Digitization Modules, Tasks, and Workflows

Leveraging Digitization Practices Across Multiple Domains
Gil Nelson
7 October 2014
Institute for Digital Information and Scientific Communication
Florida State University

Commonalities across task clusters provides an organizing 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

iDigBio
Integrated Digitized Biocollections

Tasks Common to All Domains

Workflows and protocols

Selecting and installing a database
Specify
Symbiota
Emu
Custom

Configuring and purchasing an imaging station
Copy stand and lighting
Light box

Searching and selecting image capture, workflow, and processing software

Preparing for digitization
Pre-digitization curation

Considering and planning for data enhancement/enrichment activities
Georeferencing
Digitizing source materials

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

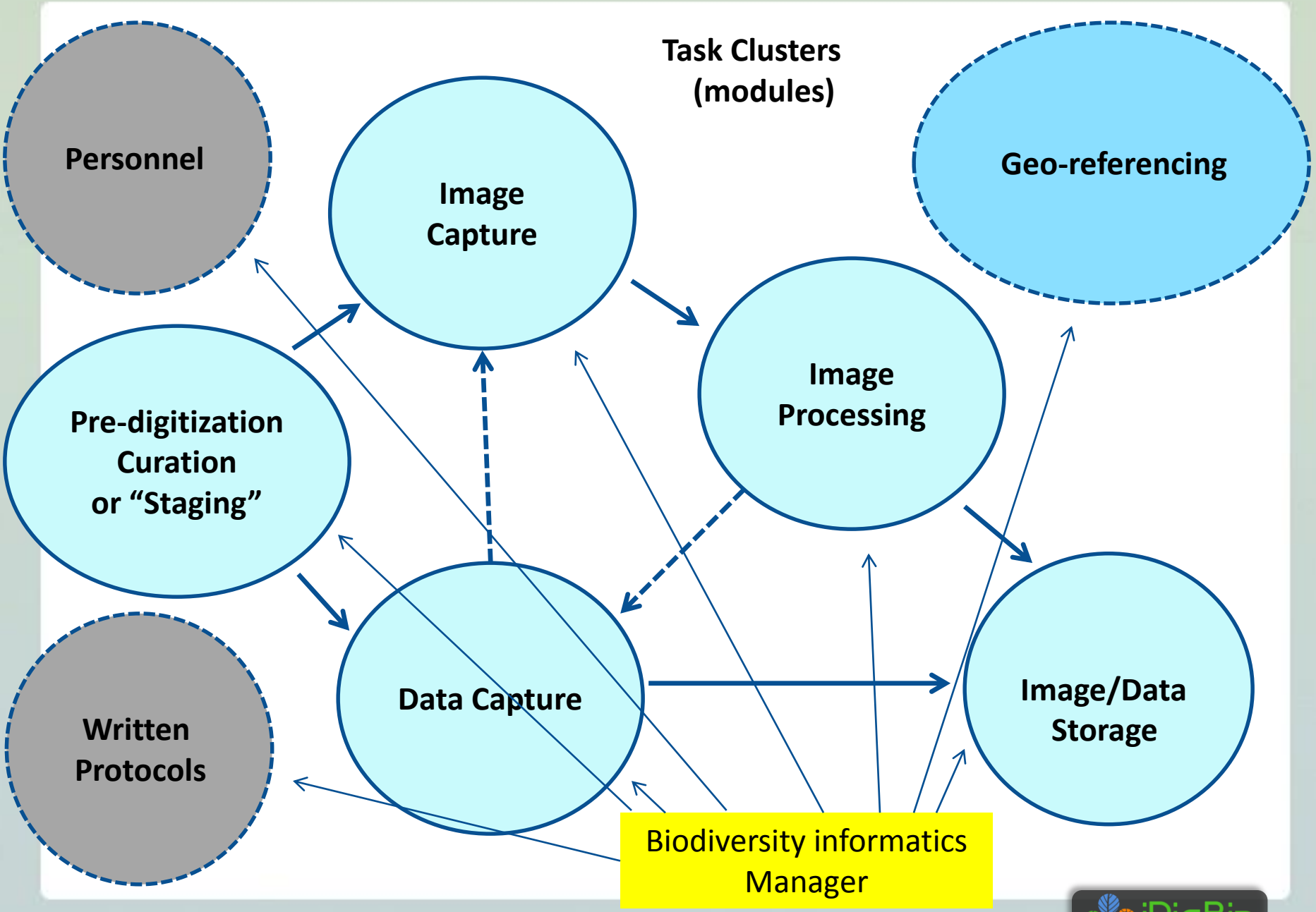
A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

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

iDigBio
Integrated Digitized Biocollections

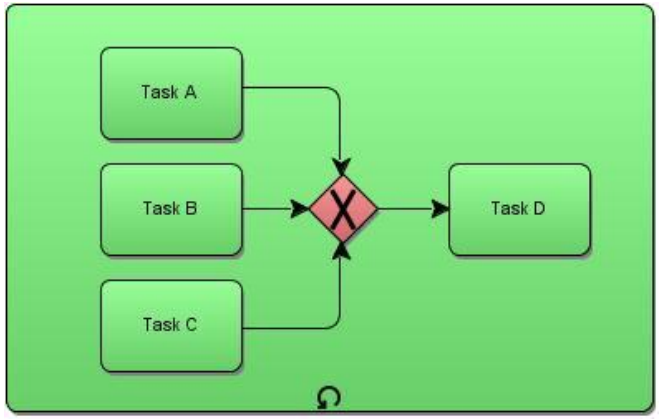

Task Clusters
(modules)
Personnel
Image
Capture
Geo-referencing
Image
Processing
Pre-digitization
Curation
or "Staging"
Written
Protocols
Data Capture
Image/Data
Storage
Biodiversity informatics
Manager

Processes that have gained definition and currency in digitization workflows

- Linking genomic and other data to specimen records
- Linking original source materials to specimen records
- Public participation (crowd-sourcing, citizen science)
- Remote annotation of specimen records
- Using digitized data for research
- Optical Character Recognition

iDigBio
Integrated Digitized Biocollections

Values of defined workflows



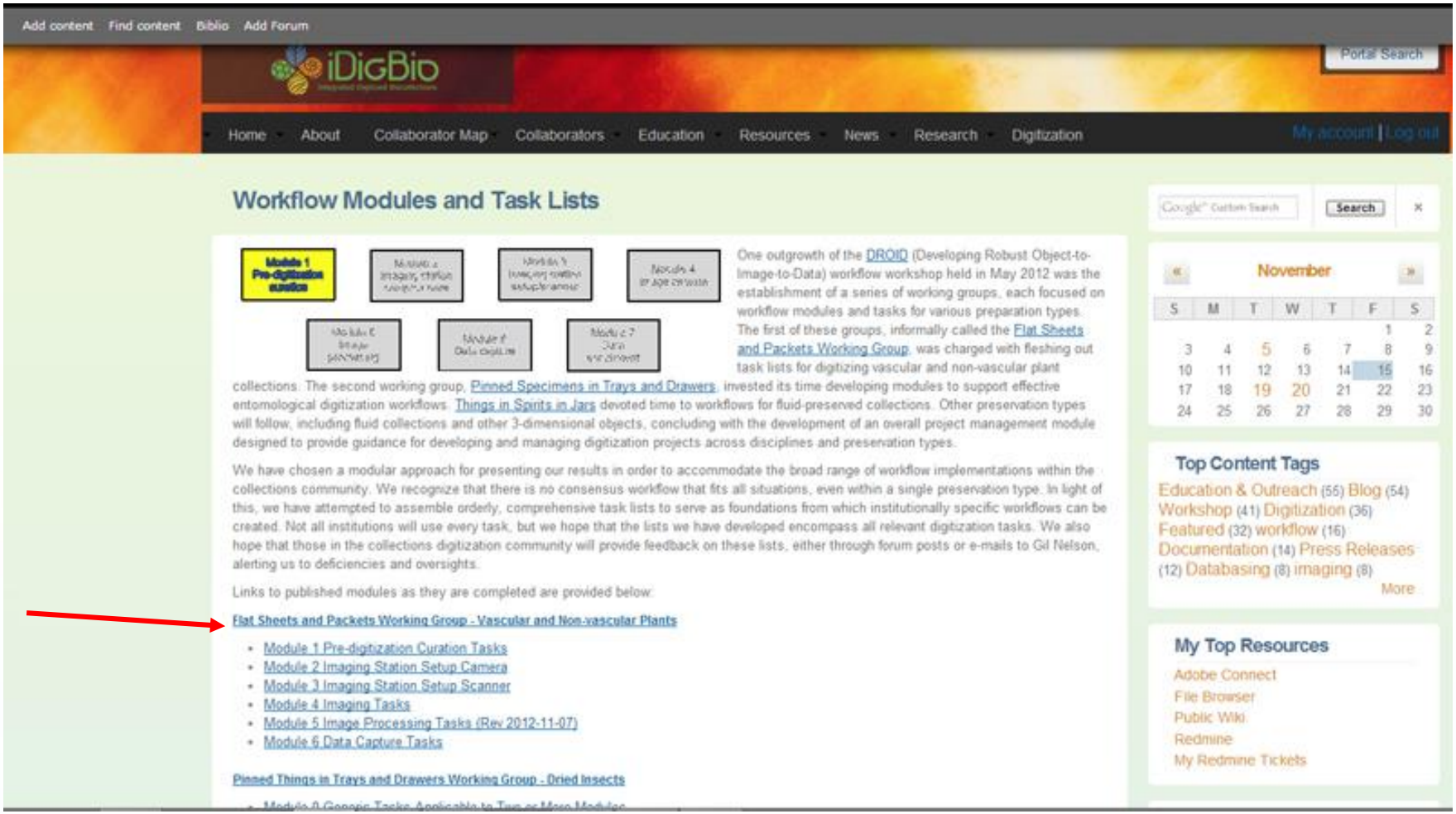
- Facilitate written protocols
- Provide for a modular approach
- Promote efficiency and automation of processes
- Facilitate routing and scheduling of activities
- Provide for balancing workloads
- Facilitate assignment of tasks to technicians
- Ensure that processes are visible and predictable
- Allow 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

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

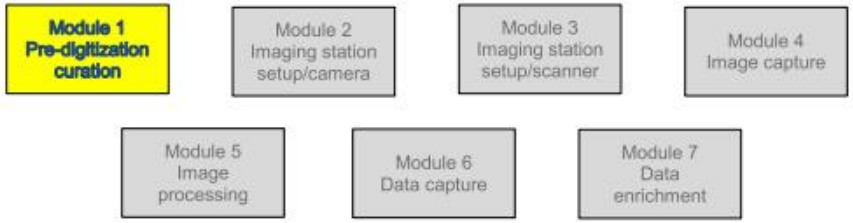
Workflow Modules and Task Lists

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


Add content
Find content
Biblio
Add Forum
iDigBio
Portal Search
Home
About
Collaborator Map
Collaborators
Education
Resources
News
Research
Digitization
My account | Log out
Workflow Modules and Task Lists
Module 1 Pre-digitization curation
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. 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
Search
November
S M T W T F S
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).	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.

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