Digitization Modules, Tasks, and Workflows

Scripps
Gil Nelson
3 February 2016
Integrated Digitized Biocollections
Florida State University
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 Nelson, Deborah Paul, Gregory Riccardi, Austin R. Mast
Pre-digitization Curation or “Staging”

Image Capture

Data Capture

Image Processing

Image/Data Storage

Geo-referencing

Task Clusters (modules)

Personnel

Written Protocols

Biodiversity informatics Manager
Preparing Infrastructure

Develop workflows and protocols

Select and install an institutional database
- Specify
- Symbiota
- EMu
- Arctos
- Custom

Design and purchase an imaging station
- Copy stand and lighting
- Light box

Search and select imaging workflow and processing software

Prepare for digitization
Pre-digitization curation

Consider and plan for data enhancement activities
- e.g. Georeferencing
Tracks to Digitization

• **Taking the inside track** is often based on stretching the institution’s resources. Decisions are made to maximize resources available for user-initiated digitization by using solid baseline practices. The primary focus on the inside track is to get the job done quickly and to fill the user’s request.
• **Taking the middle track** has the widest range of options, standards, and results. This is the most flexible of the tracks, where decisions often fall in gray areas.
• **Taking the outside track** focuses on the collections themselves. While users may initiate digitization, it is undertaken to deliver materials to a greater public. These decisions may lead to comprehensive digitization, such as an entire book, series, or collection. The goal is to create maximum access to special collections, using preservation and archival standards. This track usually involves a level of thought and planning that is more in-depth than the fulfillment of day-to-day digitization requests.
Digitization Decision Continua that Influence Data Gaps

Current Tools ↔ Future Tools
Robustness ↔ Quantity
High cost ↔ Low cost
Fitness ↔ Speed
Image everything ↔ Image exemplars ↔ Image nothing
Ancillary materials ↔ Specimens only
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
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.
Documentation and Instructions

• **Written Protocols**
  • Essential!
  • Include screen shots and pictures.
  • 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

Develop written workflows that reflect actual practice

Continuous evaluation of written and actual workflows by:

- Technicians
- Workflow managers
- Collections managers

With particular attention to:

- Bottlenecks
- Redundancy
- Handling time
- Varying rates of productivity
Imaging Equipment List

- Copy stand
- Camera
- Lens
- Cables
  - AC adapter
  - Long USB cable
- Computer
- Software
  - Camera control software
  - Image management software
- Lights OR Photo-eBox BIO
- Miscellaneous
Camera
Canon - Nikon

Best quality you can afford

Resolution
Full frame vs. cropped frame
Imaging station
Unequalled opportunity for completeness and detail
Photo eBox
Photo eBox Plus
Helicon Focus
Workflow Modules and Task Lists

One outgrowth of the DROID (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 Flat Sheets and Packets Working Group, was charged with fleshing out task lists for digitizing vascular and non-vascular plant collections. The second group, Pinned Specimens in Trays and Drawers, is investing its time developing modules to support effective entomological digitization workflows. Other preservation types will follow, 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.
Workflow Modules and Task Lists

One outgrowth of the DROID (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 Flat Sheets and Packets Working Group, was charged with fleshing out task lists for digitizing vascular and non-vascular plant collections. A reconstitution of this working group, convened in January 2015, added 8 modules to this set of workflows and updated the existing ones. The second working group, Pinned Specimens in Trays and Drawers, invested its time developing modules to support effective entomological digitization workflows. Things in Jars devotes time to workflows for fluid-preserved collections. The 3D Objects in Trays and boxes completed its work in spring 2015 and focused mostly on paleontological specimens.

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 Gil Nelson, alerting us to deficiencies and oversights.

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

Flat Sheets and Packets Working Group - Vascular and Non-vascular Plants

- Module 1 Pre-digitization Curation Tasks
- Module 2 Selecting Components for an Imaging Station
- Module 3 Imaging Station Setup Camera/Copy Stand
- Module 4 Imaging Station Setup Light box
- Module 5 Image Station Setup Scanner
- Module 6 Imaging

Researchers
Browse our specimen portal

Collections Staff
Learn how your collection can benefit from our work

Teachers & Students
Learning resources & opportunities to engage
# Workflow Detail: Pre-digitization Curation (for flat sheets and packets)

## Module 1: Pre-digitization Curation Task List

<table>
<thead>
<tr>
<th>Task ID</th>
<th>Task Description</th>
<th>Explanations and Comments</th>
<th>Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Apply storage locator barcodes to storage locations (rooms, cabinets, shelves, folders, drawers, etc.).</td>
<td>Most useful when systematically digitizing an entire collection. Otherwise potentially helpful with herbarium inventory. May be less helpful for collections that are digitizing in random order or only portions of the collection related to specific projects, or with significant separation between the pre-digitization curation, databasing, and image capture modules.</td>
<td>Barcodes, QRcode, DataMatrix.</td>
</tr>
<tr>
<td>T2</td>
<td>Select specimens to digitize.</td>
<td>For herbaria, this often includes all specimens. Where this is not the case, selection should follow the institution’s predetermined digitization policies or project management plan.</td>
<td>Digitization policy manual or project management plan.</td>
</tr>
<tr>
<td>T3</td>
<td>Associate/insert machine readable barcodes/documents with/into folders.</td>
<td>Some institutions create machine readable documents to gather data at the cabinet and/or folder level. Documents might contain such information as family, higher geography, and current identification (“live-as-name”). These data will be read and associated with individual collection records in Module 4, T1 or Module 7. Tasks T2 or T3 might also include determining whether specimens are out on loan or...</td>
<td>QRcodes, DataMatrix, 1D barcode, or OCR-readable documents for insertion into specimen folders.</td>
</tr>
</tbody>
</table>
This folder contains the original set of 14 workflow modules published with the paper Digitization workflows for flat sheets and packets of plants, algae, and fungi, Nelson, G. P., Sweeney, L. E. Wallace, R. K. Rabeler, D. Allard, H. Brown, J. R. Carter, et al., Applications in Plant Sciences 3(9): 1500065. doi:10.3732/apps.1500065 (http://www.bione.org/doi/pdf/10.3732/apps.1500065). Files in this folder are linked to the published paper and will not be edited or revised. Future revisions will be stored in a separate directory. PDF and Word versions are provided in separate folders.

These workflows are being distributed under a Creative Commons CC BY license.