

Digitizing Dried Insects Workshop

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

24 April 2013

Field Museum

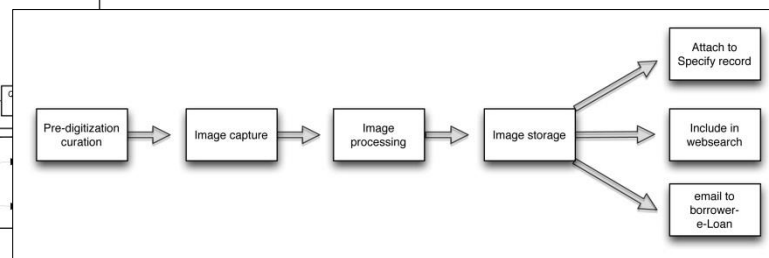
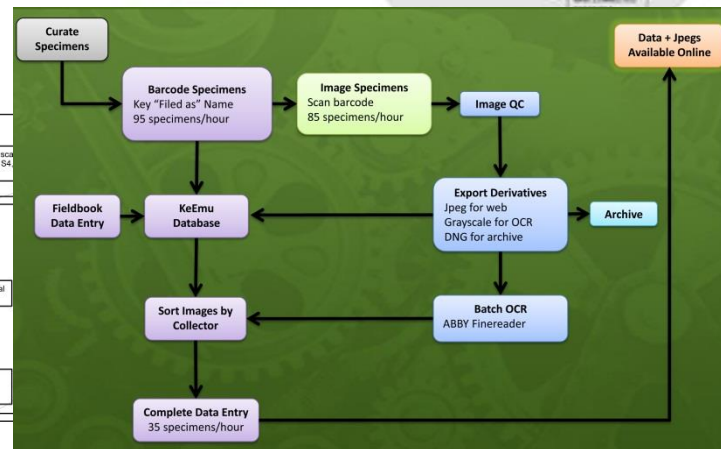
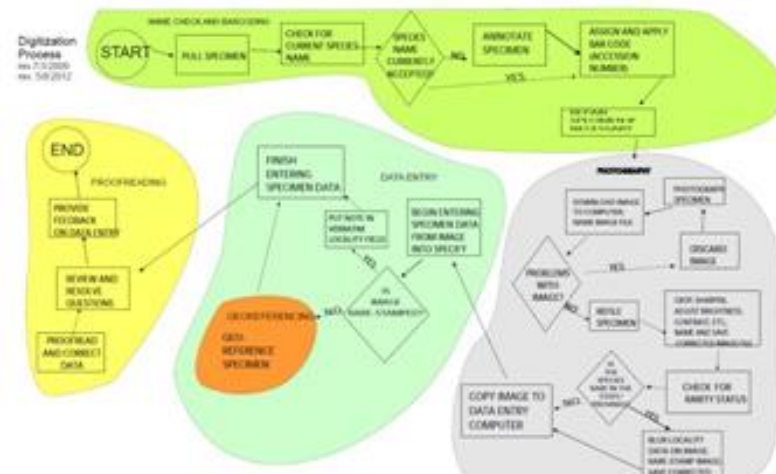
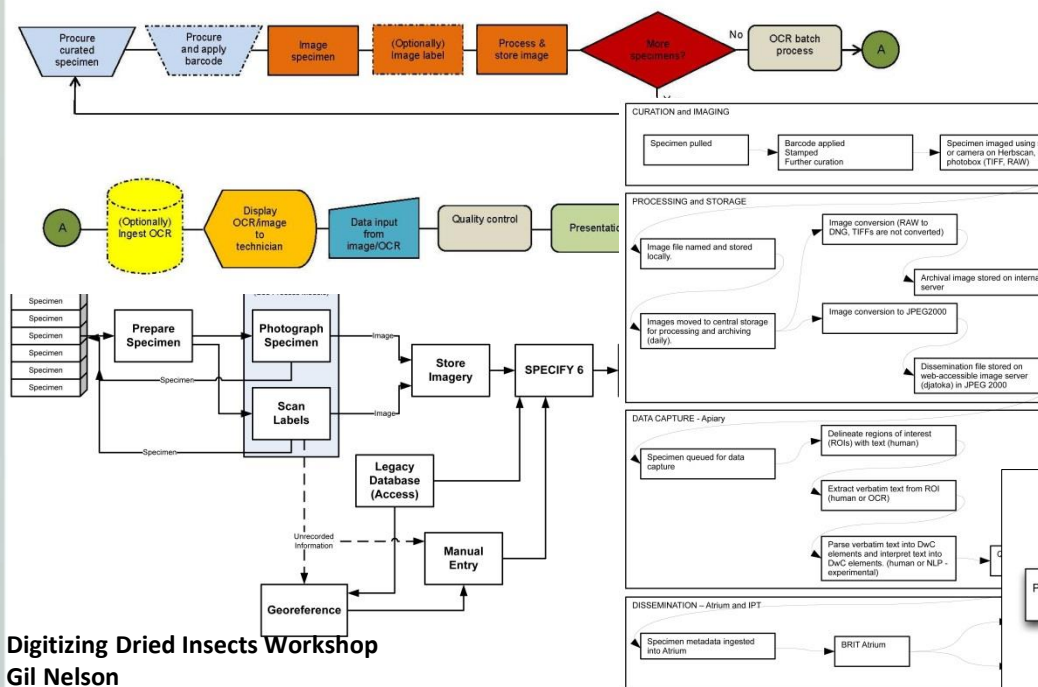
Chicago



This material is based upon work supported by the National Science Foundation under Cooperative Agreement EF-1115210. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



Digitization Workflows



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



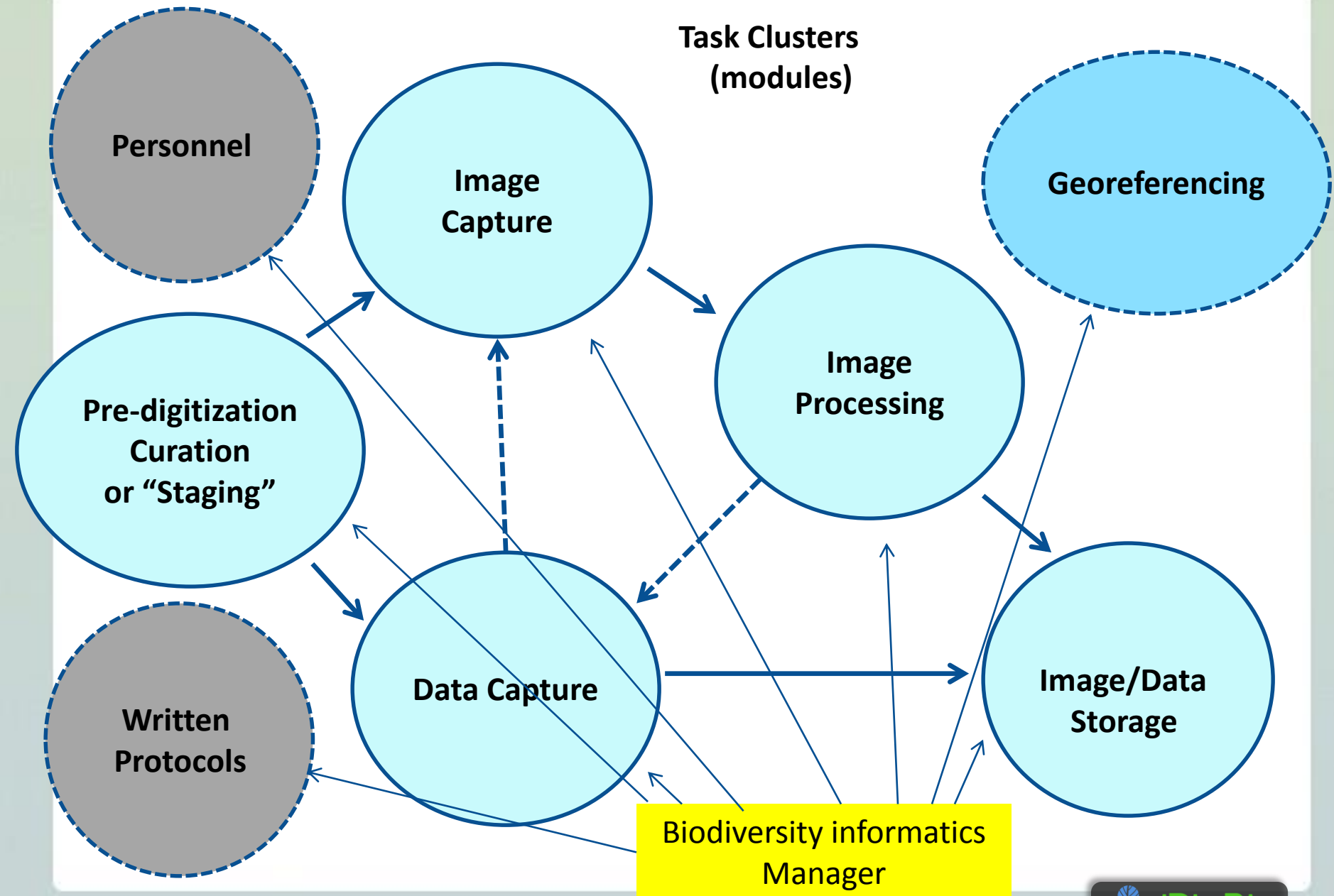
Five task clusters that enable efficient and effective digitization of biological collections

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)

**Task Clusters
(modules)**

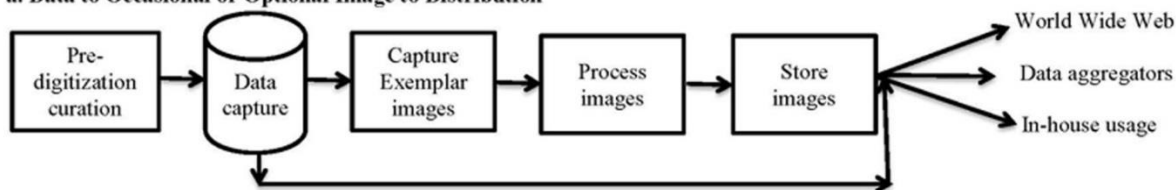


Workflows Patterns Observed

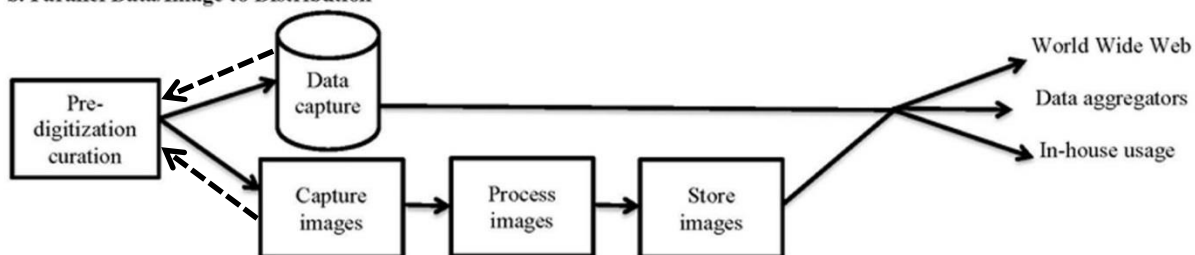
Dominant Digitization Patterns Observed

Figure 1: Dominant Digitization Workflows

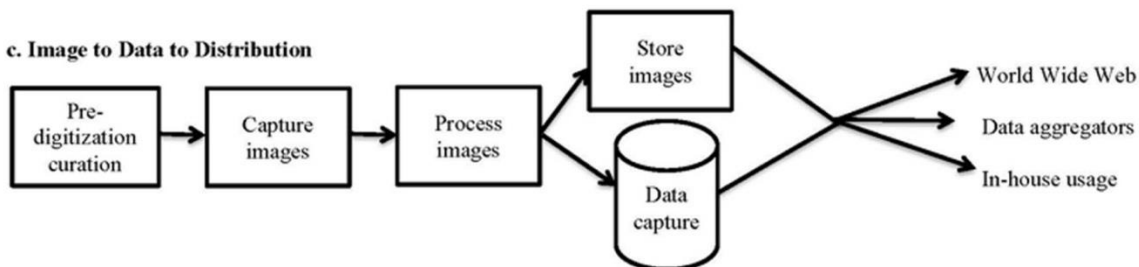
a. Data to Occasional or Optional Image to Distribution



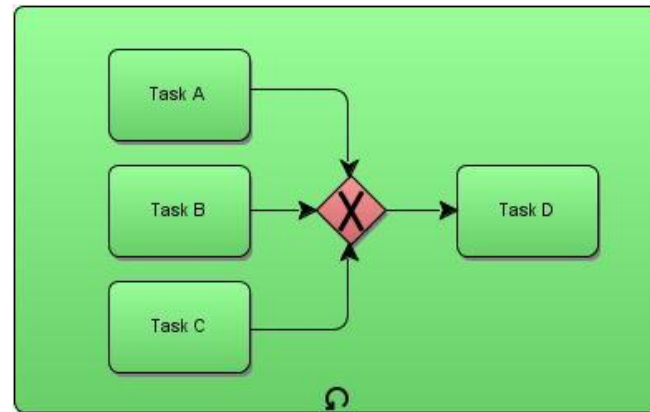
b. Parallel Data/Image to Distribution



c. Image to Data to Distribution



Values of defined workflows



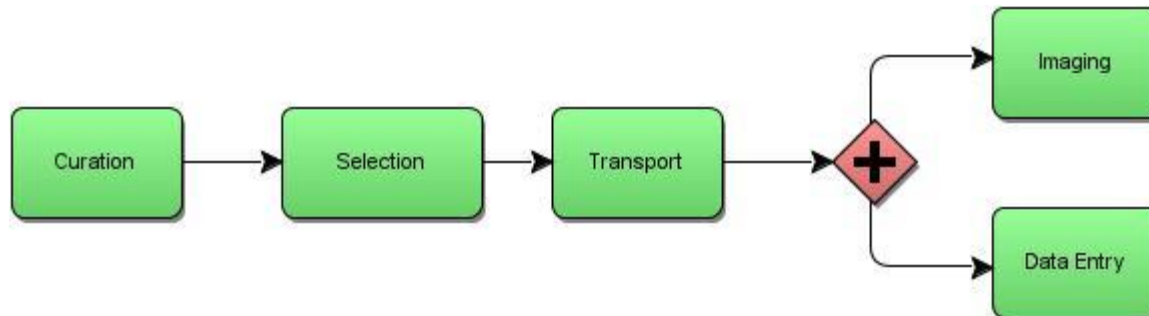
- Promote efficiency and automation of processes
- Facilitate routing and scheduling of activities
- Provide for balancing workloads
- Ensure that processes are visible and predictable
- Allow for escalations and notifications
- Enhance tracking of tasks
- Foster collaboration of all parties involved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

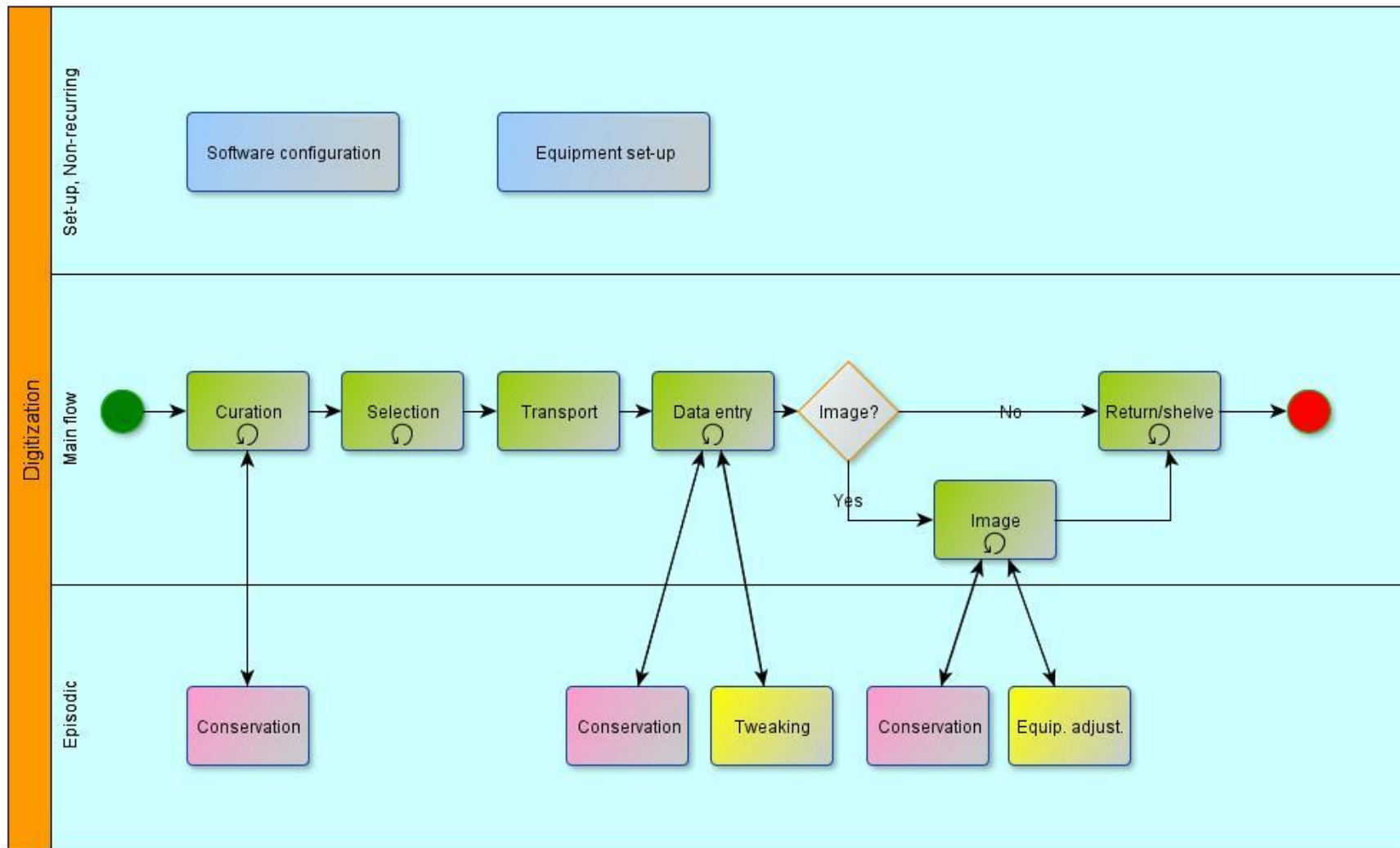
Pre-planning a Workflow Process

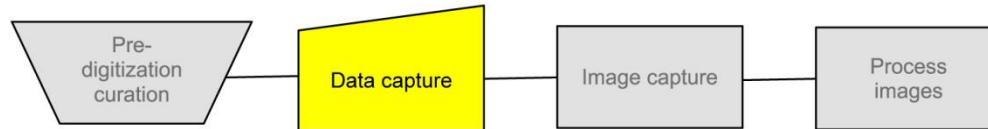
- Identify the database management system and imaging equipment to be used.
- Identify the processes/modules for which workflows will be designed.
- Identify (*in excruciating detail!*) the tasks (or task clusters) that constitute the processes/modules.
- Identify the specific actions to be taken and 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

Process	Cycle	Dependency
Software configuration	Once/non-recurring	
Equipment set-up	Once/non-recurring	
Specimen curation	Recurring	
Specimen selection	Recurring	Pre-digitization curation
Specimen transport	Recurring	Specimen selection, imaging, data entry
Conservation	Episodic	Curatorial processes, imaging, data entry
Data entry	Recurring/tasks iterative	Specimen transport
Imaging	Recurring/tasks iterative	Specimen transport
Equipment adjustments	Episodic	Technician escalation
Software update/tweaking	Episodic	Technician escalation
Specimen return/shelving	Recurring	Imaging or data entry







Guiding Principles

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.

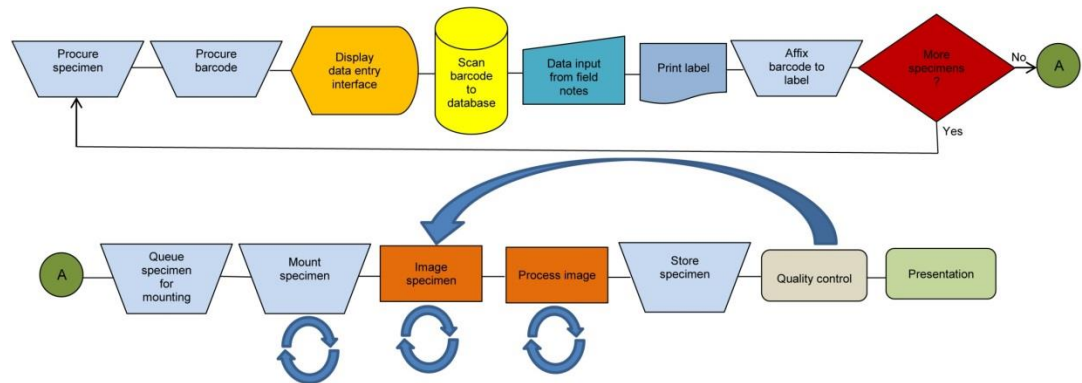
1. Open Capture NX2 and View NX2.
2. Open Camera Control Pro 2.
3. Open default.ncc as settings file:
 - Settings->Load Control Settings
 - >My Documents->CameraSettings->default.ncc.
4. Create a folder in X:\SpecimenImages\NEF, using the current date as the folder name, as 2013-04-14.
5. Retrieve next 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.
16. Close Live View.
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.

Workflows Working Groups

- 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).

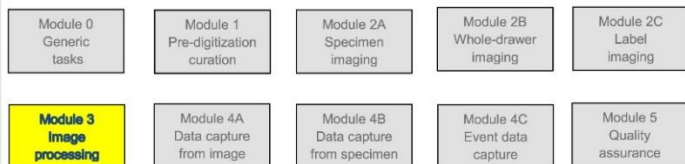
FN2D2I—New Specimen Workflow: Field notes to data to image

This workflow is designed for actively growing collections in which new specimens are regularly added. Collectors, especially in herbaria, typically keystroke label data from field notes, store the label with the specimen, and queue the specimen for mounting. Following mounting, the specimen is treated as an existing specimen with the data entered into the database by a technician, who re-keys the data previously keyed by the collector. The workflow proposed here eliminates the second keying of label data by capturing label data into the database as the label is prepared, allowing the label to be printed from the database immediately following data entry. The workflow assumes a database management system with functionality for printing labels, as well as a strategy that includes the application of bar codes to the newly printed label rather than to the specimen sheet.



Modular Approach

Workflow Detail: Specimen Image Processing (Pinned Things)



Module 3: Specimen Image Processing

Task ID	Task Name	Explanations and Comments	Resources
T1	Transfer images from camera to immediate image processing storage.	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 or network drive as images are recorded. Transfer to the image processing storage should be periodic, at least daily.	Ample storage space with backup procedures (also see T8-T9).
T2	Adjust orientation and crop images, as necessary.	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	Image management or processing software (e.g., Photoshop, Lightroom, ImageMagick, Gimp, or similar).

University of Florida • Florida Museum of Natural History • Dickinson Hall (Museum Rd. & Newell Dr.) • Gainesville, FL 32611 • 352-273-1906
iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (#EF1115210)

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 working group, [Pinned 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 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 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](#)

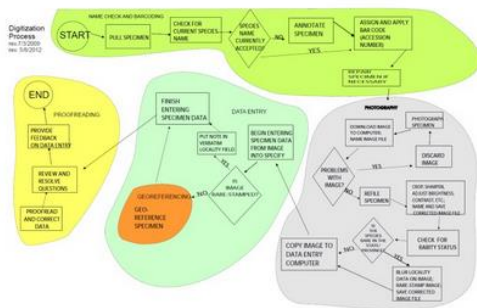
[Pinned Things in Trays and Drawers Working Group - Dried Insects](#)

- [Module 0 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 Image Tasks](#)
- [Module 4B Data Capture From Specimen Tasks](#)
- [Module 4C Event Data Capture Tasks](#)
- [Module 5 Quality Assurance Tasks](#)

Digitization Workflows

Presenter: Dorothy Allard

Digitization Workflows



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 these 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](#).

Global Plants Initiative, U. of Vermont

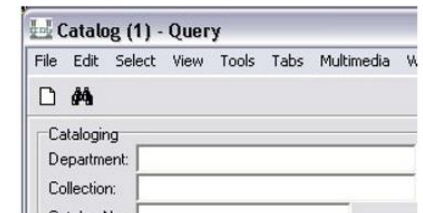
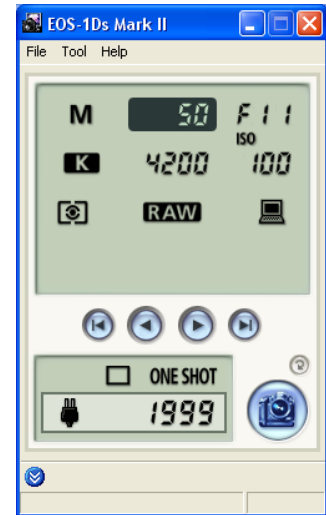
Workflow	Contributor	Workflow Documentation	Link to Public Comments (Forums)
Dominant Digitization Workflows	iDigBio	Dominant Digitization Workflows Documentation	Dominant Digitization Workflows Forum
Field Notes-to-Data-to-Image	iDigBio	Field Notes-to-Data-to-Image Documentation	Field Notes-to-Data-to-Image Forum
Specimen-to-Data-to-Exemplar Image	iDigBio	Specimen-to-Data-to-Exemplar Image Documentation	Specimen-to-Data-to-Exemplar Image Forum
Object-to-Image-to-Data (1)	iDigBio	Object-to-Image-to-Data (1) Documentation	Object-to-Image-to-Data (1) Forum
Object-to-Image-to-Data (2)	iDigBio	Object-to-Image-to-Data (2) Documentation	Object-to-Image-to-Data (2) Forum
University of Vermont Herbarium	Dorothy Allard	University of Vermont Herbarium Documentation	University of Vermont Herbarium Forum
Southwest Collections of Arthropods Network	Paul Heinrich	Southwest Collections of Arthropods Network Documentation	Southwest Collections of Arthropods Network Forum

Posted To Collaborative Workflows Page Linked to the Digitization Resources Wiki



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!