DIGITIZATION @ USNM: Where we are

Patricia Gentili-Poole & Jessica Bird
Collection Information Management Unit
Department of Entomology, NMNH, SI

• Updates to Digitization @ USNM
• Rapid Capture Workflow
EMu as the RCIS of record for NMNH

- 2001 Deployment at the Museum
- 2007 Migration of Entomology existing Types & Odonata databases
- 2009 Started updating of Type specimens records
- 2011 Deployment of Transactions Module
- 2012 Started Data clean-up and standardization
  Migration of legacy datasets
- 2013 Deployment of Collection Level Index & Taxonomic Lot
  Initiated Species Inventory
- 2014 Deployment of Genetic Samples integrated with FreezerPro
  Beta version of Field Information Management System
  Revamped the online Search interface
  Estimated # of records to represent scientific value of collection
  Development of Workflow for Rapid Capture at the specimen level
Smithsonian National Museum of Natural History

Search the Department of Entomology Collections

Bombyx (Bombus) affinis Cresson: Apidae: Hymenoptera: Insecta: Arthropoda

Barcode: USNMENT00835947
Catalog: Specimen Inventory
Order: Hymenoptera
Family: Apidae
Scientific Name: Bombyx (Bombus) affinis Cresson
Preparation: Pinned
Country: United States
Province/State: Massachusetts
District/County: Hampshire
Precise Locality: Amherst, Notch South
Collecting Date: VI, 5, 1904
Record Last Modified: 18 Jul 2014 14:34:00

Bombyx (Bombus) affinis

Title: USNMENT00835947
Creator: Pixel Acuity LLC
Format: image/tiff
Resolutions: 2500x1734, 450x348, 350x1x901
Rights Holder: This image may be protected by copyright or have other legal restrictions on use. Permission to publish should be obtained from the National Museum of Natural History, Smithsonian Institution, Washington, D.C.
Rights: Unless otherwise noted, this image or its contents may be protected by international copyright laws.

Bombyx affinis

Ulrich USDA BBSP
012770

Bombyx affinis

HOMOTYPE
H. M. A. C.
CATALOGS

- **Type Specimens Inventory:** ~120,000 expected records
  - All original information (except for Family) and images
  - Entomology Types [= updated Primary Types]
  - Inquire Types [= migrated database with Primary & Secondary Types]
  - 126,430 current records

- **Species Inventory:** ~400-450,000 expected records
  - Includes unidentified
  - 62,831 current records

- **Specimen Inventory:** ~21,000,000 expected records
  - ‘Regular Specimens’
  - Includes secondary Types, Genetic Samples
  - 171,407 current records

- **Illustration Archives:** unknown # of records
  - Published and unpublished illustrations
  - 3,893 current records
Species Inventory

It is a Management Tool

- Includes: Scientific Name, Author, Year; Specimen Count; Country level Distribution, States if United States; Bioregion; Images

- Capture by multiple passes approach:
  1) Scientific names and update classification
  2) Specimen count and distribution
  3) Images

- Migration of legacy datasets and capture *de novo*

- Involves Pre-capture preparation of the Collection
Specimen Inventory

Estimated number of records to represent scientific value of a collection of ~35 million objects is 20,835,735

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinned</td>
<td>18,813,906</td>
<td>largest problem</td>
</tr>
<tr>
<td>Slides</td>
<td>1,392,676</td>
<td>method exists</td>
</tr>
<tr>
<td>Fluid</td>
<td>522,717</td>
<td>major problem</td>
</tr>
<tr>
<td>Papered</td>
<td>106,436</td>
<td>done (Odonata)</td>
</tr>
</tbody>
</table>
The Challenge

How to get there in a reasonable length of time?
The goal is to capture the **label data**, not the specimen image per se
Bumble Bee RCPP

To develop a workflow to digitize the pinned collection

SI DPO conditions
- Medium size
- With some volume
- One shot fits all
- One camera setting fits all
- Use of SI Transcription Center

Entomology conditions
- Pre-capture ready = species inventory complete
- Important to the scientific community
- Appealing to the general public to attract digital volunteers
- Candidates: bees & wasps
The Workflow

A clear three step process:

Specimen processing → Imaging → Data processing
Step 1: Specimen processing

- Collection
- Adding barcode labels
- Removing labels from pins
- Staging
- Replacing labels on pins
- Replacing specimens in trays and drawers
- Imaging
## Results on Specimen Handling

<table>
<thead>
<tr>
<th>Personnel in staging</th>
<th>Specimens processed</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x9.5 days [228 hrs]</td>
<td>6,732</td>
<td>708/day [29/hr]</td>
</tr>
<tr>
<td>5x13 days [390 hrs]</td>
<td>15,732</td>
<td>1,210/day [40/hr]</td>
</tr>
<tr>
<td>6x10.5 days [378 hrs]</td>
<td>13,675</td>
<td>1,302/day [36/hr]</td>
</tr>
<tr>
<td>7x5 days [210 hrs]</td>
<td>7,916</td>
<td>1,583/day [37/hr]</td>
</tr>
</tbody>
</table>
Results on Specimen Handling

Traditional method
- Records in EMu with label data
- No images of labels captured
- Processed 100 specimens/person/day
- 3,800 records in 38 calendar days

RCPP Method
- Records in EMu without label data
- Images captured for preservation
- Processed 184 specimens/person/day
- 44,047 records in 201 work days (=38 calendar days)
Step 2: Imaging and post-processing

1. Capture
2. Crop
3. Rename files with Barcode Number
4. Embed Metadata
5. Generate report
6. Create folders in Staging Server
7. Quality Control
Final image ready to be uploaded into NMNH RCIS-EMu
(Image file name: USNMENT00835947)
Step 3: Data Processing

Access query to produce import file (developed by Informatics)

Creation of new records in EMu in one pass, publish to Internet

Capture by TC
Transcription by DV
Final validation by DoE

Appending data into existing records in EMu
<table>
<thead>
<tr>
<th><strong>Collection Custody</strong></th>
<th><strong>Cataloging Details</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Museum:</strong> NMNH</td>
<td><strong>Catalog number:</strong></td>
</tr>
<tr>
<td><strong>Department:</strong> Entomology</td>
<td><strong>Suffix:</strong></td>
</tr>
<tr>
<td><strong>Division:</strong></td>
<td><strong>Barcode:</strong> USNMENT01009920</td>
</tr>
<tr>
<td><strong>Catalog:</strong> Specimen Inventory</td>
<td><strong>Part Number:</strong></td>
</tr>
<tr>
<td><strong>Collection name:</strong> 1 Hymenoptera</td>
<td><strong>Whole/Part:</strong> Whole</td>
</tr>
<tr>
<td></td>
<td><strong>Cataloged by:</strong> Bird, Jessica</td>
</tr>
<tr>
<td></td>
<td><strong>Date cataloged:</strong> 21 Jul 2014</td>
</tr>
<tr>
<td></td>
<td><strong>Kind of object:</strong> Specimen/Lot</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Counts</strong></th>
<th><strong>Other Counts</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Specimen count:</strong> 1</td>
<td><strong>Kind</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Value</strong></td>
</tr>
<tr>
<td><strong>Original count:</strong> 1</td>
<td></td>
</tr>
<tr>
<td><strong>Modifier:</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Other Numbers</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Other Numbers Kind:</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Skeleton Descriptive Record in NMNH RCIS-EMu
(Unique Identifier Number highlighted)
**Skeleton Descriptive Record in NMNH RCIS-EMu**
(Note empty fields for label information)
Skeleton Descriptive Record in NMNH RCIS-EMu

(Notes indicating the record is ready for Transcription Center capture)
Multimedia Record in NMNH RCIS-EMu
(Note the File Title bearing the Unique Identifier Number)
Prior to RCPP production, a help file, template and in-line text were created in collaboration with the TC team

At time of import, a note is added to skeleton records that “tags” images for upload into transcription center

Administrators have access to data sets and publish sets as needed

Digital volunteers transcribe specimen labels into a template

Keep in touch with volunteers through email and social media
Transcription Center Data Entry
Final Data Import

- After labels are transcribed they must be validated by Collections Information Unit staff.
- Validated data is exported in .csv format.
- Data must be “massaged” in order for it to be imported back in to EMu for complete records.
  - Parse out any items in notes that have a dedicated EMu field, such as sex and the identifier.
  - Match irns for parties records (collector, identifier, etc).
  - Add note type and descriptions for all note fields.
  - Un-publish previous transcription center notes and add a new note to signify that the record was transcribed by digital volunteers.
  - Make sure that all data entered is standardized.
- Map to EMu fields, and perform final import.
**Completed Descriptive Record in NMNH RCIS-EMu**

(Note the label data transcribed and appended to the existing skeleton record)
<table>
<thead>
<tr>
<th>Note</th>
<th>Date</th>
<th>Kind</th>
<th>Attributed To</th>
<th>Publish</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMNHENTO_TR_BUMBLEBEES_SET01</td>
<td>25 Jun 2014</td>
<td>Transcription Set</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Available for Transcription</td>
<td>25 Jun 2014</td>
<td>Crowdsourcing</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>[Female symbol] 09. Bombus affinis Cresson Det. by ...</td>
<td>25 Jun 2014</td>
<td>Label</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>B.a. [female symbol]</td>
<td></td>
<td>Label</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Transcribed by digital volunteers</td>
<td>18 Jul 2014</td>
<td>Crowdsourcing</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

Completed Descriptive Record in NMNH RCIS-EMu
(Notes indicating data was transcribed by Digital Volunteers)