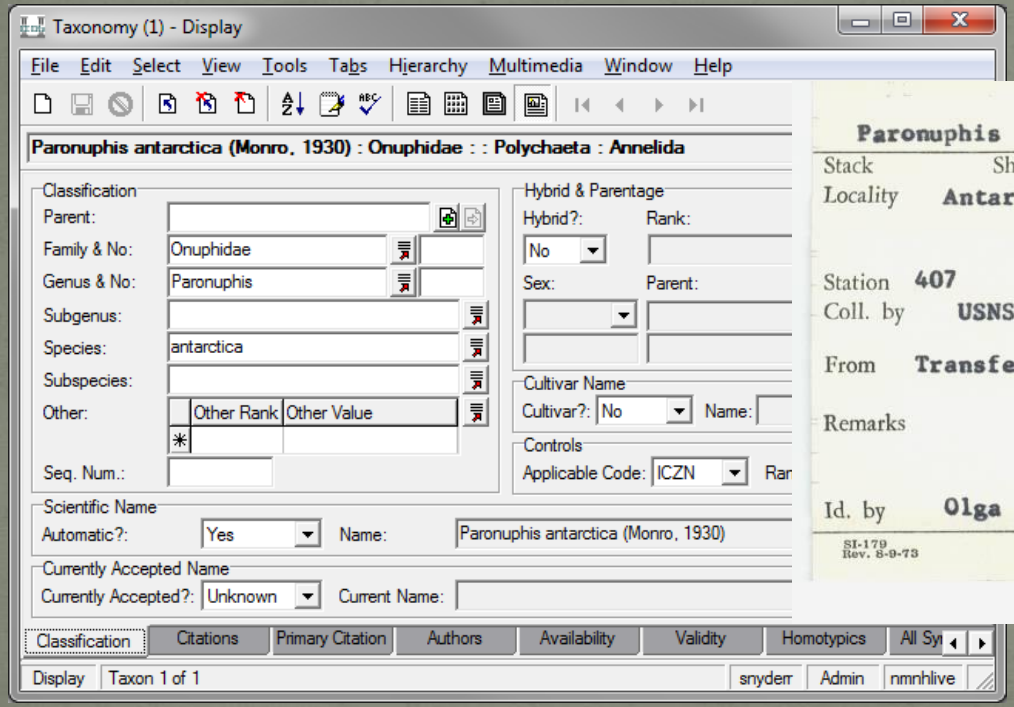


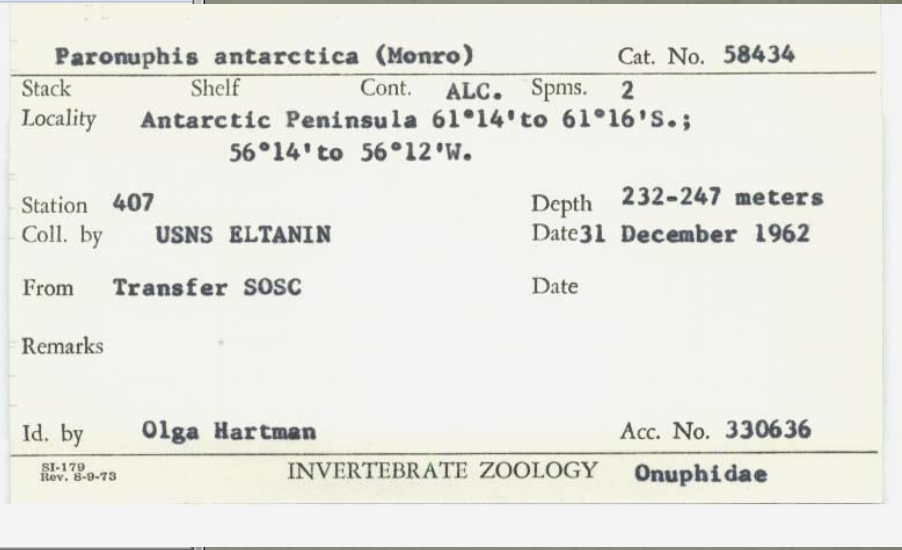
Thinking by type not department

Pinned Specimen Labels



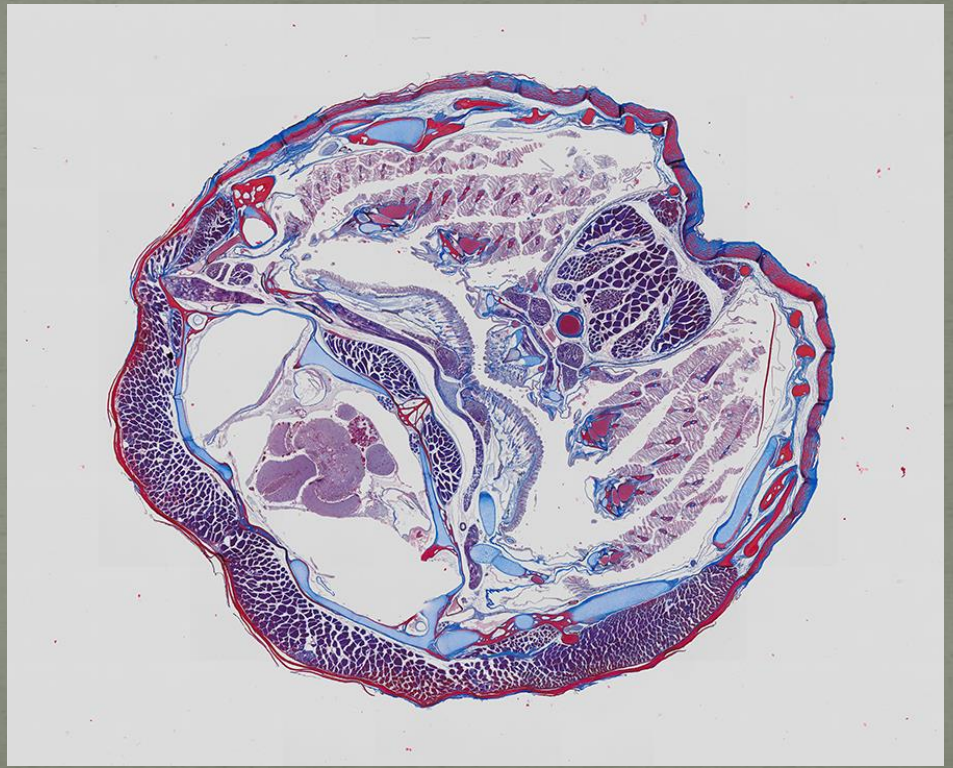
Dry Specimen Label



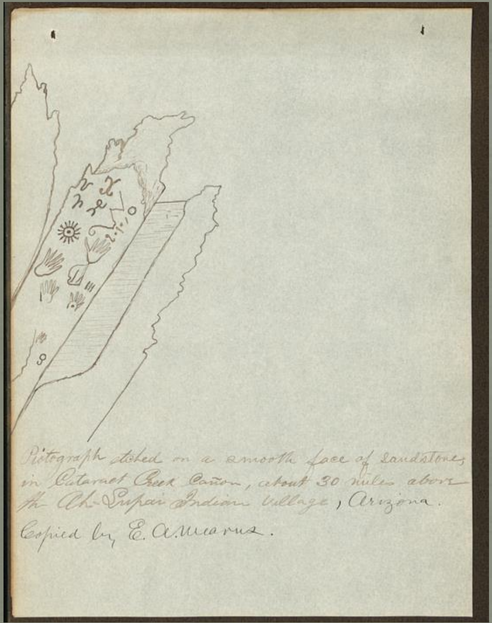


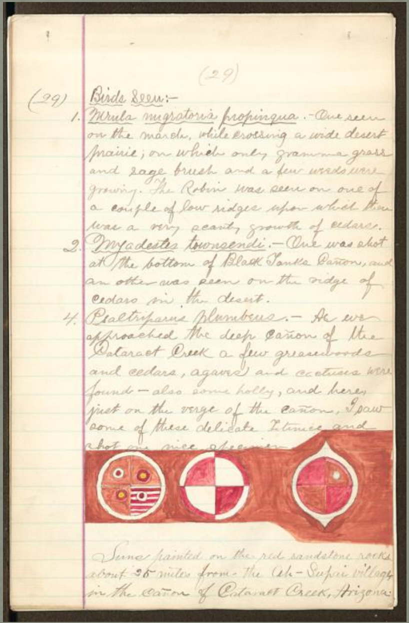
Microscope and Histology Slides

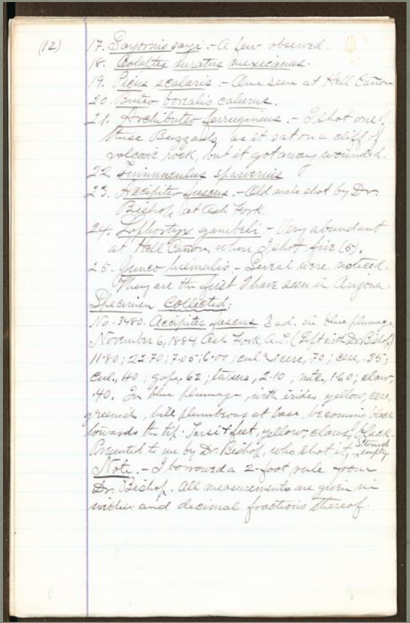
- Upwards of 5 million slides throughout the museum
- Currently testing scanning equipment
- The first scan pass captures the slide in reflected light to capture labels



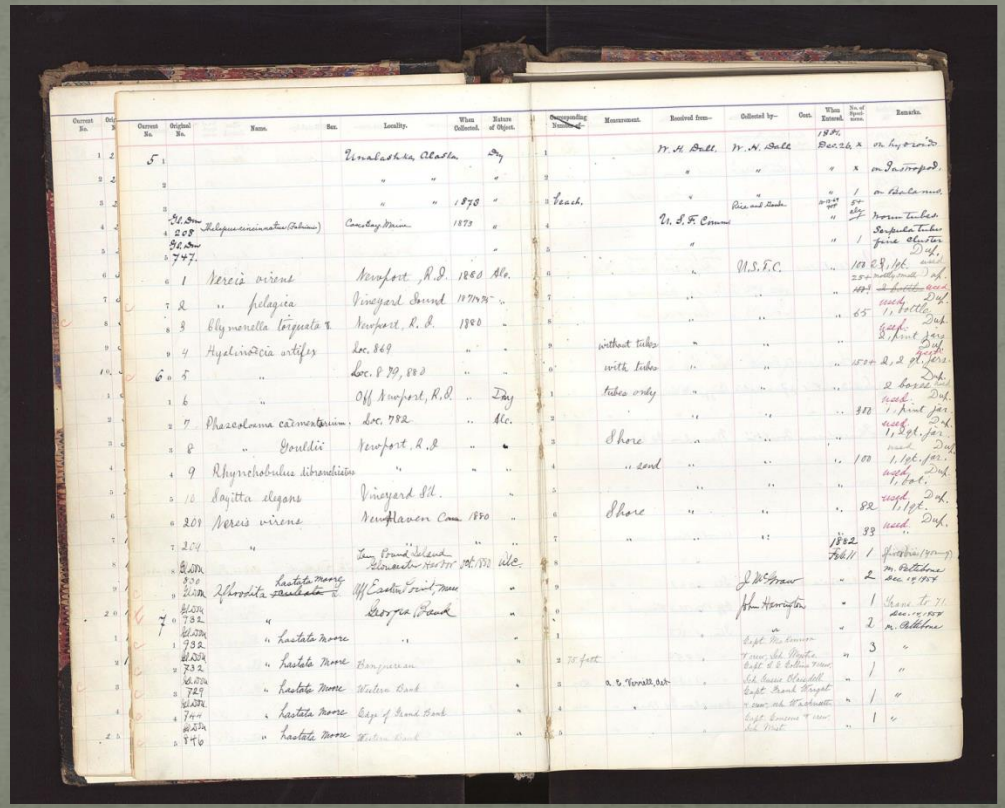
Field Notes


Pictograph etched on a smooth face of sandstone, in Cataract Creek Cañon, about 30 miles above the Ah-Supai Indian Village, Arizona.
Copied by E. A. Mearns.

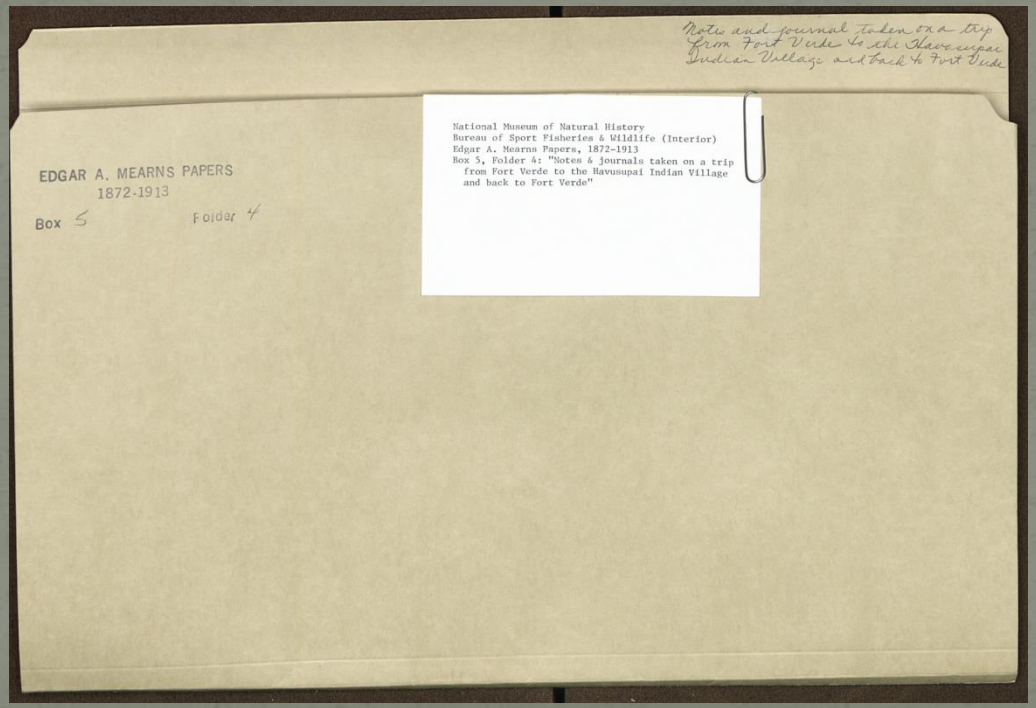

(29)
(29) Birds seen:-
1. Merula migratoria propinqua.- One seen on the march, while crossing a wide desert prairie; on which only gramma grass and sage brush and a few weeds were growing. The Robin was seen on one of a couple of low ridges upon which there was a very scanty growth of cedars.
2. Myadestes townsendi.- One was shot at the bottom of Black Tanks Cañon, and an other was seen on the ridge of cedars in the desert.
4. Psaltriparus plumbeus.- As we approached the deep Cañon of the Cataract Creek a few greasewoods and cedars, agaves and cactuses were found - also some holly, and here, just on the verge of the cañon, I saw some of these delicate Titmice and shot one nice specimen.
Suns painted on the red sandstone rocks about 25 miles from the Ah-Supai village in the cañon of Cataract Creek, Arizona.


(12)
17. Sayornis sayi.- A few observed.
18. Colaptes auratus mexicanus.
19. Picus scalaris.- One seen at Hell Cañon
20. Buteo borealis calurus.
21. Archibuteo ferrugineus.- I shot one of these Buzzards as it sat on a cliff of volcanic rock, but it got away wounded.
22. Tinnunculus sparverius
23. Accipiter fuscus.- Old male shot by Dr. Bischof at Ash Fork.
24. Lophortyx gambeli.- Very abundant at Hell Cañon where I shot five (5).
25. Junco hiemalis.- Several were noticed. They are the first I have seen in Arizona.
Specimen Collected:
No. 3480. Accipiter fuscus ♂ ad. in blue plumage, November 6, 1884, Ash Fork A.T. (Left with Dr. Bischof) 11·80; 22·70; 7·05; 6·00; cul. ? scree, .70; cere, .35; cul., .40; gape, .62; tarsus, 2·10; mid., 1·60; claw, .40. In blue plumage, with irides yellow; cere greenish, bill plumbeous at base, becoming black towards the tip. Tarsi & feet, yellow; claws, black. Presented to me by Dr. Bischof, who shot it. Stomach empty.
Note.- I borrowed a 2-foot rule from Dr. Bischof. All measurements are given in inches and decimal fractions thereof.

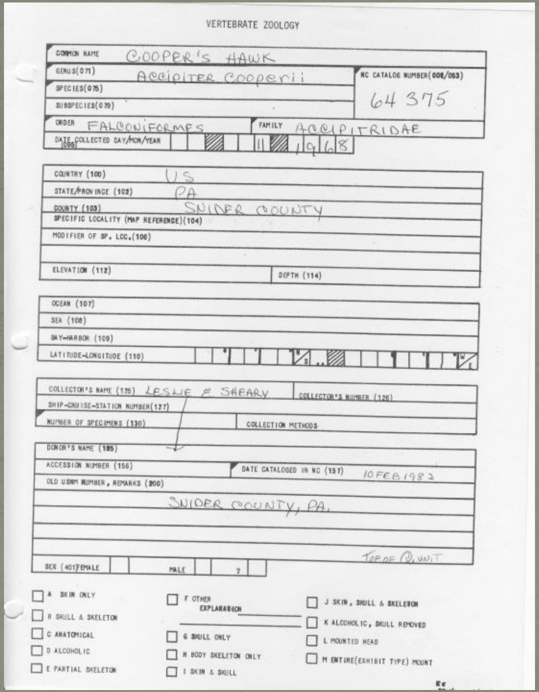
Ledgers



Archive Documentation and A/V



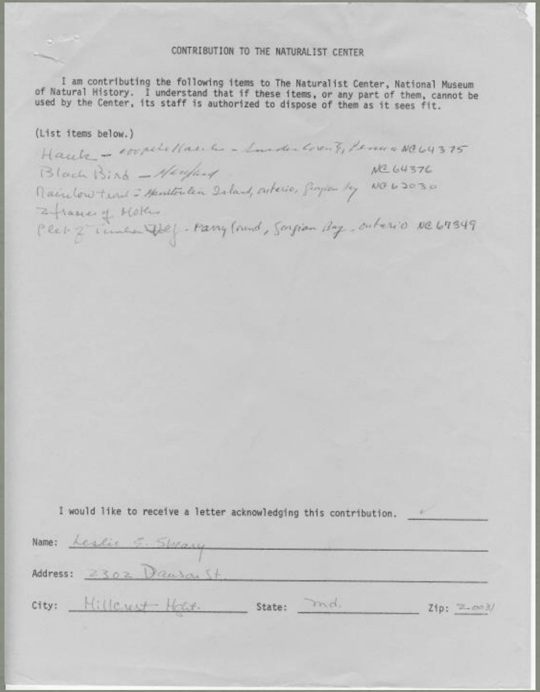
Office of the Registrar Digitization



VERTEBRATE ZOOLOGY

COMMON NAME: COOPER'S HAWK
GENUS(071): Accipiter cooperii
SPECIES(075)
SUBSPECIES(079)
NC CATALOG NUMBER(008/053): 64375
ORDER: FALCONIFORMES
FAMILY: ACCIPITRIDAE
DATE COLLECTED DAY/MON/YEAR (095): 11 / 1968
COUNTRY (100): US
STATE/PROVINCE (102): PA
COUNTY (103): SNIDER COUNTY
SPECIFIC LOCALITY (MAP REFERENCE)(104)
MODIFIER OF SP. LOC.(106)
ELEVATION (112)
DEPTH (114)
OCEAN (107)
SEA (108)
BAY-HARBOR (109)
LATITUDE-LONGITUDE (110)
COLLECTOR'S NAME (125): LESLIE E SHEARY
COLLECTOR'S NUMBER (126)
SHIP-CRUISE-STATION NUMBER(127)
NUMBER OF SPECIMENS (130)
COLLECTION METHODS
DONOR'S NAME (185)
ACCESSION NUMBER (156)
DATE CATALOGED IN NC (157): 10 FEB 1982
OLD USNM NUMBER, REMARKS (200)
SNIDER COUNTY, PA.
SEX (401) FEMALE
MALE
?

A SKIN ONLY
B SKULL & SKELETON
C ANATOMICAL
D ALCOHOLIC
E PARTIAL SKELETON
F OTHER EXPLANATION
G SKULL ONLY
H BODY SKELETON ONLY
I SKIN & SKULL
J SKIN, SKULL & SKELETON
K ALCOHOLIC, SKULL REMOVED
L MOUNTED HEAD
M ENTIRE(EXHIBIT TYPE) MOUNT



CONTRIBUTION TO THE NATURALIST CENTER

I am contributing the following items to The Naturalist Center, National Museum of Natural History. I understand that if these items, or any part of them, cannot be used by the Center, its staff is authorized to dispose of them as it sees fit.

(List items below.)

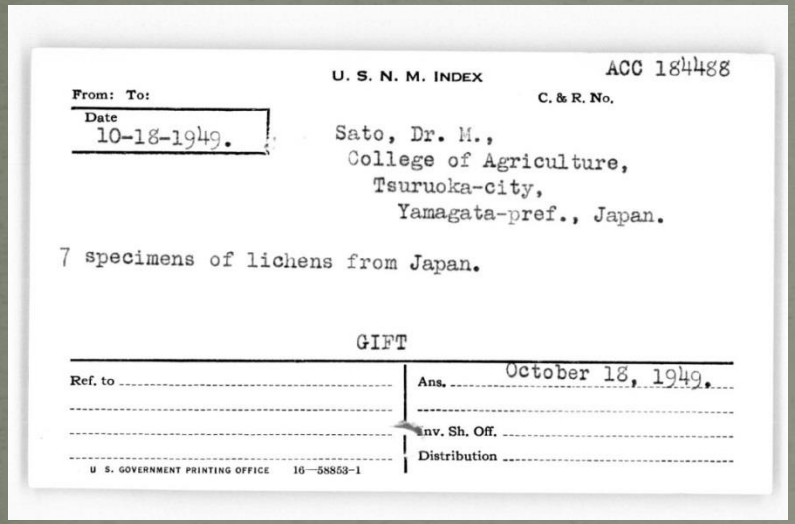
Hawk – NC 64375
Black Bird – NC 64376
NC 62030
NC 67349

I would like to receive a letter acknowledging this contribution. ✓

Name: Leslie E. Sheary
Address: 2302 ...
City: Hillcrest Hgts.
State: Md.
Zip: 20031

~500,000 documents

Office of the Registrar Digitization



U. S. N. M. INDEX ACC 184488

From: To:

C. & R. No.

Date 10-18-1949.

Sato, Dr. M.,
College of Agriculture,
Tsuruoka-city,
Yamagata-pref., Japan.

7 specimens of lichens from Japan.

GIFT

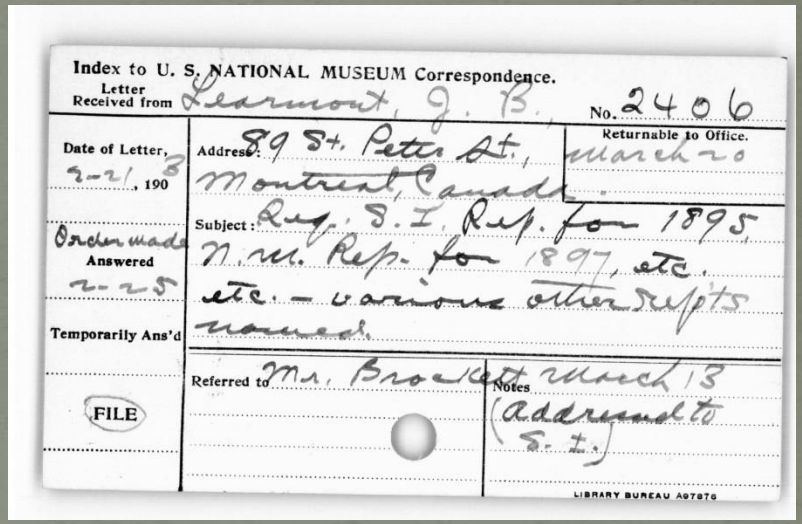
Ref. to

Ans. October 18, 1949.

Inv. Sh. Off.

Distribution

U. S. GOVERNMENT PRINTING OFFICE 16—58853-1



Index to U. S. NATIONAL MUSEUM Correspondence.

Letter Received from Learmont, J. B. No. 2406

Date of Letter, 2-21, 1895

Address: 89 St. Peter St., Montreal, Canada

Returnable to Office. March 20

Order made Answered 2-25

Subject: Reg. D. I. Rep. for 1895 N. M. Rep. for 1897, etc. etc. - various other Rep'ts named.

Temporarily Ans'd

Referred to Mr. Brockett March 13

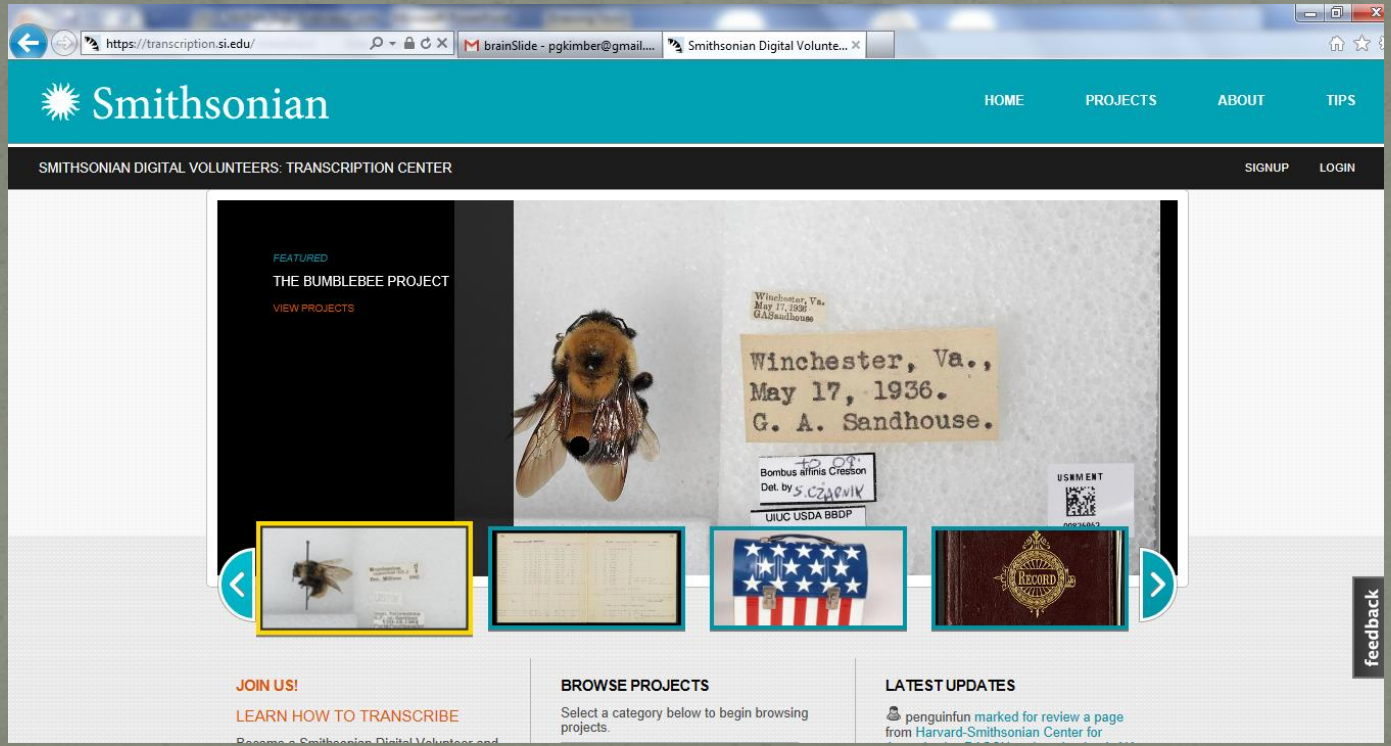
FILE

Notes (Addressed to S. I.)

LIBRARY BUREAU A97876

>350,000 index cards

Crowd Sourcing

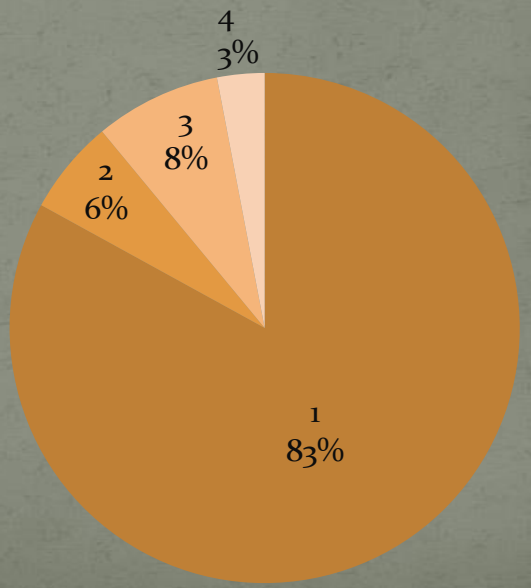


Statistics

- Yearly reporting to SI of Collection and Digitization Assessment
- Rank database records on a scale from 1-4 (Record Quality)

Record Quality

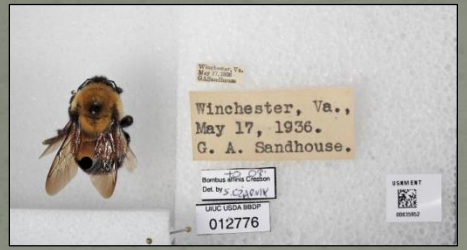
- 1 = No database record ..... 32 million
- 2 = Below standard record ..... 2 million
- 3 = Standard record ..... 3 million
- 4 = Above standard record ..... 1 million

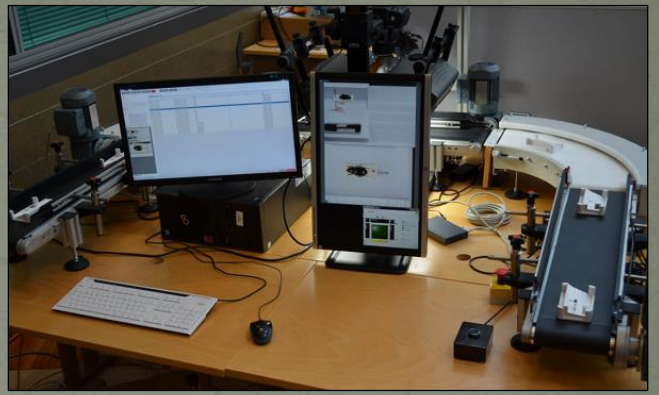


What's Left To Do

- We estimate it will take 35.5 million catalog records to represent our current collections
- That means we need at least another 26 million records… (*eek*!)
- It has taken 50 years to get to 6.8 million, though our rate has dramatically improved over the years!
- So we still have a long way to go!!!

THE FUTURE


Winchester, Va.,
May 17, 1936.
G. A. Sandhouse.
Bombus affinis Cresson
Det. by
UIUC USDA BBDP
012776



THE FUTURE

- Prioritized projects based on research question(s)
- Rapid capture – Improved workflows
- Greater automation
- Digitization and cataloging will become more dependant upon each other
- Create a digitization center
- Increased crowd sourcing – transcription centers and beyond

You Say Potato!

You Say Potato!
I Say Potato!

You Say Potato!
I Say Potato!
You Say Digitise!

You Say Potato!
I Say Potato!
You Say Digitise!
I say Digitize!

Thanks

- Rebecca Snyder
- NMNH Staff
- CSIRO
- iDigBio
- Paul Flemons