



Digitization Modules, Tasks, and Workflows

Biological Collections Digitization in the Pacific

Gil Nelson

25-27 March 2014

Institute for Digital Information and Scientific Communication

Florida State University



Preparing Infrastructure

Workflows and protocols

Selecting and installing a database

Specify

Symbiota

Custom

Design and purchase an imaging station

Copy stand and lighting

Light box

Search and select imaging workflow and processing software

Preparing for digitization

Pre-digitization curation

Consider and plan for data enhancement activities

Georeferencing

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



A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Acknowledgments

American Museum of Natural History

Botanical Research Institute of Texas

Florida Museum of Natural History

Florida State University

Harvard Herbarium

Museum of Comparative Zoology (Harvard)

New York Botanical Garden

Southeast Regional Network for Expertise and Collections

Specify Software Project (University of Kansas)

Symbiota Software Project (Arizona State University)

Tall Timbers Research Station and Land Conservancy

Tulane University Museum of Natural History

University of Kansas Insect Museum

Valdosta State University

Yale Peabody Museum

Processes that have gained definition and currency in digitization workflows

- Linking genomic and other data to vouchers
- Crowd sourcing and public participation
- Remote annotation of specimen records
- Using digitized data for research
- Optical Character Recognition

Commonalities across task clusters provide the organizational paradigm for this workshop.

A focus on distinct processes closely associated with particular preparation or collection types.

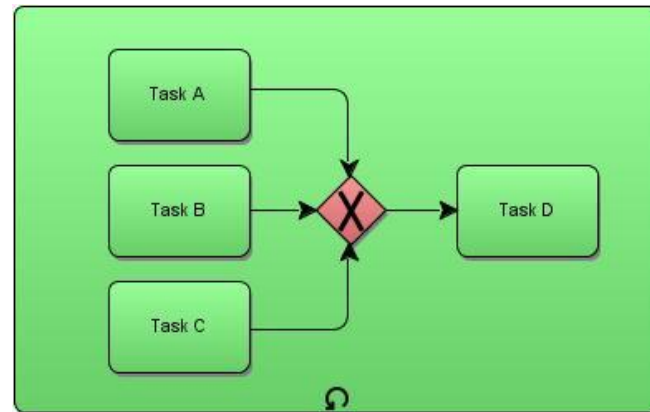
Vs.

A focus on processes common across disparate preparation and collection types to foster serendipitous discoveries and knowledge transfer across domains.

A Few Examples

Fish vs. fluid preserved arthropods
Broader applications for whole-drawer digitization
Insect soups and unsorted fossils
Flat sheets, packets, invertebrate paleontology, and Odonates
Georeferencing
Imaging workflow software, e.g. Adobe Lightroom

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

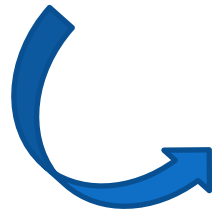
Global continua
guiding digitization

Emphasis in



Local decisions
and policies

Implementation in



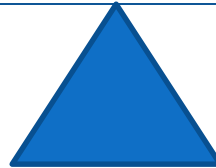
Specific
workflows

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.

Long view

Short view



Taking the long view means developing doable, effective, and sustainable strategies for balancing long term goals with short term constraints, including a commitment to implementing future enhancements.

Pressures mitigating the long view

So much data, so little time.

Our collections are not getting smaller.

The funding agencies have high output expectations.

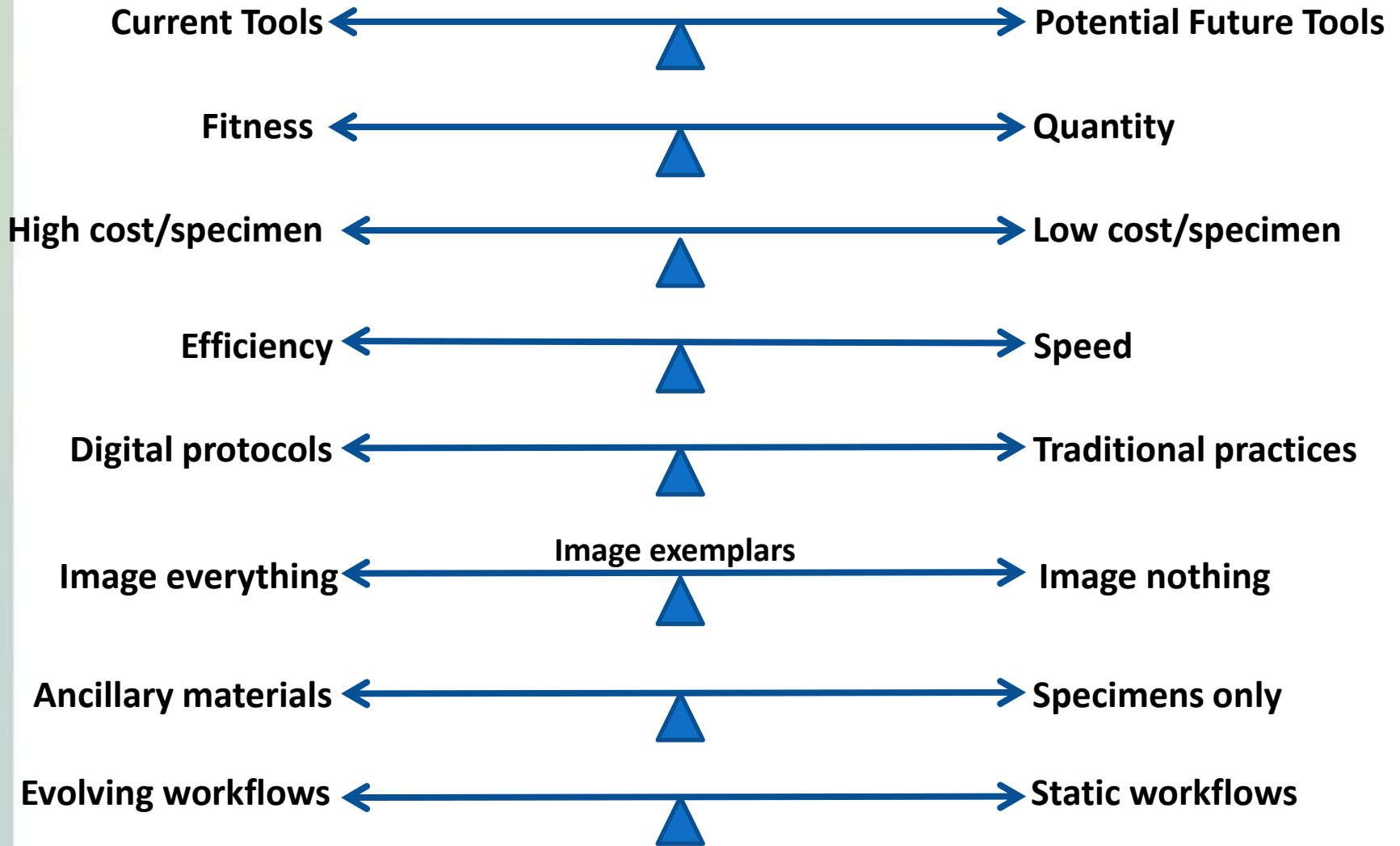
We only have 3 years to get this done.

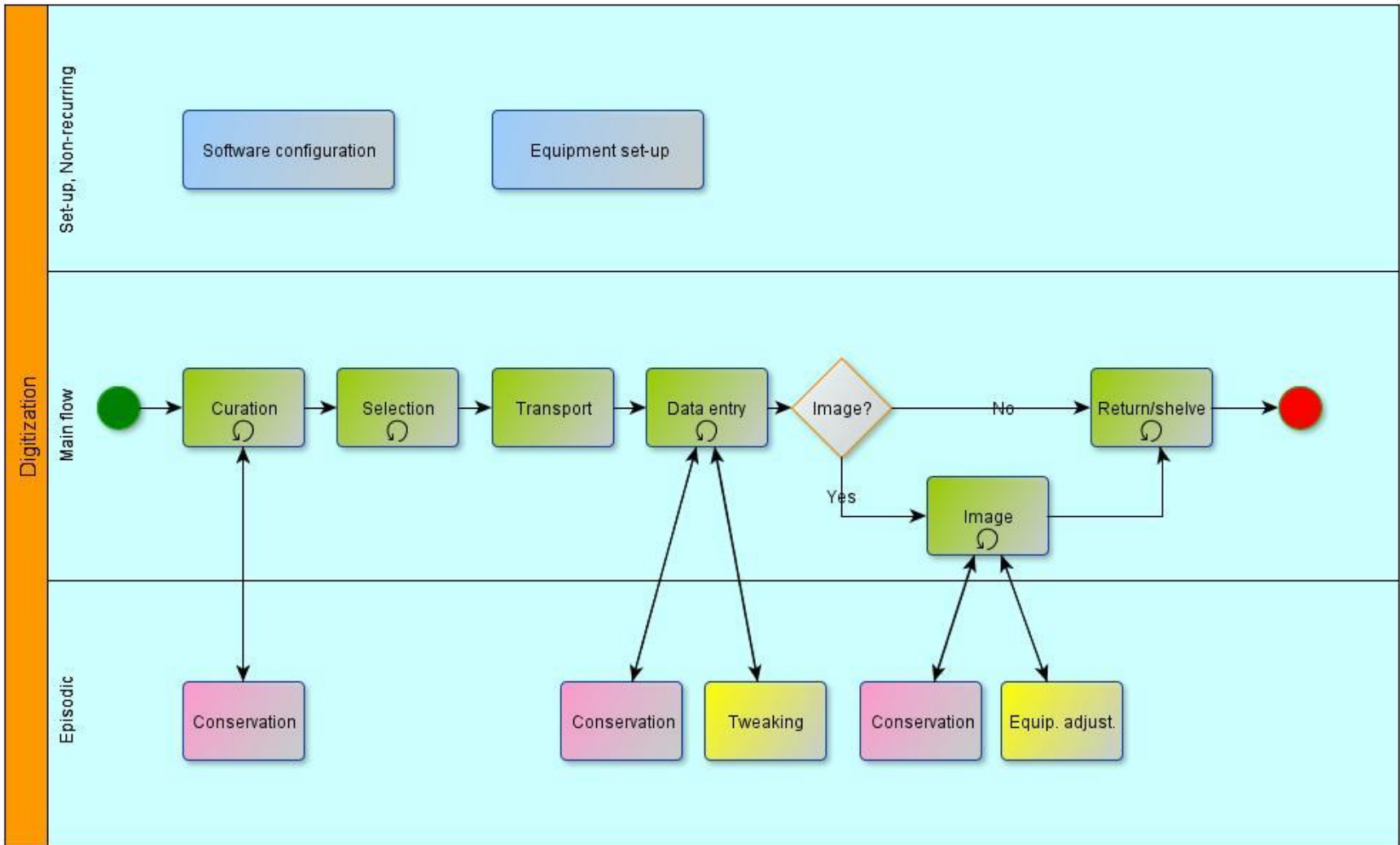
All of our data and all of our specimens are important.

Let's just use the images!

We'll do the minimum now and enhance it later.

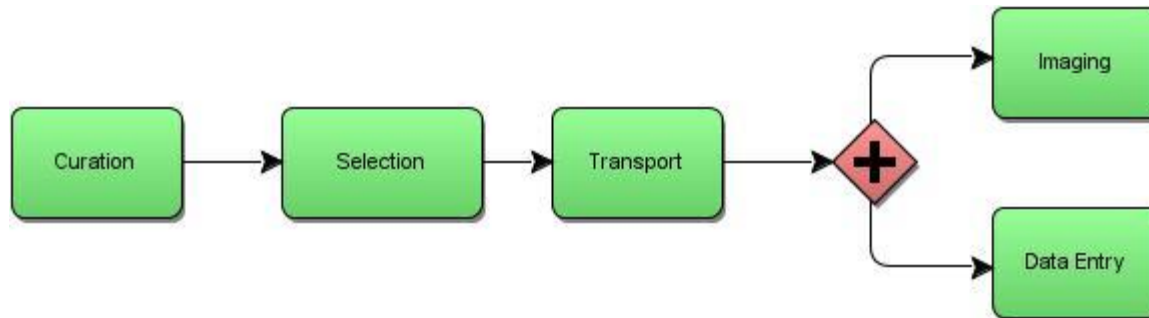
Global Digitization Continua

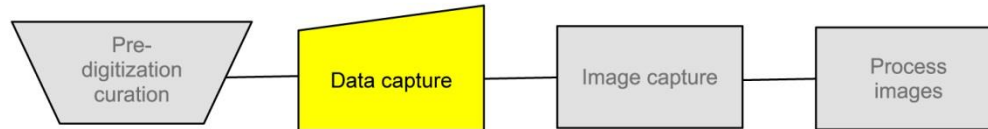




Example Processes (Modules), their Cycles and Dependencies

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 adjustment	Episodic	Data entry/imaging
Software update/tweaking	Episodic	QC
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

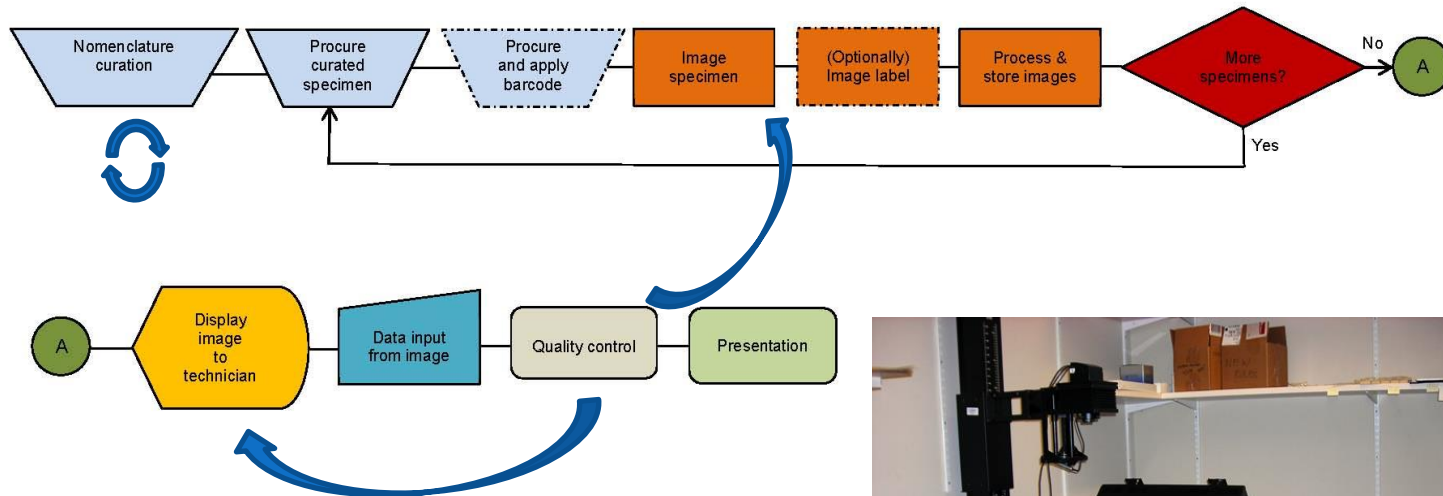
- Adjust to strengths of each technician--using students and volunteers requires flexibility in role assigned to personnel rather personnel assigned to role.

Create task lists

- Complete.
- Clear.
- Succinct.
- Ordered.
- Reusable.

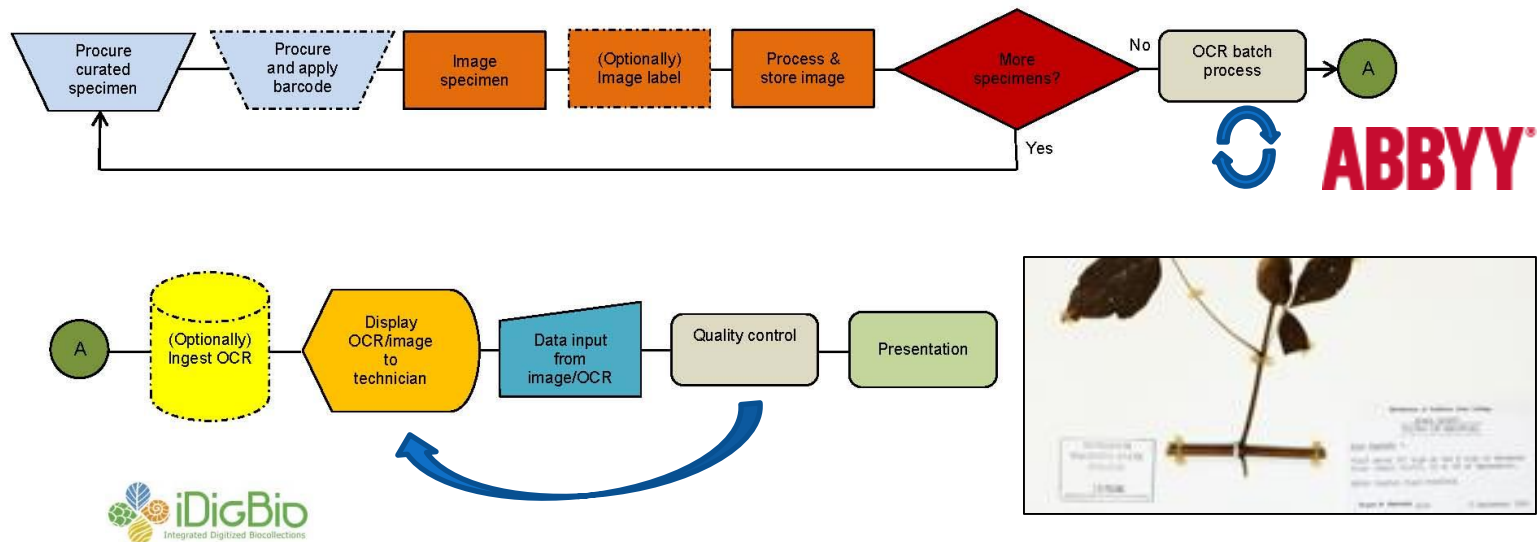
O2I2D(2)—Existing Specimen Workflow: Object to Image to Data

This workflow is designed for capturing images of existing specimens and using these images as the basis for data capture. Depending upon preparation type, barcodes are sometimes applied inline as the step immediately previous to imaging (shown optionally below) and other times en masse within an independent step during which several dozen or several hundred barcodes are applied in preparation for imaging. Pre-digitization curation and annotation is particularly important in this workflow to ensure that the current nomenclature to be used in data entry is obvious and clearly visible in the image.



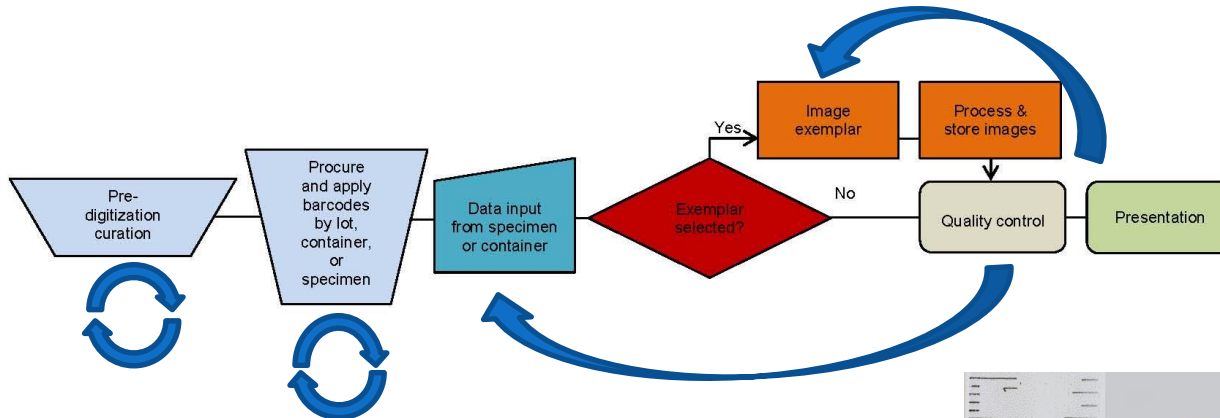
O2I2D(1)—Existing Specimen Workflow Using Optical Character Recognition: Object to Image to Data

This workflow is designed to capture images of existing specimens, pass the images through optical character recognition (OCR) software, and use the combination of image and OCR output to capture data. There are variations on this workflow. For example, depending on preparation type, barcodes are sometimes applied inline as the step immediately previous to imaging (shown optionally below) and other times en masse within an independent step during which several dozen or several hundred barcodes are applied in preparation for imaging. OCR may also occur in various ways: 1) in batch (as shown below), with numerous images being processed following the close of one or more imaging sessions, 2) "on the fly" as a record and its associated image are loaded for data entry, or 3) one image at a time as a step immediately following the imaging of each specimen. OCR output may be ingested into a field in the database (shown optionally below), stored as individual text files within the computer's file system, or virtually processed at the time the image is presented to the data entry technician. The presentation of images and OCR to data entry technicians occurs in a single interface in which database fields, OCR output, and specimen image are simultaneously visible. Pre-digitization curation and annotation is particularly important in this workflow to ensure that the current nomenclature to be used in data entry is obvious and clearly visible in the image and/or OCR output.



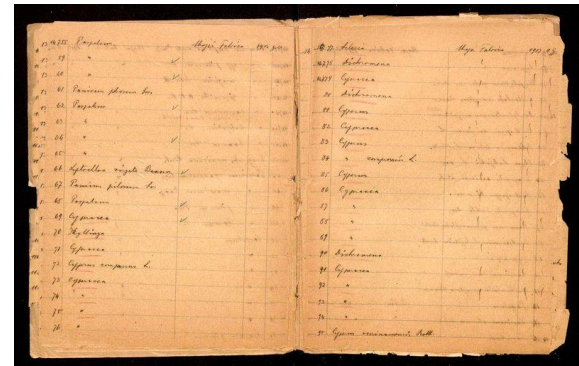
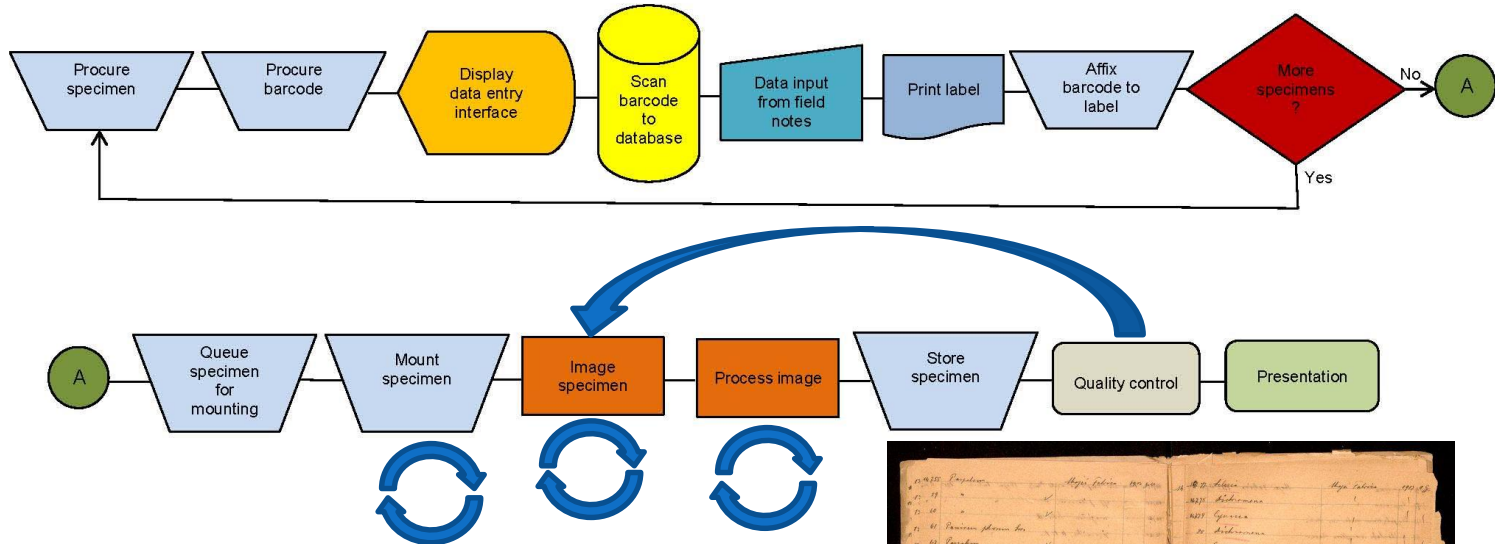
O2D2EI—Existing Specimen Workflow: Object to Data to Exemplar Images

This workflow is in use for collections that capture data in specimen lots, collecting events, taxon container, or other aggregates, but capture images only for exemplar specimens. Data capture is effected from specimen labels. Depending upon preparation type, barcodes are usually applied inline—often to the containing tray or container—as the step immediately preceding data entry. Hence, barcodes may designate a single specimen or an aggregate of specimens, such as a unit tray within an insect drawer or ethanol-filled container in a wet collection. Barcode application is executed prior data entry and image capture usually follows data entry. Pre-digitization curation, including nomenclatural annotations and specimen organization, is usually important in this workflow.



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

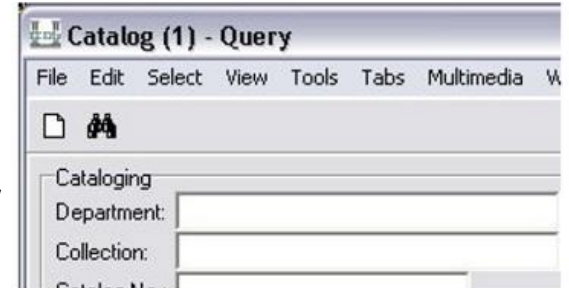
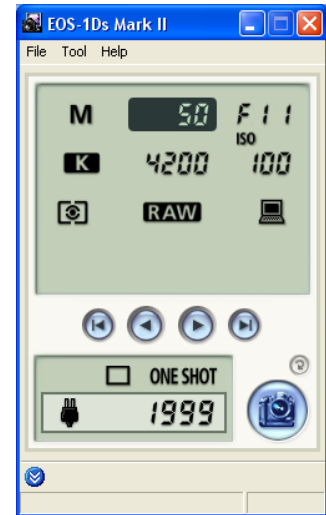
This workflow is designed for actively growing collections in which new specimens are regularly added. Collectors, especially in herbaria, typically keystroke label data from field notes, store the label with the specimen, and queue the specimen for mounting. Following mounting, the specimen is treated 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.





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.



Idigbio.org->Resources->Documentation->Workflow Modules and Task Lists

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.

[read more](#)

Workflow Modules and Task Lists



collections. The second working group, [Pinned Specimens in Trays and Drawers](#), invested its time developing modules to support effective entomological digitization workflows. [Things in Spirits in Jars](#) devoted time to workflows for fluid-preserved collections. 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](#)

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

Google Custom Search X

« November »

S	M	T	W	T	F	S
						1 2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

Top Content Tags

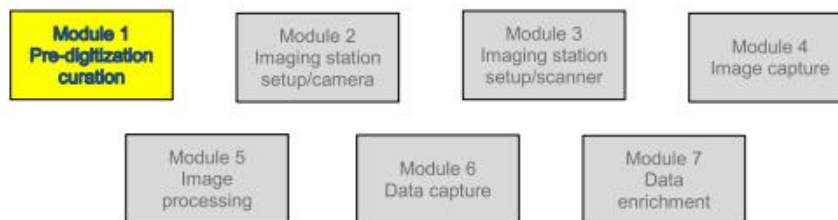
[Education & Outreach](#) (55) [Blog](#) (54)
[Workshop](#) (41) [Digitization](#) (36)
[Featured](#) (32) [workflow](#) (16)
[Documentation](#) (14) [Press Releases](#)
 (12) [Databasing](#) (8) [imaging](#) (8)
[More](#)

My Top Resources

[Adobe Connect](#)
[File Browser](#)
[Public Wiki](#)
[Redmine](#)
[My Redmine Tickets](#)



Workflow Detail: Pre-digitization Curation (for flat sheets and packets)



Module 1: Pre-digitization Curation Task List

Task ID	Task Description	Explanations and Comments	Resources
T1	Apply storage locator barcodes to storage locations (rooms, cabinets, shelves, folders, drawers, etc).	<p>Most useful when systematically digitizing an entire collection. Otherwise potentially helpful with herbarium inventory.</p> <p>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.</p>	Barcodes, QRcode, DataMatrix.
T2	Select specimens to digitize.	For herbaria, this often includes all specimens. Where this is not the case, selection should follow the institution's pre-determined digitization policies or project management plan.	Digitization policy manual or project management plan.
T3	Associate/insert machine readable barcodes/documents with/into folders.	<p>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 ("filed-as name"). These data will be read and associated with individual collection records in Module 4, T1 or Module 7.</p> <p>Tasks T2 or T3 might also include determining whether specimens are out on loan or</p>	QRcodes, DataMatrix, 1D barcode, or OCR-readable documents for insertion into specimen folders.

Collaborative Notes

- [Collaborative Notes Doc](#)

Workflow Documents

- Florida State University Herbarium Imaging Protocol
- Valdosta State University Herbarium (VSC) Vascular Plant Imaging Protocol
- Valdosta State Herbarium (VSC) Bryophyte Packet Imaging Protocol
- Valdosta Herbarium image processing with Nikon Dust Off process included
- Increasing the efficiency of digitization workflows for herbarium specimens, Tulig M, Tarnowsky N, Bevans M, et al



iDigBio's Flat Sheets and Packets Working Group Workflows

- Module 1 Pre-digitization Curation Tasks
- Module 2: Imaging station setup for camera stations
- Module 3: Imaging station setup for scanners
- Module 4: Imaging tasks
- Module 5: Image processing
- Module 6: Data capture

Imaging Equipment

- NYBG Herbarium Imaging and Equipment Specifications
- Nikon cameras and a related copy stand and lights

Imaging Procedures and Workflows

- NYBG: Standardized Digital Imaging and Archiving Procedures, Mike Bevans

Image Processing

- NYBG Image editing guidelines
- iDigBio's Recommendations for the Acquisition, Processing, Storage, and Distribution of Digital Images

Related Articles and Papers

- Streamlining Collaborative Digitization, Tulig & Watson, NYBG

Barcodes

- LBCC barcode document

Sources for Barcodes

- Computype 2285 West County Road C St. Paul, MN 55113 (800) 328-0852
- Watson Label Products 3884 Forest Park Blvd St Louis, Mo 63108 (314) 652-6715 (800) 678-6715
- University Products blank archival labels

Retrieved from "https://www.idigbio.org/wiki/index.php/Small_Herbarium_Workshop_FSU"

This page was last modified on 20 November 2012, at 12:49

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



iDigBio

Integrated Digitized Biocollections

