

Paleontology Digitization Workflows Working Group (DROID)

Gil Nelson

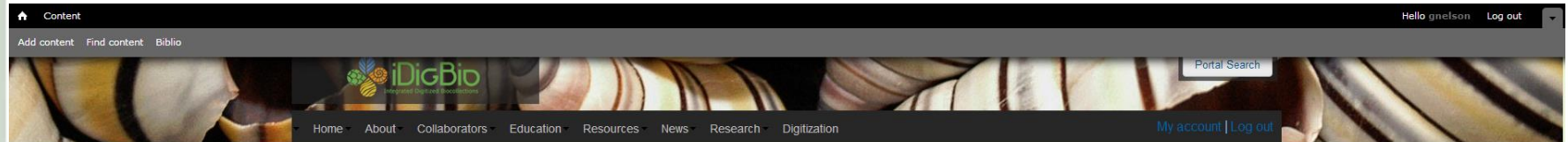
Institute for Digital Information and Scientific Communication
Integrated Digitized Biollections
Florida State University

Paleo Digitization Workshop
23-25 September 2013
Yale Peabody Museum

This material is based upon work supported by the National Science Foundation under Cooperative Agreement EF-1115210. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



Developing Robust Object to Image to Data Workflows (DROID)



- Scientific Software Innovation Institutes
- Yale Peabody Museum
- Biodiversity Institute, KU
- iDigBio

Digitization Workflow Workshop Report



Developing Robust Object-to-Image-to-Data (DROID) Workflow Workshop

30-31st May 2012, Florida Museum of Natural History, University of Florida (FLMNH)

Biological specimens document the historical and modern occurrence of plant and animal species - and most of what we know about the diversity and distribution of life on earth. The majority of collected specimens have yet to be digitized, but at the same time, current biodiversity digitization processes and technologies are often inefficient and uncoordinated, preventing timely and cost-effective digitization of these specimens. This research workshop focused on the design, documentation, and optimization of workflows necessary to transform physical specimens collected in the field into useful, shareable, and manageable

digital objects within a collection. Approximately twenty hands-on collections experts provided input during the workshop.

Why document workflows?

Workflow documentation is a powerful tool both within a collection and across the entire collections community. Internally, effective workflow documentation for a collection can highlight inefficiencies, identify bottlenecks that hinder throughput, and expose opportunities for automation. Workflow documentation also serves as initial input into the development of collections digitization training materials and checklists that improve quality and consistency. Collectively, the documentation and sharing of effective digitization workflows 1) enables collections to test and compare results in order to identify optimal processes, 2) prevents collections from investing resources in (re)designing a process that already exists within the community, 3) enhances communication and standardization by enabling agreement on a common workflow vocabulary for each task, and 4) exposes new innovations to the entire community. Additionally, comprehensive workflow documentation enables the natural history collections community to approach digitization and technology innovators from other domains, such as library sciences, robotics development, industrial workflow design, or software development, for assistance. This includes the ability to present documented workflows to collaborators to learn about improved methods as well as innovative or re-purposed tools.

But we are unique!

The workshop participants recognized that various factors impact the design of appropriate workflows for a particular collection.

- Tradeoffs must be determined at a high level (e.g., volume of objects digitized to text vs. completeness of each record). These decisions may be dependent upon grant requirements or other externally imposed requirements.
- Local decisions and policies may impact a digitization workflow, including institutional or collection policies.
- Specific workflow decisions within a collection will be based upon constraints such as the quantity of personnel, available expertise, available funds, physical layout of the collection space, the method of specimen preservation, and other factors.

To overcome these issues, the DROID workshop participants produced two recommendations. The first was to approach the challenge by developing workflows specific to three broad preservation types, including 1) objects on flat sheets (typically plant specimens), 2) objects on pins (primarily insects), and 3) larger three-dimensional objects (fossils, mammals, reptiles, etc.). Each high-level preservation type has enough similarity that workflows can be developed that have a reasonable number of common tasks. Participants then divided into groups, each focused on the requirements for a specific type.

A second recommendation was to develop more generalized, flexible workflows, with common tasks grouped into "modules" that could be inserted, removed or re-ordered within a collection's workflow based upon the factors described above. Workshop participants were quickly able to

March						
S	M	T	W	T	F	S
		5	6	7		2
3	4					8
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

My Top Resources

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

Upcoming Events

[ASB 2013: Workflows and Challenges in Digitization of Museum Specimens](#)
04-12-2013 to 04-13-2013

[iDigBio Entomology Digitization Workshop \(DROID 2\)](#)
04-24-2013 to 04-25-2013

[2013 Society for the Preservation of Natural History Collections \(SPNHC\) : DemoCamp](#)
06-17-2013 to 06-22-2013

[more events >>](#)

Blog Archives

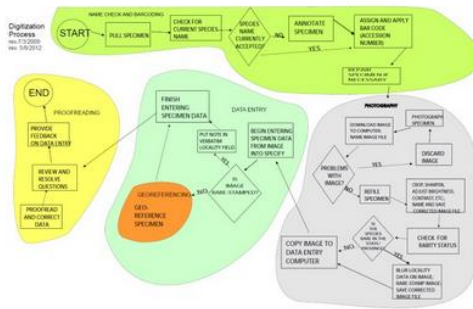
[Hackathon and iConference Update Part II](#)
Post date: 03-01-2013

[Map of Life Collaboration Meeting](#)

Digitization Workflows

Presenter: Dorothy Allard

Digitization Workflows



Efficient and effective workflows are at the heart of successful biological and paleontological collections digitization. Much work has been done with developing workflows and protocols at the museum and collections level, but few of these workflows have been documented or made available to the larger collections community. iDigBio, through its Documentation pages, is establishing an online repository for sharing existing customized workflows from as many collection types and institutions as possible, an idea that stems largely from the [Developing Robust Object-to-Image-to-Data \(DROID\)](#) workshop held May 30-31, 2012. We have assembled an initial set of workflows, including selected examples from the DROID workshop, as well as those developed by iDigBio staff. Here we offer the beginnings of the repository and encourage those in the community to both discuss the workflows via the forum links, and to contribute to this resource by adding new workflows and updating existing workflows. If you would like to submit a workflow for inclusion on this page, please [contact iDigBio](#) for instructions. We are also assembling detailed modules of tasks to be performed at each

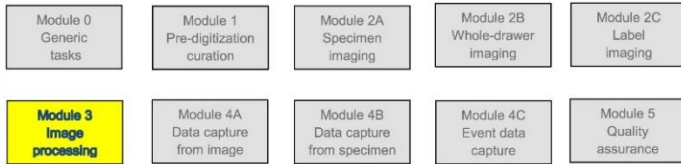
stage of the workflow, accessible on our [Workflow Modules and Tasks page](#).

Global Plants Initiative, U. of Vermont

Workflow	Contributor	Workflow Documentation	Link to Public Comments (Forums)
Dominant Digitization Workflows	iDigBio	Dominant Digitization Workflows Documentation	Dominant Digitization Workflows Forum
Field Notes-to-Data-to-Image	iDigBio	Field Notes-to-Data-to-Image Documentation	Field Notes-to-Data-to-Image Forum
Specimen-to-Data-to-Exemplar Image	iDigBio	Specimen-to-Data-to-Exemplar Image Documentation	Specimen-to-Data-to-Exemplar Image Forum
Object-to-Image-to-Data (1)	iDigBio	Object-to-Image-to-Data (1) Documentation	Object-to-Image-to-Data (1) Forum
Object-to-Image-to-Data (2)	iDigBio	Object-to-Image-to-Data (2) Documentation	Object-to-Image-to-Data (2) Forum
University of Vermont Herbarium	Dorothy Allard	University of Vermont Herbarium Documentation	University of Vermont Herbarium Forum
Southwest Collections of Arthropods Network	Paul Heinrich	Southwest Collections of Arthropods Network Documentation	Southwest Collections of Arthropods Network Forum

Posted To Collaborative Workflows Page Linked to the Digitization Resources Wiki

Workflow Detail: Specimen Image Processing (Pinned Things)



Module 3: Specimen Image Processing

Task ID	Task Name	Explanations and Comments	Resources
T1	Transfer images from camera to immediate image processing storage.	This task varies by institution. Some institutions record images to a card within the camera, others download directly to the imaging computer or an external or network drive as images are recorded. Transfer to the image processing storage should be periodic, at least daily.	Ample storage space with backup procedures (also see T8-T9).
T2	Adjust orientation and crop images, as necessary.	Images should be framed and recorded as precisely as possible to prevent the need for cropping. In cases where cropping is required, batch crop routines for processing multiple images to identical parameters are preferable. Where batch cropping is not possible due to random variation of exemplar image files, individual cropping may	Image management or processing software (e.g., Photoshop, Lightroom, ImageMagick, Gimp, or similar).

University of Florida • Florida Museum of Natural History • Dickinson Hall (Museum Rd. & Newell Dr.) • Gainesville, FL 32611 • 352-273-1906
iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (#EF1115210)

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 working group, [Pinned Specimens in Trays and Drawers](#), invested its time developing modules to support effective entomological digitization workflows. 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](#)
- [Module 1 Pre-digitization Curation Tasks](#)
- [Module 2A Specimen Imaging Tasks](#)
- [Module 2B Whole-drawer Imaging Tasks](#)
- [Module 2C Label Imaging Tasks](#)
- [Module 3 Image Processing Tasks](#)
- [Module 4A Data Capture From Image Tasks](#)
- [Module 4B Data Capture From Specimen Tasks](#)
- [Module 4C Event Data Capture Tasks](#)
- [Module 5 Quality Assurance Tasks](#)

Collections Digitization Workflows

Contents

[hide]

- 1 This Wiki includes links to preparation-specific workflows and protocols for digitizing biodiversity and paleontology collections. The page serves as a community collaboration. Contributions of existing workflows and protocols are encouraged, whether such workflows were developed by the contributor or discovered while searching the internet. Create a free iDigBio account to upload and link your own contributions, or e-mail contributions (links or documents) to Gil Nelson (gnelson@bio.fsu.edu) for uploading and linking. An initial set of stubs is provided. Please expand as needed.
- 2 Digitization Resources Home
- 3 iDigBio's Collaborative Workflows Page
- 4 Herbarium Digitization Workflows and Protocols
- 5 Invertebrate Digitization Workflows and Protocols
- 6 Vertebrate Digitization Workflows and Protocols
- 7 Paleontology Digitization Workflows and Protocols
- 8 Fluid-preserved Specimen Digitization Workflows and Protocols

This Wiki includes links to preparation-specific workflows and protocols for digitizing biodiversity and paleontology collections. The page serves as a community collaboration. Contributions of existing workflows and protocols are encouraged, whether such workflows were developed by the contributor or discovered while searching the internet. Create a free iDigBio account to upload and link your own contributions, or e-mail contributions (links or documents) to Gil Nelson (gnelson@bio.fsu.edu) for uploading and linking. An initial set of stubs is provided. Please expand as needed. [\[edit\]](#)

[Digitization Resources Home](#) [\[edit\]](#)

[iDigBio's Collaborative Workflows Page](#) [\[edit\]](#)

[Herbarium Digitization Workflows and Protocols](#) [\[edit\]](#)

- 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
- Consortium of Pacific Northwest Herbaria imaging workflows
- Imaging Plants, E-Type Initiative, Harvard
- Bryophyte/Lichen Data and Image Capture Workflows (LBCC Thematic Collections Network)

[Invertebrate Digitization Workflows and Protocols](#) [\[edit\]](#)

- A Guide to Digitizing Insect Collections (MCZ Entomology Type Image Project)
- South Australian Museum Procedures Manual Supplement: Macrophotography
- South Australian Museum Procedures Manual Supplement: Microphotography
- Preparing Insect Specimens, E-type Initiative at Harvard
- Imaging Insect Specimens, E-Type Initiative at Harvard

[Vertebrate Digitization Workflows and Protocols](#) [\[edit\]](#)

[Paleontology Digitization Workflows and Protocols](#) [\[edit\]](#)

Community-based Workflow Wiki for sharing workflows across prep types and institutions.

DROID 4 Working Group: 3D objects in boxes and drawers

- **Roger Burkhalter (U. Oklahoma)**
- **Susan Butts (Peabody Museum, Yale)**
- **Una Farrell (U. Kansas)**
- **Talia Karim (Museum of Natural History, U. Colorado)**
- **Ann Molineux (U. Texas)**
- **Gil Nelson (iDigBio)**
- **Jessica Utrup (Peabody Museum, Yale)**

Link Available on the Wiki

https://www.idigbio.org/wiki/images/a/a4/Droid4_Workflow.pdf



iDigBio

Integrated Digitized Biocollections