

# **DROID: Developing Robust Object-to-Image-to-Data Workflows**

A Workshop on the Digitization of Biodiversity Collections

30th - 31st May 2012

*The DROID workshop is organized by Integrated Digitized Biocollections (iDigBio), a National Resource Center at the University of Florida and Florida State University, in collaboration with the Botanical Research Institute of Texas, Yale University, and the University of Kansas. The workshop is supported by the U.S. National Science Foundation's Office of Cyberinfrastructure and Directorate for Biological Sciences, through the Scientific Software Innovation Institutes (S2I2) and Advancing Digitization of Biodiversity Collections (ADBC) Programs.*

## **Overview:**

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. This research workshop addresses the design, documentation, and optimization of Object-to-Image-to-Data workflows for digitizing biological specimens which are curated in thousands of museum and herbarium collections worldwide.

Documenting digitization workflows begins with the recognition of differences that exist between specimen preparation types due to their physical properties and discipline-specific handling, collecting and preservation methods, curatorial and conservation practice, storage environments, data conceptualizations, and data label techniques. Digitizing data recorded on tags tied to vertebrate skins, on labels encircling snakes submerged in solutions of alcohol, on the lilliputian labels of pinned insects, and on the large, verbose labels glued on flat sheets of plant specimens, presents specific constraints and opportunities in each case for efficient digitization workflow design.

## **Goals of the Workshop:**

1. To illustrate and analyze a diversity of existing biological specimen digitization workflows with the aim of gaining a deeper and broader understanding of the practical logistics and efficiencies involved in the handling of biological specimens for the purpose of creating digital database records for publication and for new research applications of the biological, geospatial, and temporal information associated with specimens.
2. To discuss and dissect the dimensions of: digitization project goal definition, the choice of project outcomes and metrics for their assessment, curatorial practice and technology application, human resource and training issues, social and professional values, and the promised deliverables which impact digitization project definition, processes, and success.
3. To engage in the application of lightweight business process modeling (BPM) to create and document reference workflow models for representative disciplines or specimen preservation types with the aim of enabling biological collection curators to implement efficient data capture workflow through comparative analysis and quantitative evaluation of individual workflow steps and tasks.

## **Workshop Objectives:**

1. To review and examine a resplendent set of existing participant collections workflows as case studies, observing constraints, local optimizations, and creative solutions.
2. To gain exposure to workflow design and implementation techniques from libraries and business.
3. To consider how existing or proposed workflows could be enhanced or extended to gain cost efficiency, scalability, and generality (for implementing across additional collections).
4. To identify critical constraints to digitization by discipline or preservation type which represent serious throughput bottlenecks and which may require out-of-the-box solutions and/or redefining digitization project goals or outputs.
5. To identify opportunities for existing or new technology to address costly labor-intensive steps or processing gaps.
6. To examine workflow goals, scope, and procedures for efficiencies of cost, staff utilization, technology, and outputs, and to propose general guidelines for evaluating workflow designs and workflow project success.
7. To identify the synergies of collaborative digitization within TCN workgroups or across innumerable collections within a discipline.
8. To propose near-term project design research and technology development priorities for accelerating the rate of specimen digitization and data publishing.

## **Desired Outcomes:**

1. Formation of a working group to collate work done at this workshop and advance the desired outcomes listed here.
2. To identify best ways, based on existing human resources and technologies, for implementing scalable, efficient solutions for image capture and the integration of label images into data authoring workflows.
3. To document methods for evaluating and quantifying the efficiency of workflow components and tasks, and their suitability/relevance/necessity to the core digitization project goals.
4. To contribute to an annotated web resource illustrating common and divergent digitization tasks, issues, and constraints across disciplines/preparation types.
5. To issue a call to action to identify resources, social and technical approaches, and hardware and software tools to bridge gaps in existing workflow end-to-end integrity.
6. To produce a publication of Workshop findings in Collection Forum, PLoS ONE, and/or appropriate society/discipline journals.

**Schedule:****Day 1, Wednesday, 30 May 2012**

<b>Time</b>	<b>Activity</b>	<b>Owner(s)</b>
9:30AM	Welcome, overview, and brief participant introductions	Jason Grabon Amanda Neill
9:45 AM	Workshop goals and agenda run-through	Chris Norris Jim Beach Deb Paul
10:00 AM	Lightning Round of workflow summaries  5 minutes and 1 slide per presenter (~18 presenters)	Participants
10:30 AM	Coffee break	Pascal's
11:00 AM	Continuation of Lightning Round Group discussion Breakout group definition and self-assignment	Participants
12:30 PM	Box lunch	
1:15 PM	Training session: Workflow Core Concepts (level-set practices, processes, and developing a common terminology)  Q&A Session	Laurie Taylor Mark Sullivan
2:00 PM	Presentation: Workflow Elements and Concepts - Common Practices	Gil Nelson
3:00 PM	Coffee break	Pascal's
3:30 PM	Presentation: Social Issues in Collaborative Digitization	Deb Paul
4:00 PM	Breakout Groups: small groups self-assigned by disciplinary interest to identify and record commonalities and divergences in: <ul style="list-style-type: none"> <li>● Current workflow constraints</li> <li>● Workflow processes across institutions</li> <li>● Metrics &amp; how to measure success</li> <li>● Workflow evaluation matrix</li> </ul>	Breakout Groups & Moderators
5:45 PM	Review of evening activity and Day 2 agenda	Amanda Neill
6:00 PM	Group photo, dinner, and team building activities	

## Day 2, Thursday, 31 May 2012

Time	Activity	Owner(s)
9:30 AM	Review of Day 1, Day 2 agenda summary	Amanda Neill
9:45 AM	Breakout Group reports to the re-assembled Plenary Group	Breakout Groups
10:30 AM	Coffee break	Pascal's
11:00 AM	Pre-Workshop Survey results and discussion	Shari Ellis
11:30 AM	Training session: Business Process Modeling	Brian Anthony
12:30 PM	Breakout Groups reconvene for box lunch and generate one or more redesigned workflows by addressing: <ul style="list-style-type: none"> <li>• What would you change now?</li> <li>• Is a consensus workflow possible for your group?</li> <li>• Is a consensus workflow possible for a single preparation type? For a taxon?</li> <li>• What would you do to optimize these now?</li> </ul>	Breakout Groups & Moderators
3:00 PM	Coffee break	Pascal's
3:30 PM	Plenary: reports back from the Breakout Groups and discussion	Participants
4:30 PM	Vision for the Future. Minority Reports. Out-of-the-box ideas.	Jim Beach
5:00 PM	Plenary wrap-up discussion. DROID Working Group strategy for polishing and dissemination of workshop products.	Amanda Neill Deb Paul Gil Nelson
5:30 PM	Adjourn	