Digitizing Dried Insects Workshop

• Gil Nelson (gnelson@bio.fsu.edu)
  (Florida State University)

24 April 2013
Field Museum
Chicago
Digitization Workflows

Digitizing Dried Insects Workshop
Gil Nelson
24 April 2013
Field Museum, Chicago
Assessing Digitization Practices in Biological and Paleontological Collections

28 Collections
10 Museums
Spanning biological and paleontological collections
Insects and other invertebrates, plants, birds, mammals
Wet, dry
Five task clusters that enable efficient and effective digitization of biological collections

Gil Nelson¹, Deborah Paul¹, Gregory Riccardi¹, Austin R. Mast²

¹ Institute for Digital Information, Florida State University, Tallahassee, FL 32306-2100, United States ² Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, United States

Corresponding author: Gil Nelson (gnelson@bio.fsu.edu)
Workflows Patterns Observed

Dominant Digitization Patterns Observed

Figure 1: Dominant Digitization Workflows

a. Data to Occasional or Optional Image to Distribution

Pre-digitization curation → Data capture → Capture Exemplar images → Process images → Store images → World Wide Web, Data aggregators, In-house usage

b. Parallel Data/Image to Distribution

Pre-digitization curation → Data capture → Capture images → Process images → Store images → World Wide Web, Data aggregators, In-house usage

c. Image to Data to Distribution

Pre-digitization curation → Capture images → Process images → Store images → Data capture → World Wide Web, Data aggregators, In-house usage
Values of defined workflows

- Promote efficiency and automation of processes
- Facilitate routing and scheduling of activities
- Provide for balancing workloads
- Ensure that processes are visible and predictable
- Allow for escalations and notifications
- Enhance tracking of tasks
- Foster collaboration of all parties involved
- Stimulate the convergence of process and information
- Promote continuous evaluation and redesign
Pre-planning a Workflow Process

- Identify the database management system and imaging equipment to be used.
- Identify the processes/modules for which workflows will be designed.
- Identify *(in excruciating detail!)* the tasks (or task clusters) that constitute the processes/modules.
- Identify the specific actions to be taken and the attributes (if any) associated with these actions.
- Identify roles (and only secondarily the people who will fill them).
- Identify points/processes/parameters for notifications and escalations.
- Identify dependencies, transitions, and iterations.
- Determine minimal data requirements for defining a complete record.
- Determine how records and objects will be uniquely identified in a global environment.
- Determine how these identifiers will be assigned.
- Determine how/if these identifiers will be affixed.
- Determine a consistent file naming strategy for images, attachments, and other related materials.
- Determine file storage needs and location for data, images, and ancillary materials.
- Define and diagram flow.
Example Processes (Modules), their Cycles and Dependencies

<table>
<thead>
<tr>
<th>Process</th>
<th>Cycle</th>
<th>Dependency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software configuration</td>
<td>Once/non-recurring</td>
<td></td>
</tr>
<tr>
<td>Equipment set-up</td>
<td>Once/non-recurring</td>
<td></td>
</tr>
<tr>
<td>Specimen curation</td>
<td>Recurring</td>
<td></td>
</tr>
<tr>
<td>Specimen selection</td>
<td>Recurring</td>
<td>Pre-digitization curation</td>
</tr>
<tr>
<td>Specimen transport</td>
<td>Recurring</td>
<td>Specimen selection, imaging, data entry</td>
</tr>
<tr>
<td>Conservation</td>
<td>Episodic</td>
<td>Curatorial processes, imaging, data entry</td>
</tr>
<tr>
<td>Data entry</td>
<td>Recurring/tasks iterative</td>
<td>Specimen transport</td>
</tr>
<tr>
<td>Imaging</td>
<td>Recurring/tasks iterative</td>
<td>Specimen transport</td>
</tr>
<tr>
<td>Equipment adjustments</td>
<td>Episodic</td>
<td>Technician escalation</td>
</tr>
<tr>
<td>Software update/tweaking</td>
<td>Episodic</td>
<td>Technician escalation</td>
</tr>
<tr>
<td>Specimen return/shelving</td>
<td>Recurring</td>
<td>Imaging or data entry</td>
</tr>
</tbody>
</table>
Follow a modular approach

• “Plug and play” modules are preferred.
• Simple modules involving a limited number of tasks are easier to troubleshoot and maintain.
• Divide large modules into sub-modules.
• Modules are generally self-contained but tangential.
• There is no consensus workflow, virtually all workflows are customized.

Assign roles deliberately

• Adjust to strengths of each technician--using students and volunteers requires flexibility in role assigned to personnel rather personnel assigned to role.

Create task lists

• Complete.
• Clear.
• Succinct.
• Ordered.
• Reusable.
A sample, detailed task list.

1. Open Capture NX2 and View NX2.
2. Open Camera Control Pro 2.
3. Open default.ncc as settings file:
   - Settings->Load Control Settings
   - My Documents->CameraSettings->default.ncc.
4. Create a folder in X:\SpecimenImages\NEF, using the current date as the folder name, as 2013-04-14.
5. Retrieve next specimens to image from cabinet.
6. Insert Image “From Here” tag to proper place in cabinet.
7. Set image number in Camera Control 2 to next bar code:
   - tools->download options
   - Edit
   - Start numbering at: <Enter next bar code number; no leading zeros>.
8. In Download Options, set the default folder to the one you created in step 4.
9. Position specimen in frame, ensuring complete specimen is visible.
10. Open Live View, position the focus square on specimen.
11. Click AF to test.
12. Click AF and Shoot.
13. Once the first image loads, navigate to it in Capture NX2 or View NX2.
14. Open the image, zoom in and check margins to ensure all of the specimen is visible.
15. Repeat 9-14 until satisfied, resetting image number each time.
17. Load next specimen in frame.
18. Use remote release on camera and record the images.
19. As you shoot, check each image bar code to ensure it is in sequence with the one preceding it and matches the next one in the series.
20. For out-of-sequence bar codes, change the number in the download options.
21. Repeat 17-20 until all specimens are imaged.
• The Flat Sheets and Packets Working Group has completed modules and associated tasks for herbarium and related collections (October 2012).
• The Pinned Things in Trays and Drawers has finished and posted its work for entomology (January 2013).
• 3D Objects in Spirits in Jars is nearing completion of its workflows for fluid-preserved specimens (May 2013).
• 3D Objects in Drawers and Trays workflows group to start work in April (June 2013).
• Preparation-independent workflows to follow (2013).
**Workflow Modules and Task Lists**

One outcome of the [BISON](https://www.bison.org) (Developing Robust Object-to-Image-to-Data) workflow workshop held in May 2012 was the establishment of a series of working groups, each focused on workflow modules and tasks for various preparation types. The first of these groups, informally called the **Flats Sheets and Packets Working Group**, was charged with fleshing out task lists for digitizing vascular and non-vascular plant collections. The second working group, **Planted Specimens in Trays and Drawers**, invested its time developing modules to support effective entomological digitization workflows. Other preservation types will follow, including fluid collections and other 3-dimensional objects, concluding with the development of an overall project management module designed to provide guidance for developing and managing digitization projects across disciplines and preservation types.

We have chosen a modular approach for presenting our results in order to accommodate the broad range of workflow implementations within the collections community. We recognize that there is no consensus workflow that fits all situations, even within a single preservation type. In light of this, we have attempted to assemble orderly, comprehensive task lists to serve as foundations from which institutionally specific workflows can be created. Not all institutions will use every task, but we hope that the lists we have developed encompass all relevant digitization tasks. We also hope that those in the collections digitization community will provide feedback on these lists, either through forum posts or e-mails to Gill Nelson, alerting us to deficiencies and oversights.

Links to published modules as they are completed are provided below:

**Flats Sheets and Packets Working Group - Vascular and Non-vascular Plants**
- Module 1 Pre-digitization Curation Tasks
- Module 2 Imaging Station Setup Camera
- Module 3 Imaging Station Setup Scanner
- Module 4 Imaging Tasks
- Module 5 Image Processing Tasks (Rev. 2012-11-07)
- Module 6 Data Capture Tasks

**Planted Things in Trays and Drawers Working Group - Insecta**
- Module 9 Generic Tasks Applicable to Two or More Modules
- Module 1 Pre-digitization Curation Tasks
- Module 2A Specimen Imaging Tasks
- Module 2B Whole-drawer Imaging Tasks
- Module 2C Label Imaging Tasks
- Module 3 Image Processing Tasks
- Module 4A Data Capture From Specimen Tasks
- Module 4B Data Capture From Image Tasks
- Module 4C Event Data Capture Tasks
- Module 5 Quality Assurance Tasks

---

**Workflow Detail: Specimen Image Processing (Pinned Things)**

<table>
<thead>
<tr>
<th>Module ID</th>
<th>Task Name</th>
<th>Explanations and Comments</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 3</td>
<td>Specimen Imaging Processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module 4A</td>
<td>Data capture from image</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module 4B</td>
<td>Data capture from specimen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module 5</td>
<td>Quality assurance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Task 1: Transfer images from camera to immediate image processing storage.

- **Task ID**: T1
- **Task Name**: Transfer images from camera to immediate image processing storage.
- **Explanations and Comments**: This task varies by institution. Some institutions record images to a card within the camera, others download directly to the imaging computer or an external network drive as images are recorded.
- **Resources**: Ample storage space with backup procedures (also see T8-T9).

### Task 2: Adjust orientation and crop images, as necessary.

- **Task ID**: T2
- **Task Name**: Adjust orientation and crop images, as necessary.
- **Explanations and Comments**: Images should be framed and recorded as precisely as possible to prevent the need for cropping. In cases where cropping is required, batch crop routines for processing multiple images to identical parameters are preferable. Where batch cropping is not possible due to random variation of exemplar image files, individual cropping may be necessary.
- **Resources**: Image management or processing software (e.g., Photoshop, Lightroom, ImageMagick, Gimp, or similar).
Digitization Workflows

Presenter: Dorothy Alland

Efficient and effective workflows are at the heart of successful biological and paleontological collections digitization. Much work has been done with developing workflows and protocols at the museum and collections level, but few of those workflows have been documented or made available to the larger collections community. iDigBio, through its Documentation pages, is establishing an online repository for sharing existing customized workflows from as many collection types and institutions as possible, an idea that stems largely from the Developing Robust Object-to-Image-to-Data (DROID) workshop held May 30-31, 2012. We have assembled an initial set of workflows, including selected examples from the DROID workshop, as well as those developed by iDigBio staff. Here we offer the beginnings of the repository and encourage those in the community to both discuss the workflows via the forum links, and to contribute to this resource by adding new workflows and updating existing workflows. If you would like to submit a workflow for inclusion on this page, please contact iDigBio for instructions.

We are also assembling detailed modules of tasks to be performed at each stage of the workflow, accessible on our Workflow Modules and Tasks page.

<table>
<thead>
<tr>
<th>Workflow</th>
<th>Contributor</th>
<th>Workflow Documentation</th>
<th>Link to Public Comments (Forums)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant Digitization Workflows</td>
<td>iDigBio</td>
<td>Dominant Digitization Workflows Documentation</td>
<td>Dominant Digitization Workflows Forum</td>
</tr>
<tr>
<td>Field Notes-to-Data-to-Image</td>
<td>iDigBio</td>
<td>Field Notes-to-Data-to-Image Documentation</td>
<td>Field Notes-to-Data-to-Image Forum</td>
</tr>
<tr>
<td>Specimen-to-Data-to-Exemplar Image</td>
<td>iDigBio</td>
<td>Specimen-to-Data-to-Exemplar Image Documentation</td>
<td>Specimen-to-Data-to-Exemplar Image Forum</td>
</tr>
<tr>
<td>Object-to-Image-to-Data (1)</td>
<td>iDigBio</td>
<td>Object-to-Image-to-Data (1) Documentation</td>
<td>Object-to-Image-to-Data (1) Forum</td>
</tr>
<tr>
<td>Object-to-Image-to-Data (2)</td>
<td>iDigBio</td>
<td>Object-to-Image-to-Data (2) Documentation</td>
<td>Object-to-Image-to-Data (2) Forum</td>
</tr>
<tr>
<td>University of Vermont Herbarium</td>
<td>Dorothy Alland</td>
<td>University of Vermont Herbarium Documentation</td>
<td>University of Vermont Herbarium Forum</td>
</tr>
<tr>
<td>Southwest Collections of Arthropods Network</td>
<td>Paul Heinrich</td>
<td>Southwest Collections of Arthropods Network Documentation</td>
<td>Southwest Collections of Arthropods Network Forum</td>
</tr>
</tbody>
</table>
Documentation and Instructions

• **Written Protocols**
  - Essential!
  - Include illustrations/screen shots.
  - Attention to detail (leave nothing to the imagination).
  - Express limits on technician authority.

• **Feedback Loops**
  - Technicians: best source of efficiency adaptations, either by show or tell.
  - Easy methods for receiving feedback.
  - Personal copies of the protocol.
  - Master copy available via Google docs or other shared storage for updates and suggestions.
Continuous Workflow Improvement

Continuously evaluate written against actual workflows.

With input from:
- Technicians
- Workflow managers
- Collections managers

With particular attention to:
- Bottlenecks
- Redundancy
- Handling time
- Varying rates of productivity
Thank you!