Digitization Modules, Tasks, and Workflows

Vertebrate Digitization Workshop
Gil Nelson
4-6 April 2016
Museum of Vertebrate Zoology
University of California, Berkeley

Preparing Infrastructure

Develop workflows and protocols

Select a database
Specify, Symbiota, Arctos, Emu, Custom

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

Select imaging workflow and processing software
Lightroom, GIMP, Canon Digital Pro, Nikon Capture NX2, Nikon Camera Control Pro,
Helicon Focus, Zerene, DigiCamControl

Preparing for digitization
Pre-digitization curation

Consider and plan for data enhancement activities
Georeferencing

iDigBio
Integrated Digitized Biocollections

Assessing Digitization Practices in Biological and Paleontological Collections

28 Collections
10 Museums
Spanning biological and paleontological collections
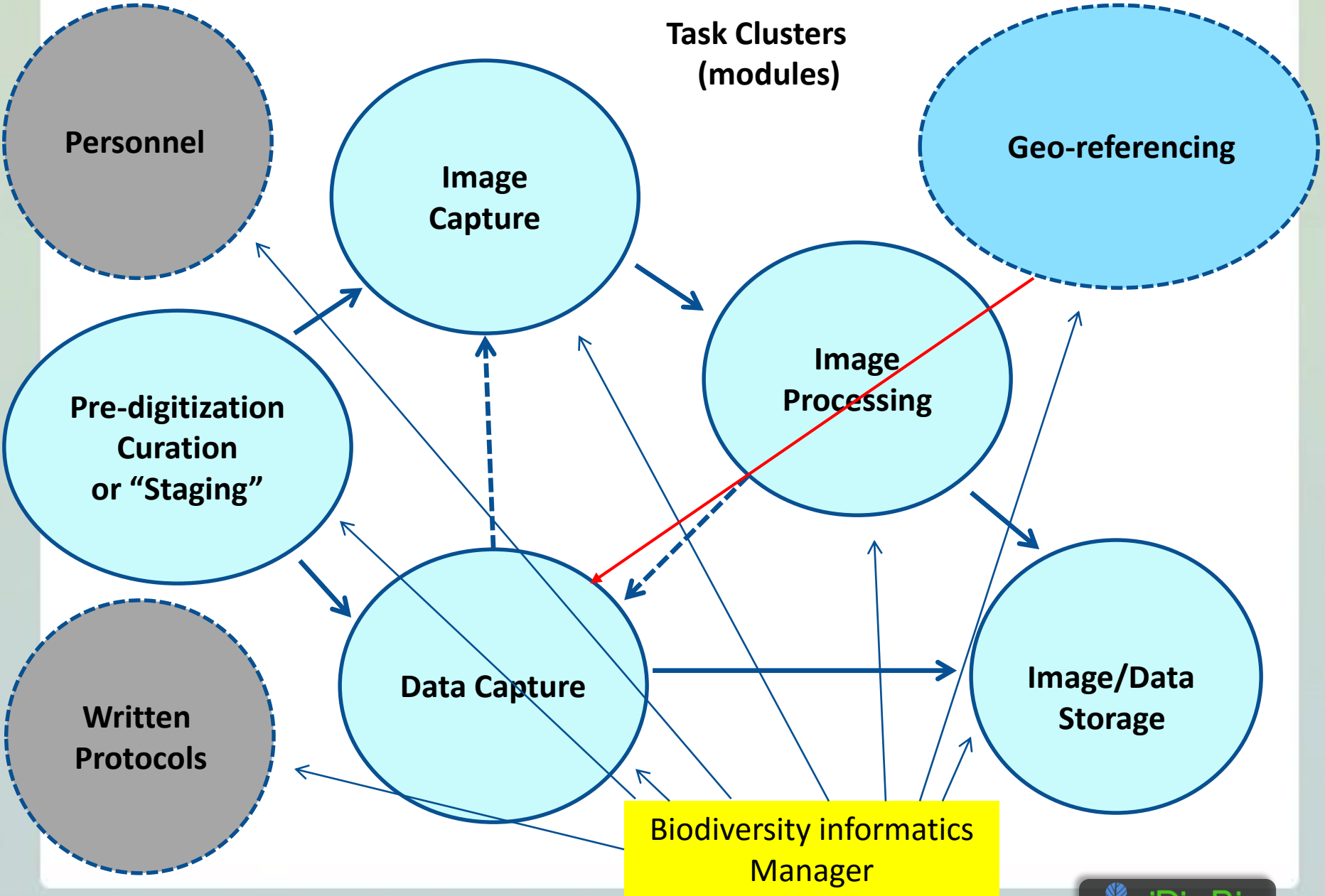Insects and other invertebrates, plants, birds, mammals
Wet, dry

Five task clusters that enable efficient and effective digitization of biological collections
Gil Nelson, Deborah Paul, Gregory Riccardi, Austin R. Mast

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

iDigBio
Integrated Digitized Biocollections

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

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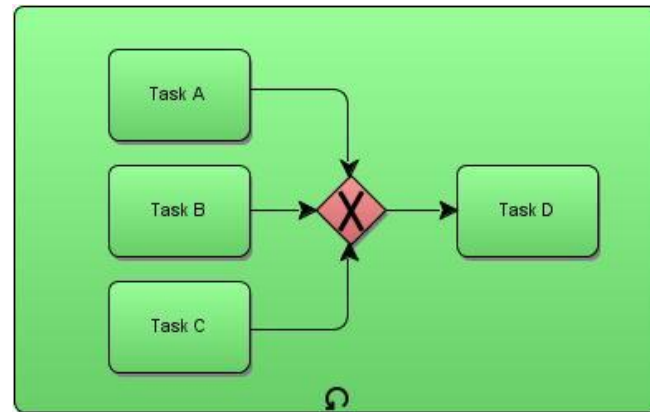
- Promote efficiency and automation of processes
- Facilitate routing and scheduling of activities
- Provide for balancing workloads
- Ensure that processes are visible and predictable
- Allow for escalations and notifications
- Enhance tracking of tasks
- Foster collaboration of all parties involved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

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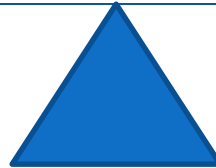
Taking the long view means developing doable, effective, and sustainable strategies for balancing long term goals with short term constraints, including a commitment to implementing future enhancements.

Pressures mitigating the long view
So much data, so little time.
Our collections are not getting smaller.
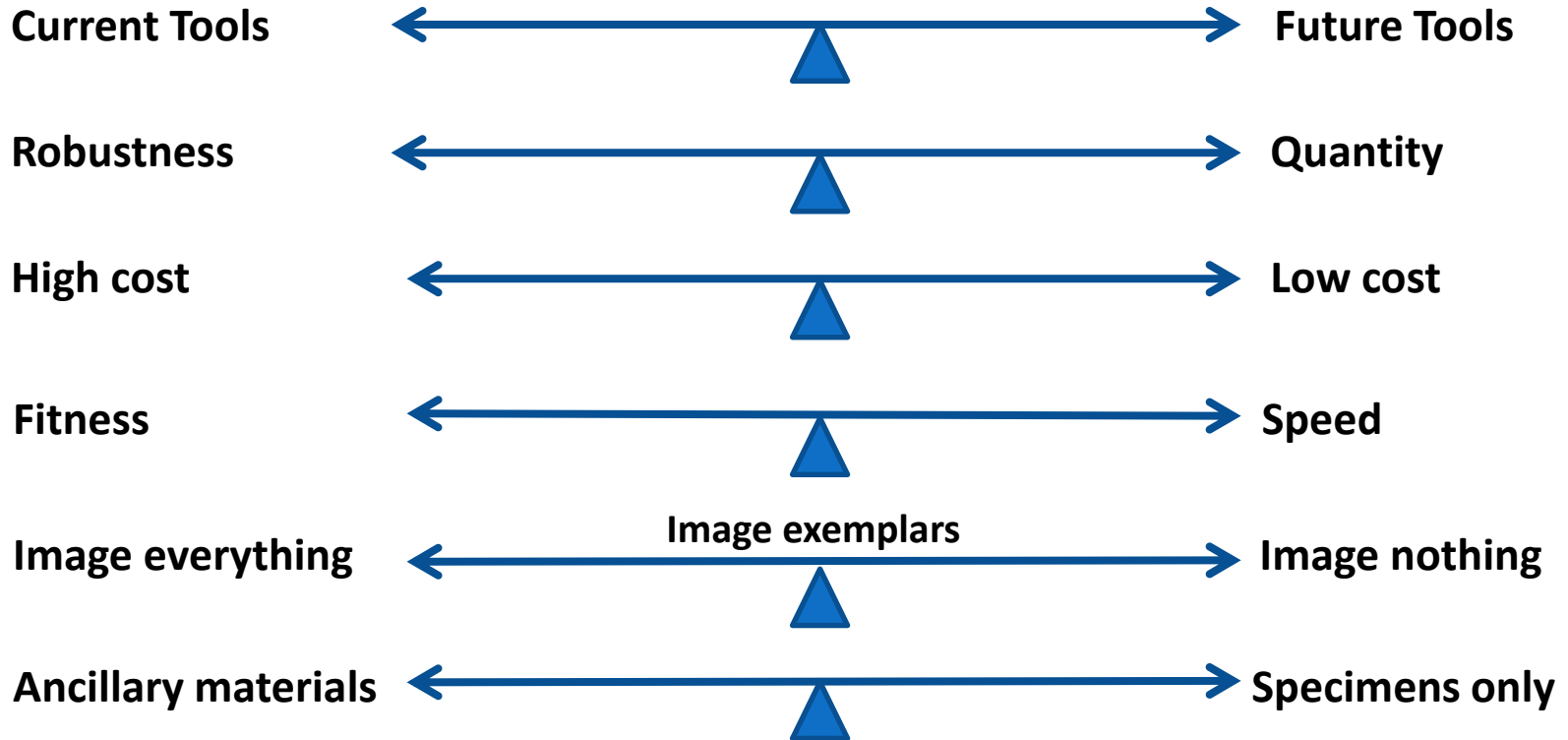The funding agencies have high output expectations.
We only have 3 years to get this done.
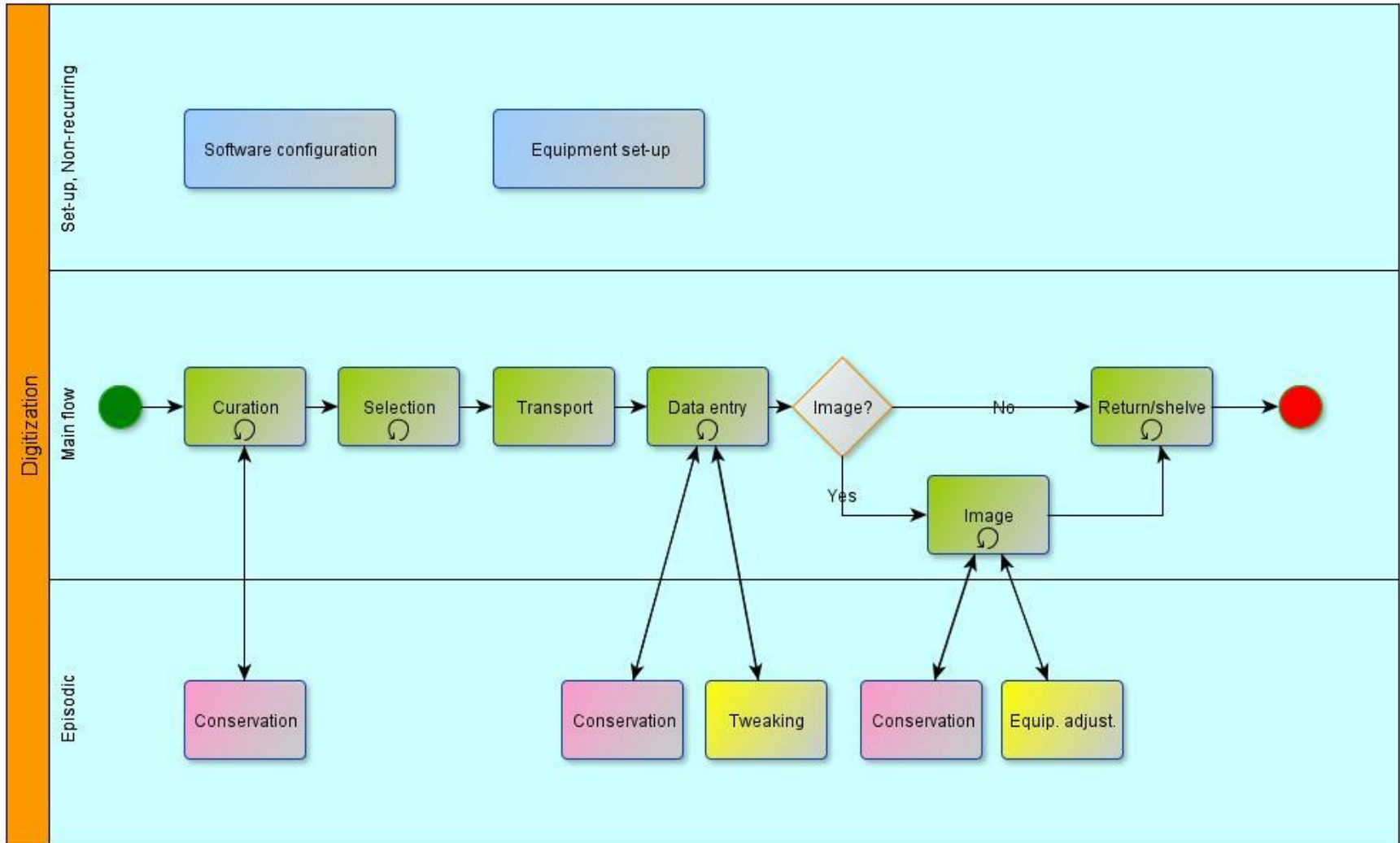All of our data and all of our specimens are important.
Let's just use the images!
We'll do the minimum now and enhance it later.

# Digitization Decision Continua that Influence Data Gaps

Current Tools ←————————————————————→ Future Tools

Robustness ←————————————————————→ Quantity

High cost ←————————————————————→ Low cost

Fitness ←————————————————————→ Speed

Image exemplars

Image everything ←————————————————————→ Image nothing

Ancillary materials ←————————————————————→ Specimens only

Business Process Modeling Approach

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
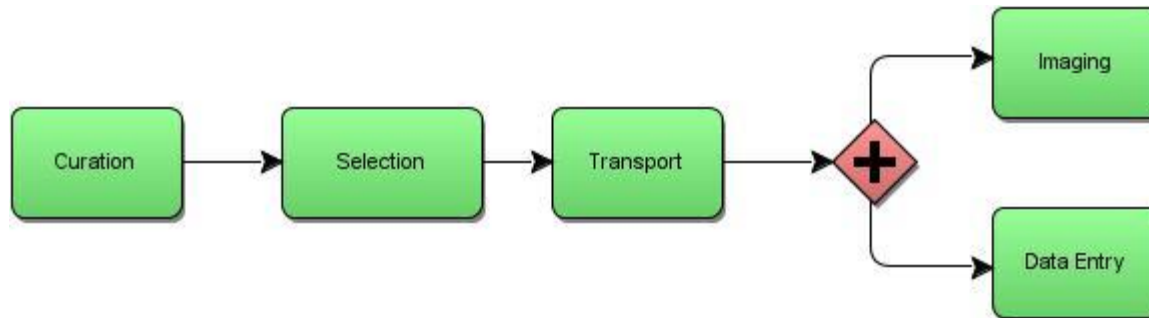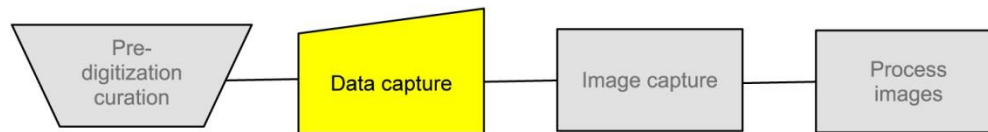
Guiding Principles

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.
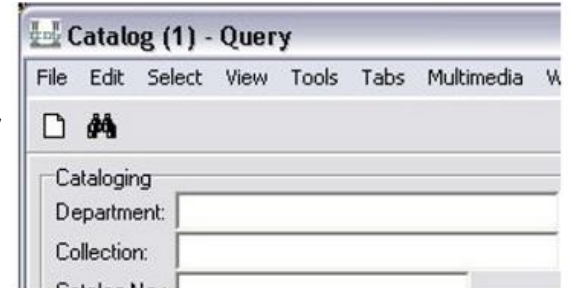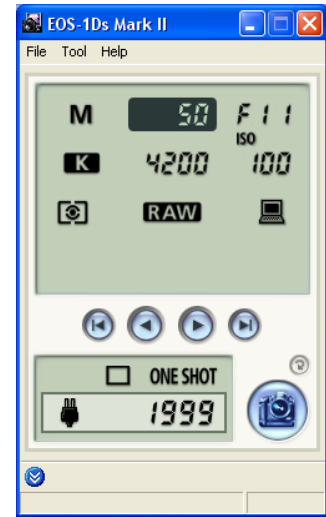
Documentation and Instructions

- **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
Integrated Digitized Biocollections

Google™ Custom Search   **Search**   My account   Log out

Technical Info   **Working Groups**   **Workshops**   **Digitization**   **Proposals**   **Bibliography**   **Wiki**

Workflow Modules and Task Lists
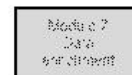
One outgrowth of the DROID (Developing Robust Object-to-Image-to-Data) workflow workshop held in May 2012 was the establishment of a series of working groups, each focused on workflow modules and tasks for various preparation types. The first of these groups, informally called the Flat Sheets and Packets 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 Spirits in Jars devoted 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
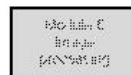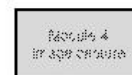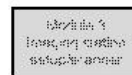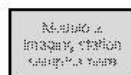
Researchers
Browse our specimen portal

Collections Staff
Learn how your collection can benefit from our work

Teachers & Students
Learning resources & opportunities to engage

iDigBio
Integrated Digitized Biocollections

- Module 4B Data Capture From Specimen Tasks
- Module 4C Event Data Capture Tasks
- Module 5 Quality Assurance Tasks

Things in Spirits in Jars

- Module 0 Pre-digitization Curation Tasks
- Module 1A Imaging Ledgers, Cards, Field Notes
- Module 1B Imaging Specimen Labels
- Module 1C Specimen Imaging
- Module 1D Image Processing
- Module 1E Phototank Immersion Imaging Seteup
- Module 1F Phototank Immersion Specimen Prep
- Module 1G Phototank Immersion Image Capture
- Module 1F Phototank Immersion Image Processing
- Module 2 Data Entry
- Module 3 Proactive Digitization

Three-dimensional Objects in Trays and Boxes

- Module 0 Predigitization Curation and Setup
- Module 1 Data Entry from Ledger-Card-Label-Catalog-Images
- Module 2A Ledger-card catalog-field notebook imaging (materials not stored with specimens)
- Module 2B Imaging Labels Associated with Collection Objects
- Module 2C Imaging Three Dimensionally Preserved Specimens
- Module 2D Imaging Two Dimensional Compressed Fossils
- Module 2E Image Processing
- Module 3 Proactive Digitization

Module 1: Imaging, Fluid-preserved

Module 1C: Recording images of specimens

Task ID	Task Name	Explanations and Comments	Resources
T1	Select and retrieve specimens, lots, and/or containers from storage location.	Some institutions record images of labels and specimens simultaneously, combining relevant tasks from M1B and M1C.	• Institutional specimen imaging policy or project guidelines. • Technician.
T2	Transport selected collection objects to staging or preparation area.	A staging or preparation area should be of adequate size and proximity to the imaging station.	Cart.
T3	Find specimen(s) in lot or container	Institutional strategies vary. In some instances, specimens are ordered by size to optimize imaging efficiency by reducing or eliminating frequent lens changes and copy stand and lighting adjustments. In other instances, specimens are selected by taxonomic group. Decisions to be made include: • whether to image multiple or single specimens from a single lot, • selecting the best quality specimen for exemplar images, • deciding the size(s) of specimens to image from a single lot (e.g., smallest, largest, average,	Institutionally specific guidelines clearly defining specimen selection criteria.

Seeking interested collections folks who want to revise the wet collections workflows for vertebrates

And, those who want to create workflows for dry-preserved vertebrate specimens