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
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
Anthropology: 2%
Botany: 3%
Education & Outreach: 0%
Vertebrate Zoology: 7%
Paleobiology: 30%
Mineral Sciences: 1%
Entomology: 23%
Invertebrate Zoology: 34%
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
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
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

Paronuphis antarctica (Monro, 1930)  Cat. No. 58434

Classification:
- Parent: Onuphididae
- Genus: Paronuphis
- Species: antarctica

Locality:
- Antarctic Peninsula 61°14' to 61°16'S; 56°14' to 56°12'W.

Station: 407  Depth: 232-247 meters
Coll. by: USNS ELTANIN  Date: 31 December 1962
From: Transfer SOSC

Id. by: Olga Hartman  Acc. No. 330636

Remarks:
- INVERTEBRATE ZOOLOGY
- Onuphididae
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
Archive Documentation and A/V
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
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