An Introduction to Workflows

Digital Data in Paleontological Research Workshop
University of California Museum of Paleontology
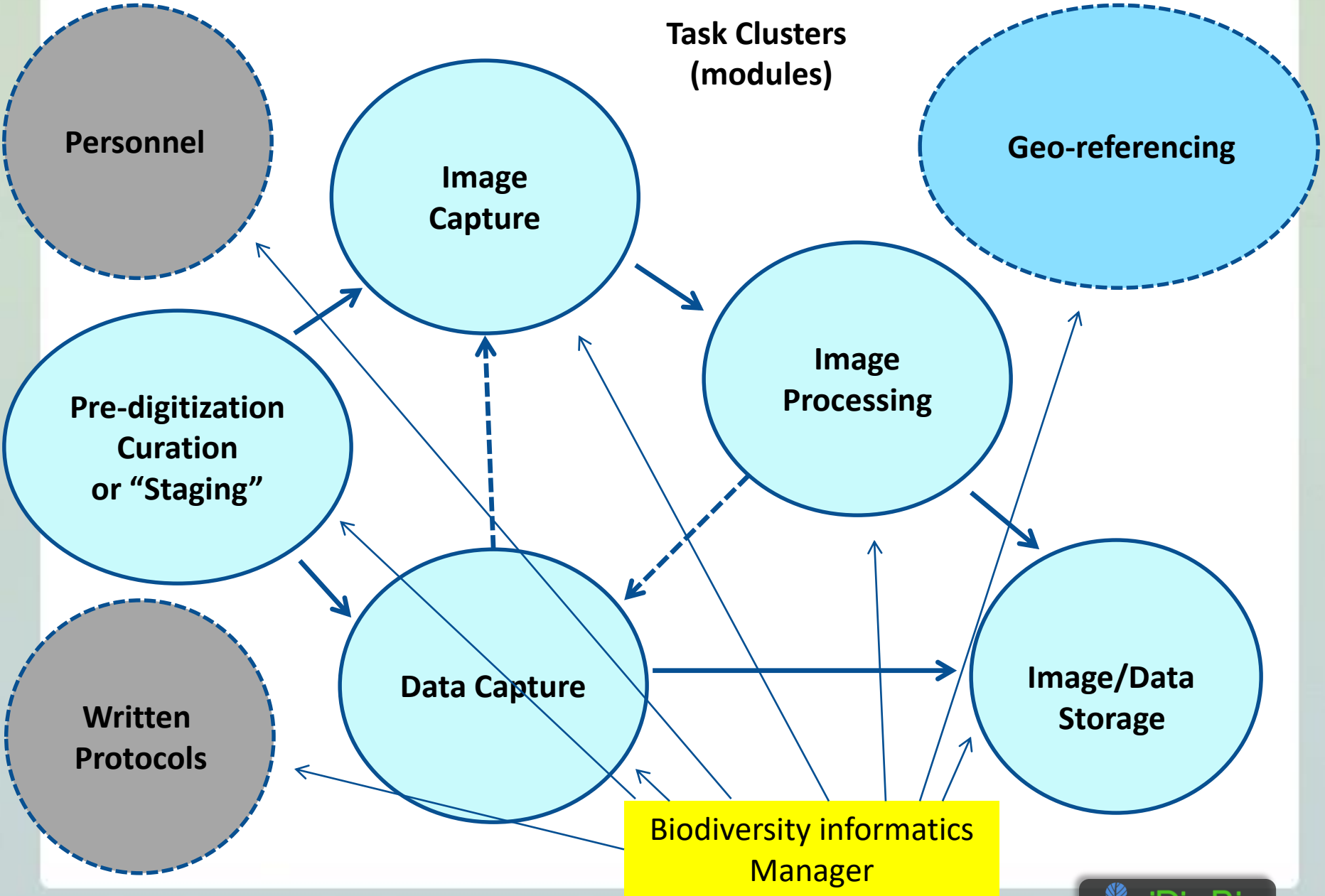**Gil Nelson**
27-28 March 2017
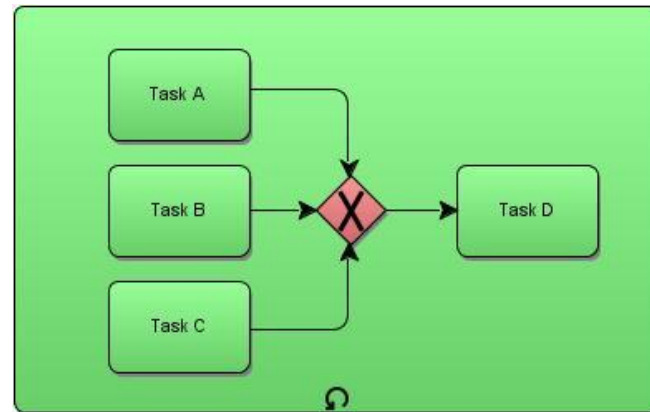**iDigBio/Institute for Digital Information and Scientific Communication**
Florida State University

Workflow Modules

Task Clusters (modules)

Personnel

Image Capture

Geo-referencing

Pre-digitization Curation or "Staging"

Image Processing

Written Protocols

Data Capture

Image/Data Storage

Biodiversity informatics Manager

iDigBio
Integrated Digitized Biocollections

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
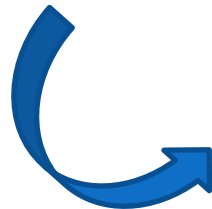- Promote continuous evaluation and redesign

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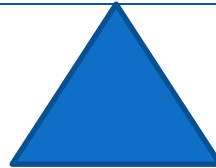
Taking the long view means developing doable, effective, and sustainable strategies for balancing long term goals with short term constraints, including a commitment to implementing future enhancements.

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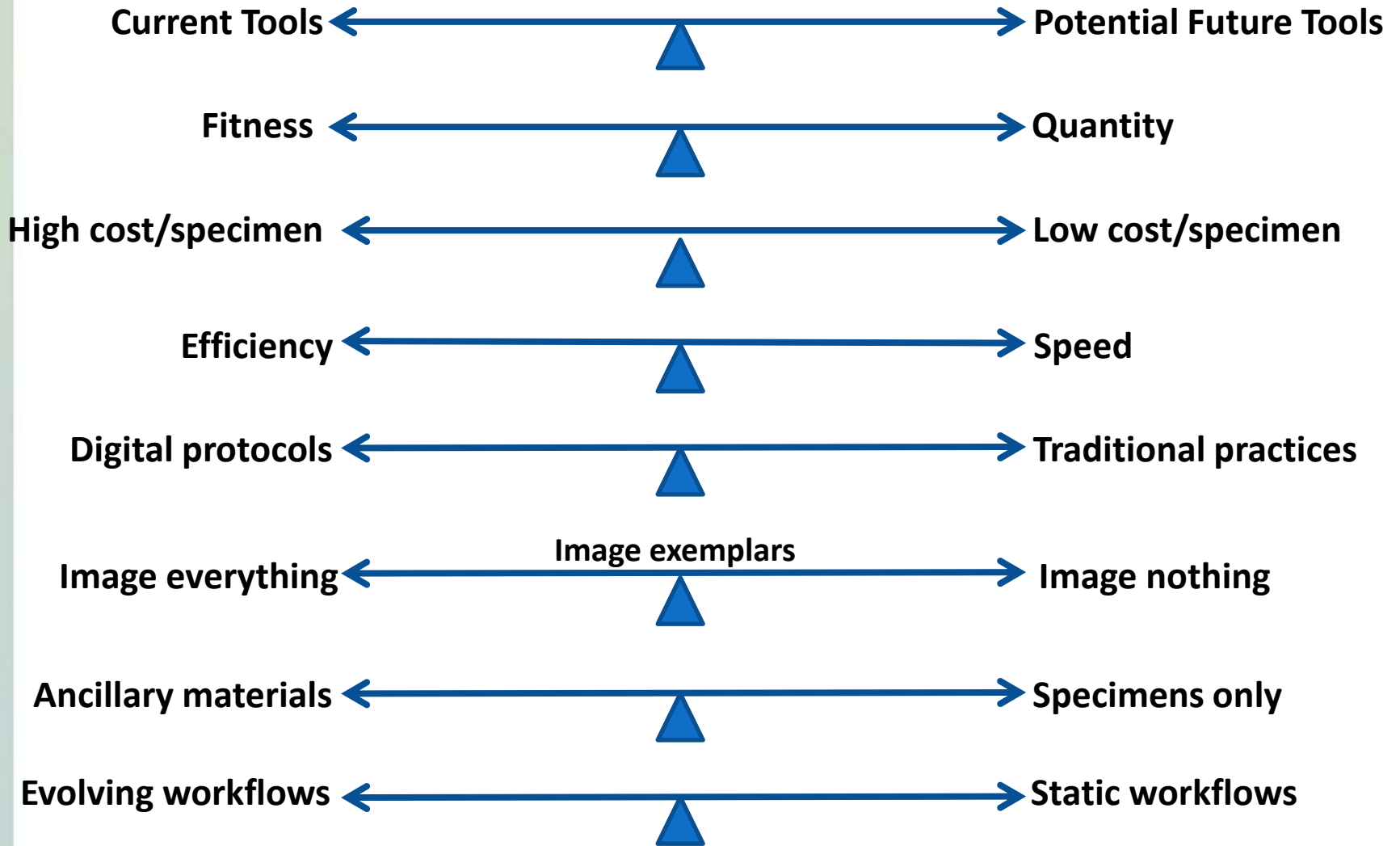The funding agencies have high output expectations.
We only have 3 years to get this done.
All of our data and all of our specimens are important.
Let's just use the images!
We'll do the minimum now and enhance it later.

Global Digitization Continua

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

Fitness ←————————————————→ Quantity

High cost/specimen ←————————————————→ Low cost/specimen

Efficiency ←————————————————→ Speed

Digital protocols ←————————————————→ Traditional practices

Image exemplars

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

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

Evolving workflows ←————————————————→ Static workflows

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
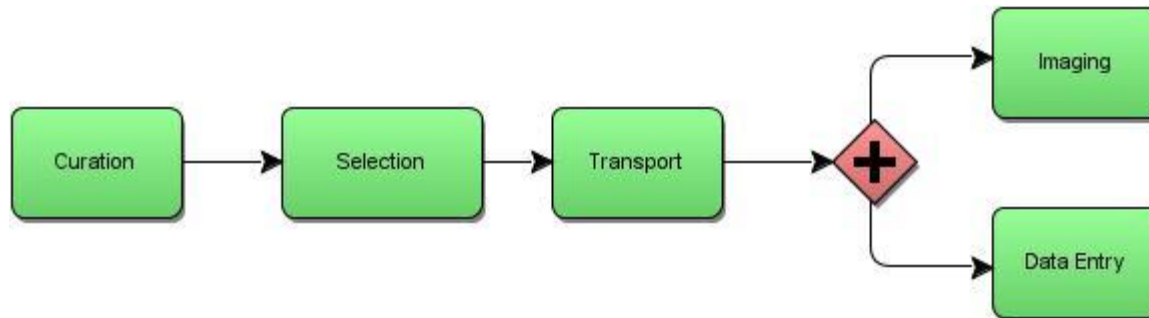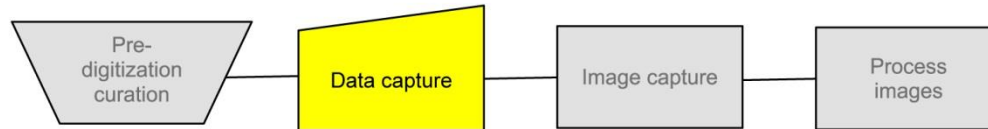
Guiding Principles

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.
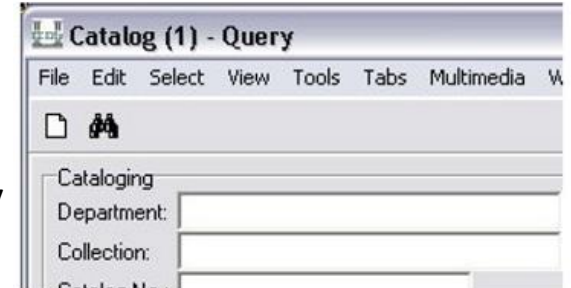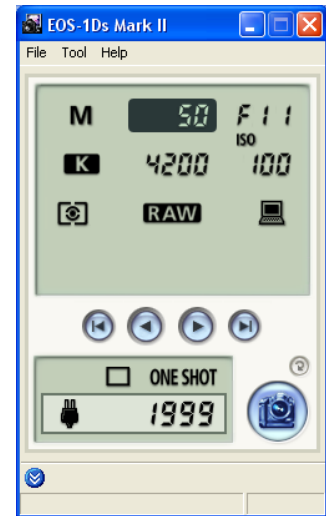
Continuous Workflow Improvement

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

Palaeontologia Electronica
palaeo-electronica.org

19 October 2016

Digitization workflows for paleontology collections

Talia S. Karim, Roger Burkhalter, Úna C. Farrell, Ann Molineux, Gil Nelson, Jessica Utrup, and Susan H. Butts

ABSTRACT

The development of digitization workflows is an essential part of any formalized large-scale digitization program. Paleontological collections literature has addressed the need for, and utility of, digitized collections for nearly four decades, but no modern, community-vetted set of digitization workflows to accomplish this goal has been widely adopted. With the advent of the U.S. National Science Foundation's (NSF) Advancing the Digitization of Biodiversity Collections (ADBC) program in 2011, iDigBio, NSF's national coordinating center for facilitating digitization, in collaboration with broad community representation from numerous institutions, launched a series of working groups to address workflow development across all major preparation types. Workflow modules have been developed for pre-digitization curation, data entry, imaging objects (catalogs, field notes and other materials not stored with specimens, labels, two- and three-dimensionally preserved specimens), image processing, and proactive digitization. Modules and the tasks they include may be implemented in any order and customized for specific configurations and institutional parameters. The workflows are made publicly available for download and customization at GitHub and via the iDigBio documentation pages. A review of platforms for electronic data publishing through online aggregators, a crucial step in any digitization program, is also provided.

Talia S. Karim. University of Colorado Museum of Natural History, Boulder, Colorado 80503, USA. talia.karim@colorado.edu
Roger Burkhalter. Sam Noble Museum, University of Oklahoma, 2401 Chautauqua Avenue, Norman, Oklahoma 73072, USA. rjb@ou.edu
Úna C. Farrell. Department of Geological Sciences, Stanford University, 450 Serra Mall, Stanford, California 94305, USA. ufarrell@stanford.edu
Ann Molineux. Non-vertebrate Paleontology Lab, University of Texas, 10100 Burnet Road, Austin, Texas 78758, USA. annm@austin.utexas.edu
Gil Nelson. Department of Biological Sciences, Florida State University, Tallahassee, Florida 32303, USA. gnelson@bio.fsu.edu
Jessica Utrup. Yale University, Peabody Museum of Natural History, Division of Invertebrate Paleontology, 170 Whitney Avenue, PO Box 208118, New Haven, Connecticut 06520, USA. jessica.utrup@yale.edu
Susan H. Butts. Yale University, Peabody Museum of Natural History, Division of Invertebrate Paleontology, 170 Whitney Avenue, PO Box 208118, New Haven, Connecticut 06520, USA. susan.butts@yale.edu

iDigBio
Integrated Digitized Biocollections

Workflow Modules

Module 0: Pre-Digitization Curation
Module 1: Data Entry
Module 2: Imaging
    Module 2A: Imaging catalogs, field notes, and other materials not stored with specimens
    Module 2B: Imaging labels associated with specimens
    Module 2C: Imaging three-dimensional preserved specimens
    Module 2D: Imaging two-dimensionally compressed specimens
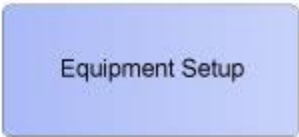   Module 2E: Image processing
Module 3: Proactive Digitization
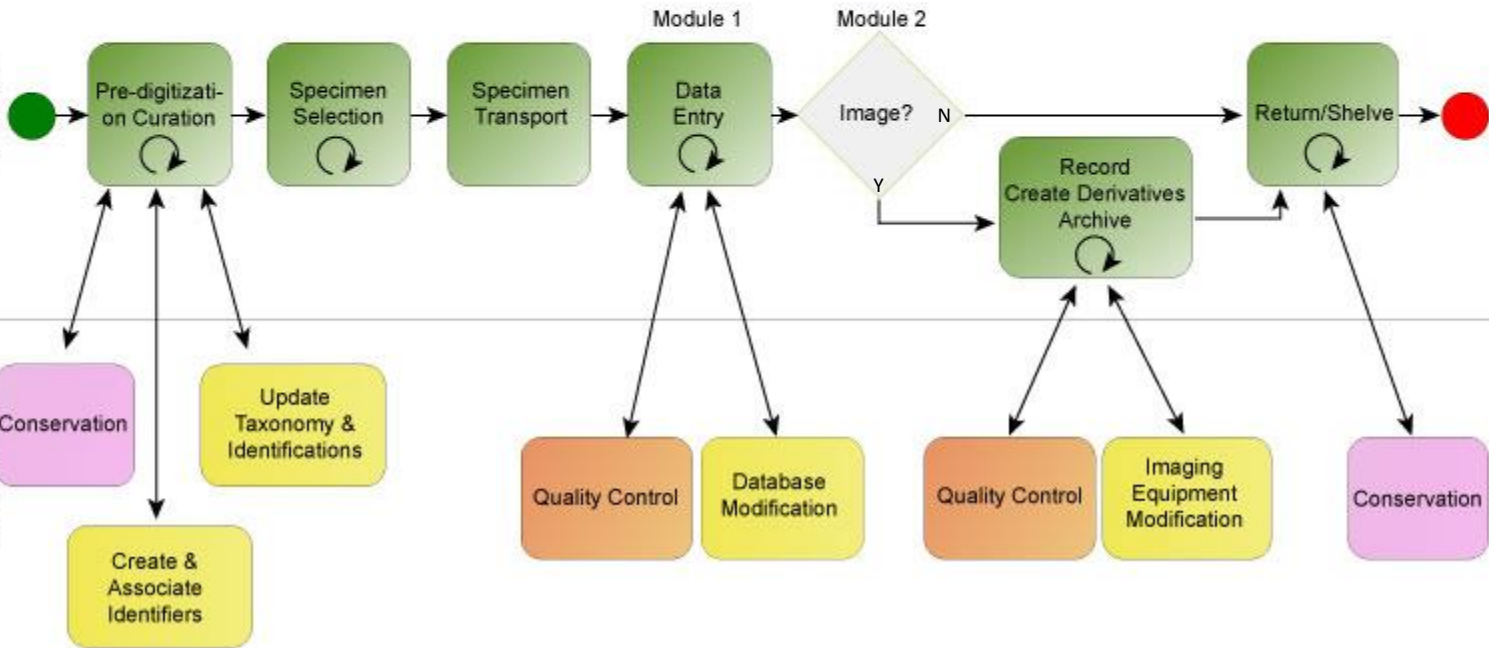
Module 0: Pre-Digitization Curation and Setup

Task ID	Task Name	Explanations and Comments	Resources
T1	Prioritize specimens, collection objects, ledgers, field notes, and catalogs to digitize.	Varies by institution. Should follow institutional digitization policies and guidelines.	Institutional policy, project guidelines, active research criteria, etc.
T2	Note damage to object to be digitized that needs immediate attention.	Route to conservation workflow as necessary, based on institutional policy or curatorial practices.	Institutionally specific curation guidelines.
T3	Update specimen taxonomy (and related authority files) as necessary.	This may happen prior to the digitization of any taxonomic group. Some institutions update specimens with expert determinations prior to digitization. Others record determinations from the label in anticipation of community involvement in helping correct determinations.	According to institutional protocol and procedures and/or project requirements.