Digitization Workflows

American Society of Mammalogists

Gil Nelson
12 June 2015

iDigBio
Institute for Digital Information and Scientific Communication
Florida State University
Tall Timbers Research Station and Land Conservancy

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

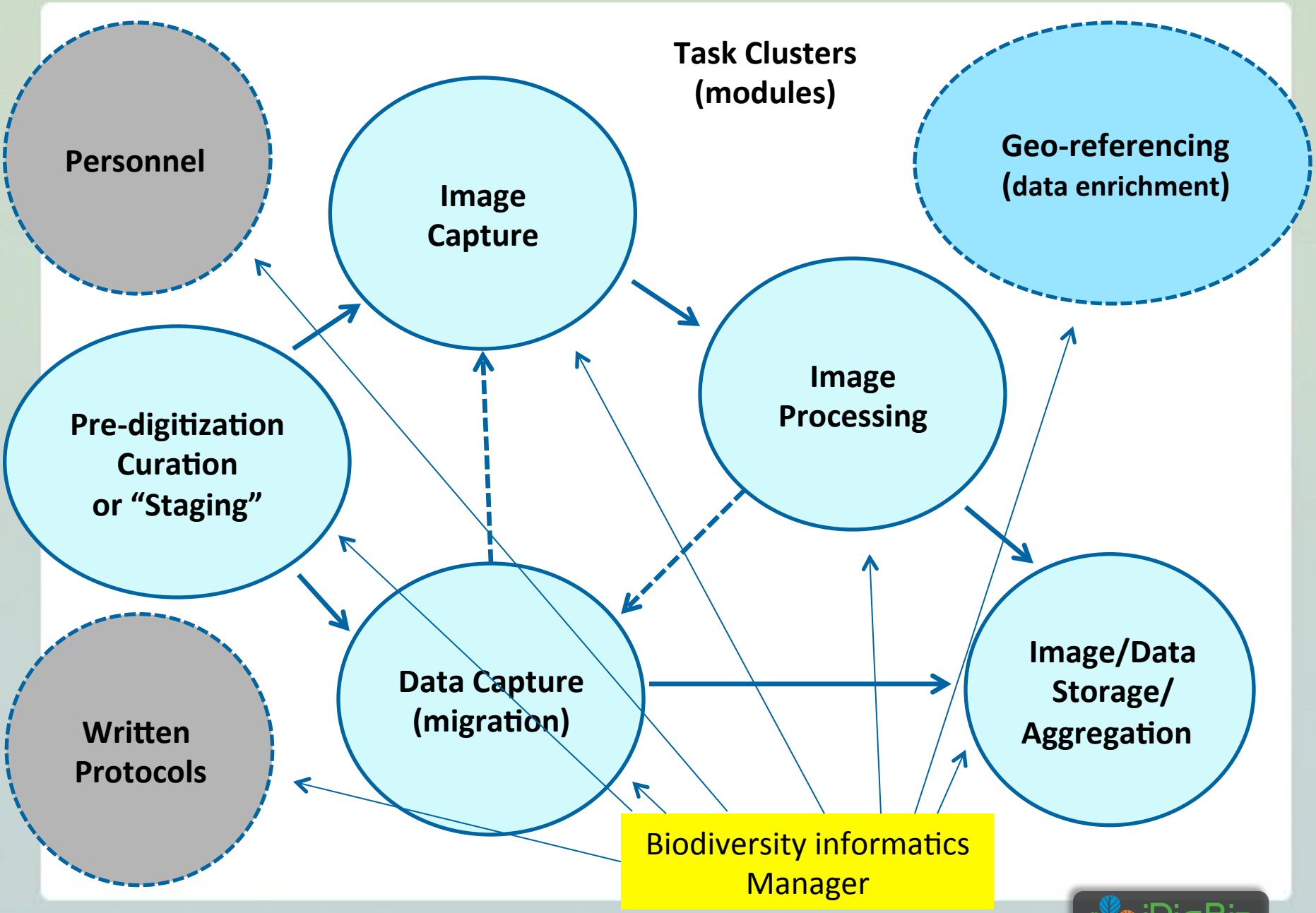
Gil Nelson, Deborah Paul, Gregory Riccardi, Austin R. Mast

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

iDigBio
Integrated Digitized Biocollections

Acknowledgments

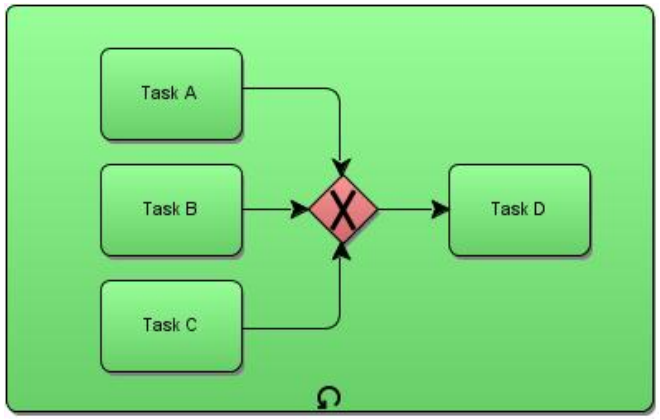
American Museum of Natural History
Botanical Research Institute of Texas
Florida Museum of Natural History
Florida State University
Harvard Herbarium
Museum of Comparative Zoology (Harvard)
New York Botanical Garden
Southeast Regional Network for Expertise and Collections
Specify Software Project (University 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


Task Clusters
(modules)
Personnel
Image
Capture
Geo-referencing
(data enrichment)
Pre-digitization
Curation
or "Staging"
Image
Processing
Written
Protocols
Data Capture
(migration)
Image/Data
Storage/
Aggregation
Biodiversity informatics
Manager

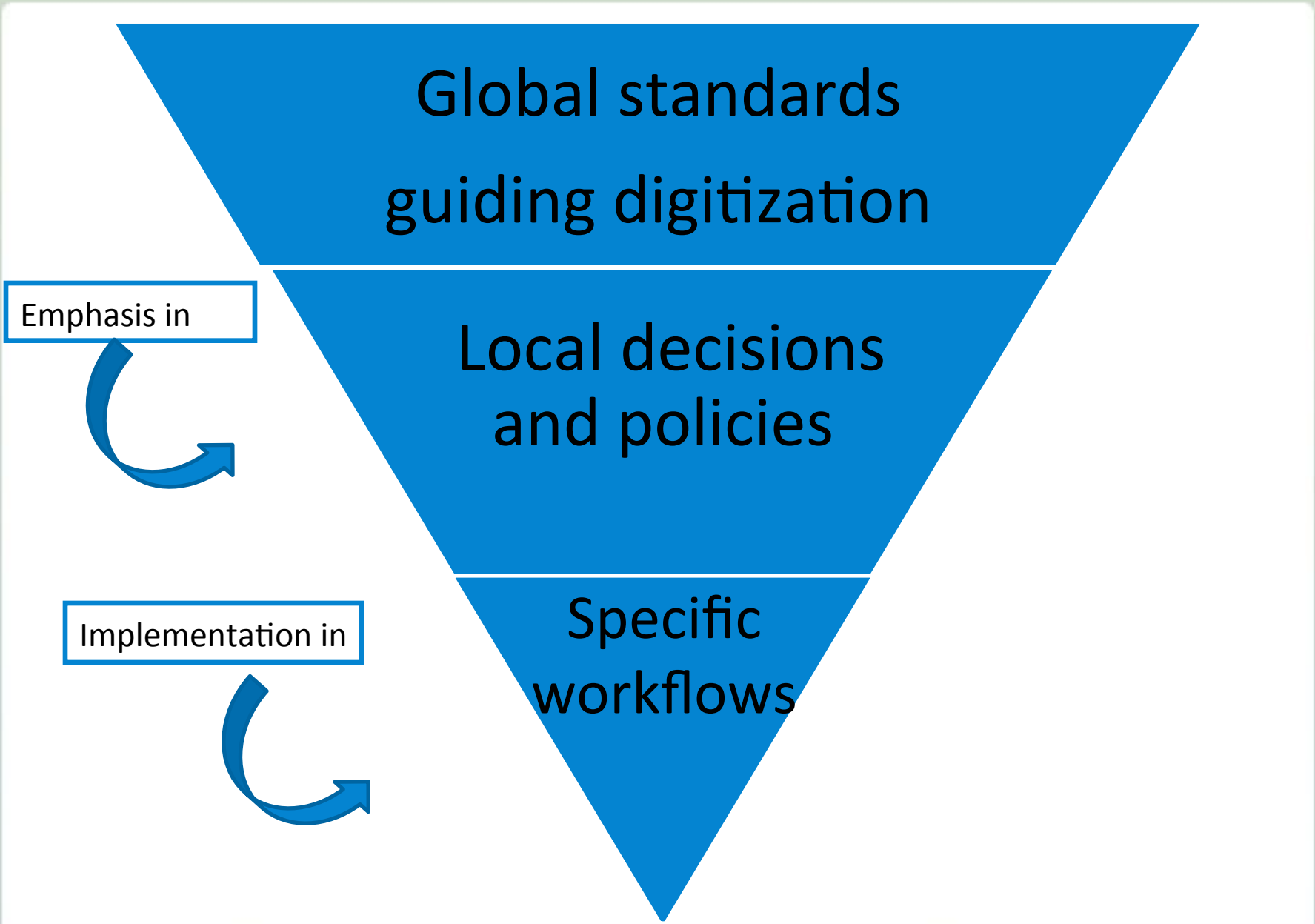
Processes that have gained definition and currency in digitization workflows

- 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

Values of defined workflows



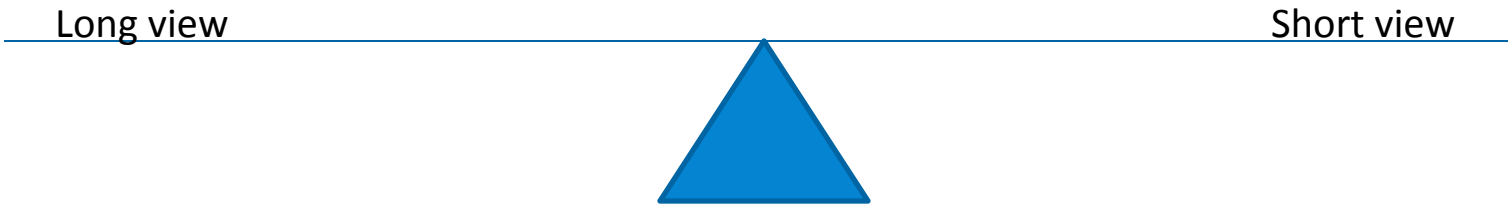
- Promote efficiency and automation of processes
- Facilitate routing and scheduling of activities
- Provide for balancing workloads
- Ensure that processes are visible and predictable
- Allow for escalations and notifications
- Enhance tracking of tasks
- Foster collaboration of all parties involved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


Global standards guiding digitization
Local decisions and policies
Specific workflows
Emphasis in
Implementation in

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.

Scan and Deliver: Managing User-initiated Digitization in Special Collections and Archives, 2011
J. Schaffner, F. Snyder. S. Supple

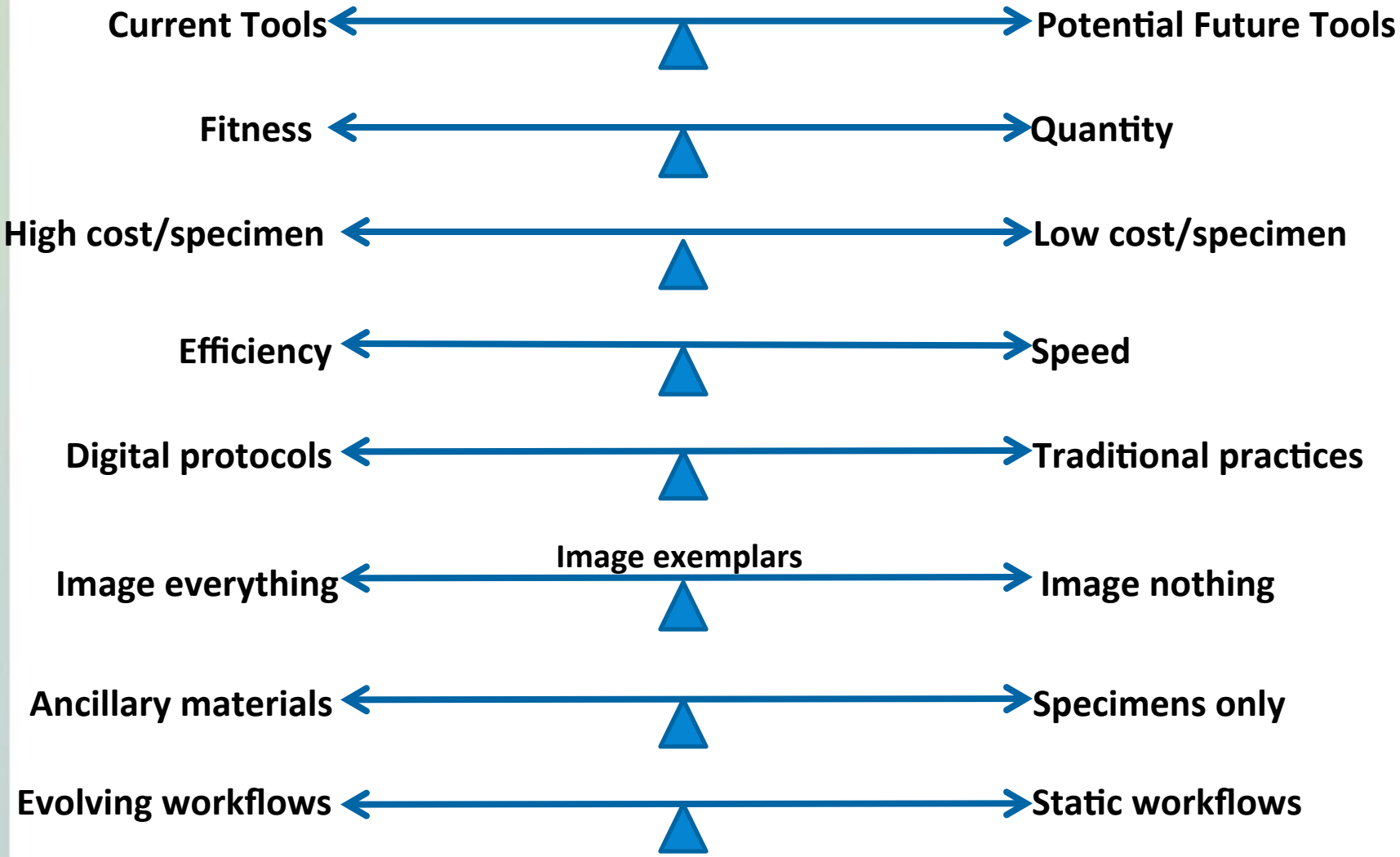


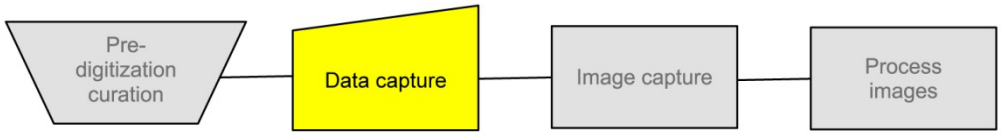
Taking the long view means developing doable, effective, and sustainable strategies for balancing long term goals with short term constraints, including a commitment to implementing future enhancements.

Pressures mitigating the long view

So much data, so little time, so much to share.
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





Workflow Design 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

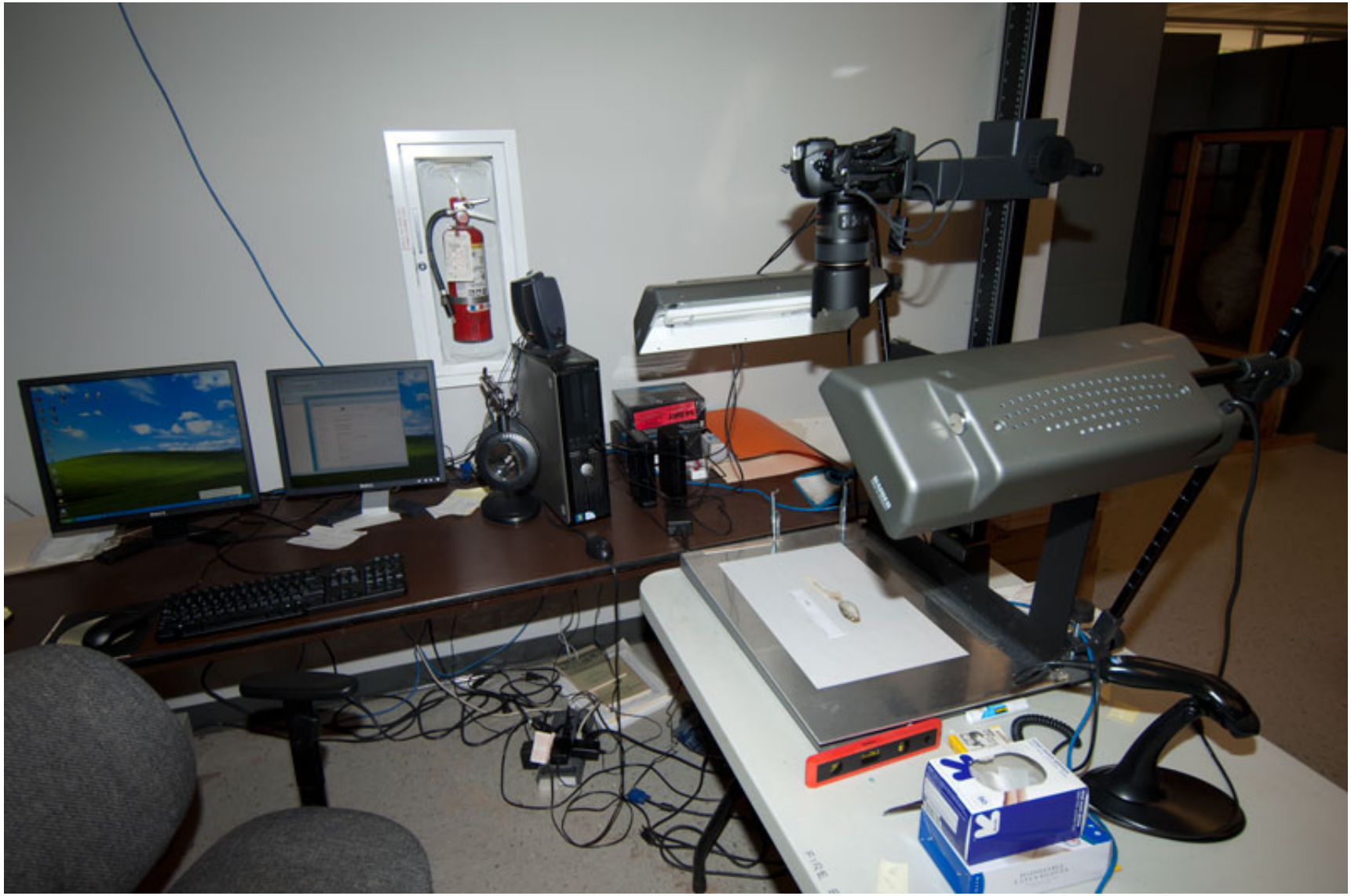
- Adjust to strengths of each technician--using students and volunteers requires flexibility in role assigned to personnel rather personnel assigned to role.

Create task lists

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

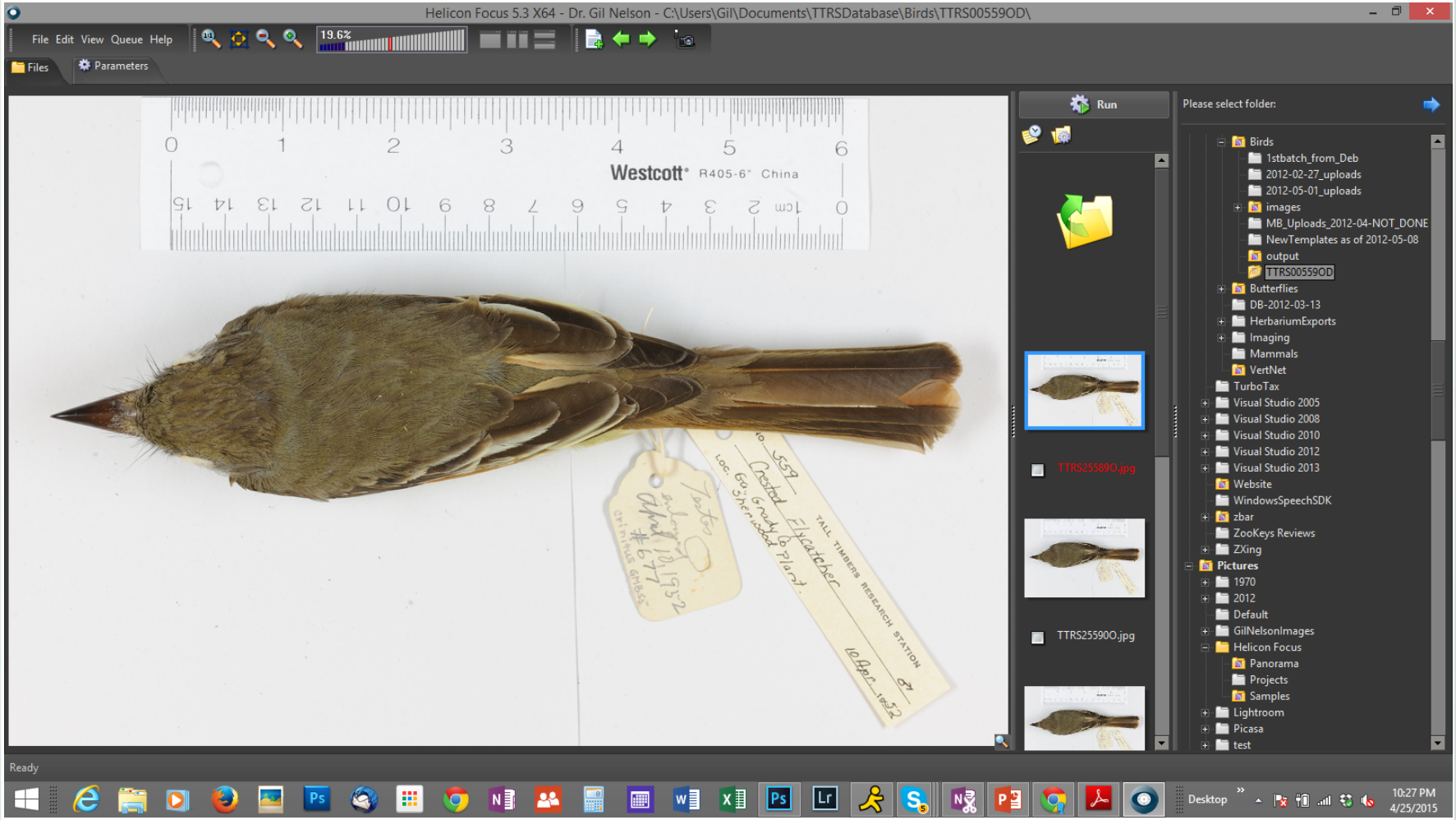
Detailed task list

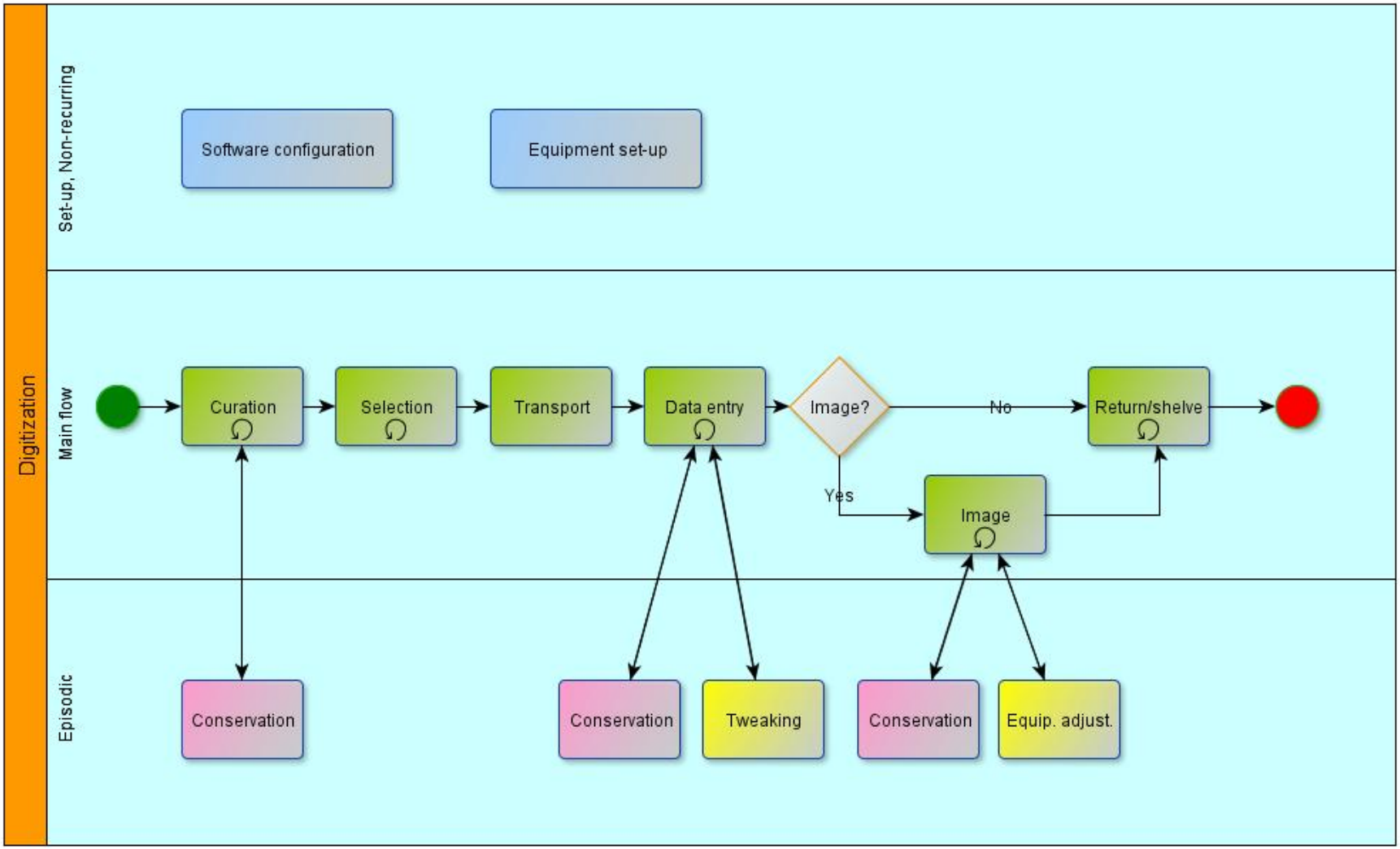
1. Position specimen
2. Take a batch of images
3. In the Capture NX2 browser, select all images, rotate to vertical
4. In Capture NX2 select one image and double click it
5. Click Adjust->Light->Autolevels
6. Crop image to size
7. Click Batch->Save Adjustments
8. Select a name to save the batch as: as MakeTiff
9. Browse to the TIFF destination folder, click OK
10. Click Batch->Run Batch Process
11. Ensure that you are applying settings from your settings file, that the file format is TIF, and that your destination folder is your TIFF folder
12. Click Start
13. When complete, close the Processing Queue dialog
14. Click Batch->Save Adjustments
15. Select a name to save the batch as: as MakeJpeg
16. Select your JPEG folder as the destination
17. Click OK
18. Click Batch->Run Batch Process
19. Select your MakeJpeg.set file
20. Set your destination folder to your JPEG folder
21. Set your file format to JPG
22. Close the Processing Queue dialog
23. Close the image you performed the activites on, do not save
24. To make a thumbnail, open an image in Capture NX2
25. Click Edit->Size/Resolution
26. Crop to remove color bar
27. Set Output Size Width to 0.3, Height to 0.43, Resolution to 350
28. Click Batch->Save Adjustments
29. Save as: MakeThumb
30. Select your Thumbs folder as your Destination
31. Select Batch->Run Batch Process
32. Select the MakeThumb.set file
33. Select the thumbs folder as the destination
34. Close but do not save your adjusted image
35. Close the Processing Queue dialog
36. Navigate to the Thumbs folder
37. Run RN and navigate to the Thumbs folder
38. Select Insert/append
39. For Pos: enter 15
40. In the blank, enter -thumbnail
41. Check the New name (Preview) column to ensure your entries are correct
42. Click Start



Imaging station

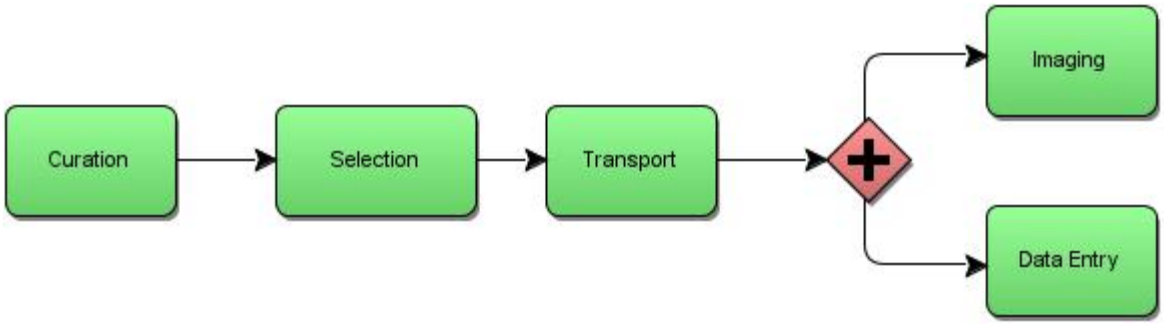
Helicon Focus




Digitization
Set-up, Non-recurring
Software configuration
Equipment set-up
Main flow
Curation
Selection
Transport
Data entry
Image?
No
Return/shelve
Yes
Image
Episodic
Conservation
Conservation
Tweaking
Conservation
Equip. adjust.

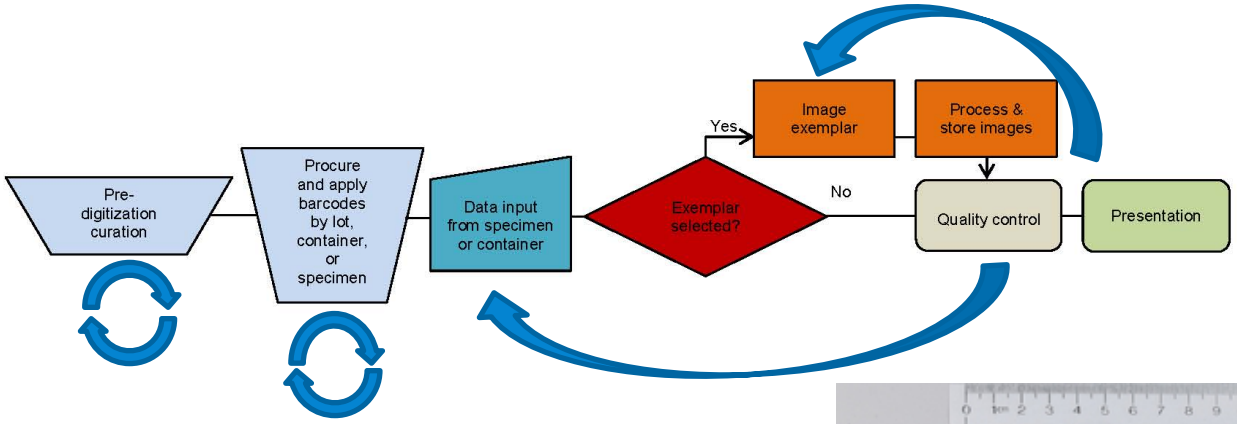
Example Processes (Modules), their Cycles and Dependencies

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



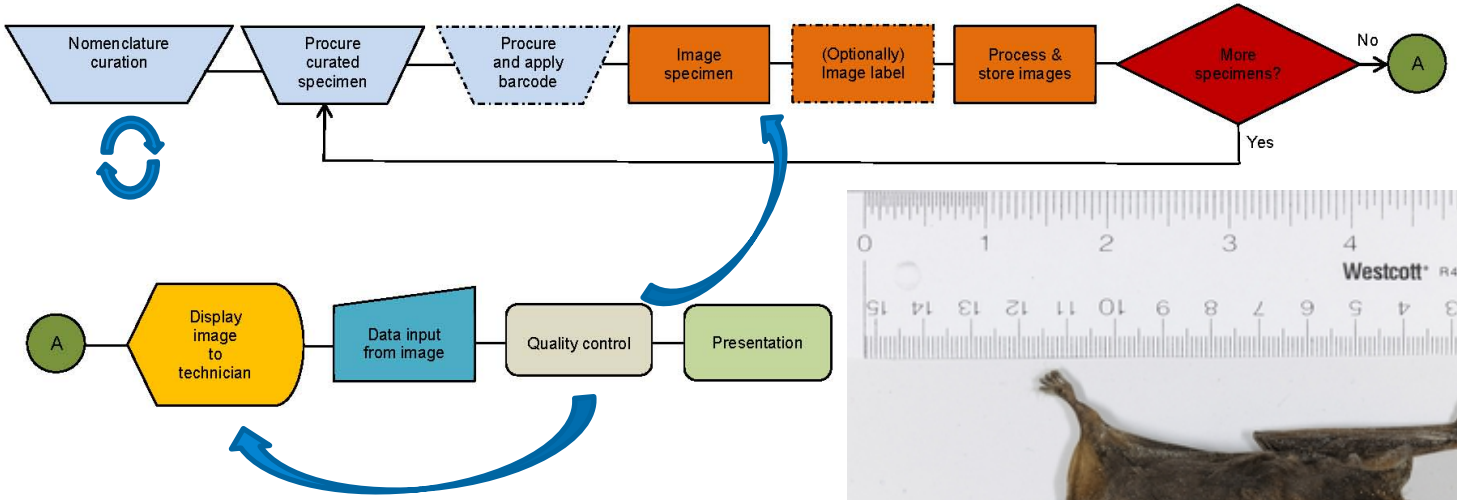
O2D2EI—Existing Specimen Workflow: Object to Data to Exemplar Images

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.



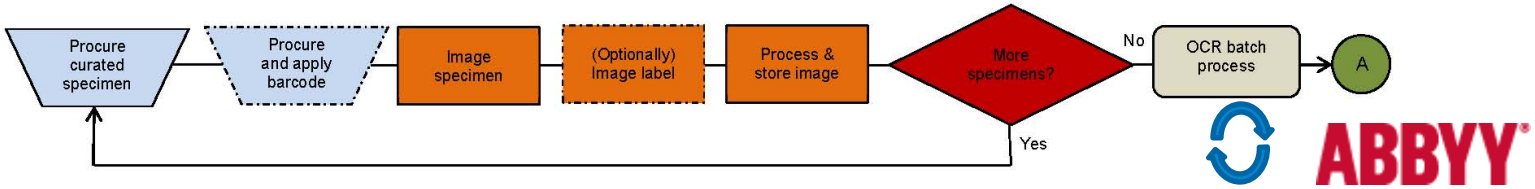
O2I2D(2)—Existing Specimen Workflow: Object to Image to Data

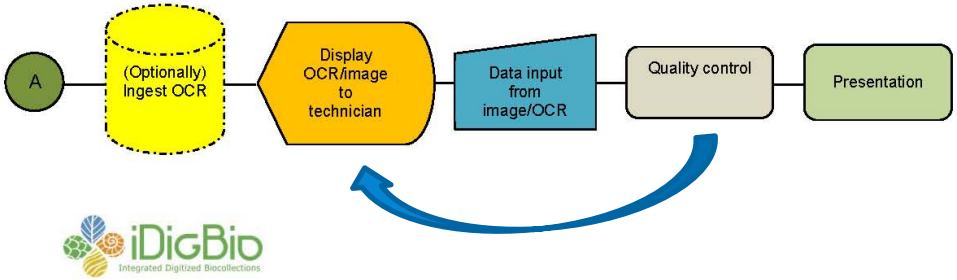
This workflow is designed for capturing images of existing specimens and using these images as the basis for data capture. Depending upon preparation type, barcodes are sometimes applied inline as the step immediately previous to imaging (shown optionally below) and other times en masse within an independent 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

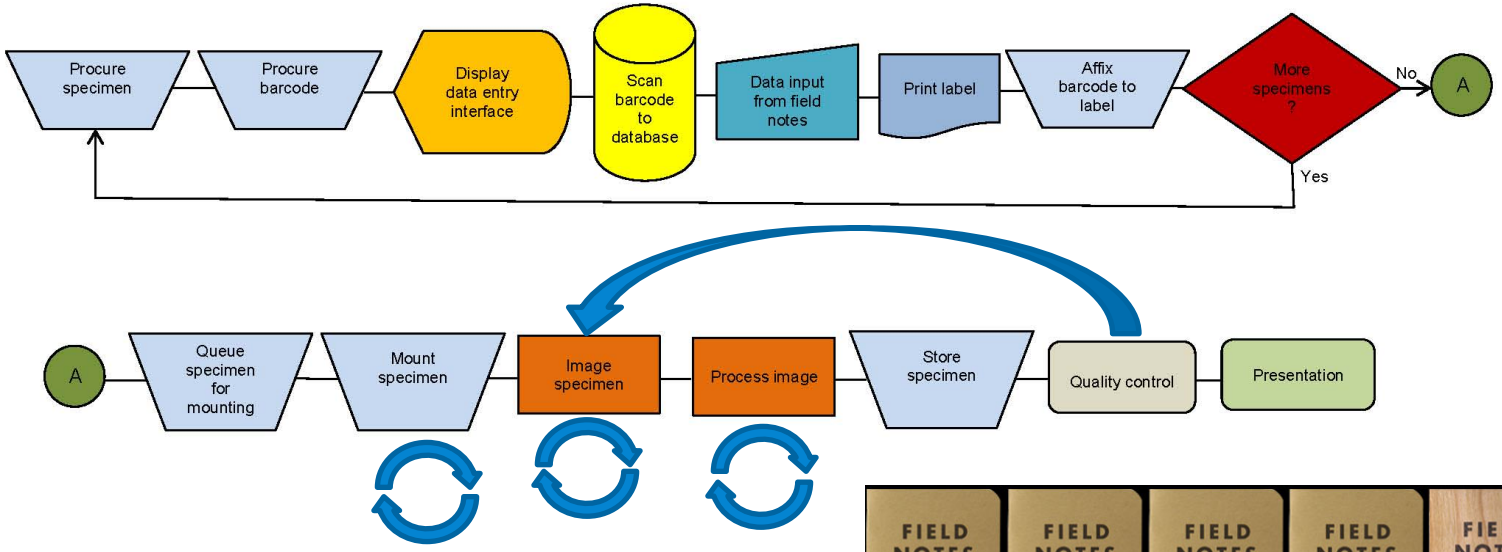
This workflow is designed to capture images of existing specimens, pass the images through optical character recognition (OCR) software, and use the combination of image and OCR output to capture data. There are variations on this workflow. For example, depending on preparation type, barcodes are sometimes 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.





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

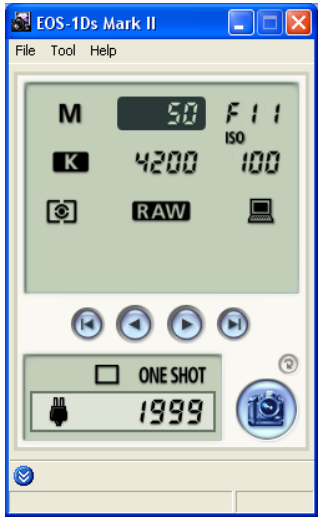
This workflow is designed for actively growing collections in which new specimens are regularly added. Collectors, especially in herbaria, typically keystroke label data from field notes, store the label with the specimen, and queue the specimen for mounting. Following mounting, the specimen is treated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



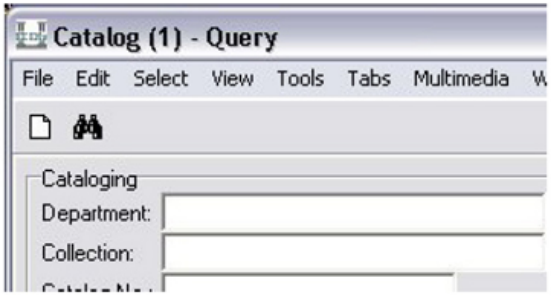
iDigBio
Integrated Digitized Biocollections

Documentation and Instructions

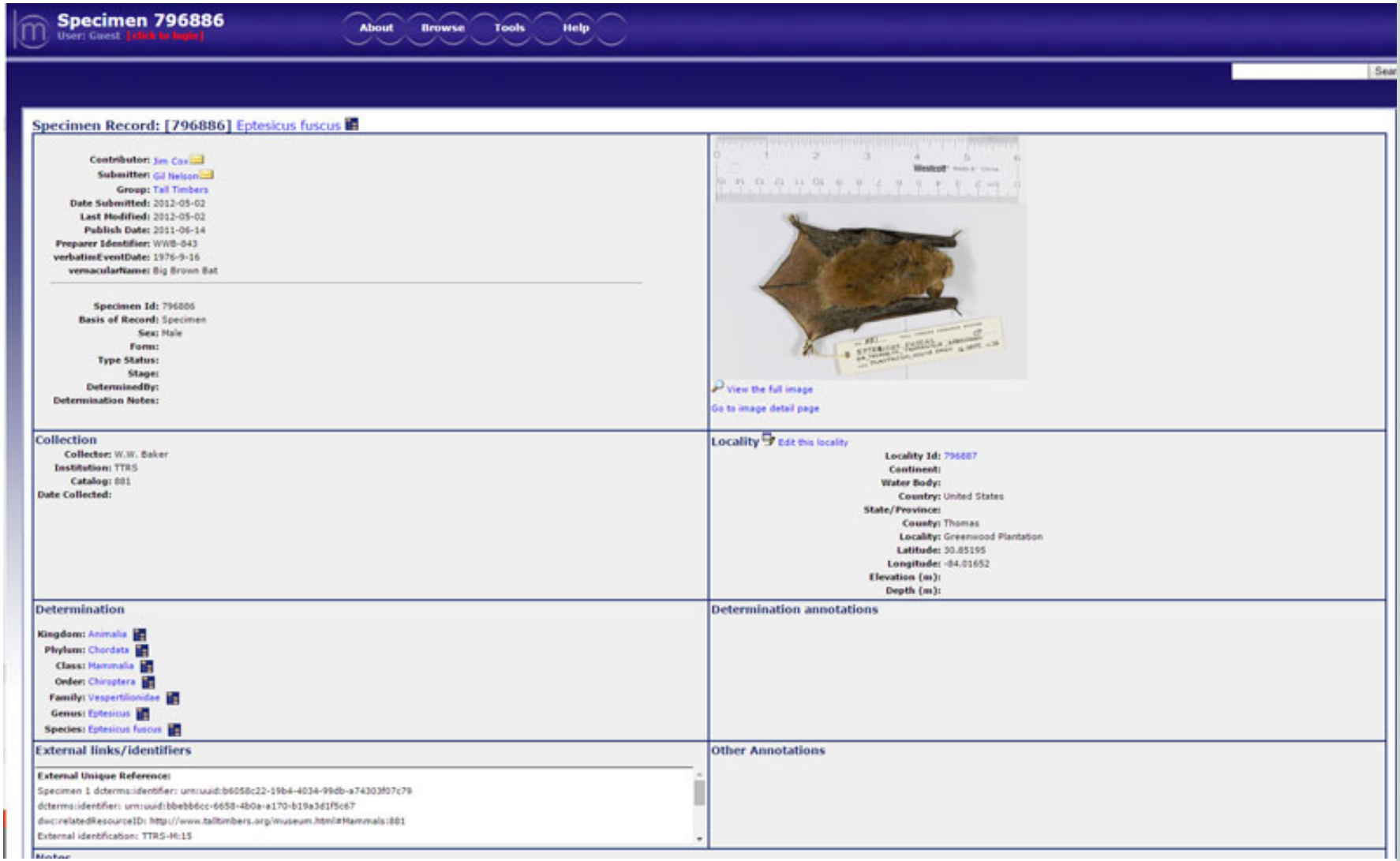
- **Written Protocols**
  - Essential!
  - Include screen shots and pictures.
  - Attention to detail (leave nothing to the imagination).
  - Express limits on technician authority.



- **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.



Morphbank.net



VertNet

Search | Publishers | Join | Resources | About | Feedback | Login

TTRS Mammals 881 *Eptesicus fuscus*

Submit data issue

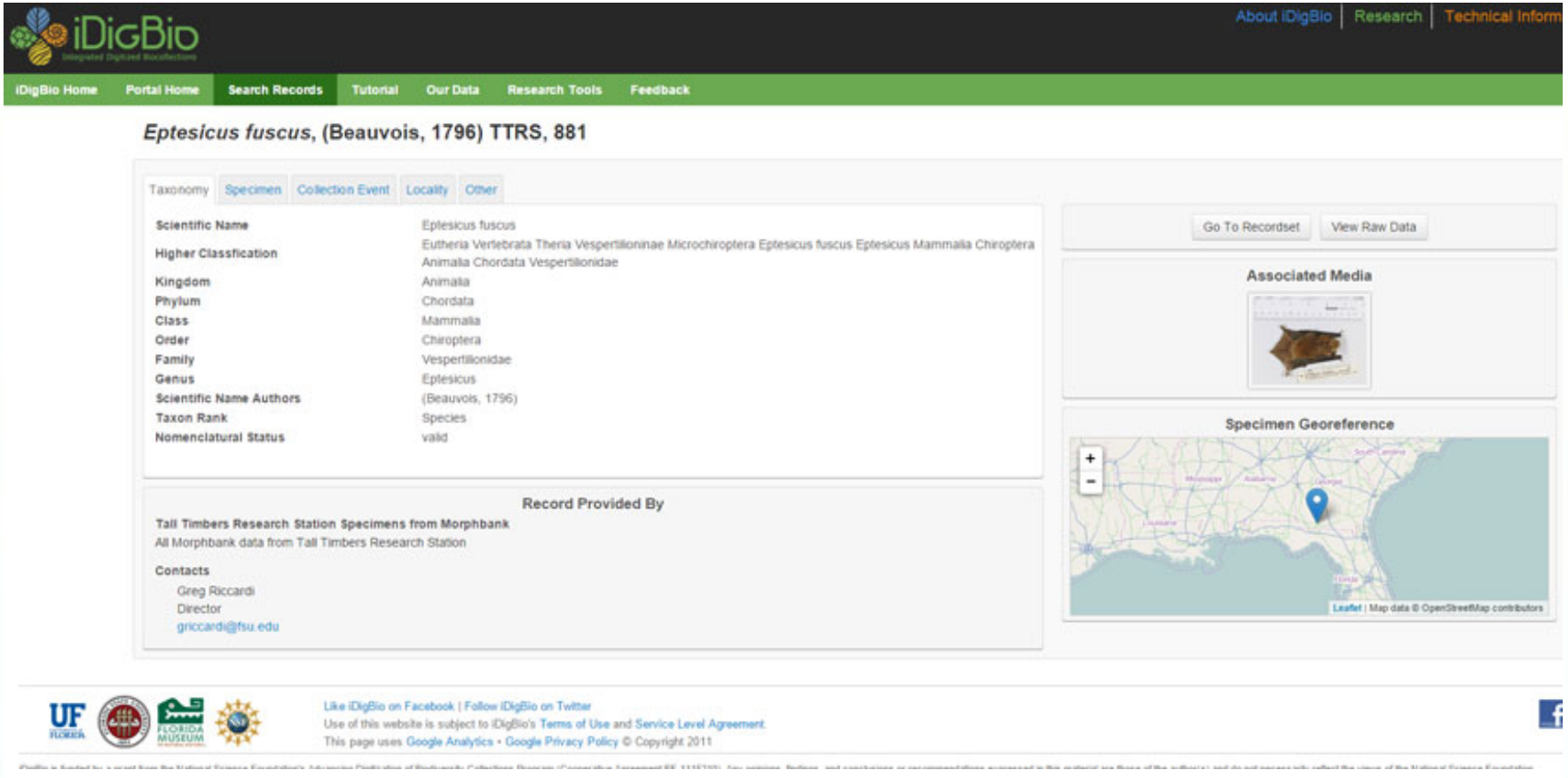
Darwin Core | Rights | Spatial quality | Index

View media

Term	Value
InstitutionCode	TTRS
CollectionCode	Mammals
BasisOfRecord	PreservedSpecimen
OccurrenceID	bbebb6cc-6658-4b0a-a170-b19a3d1f5c67
CatalogNumber	881
OccurrenceRemarks	Preparer's identifier WWB-843
RecordedBy	W.W. Baker
IndividualCount	1
Sex	Male
Preparations	Skin
AssociatedMedia	http://www.morphbank.net/Browse/ByImage/?specimenId=796886
Year	1976
Month	9
Day	16

VertNet © version 2015-05-07T11:17:00 | National Science Foundation WHERE DISCOVERIES BEGIN

iDigBio Integrated Digitized Biocollections


iDigBio
About iDigBio | Research | Technical Inform
iDigBio Home
Portal Home
Search Records
Tutorial
Our Data
Research Tools
Feedback
Eptesicus fuscus, (Beauvois, 1796) TTRS, 881
Taxonomy
Specimen
Collection Event
Locality
Other
Scientific Name
Eptesicus fuscus
Higher Classification
Eutheria Vertebrata Theria Vespertilioninae Microchiroptera Eptesicus fuscus Eptesicus Mammalia Chiroptera Animalia Chordata Vespertilionidae
Kingdom
Animalia
Phylum
Chordata
Class
Mammalia
Order
Chiroptera
Family
Vespertilionidae
Genus
Eptesicus
Scientific Name Authors
(Beauvois, 1796)
Taxon Rank
Species
Nomenclatural Status
valid
Record Provided By
Tall Timbers Research Station Specimens from Morphbank
All Morphbank data from Tall Timbers Research Station
Contacts
Greg Riccardi
Director
griccardi@fsu.edu
Go To Recordset
View Raw Data
Associated Media
Specimen Georeference
Leaflet | Map data © OpenStreetMap contributors
Like iDigBio on Facebook | Follow iDigBio on Twitter
Use of this website is subject to iDigBio's Terms of Use and Service Level Agreement.
This page uses Google Analytics • Google Privacy Policy © Copyright 2011

ark:/99999/bbebb6cc-6658-4b0a-a170-b19a3d1f5c67

Specimen of **Eptesicus fuscus (Beauvois, 1796)** recorded on Sep 16, 1976

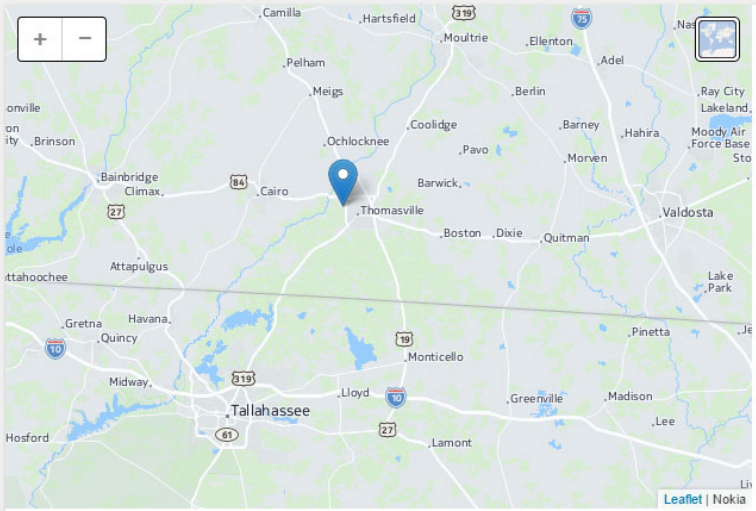
mals dataset

Information | Verbatim

INTERPRETATION ISSUES

GBIF found issues interpreting the verbatim content of this record:

- Coordinate rounded



Location

LOCALITY

Greenwood Plantation, United States

-84.01652, 30.85195

GEOGRAPHIC CLASSIFICATION

North America > United States > Georgia > Thomas

GEOREFERENCING

STATUS requires verification

Identification details According to GBIF Backbone Taxonomy

IDENTIFIED AS SPECIES

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

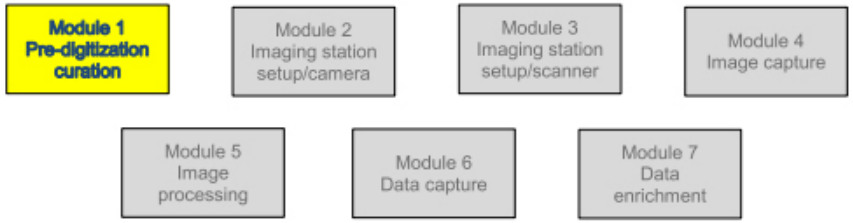
Workflow Modules and Task Lists

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

Vertebrate Digitization Workflow Working Group

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).	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.	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.	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. Tasks T2 or T3 might also include determining whether specimens are out on loan or	QRcodes, DataMatrix, 1D barcode, or OCR-readable documents for insertion into specimen folders.

iDigBio

Integrated Digitized Biocollections