Integrated Digitized Biocollections (iDigBio) An Introduction

Gil Nelson
Institute for Digital Information and Scientific Communication
Integrated Digitized Biocollections
Florida State University

CSIRO
16 July 2013

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.

The U.S. National Science Foundation estimates there may be as many as 1.8 billion biological and paleontological specimens stored in U. S. museums and academic institutions (perhaps as many as 3 billion worldwide). But, no one really knows!

In an effort to make these collections universally accessible to taxonomists, ecologists, researchers, and the general public, in 2011 NSF launched a $100 million, 10-year Advancing Digitization of Biodiversity Collections program and named Florida State University and University of Florida jointly as the national resource for digitization.

Advancing Digitization of Biodiversity Collections

Integrated Digitized Biocollections (iDigBio)
University of Florida
Florida State University
Florida Museum of Natural History

The goal is to digitize and make available via the Web at least 1 billion biological and paleontological records over the 10-year life of the project.

iDigBio
Integrated Digitized Biocollections

Mandate and Responsibility

- Provide/facilitate portal access to collections data
  - Make information available and discoverable
  - Label Data and images
- Enable digitization and research
  - Facilitate digitization workflows
  - Oversee implementation of standards and best practices for digitization
  - Allow for data discovery across organismal groups
- Be a client of digitization projects/networks
  - Actively seek partners and data sources
  - Respond to cyberinfrastructure needs
- Engage communities
  - Collections
  - Research
  - Citizen science and education
- Support ADBC goals
  - Access to information
  - Support for collections
  - Sustainability

Mandate and Responsibility

- Provide/facilitate portal access to collections data
  -
  -
- Enab
  -
  -
  -
- Be a
  -
  -
- Enga
  -
  - Research
  - Citizen science and education
- Support ADBC goals
  - Access to information
  - Support for collections
  - Sustainability

Grand Challenge

Develop a cloud computing infrastructure that links biological data from collections across the U.S. through one or more unified web interfaces to overcome the limitations of "data silos."

Mandate and Responsibility

- Provide/facilitate portal access to collections data
  -
  -
- Enab
  -
  -
  -
- Be a
  -
  -
- Enga
  -
  - Research
  - Citizen science and education
- Support ADBC goals
  - Access to information
  - Support for collections
  - Sustainability

Grand Challenge

Devel ture
that li ons
across
unifie ne
limitations of "data silos."

More recently, we have been encouraged by NSF to establish international collaboration and sharing.

The challenges being pursued by iDigBio are reflective of worldwide trends in digitization

- Global Biodiversity Informatics Facility (GBIF)
- OpenUp! (European Union)
- Atlas of Living Australia (ALA)
- SYNTHESYS (20 European natural museums)
- CRIA (Brazil)

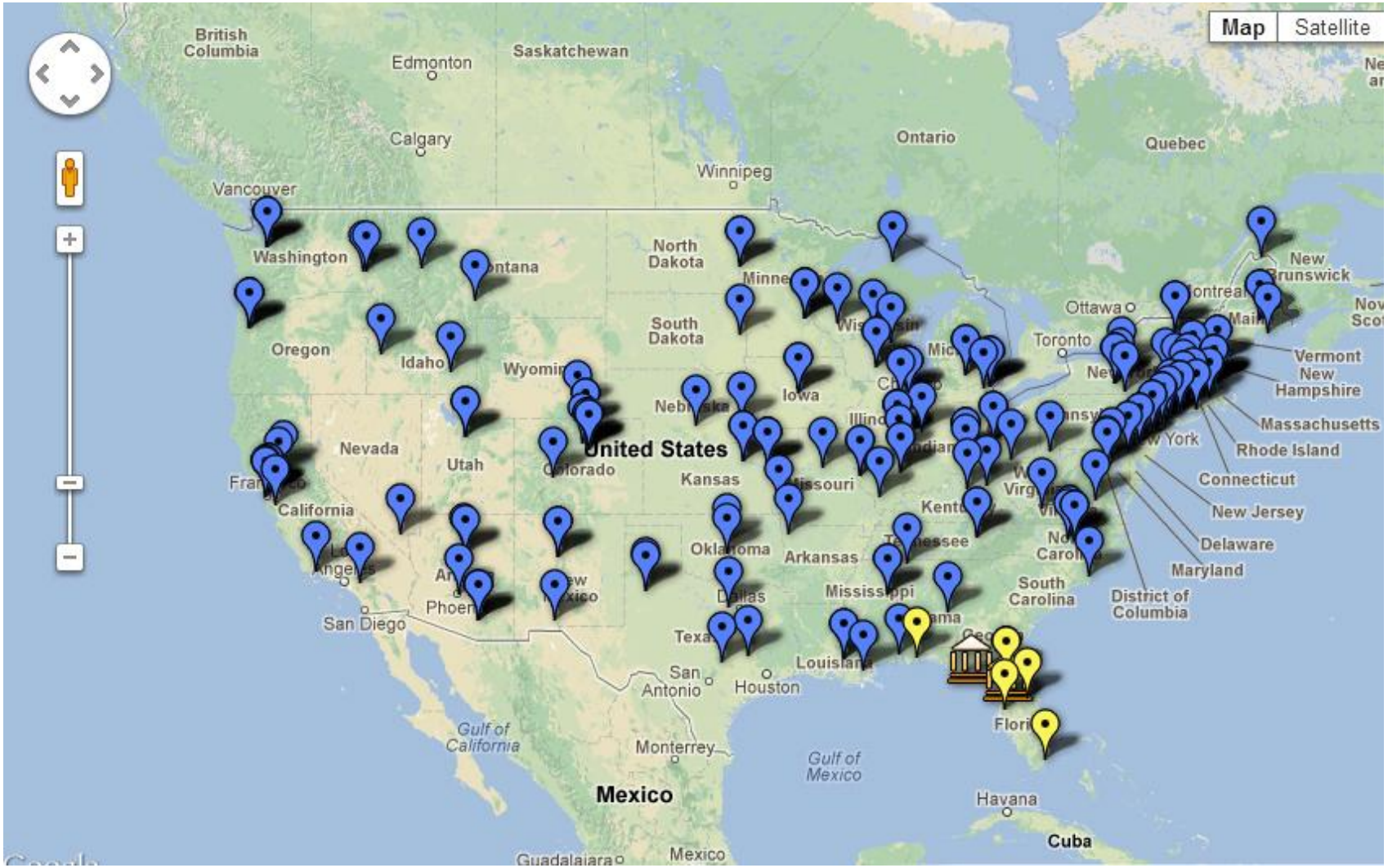
Ten Thematic Collections Networks (TCNs) plus 2 Partner to Existing Networks (PENs)

- InvertNet: An Integrative Platform for Research on Environmental Change, Species Discovery and Identification (*Illinois Natural History Survey, University of Illinois*) http://invertnet.org
- Plants, Herbivores, and Parasitoids: A Model System for the Study of Tri-Trophic Associations (*American Museum of Natural History*) http://tcn.amnh.org
- North American Lichens and Bryophytes: Sensitive Indicators of Environmental Quality and Change (*University of Wisconsin – Madison*) http://symbiota.org/nalichens/index.php http://symbiota.org/bryophytes/index.php (plus 2 PENs)
- Digitizing Fossils to Enable New Syntheses in Biogeography - Creating a PALEONICHES-TCN (*University of Kansas*)
- The Macrofungi Collection Consortium: Unlocking a Biodiversity Resource for Understanding Biotic Interactions, Nutrient Cycling and Human Affairs (*New York Botanical Garden*)
- Mobilizing New England Vascular Plant Specimen Data to Track Environmental Change (*Yale University*)
- Southwest Collections of Anthropods Network (SCAN): A Model for Collections Digitization to Promote Taxonomic and Ecological Research (*Northern Arizona University*) http://hasbrouck.asu.edu/symbiota/portal/index.php

<u>New as of 1 July 2013</u>

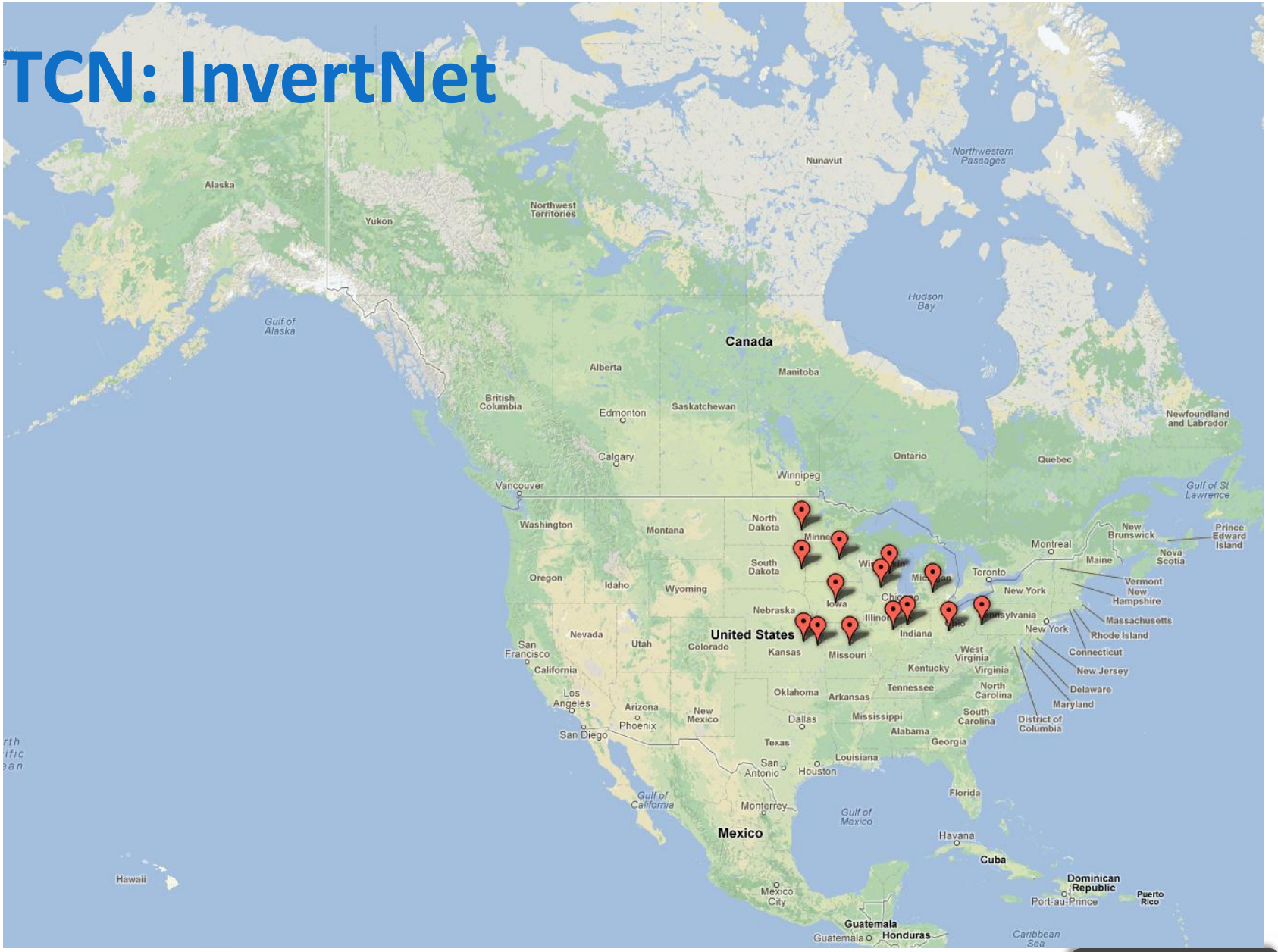
- ?
- ?
- ?

National Resource (iDigBio), Thematic Collection Networks (TCNs)

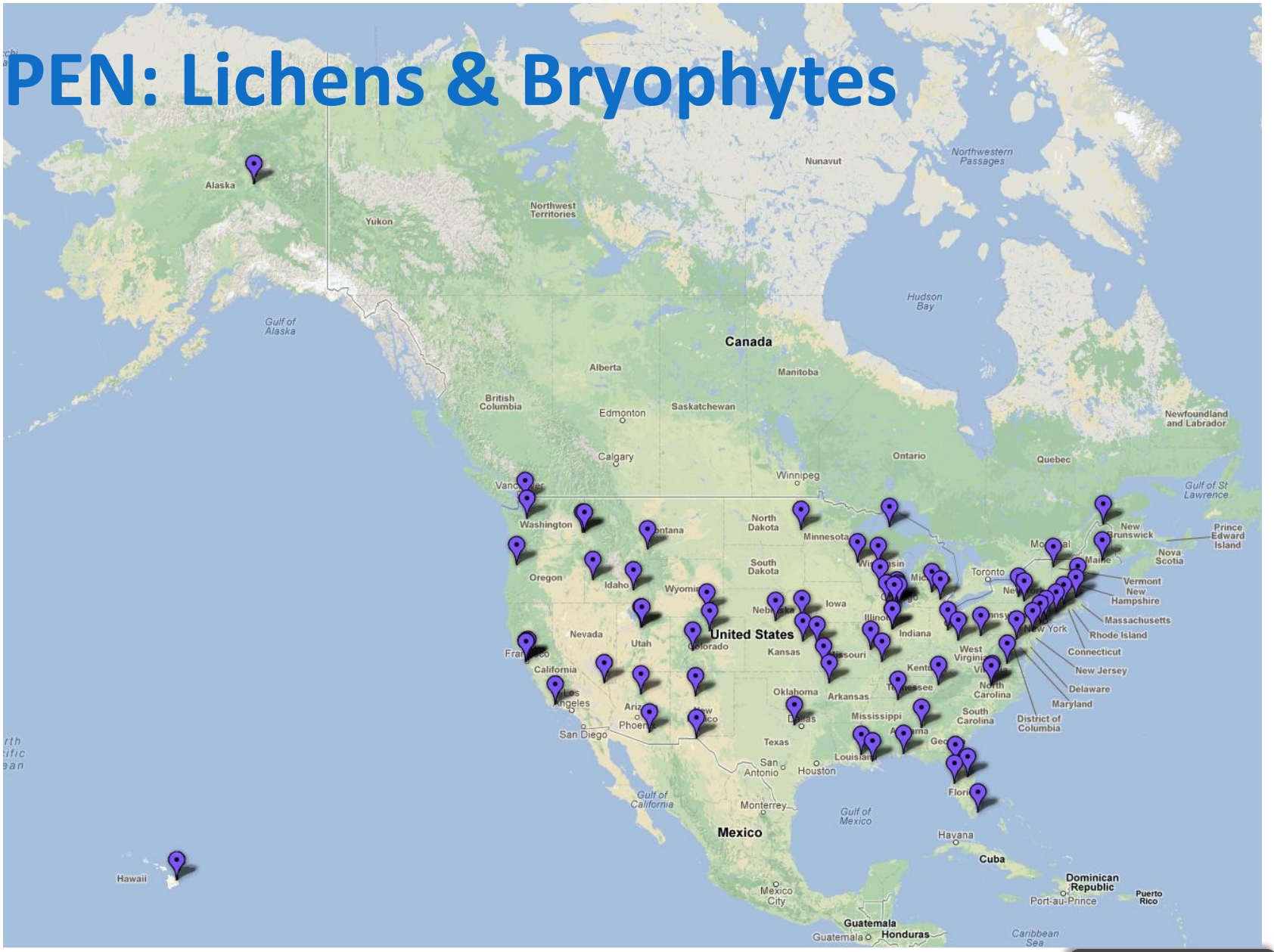


To date: 10 TCNs, 2 PENs, 160+ participating institutions, 49 states

TCN: InvertNet

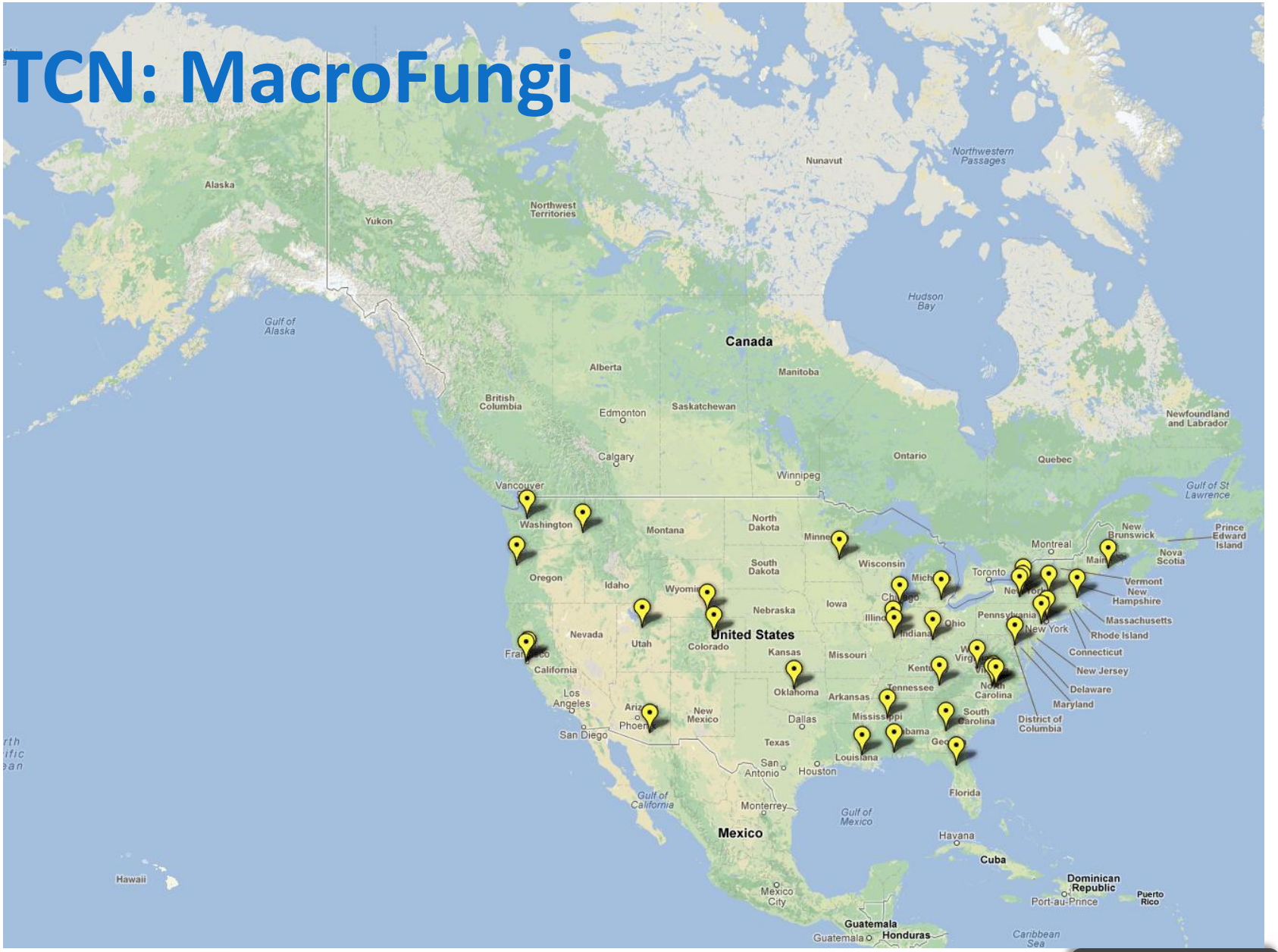


PEN: Lichens & Bryophytes



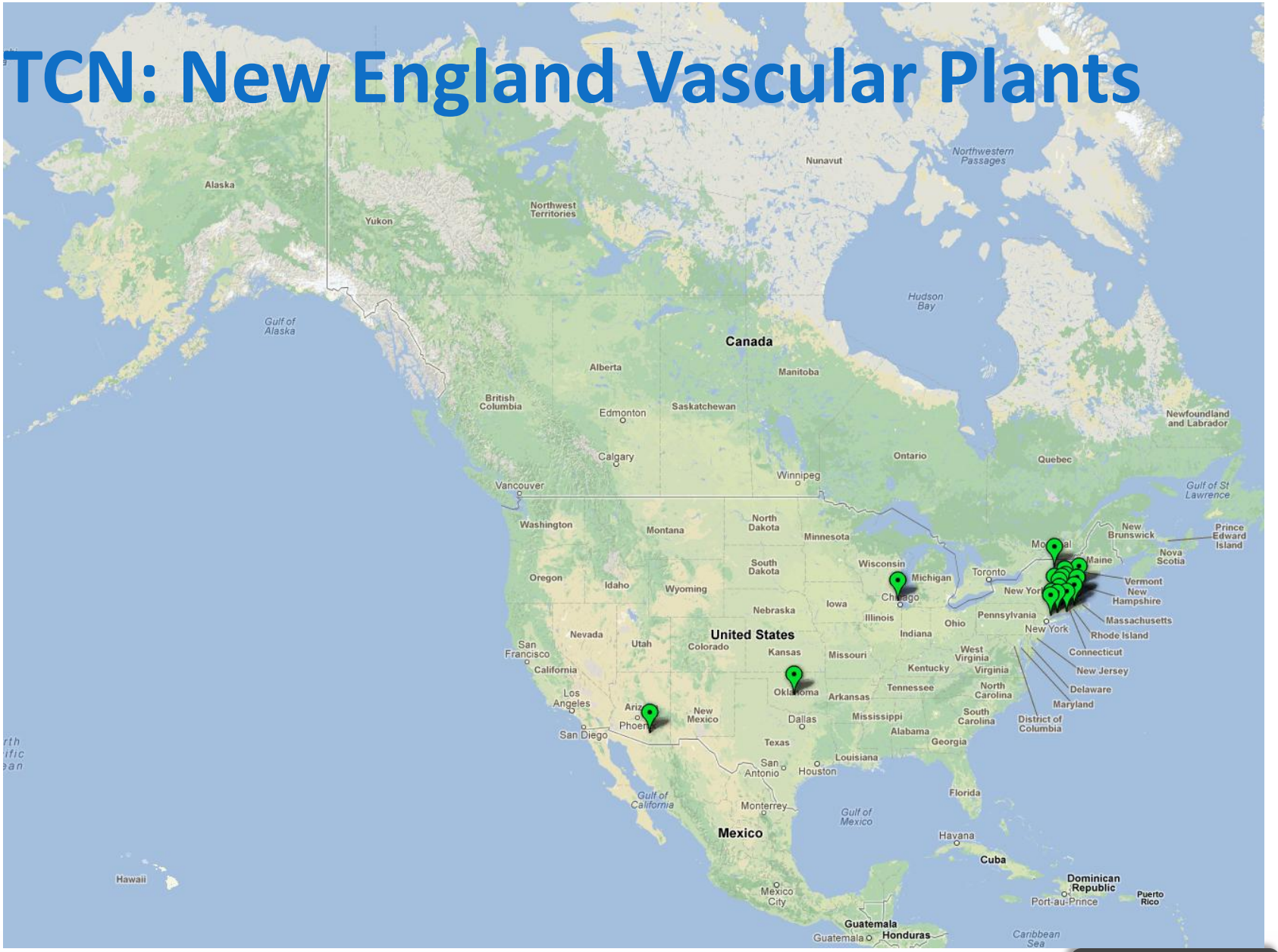
iDigBio
Integrated Digitized Biocollections

TCN: MacroFungi



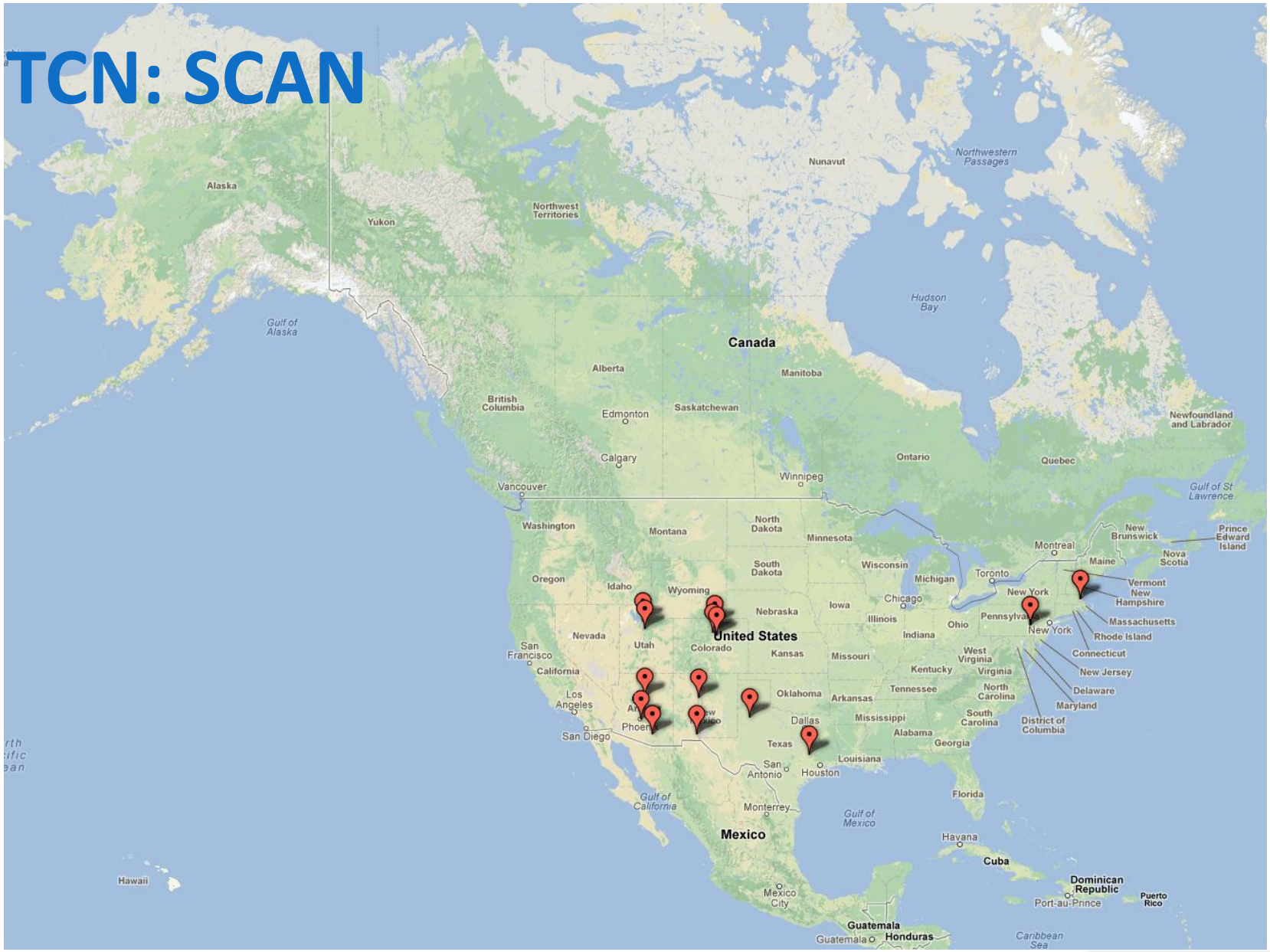
iDigBio
Integrated Digitized Biocollections

TCN: New England Vascular Plants



iDigBio
Integrated Digitized Biocollections

TCN: SCAN


Alaska
Yukon
Northwest Territories
Nunavut
Northwestern Passages
Hudson Bay
Gulf of Alaska
Canada
Alberta
Manitoba
British Columbia
Edmonton
Saskatchewan
Calgary
Ontario
Quebec
Newfoundland and Labrador
Vancouver
Winnipeg
Gulf of St Lawrence
Washington
Montana
North Dakota
Minnesota
New Brunswick
Prince Edward Island
Montreal
Nova Scotia
Maine
Oregon
Idaho
Wyoming
South Dakota
Wisconsin
Michigan
Toronto
Vermont
New Hampshire
New York
Chicago
Iowa
Nebraska
Illinois
Pennsylvania
Massachusetts
Ohio
Nevada
Utah
Colorado
United States
Kansas
Missouri
Indiana
Rhode Island
San Francisco
California
West Virginia
Connecticut
Kentucky
Virginia
New Jersey
Los Angeles
Oklahoma
Arkansas
Tennessee
North Carolina
Delaware
Maryland
San Diego
Dallas
Mississippi
South Carolina
District of Columbia
Alabama
Georgia
Texas
Louisiana
San Antonio
Houston
Florida
Gulf of California
Monterrey
Gulf of Mexico
Mexico
Havana
Cuba
Hawaii
Dominican Republic
Port-au-Prince
Puerto Rico
Mexico City
Guatemala
Honduras
Guatemala
Caribbean Sea

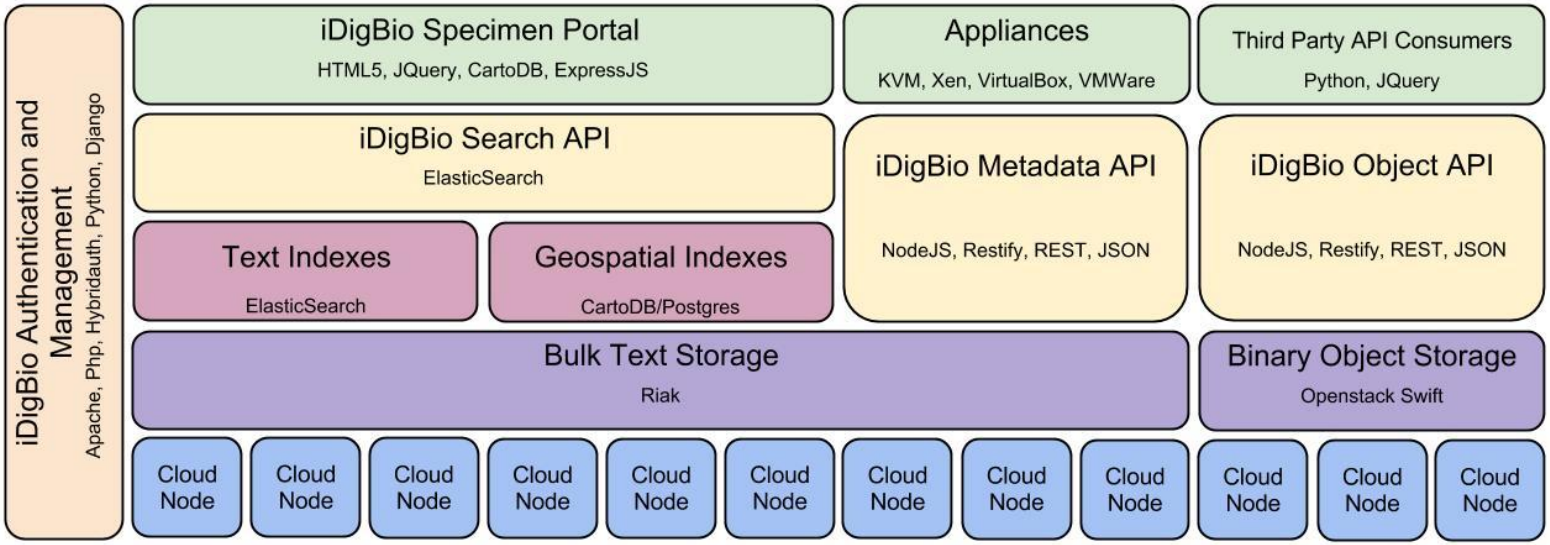
TCN: PALEONICHES

iDigBio
Integrated Digitized Biocollections

TCN: Tri-Trophic

Building the iDigBio Cloud

- Cloud-based strategy
  - Providing useful services/APIs (programmatic and web-based Application Programming Interface)
  - Federated scalable object storage and information processing
  - Digitization-oriented virtual appliances
  - Reliance on standards, proven solutions, and sustainable software
- Continuous consultation with stakeholders
  - Surveys, working groups, interest groups, workshops, person-to-person



Key Features of iDigBio

- Ingest all contributed data with emphasis on use of GUIDs, no restrictions
- Maintain persistent datasets and versioning, allowing new and edited records to be uploaded as needed while preserving existing records
- Ingest textual specimen records, plus associated still images, video, audio, and other media (or links to these resources as determined by the provider)
- Ingest linked documents and associated literature, including field notes, ledgers, monographs, related specimen collections, etc.
- Provide virtual annotation capabilities and track annotations back to the originating collection (collaborating with FilteredPush)
- Facilitate sharing and integration of data relevant to biodiversity research
- Provide computational services for biodiversity research

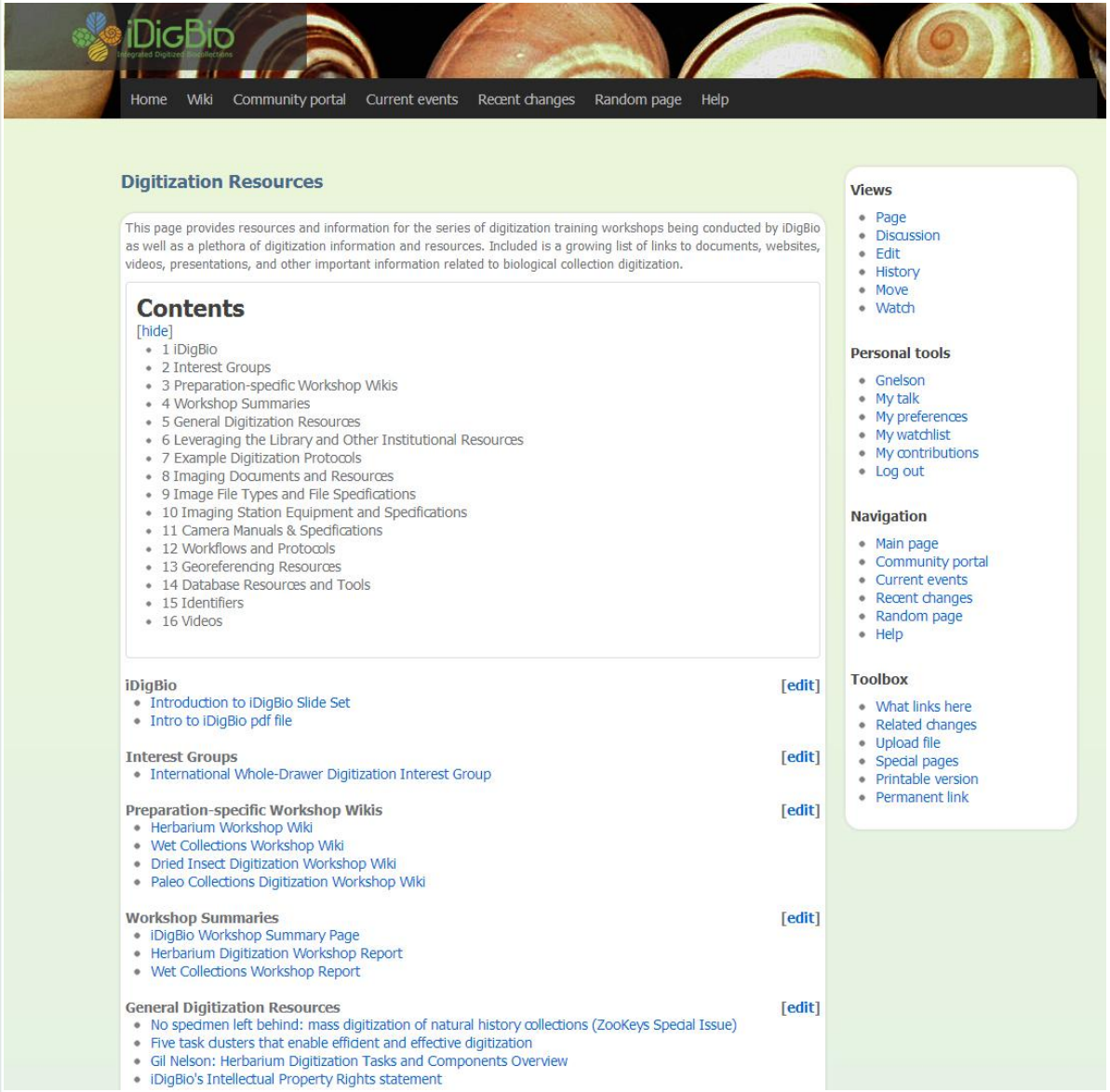
Recent, Ongoing, Upcoming Activities

- Assessment of common and effective digitization practices (paper in *ZooKeys*)
- Working groups
  - Minimum information for scientific collections working group (MISC)
  - Digitization workflows working groups
  - Georeferencing
  - Optical character recognition (OCR)
  - Biodiversity Informatics Manager working group
- Workshops - year 2:
  - > 150 institutions, 9 workshops, 3 symposia
  - 368 sponsored participants
  - Video archives on Vimeo, live streaming for remote participation
  - New model this year: train the trainer
  - Series of digitization training workshops (herbaria, wet collections, entomology, paleontology, fluid-preserved invertebrate imaging, small herbaria,)
- Server hosting: 8 virtual machines, TCN support
- Specimen data portal and website – continuous improvements
- Call for appliances, frequent opinion surveys

Digitization Workshops

In March 2012, the Steering Committee established a series of preparation-specific digitization training workshops focused on helping collections managers get started with and/or enhance local digitization programs, all to be held at host institutions.

- DROID (Developing Robust Object->Image->Data, May 2012)
- Herbarium digitization (Valdosta State, September 2012)
- Fluid-preserved collections digitization (U. Kansas, March 2013)
- Dried insect collections digitization (Field Museum, April 2013)
- Collections Digitization (West Virginia, ASB, April 2013)
- Imaging fluid-preserved invertebrates (U. Michigan, September 2013)
- Paleontology digitization (Yale Peabody Museum, September 2013)
- Small Herbarium Digitization (Florida State University, December 2013)
- Broadening Biodiversity in the Biodiversity Sciences (January, 2014)
- Original Source Materials Digitization (Yale Peabody, March 2014)

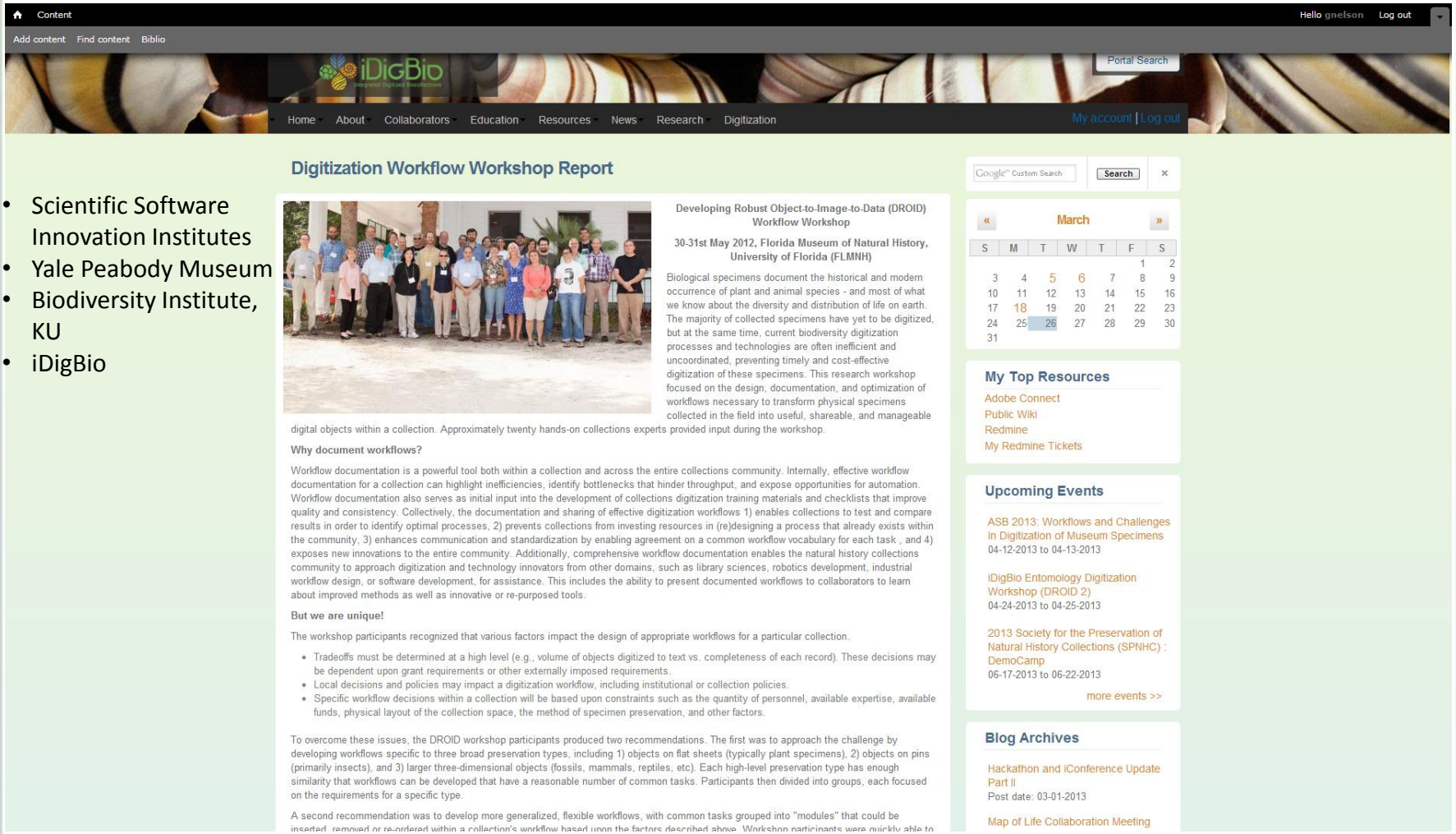


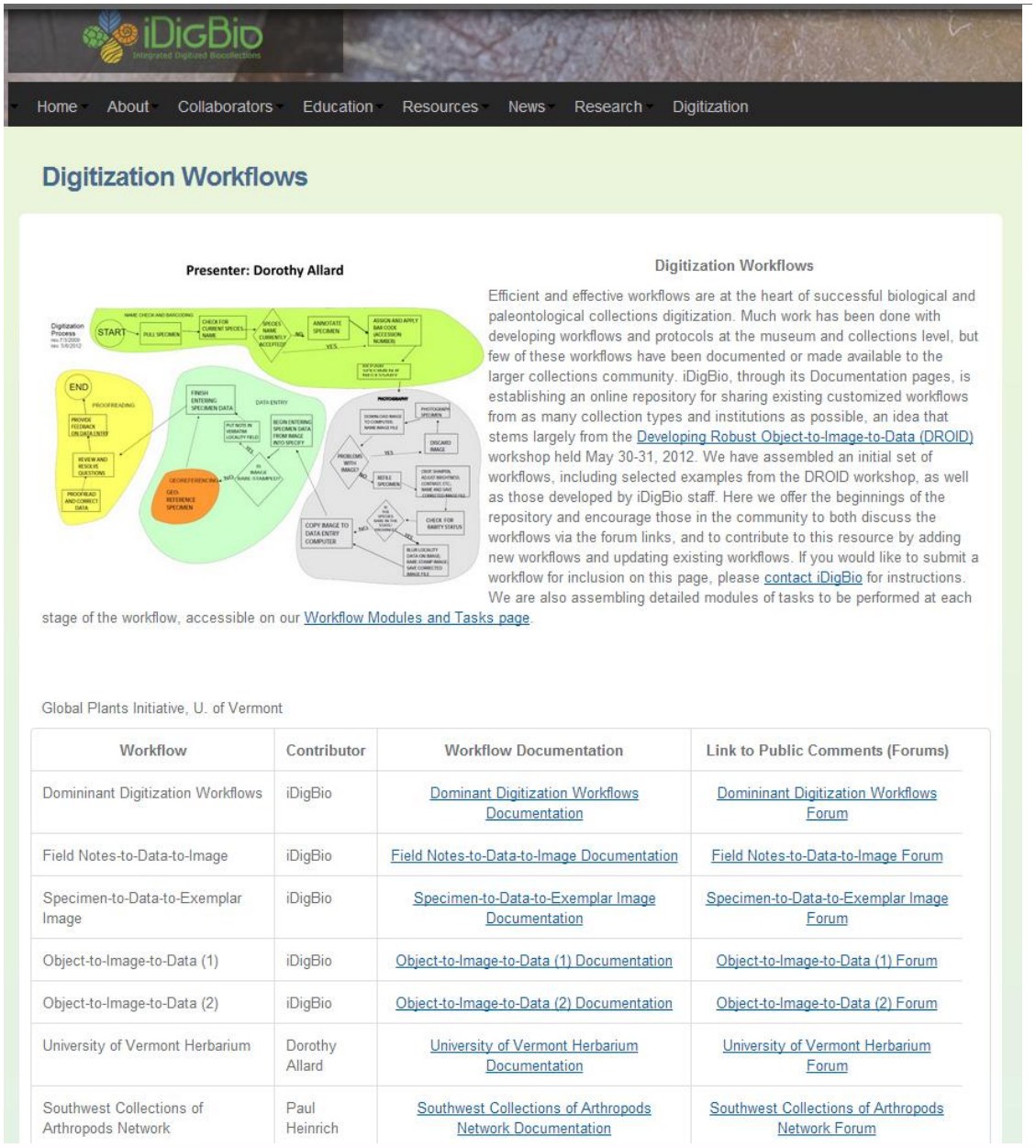
Developed a community-oriented digitization resources wiki in support of our workshops and to serve digitization-related information across all preparation types.

Established a digitization list serv to promote workshop follow-up as well as community discussion and sharing.

Developing Robust Object to Image to Data Workflows (DROID)

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

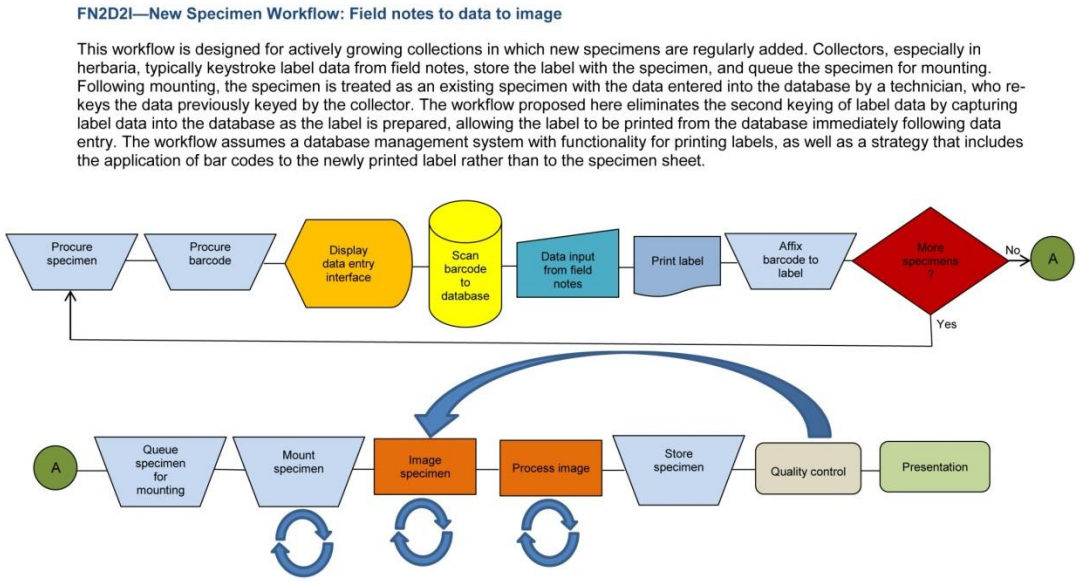




Posted To Collaborative Workflows Page Linked to the Digitization Resources Wiki

DROID Workflows Working Groups

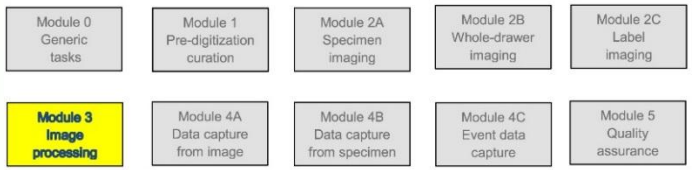
- The Flat Sheets and Packets Working Group completed modules and associated tasks for herbarium and related collections (October 2012).
- The Pinned Things in Trays and Drawers finished and posted its work for entomology (January 2013).
- 3D Objects in Spirits in Jars and Vials completed and posted its workflows for fluid-preserved specimens (May 2013).
- 3D Objects in Drawers and Trays workflows group in process now.
- Preparation-independent workflows to follow (2013).



Modular Approach

iDigBio

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)

FLORIDA MUSEUM OF NATURAL HISTORY · THE FLORIDA STATE UNIVERSITY · UF UNIVERSITY of FLORIDA

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:

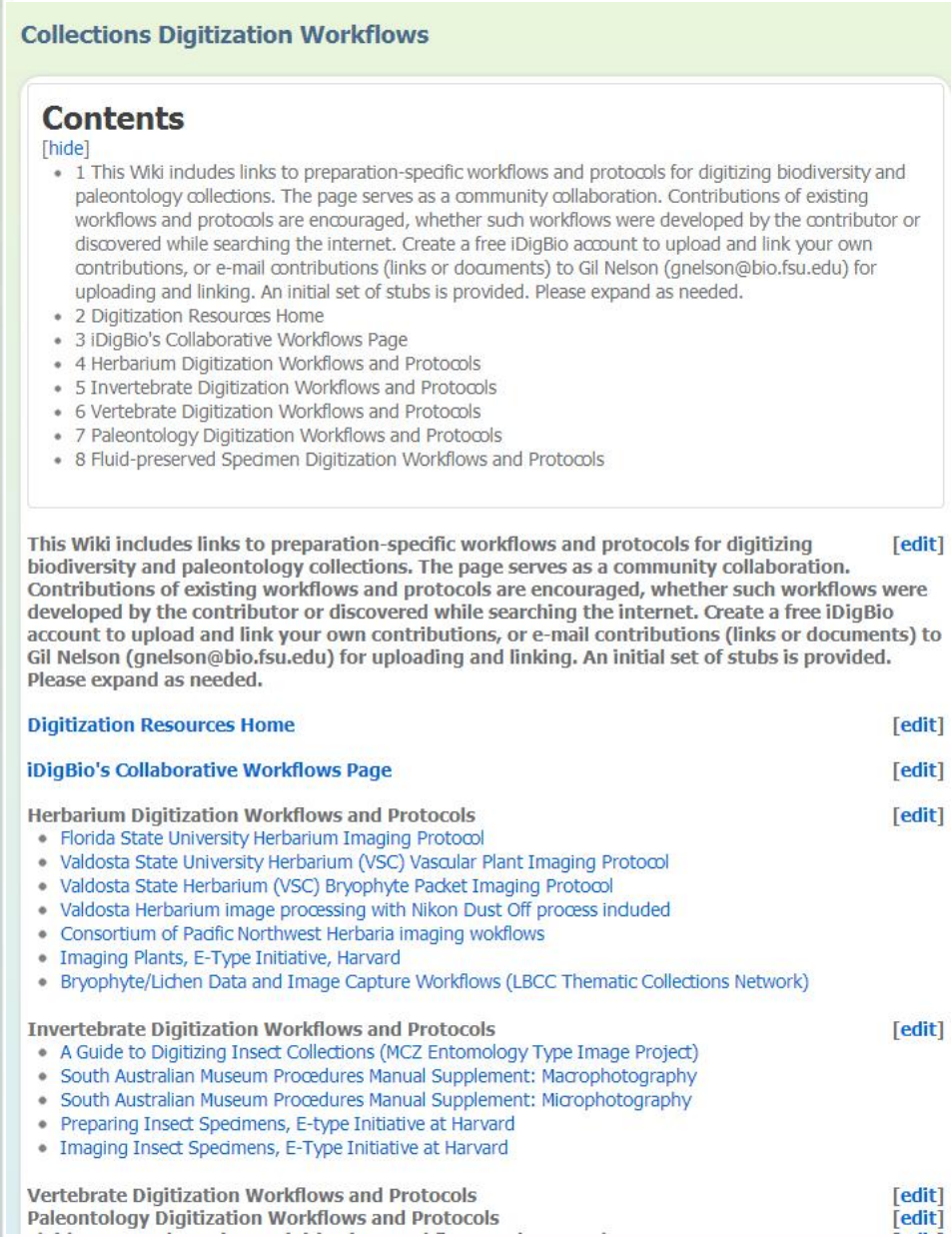
Flat Sheets and Packets Working Group - Vascular and Non-vascular Plants

- 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

iDigBio Integrated Digitized Biocollections



Collections Digitization Workflows

Contents

[hide]

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 wokflows
- Imaging Plants, E-Type Initiative, Harvard
- Bryophyte/Lichen Data and Image Capture Workflows (LBCC Thematic Collections Network)

Invertebrate Digitization Workflows and Protocols [edit]

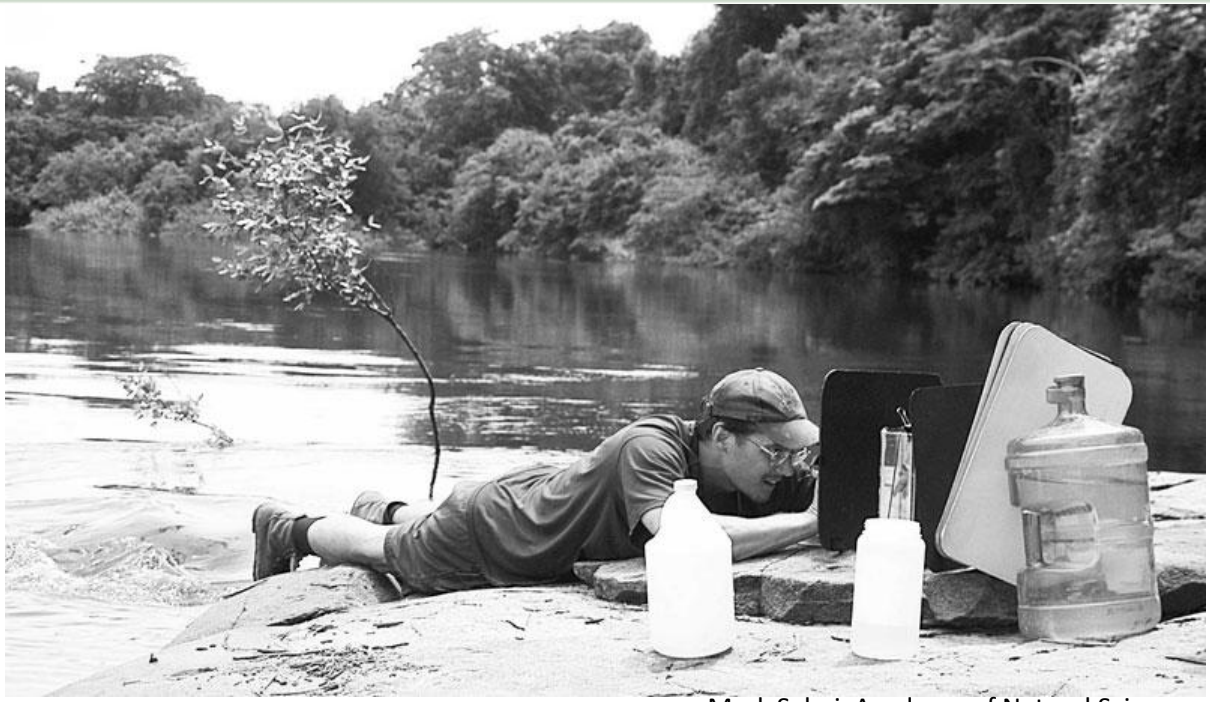
- A Guide to Digitizing Insect Collections (MCZ Entomology Type Image Project)
- South Australian Museum Procedures Manual Supplement: Macrophotography
- South Australian Museum Procedures Manual Supplement: Microphotography
- Preparing Insect Specimens, E-type Initiative at Harvard
- Imaging Insect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

Proactive Digitization



Mark Sabaj, Academy of Natural Sciences

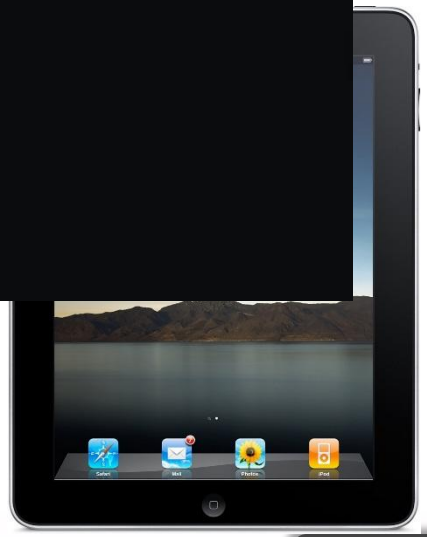
With maturity of experience (including collectors) and technology, digitization activities will move to the field—to include electronic data gathering, Georeferencing, field imaging, etc.—which may reduce the workload for collections managers.

Proactive Digitization

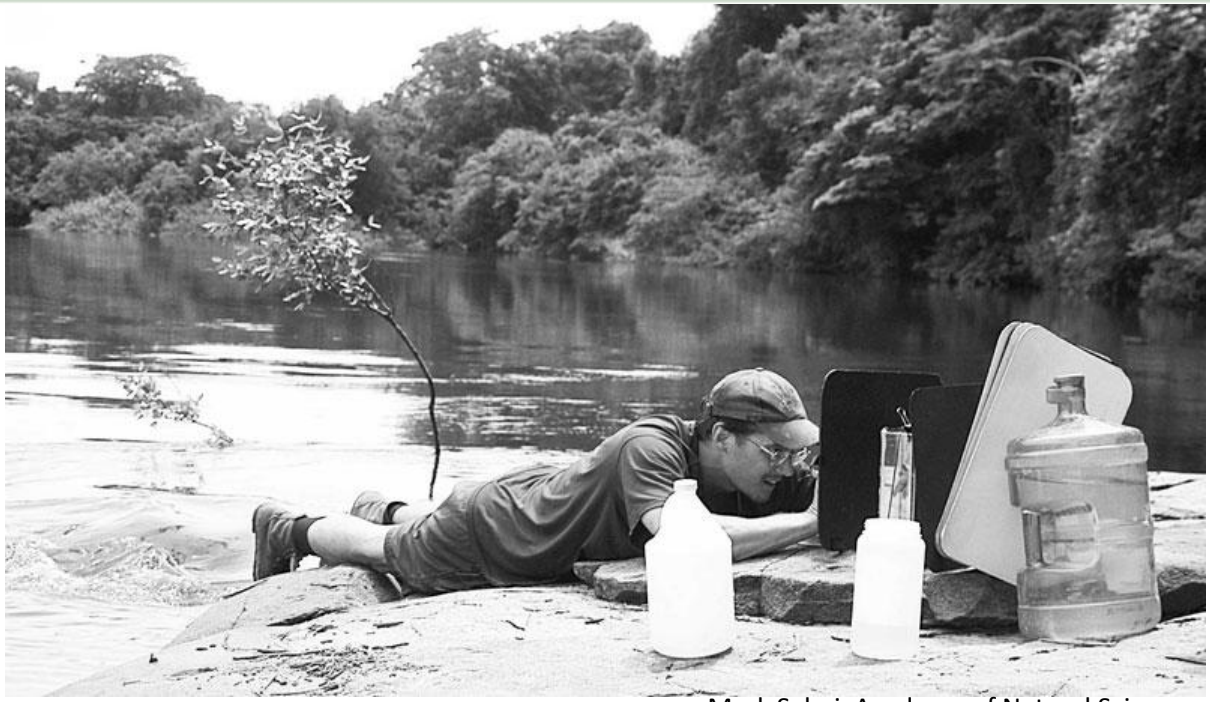
With mat
(including
technolog
activities
field—to include electronic data gathering, Georeferencing, field imaging, etc.—which may reduce the workload for collections managers.

Proactive Digitization

With mat
(including
technolog
activities
field—to i
data gath
Georeferencing, field imaging, etc.—which may reduce the workload for collections managers.



Proactive Digitization



Mark Sabaj, Academy of Natural Sciences

With maturity of experience (including collectors) and technology, digitization activities will move to the field—to include electronic data gathering, Georeferencing, field imaging, etc.—which may reduce the workload for collections managers.

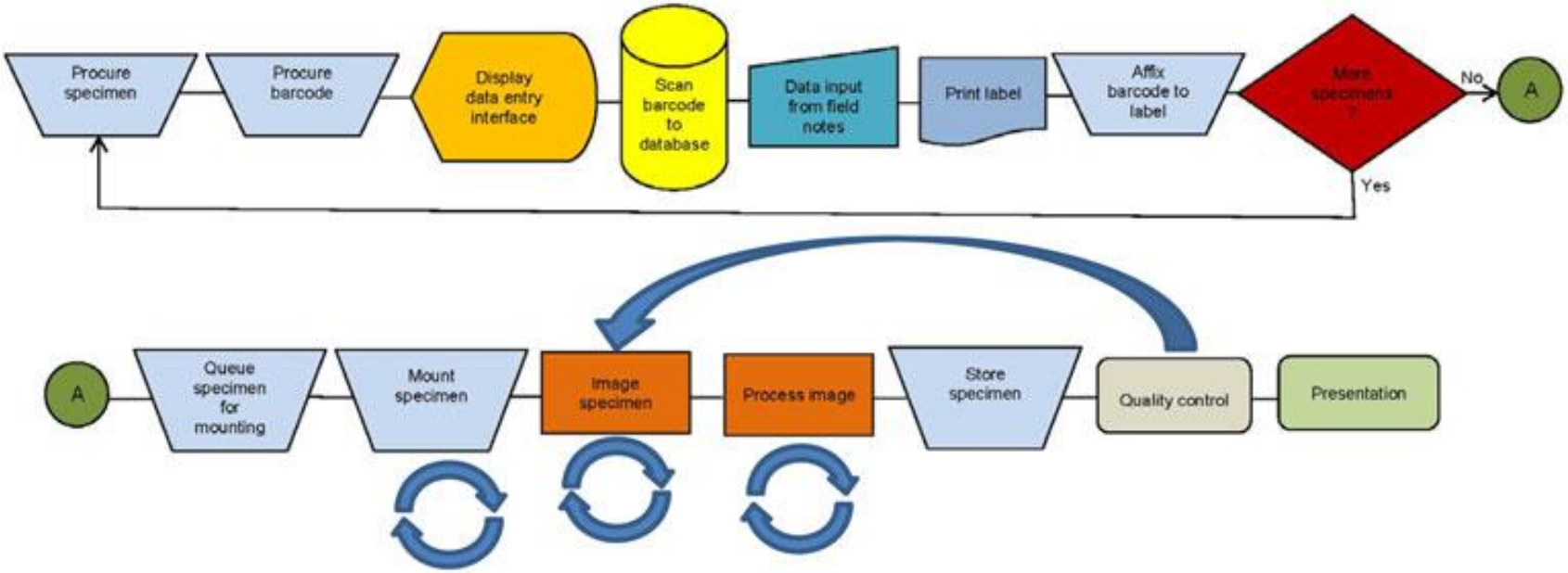
Module 3: Proactive Digitization, Fluid Preserved

Task ID	Task Name	Explanations and Comments	Resources
T1	Collect object in field.	Much of current digitization practice focuses on the conversion of existing (often called legacy) data and specimens to digital format. More recently, collectors have begun to capture at least some digital data in the field to submit along with newly collected specimens, resulting in the existence of digitized information prior to accessioning and cataloging. Though this "proactive digitization" process is likely to be modified and improved over time, the current module addresses tasks for dealing with these data. It should be noted that standards and processes for electronic field data collection are often beyond the control of curators and collections managers. Nevertheless, the establishment of such standards and processes is important for rapidly processing pre-existing digital data to ensure smooth transition and synchronization of these data with collections databases and image repositories.	Institutionally specific digital field data collection guidelines.

T2	Utilize pre-formatted spreadsheet for field data collection.	This process requires negotiation and discussion with and potentially individualized training and orientation for collectors. Utilizing pre-formatted data-input instruments will benefit the collector and collections managers by ensuring faster processing of newly received collections and shorter times from date-of-collection to online availability of images and data. Pre-formatted spreadsheets should be developed by collections staff for dissemination to cooperating collectors to ensure automation and synchronization between field data instruments and database importing. Collectors electronically populate partial or complete digital records while in field, potentially to include: • locality description, • field identifier, • date collected, • collector name, • georeference (or related identifier from GPS device for later repatriation), • datum, and • taxonomic determination. Note: georeference coordinates might exist within a GPS device separate from the handheld computer or tablet into which other collection data are entered,	• Institutionally specific field collection protocol. • Pre-formatted field data collection instrument. • Method for receiving and linking GPS waypoint data.

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.



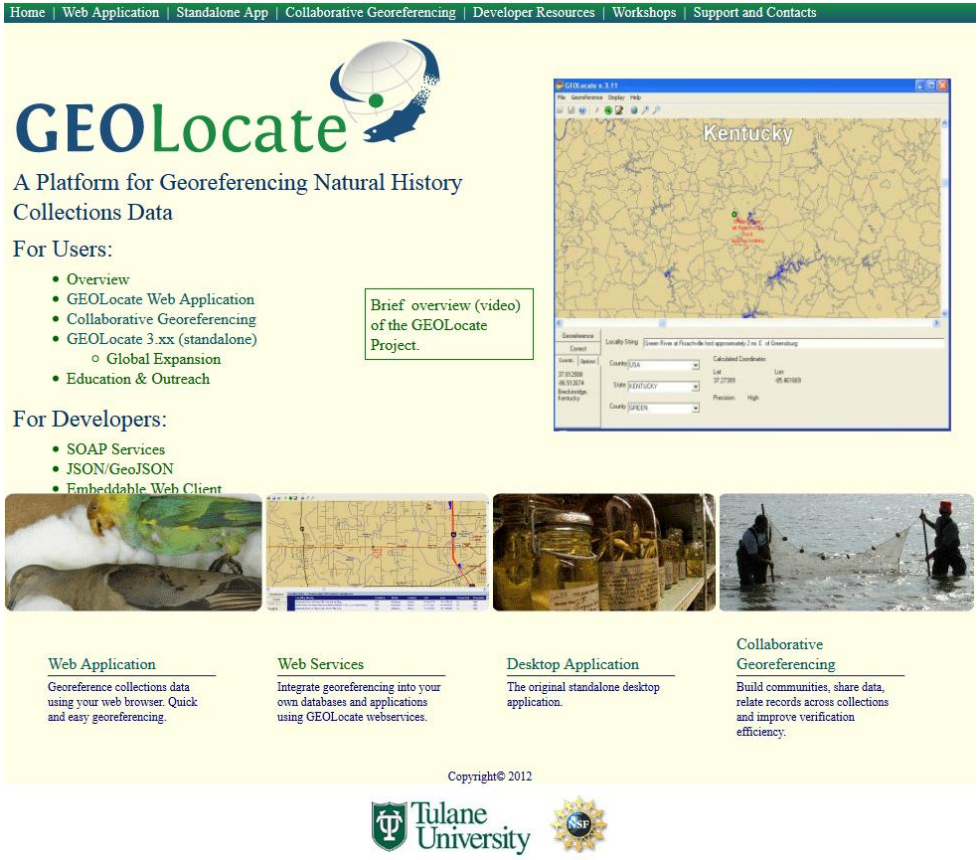
Launched the **Biodiversity Informatics Managers Working Group** to focus on the role of biodiversity informatics manager as an essential component underpinning the successful digitization enterprise, including the definition and delineation of career path dimensions, skill sets, academic training requirements, and recommendations about the placement of this role within the organizational structure of museums and academic institutions.

Launched the **International Whole-Drawer Digitization Interest Group** in collaboration with partners at CSIRO, with representatives from Australia, Germany, The Netherlands, the United Kingdom, and the United Kingdom.

Augmenting OCR Working Group

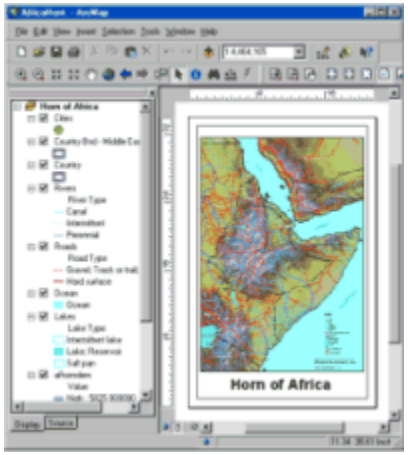
- Born of community interest.
- Designed to explore and enable uses of OCR for specimen digitization.

Georeferencing Working Group
Train the Trainers



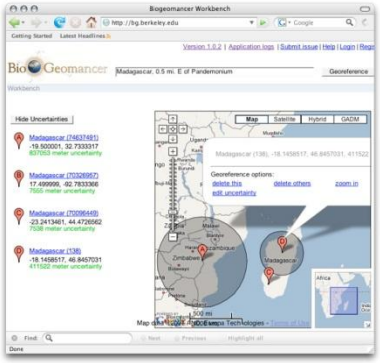
Georeferencing Methods and Tools

ArcGIS



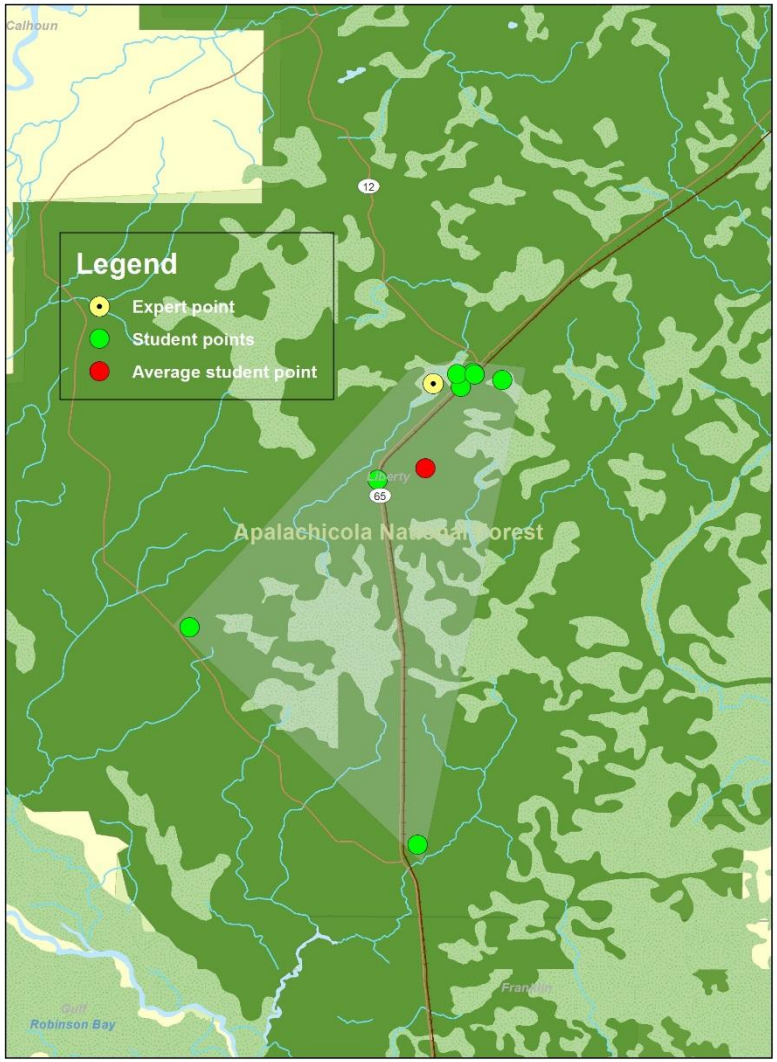
Google Earth/Maps

Biogeomancer



Geolocate

GEOLocate



Experiment under way at Florida State University:

How much training does it take to accurately georeference a locality?

1. Trained techs with regular communication.

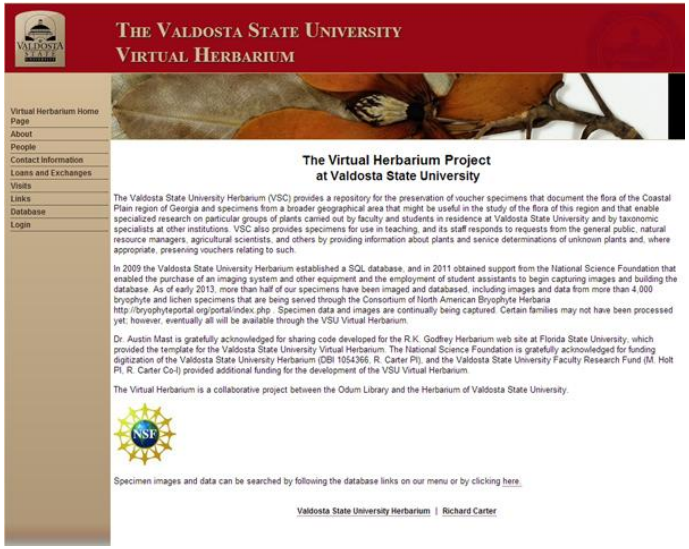
2. Plant biology students with about 30 minutes of training.

Apalachicola National Forest.
Verbatim Locality: titi bog, Appalachia National Forest, near Wilma.
Habitat: in a sphagnous area, presently dry, titi bog.

iDigBio
Integrated Digitized Biocollections

Prompted by a workshop at the Association of Southeastern Biologists, established the **Small Herbarium Working Group** (< 100k specimens) to focus on coordination, collaboration, and common issues related to digitizing small herbaria. We view these as important, often singular collections of great value.

In the process of establishing the **Library Sciences Collaborative Working Group** to explore how the information management and library communities can coordinate and collaborate with the biodiversity collections community in digitization and data distribution.



Minimum Standards for Scientific Collections (MISC)

Gil Nelson
Institute for Digital Information and Scientific Communication
Integrated Digitized Biocollections
Florida State University

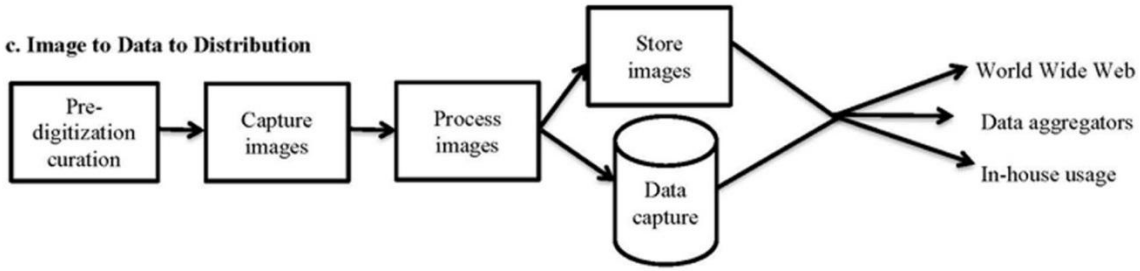
CSIRO
16 July 2013

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.

Two things we knew we needed at the outset:

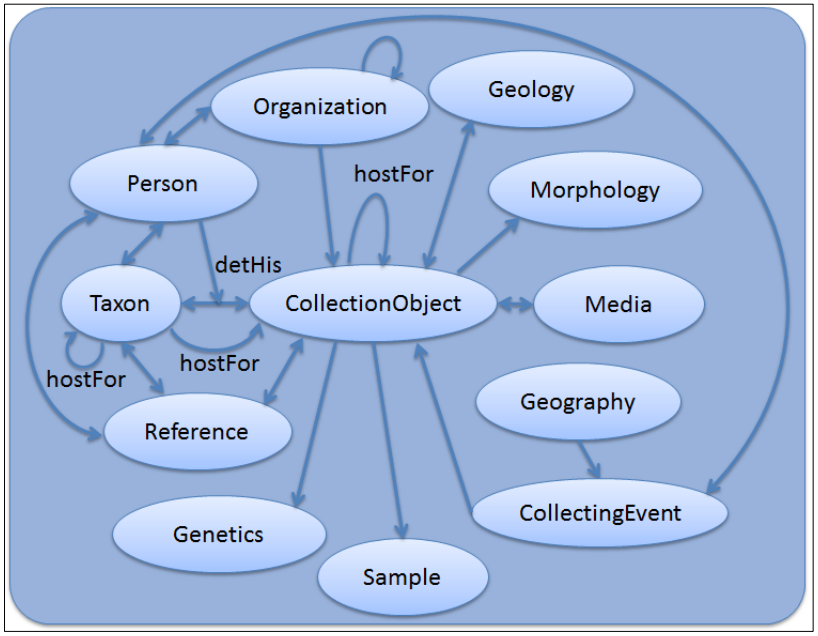
1. Clear, biologically relevant standards to guide data acquisition and distribution.

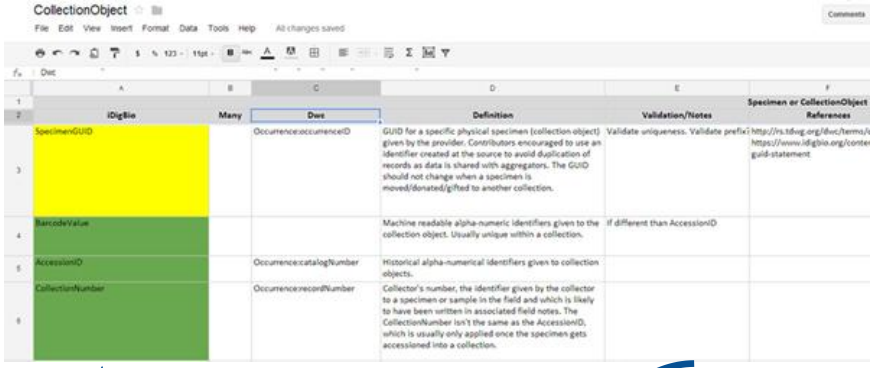
2. A strong sense of effective, community-based digitization workflows and practices.



iDigBio Informatics and Cyberinfrastructure Workshop
28–30 March 2012

Minimum Information for Scientific Collections Working Group





Originally two working groups:

MISC
Authority Files

The two were combined into a single effort. Essentially, iDigBio does not want to be the arbiter or resolver of names. More advantageous for us to partner with other projects for names, including iPlant and the Taxonomic Names Resolution Service (TNRS).

Goal

Taxonomic trees should reflect the world view of individual providers, not iDigBio.

Working Group – Mission and Scope

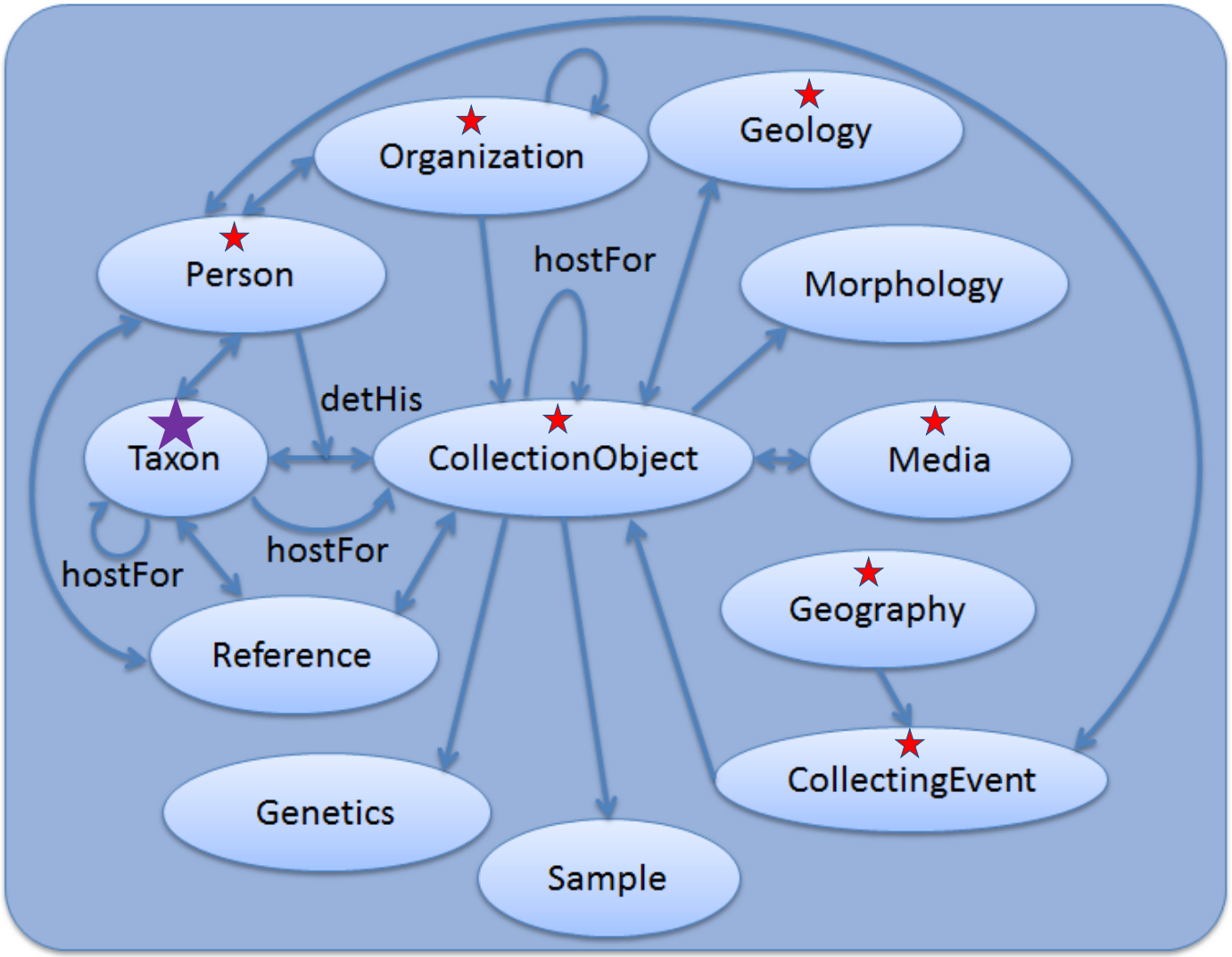
- Document basic assumptions underpinning a minimal information for scientific collections expectation
- Develop a document outlining and providing a rationale for required, minimal, desired, and supplementary data elements to be provided by iDigBio contributors, including an enumeration of these elements with user friendly labels
- Develop guidelines for algorithmic data validation and cleaning tools that can be distributed as an appliance or via an API or web service
- Develop guidelines for offering training and assistance to iDigBio contributors for configuring data for ingestion into the iDigBio portal
- Make all work of the committee available for comment to the community

Original Goals

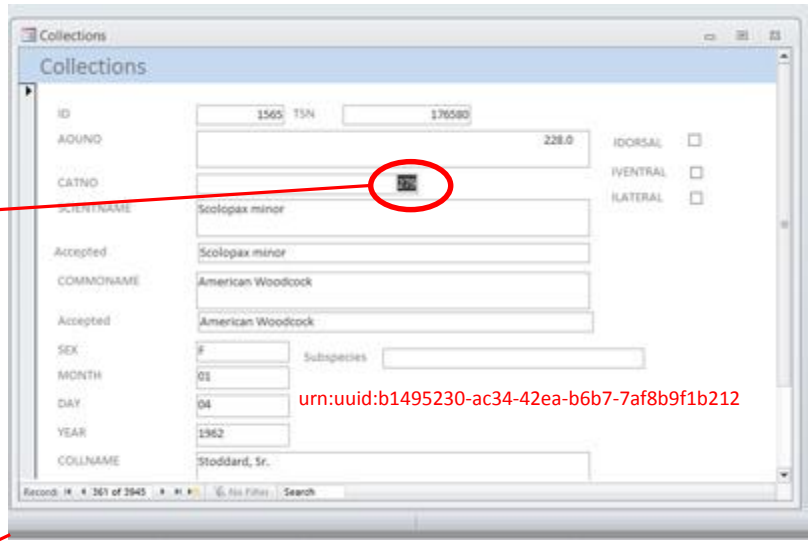
- Ensure that data elements ingested by iDigBio are relevant and understandable to biologists and collections managers.
- Ensure that iDigBio data elements reflect, as much as possible, the content expressed by terms common to widely used biodiversity databases, schemas, and standards.
- Categorize data elements into those that are:
  - required by iDigBio,
  - highly desired for maximum scientific value,
  - Complementary/supplementary (but important).
- Ensure flexibility by:
  - being open to all contributed data, regardless of whether currently included in MISC, DwC, AC, or other standards,
  - preserving opportunities to expand and refine MISC and the elements we ingest in the face of changing needs, standards, and contributions.

MISC Phase I

https://www.idigbio.org/wiki/index.php/MISC-Authority-File-Working-Group



Identifying Objects



Add column to data record for a globally unique, persistent identifier.

http://www.talltimbers.org/museum.html#Birds:279
urn:uuid:3Ab1495230-ac34-42ea-b6b7-7af8b9f1b212

Resolver

UUID or GUID does not have to appear on the specimen itself.

iDigBio Guidelines for Managing Persistent Identifiers

- Requirement: TCNs and other institutions ensure that all identifiers provided to iDigBio are unique and persistent.
- Persistent identifiers are those that are assigned once, only once, and are associated with a single object. Once assigned to an object, an identifier cannot be assigned to a different object.
- Sample URI pattern using a domain name:
  http://ids.flmnh.ufl.edu/herb/co/abcd12345678

  Domain name registered and owned by the TCN or institution

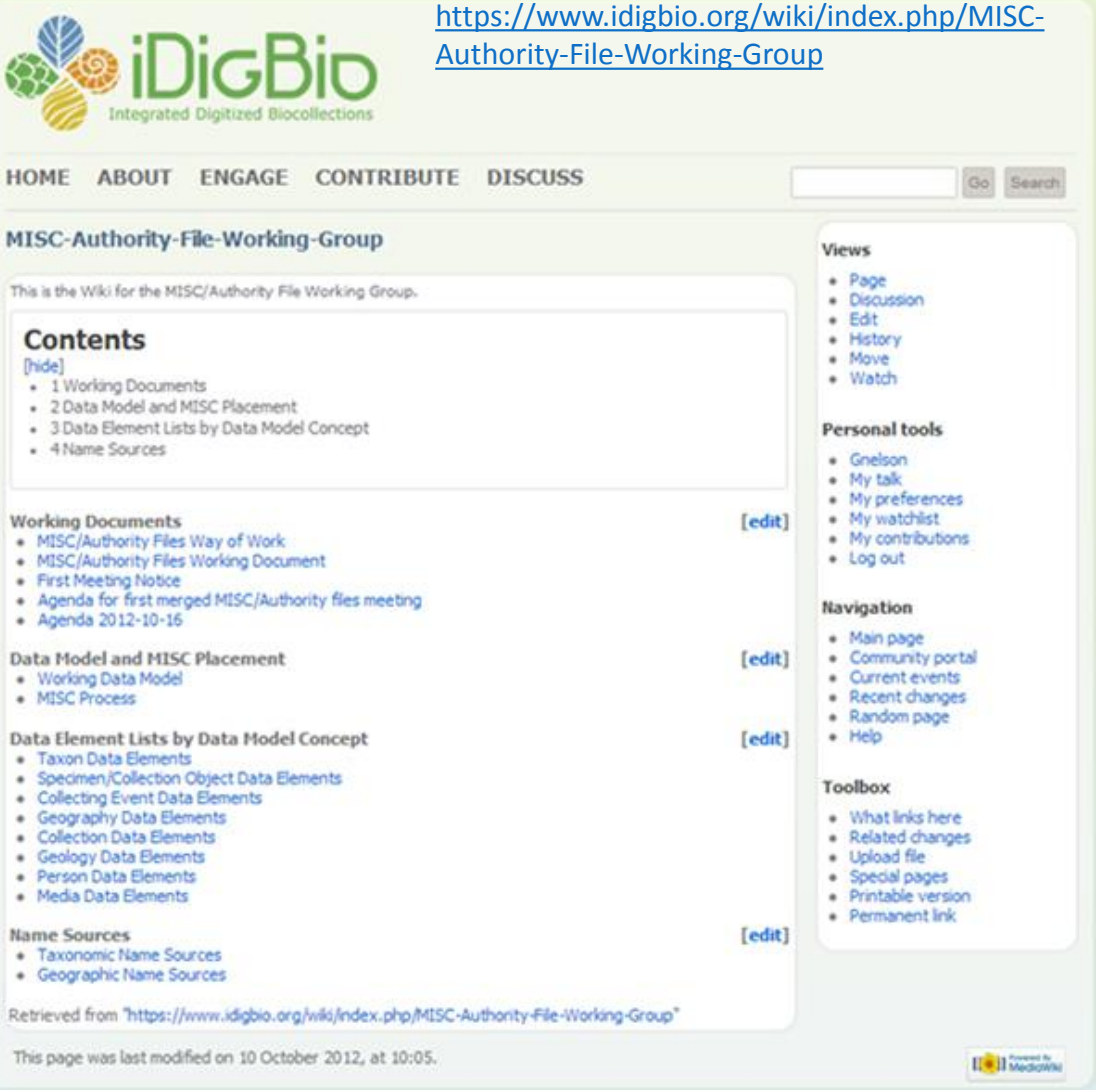
  Locally unique identifier
- We also recommend UUID identifiers appended to a prefix.

MISC Product

At its core, the product of the MISC working group is iDigBio's attempt to:

- put flesh on the bones of the data model presented earlier,
- bring a biologist's or collection manager's perspective to the data elements iDigBio ingests,
- ensure that we account for all data currently or potentially stored in collections databases (hence, MISC may be a misnomer),
- narrowly (and perhaps selfishly?) focus on data elements iDigBio should be prepared to ingest over the long haul, to prioritize these elements with respect to whether they should be treated as required, highly desired, or supplementary, and to recognize that the list of these elements might grow over time,
- take a scientific perspective on data fitness,
- start with Darwin Core as a foundation and augment this standard from the many other schemas currently in use in our community,
- map MISC data elements to as many existing vocabularies/schemas as possible to facilitate ingestion.

https://www.idigbio.org/wiki/index.php/MISC-Authority-File-Working-Group



iDigBio
Integrated Digitized Biocollections

HOME ABOUT ENGAGE CONTRIBUTE DISCUSS

Go Search

MISC-Authority-File-Working-Group

This is the Wiki for the MISC/Authority File Working Group.

Contents

[hide]

Working Documents [edit]
- MISC/Authority Files Way of Work
- MISC/Authority Files Working Document
- First Meeting Notice
- Agenda for first merged MISC/Authority files meeting
- Agenda 2012-10-16

Data Model and MISC Placement [edit]
- Working Data Model
- MISC Process

Data Element Lists by Data Model Concept [edit]
- Taxon Data Elements
- Specimen/Collection Object Data Elements
- Collecting Event Data Elements
- Geography Data Elements
- Collection Data Elements
- Geology Data Elements
- Person Data Elements
- Media Data Elements

Name Sources [edit]
- Taxonomic Name Sources
- Geographic Name Sources

Retrieved from "https://www.idigbio.org/wiki/index.php/MISC-Authority-File-Working-Group"

This page was last modified on 10 October 2012, at 10:05.

Views
- Page
- Discussion
- Edit
- History
- Move
- Watch

Personal tools
- Gnelson
- My talk
- My preferences
- My watchlist
- My contributions
- Log out

Navigation
- Main page
- Community portal
- Current events
- Recent changes
- Random page
- Help

Toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

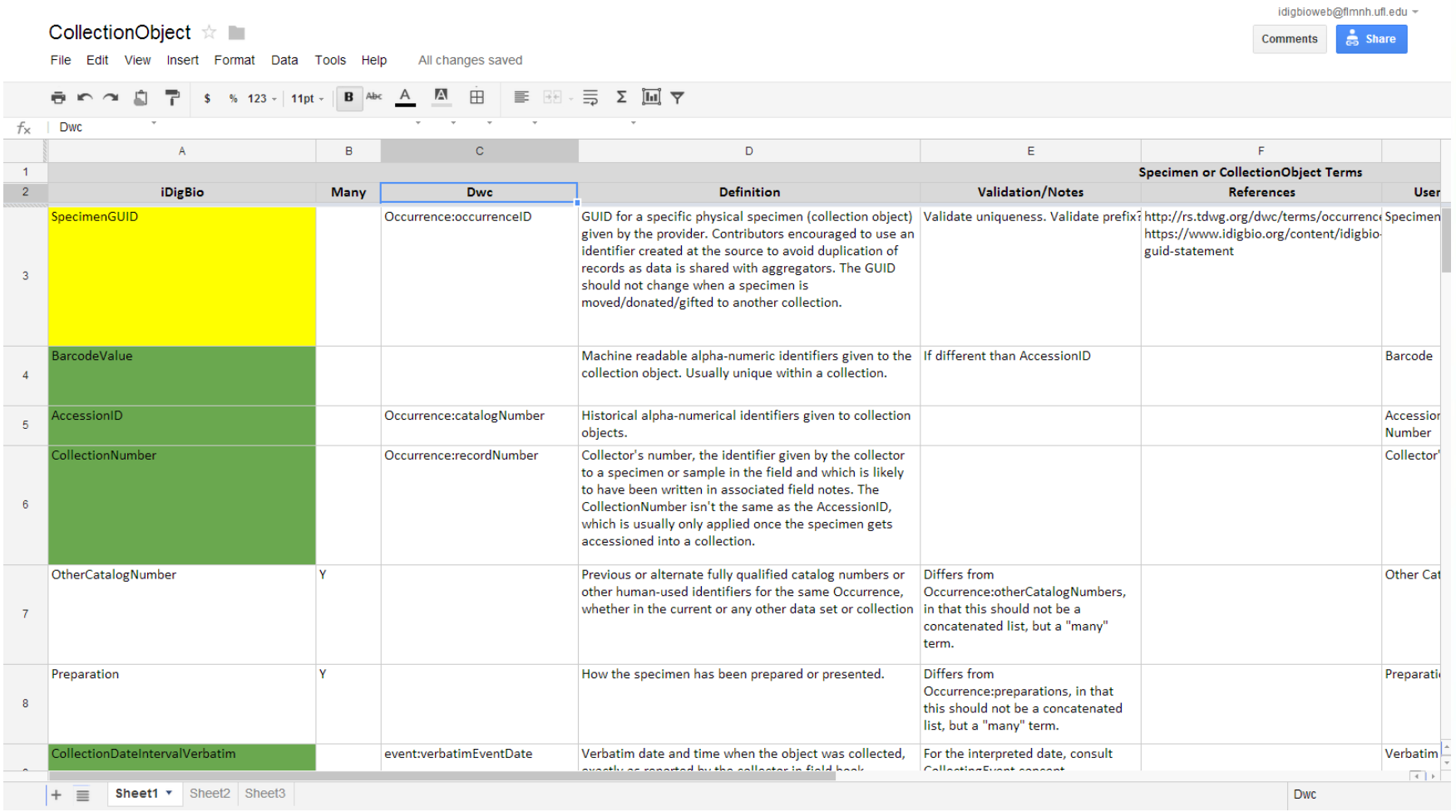
David Jennings, Project Manager • Cathy Bester, Program Assistant • Dickinson Hall • Gainesville, FL 32611 • 352-273-1906

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biological Collections Program (#EF1115210)

FLORIDA MUSEUM OF NATURAL HISTORY · THE FLORIDA STATE UNIVERSITY · UF UNIVERSITY of FLORIDA

iDigBio Integrated Digitized Biocollections

MISC Working Group – Terms for Data Model Concepts



CollectionObject

File Edit View Insert Format Data Tools Help All changes saved

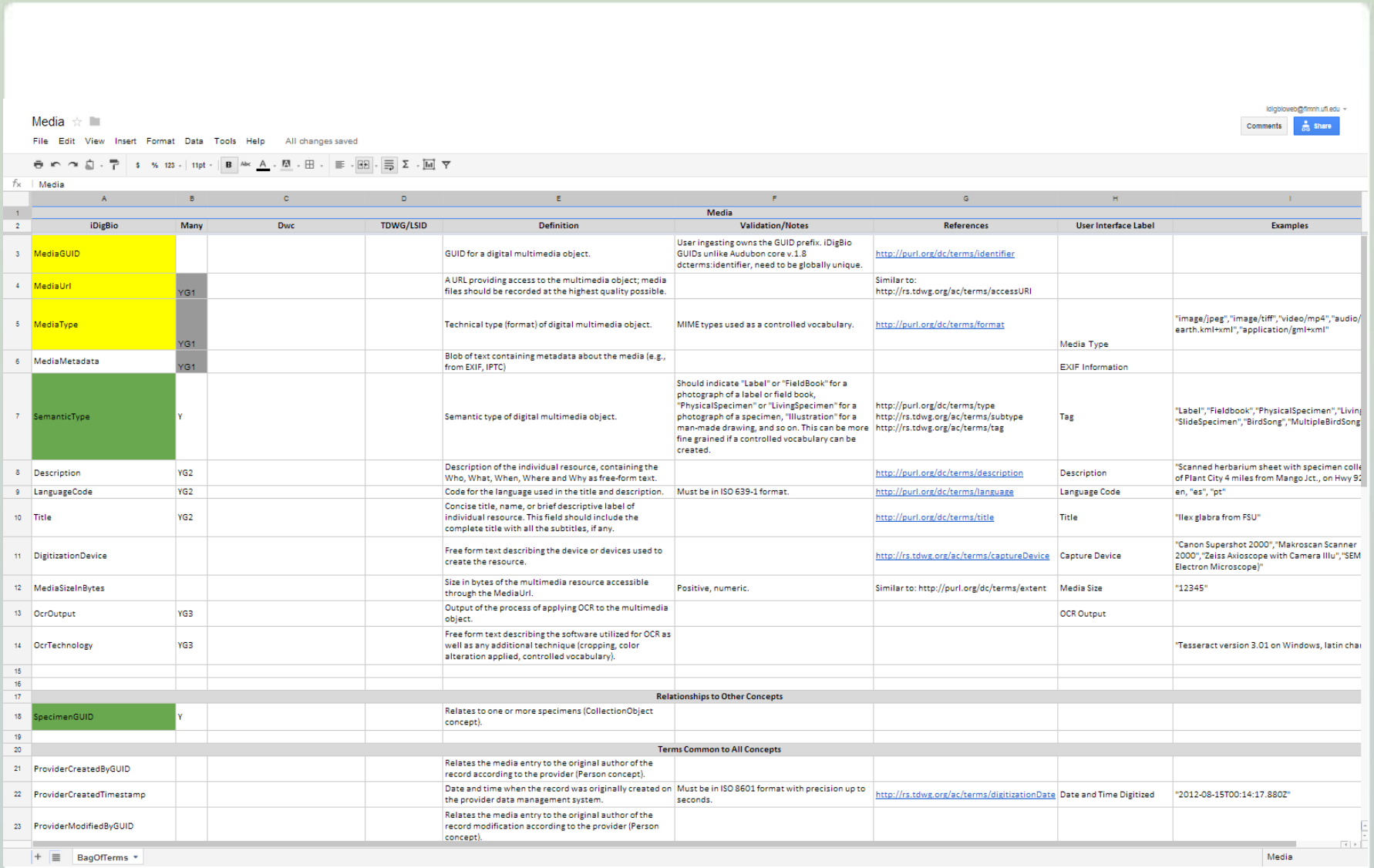
idigbioweb@flmnh.ufl.edu

Comments Share

Dwc

	A	B	C	D	E	F	
1						Specimen or CollectionObject Terms	
2	iDigBio	Many	Dwc	Definition	Validation/Notes	References	User
3	SpecimenGUID		Occurrence:occurrenceID	GUID for a specific physical specimen (collection object) given by the provider. Contributors encouraged to use an identifier created at the source to avoid duplication of records as data is shared with aggregators. The GUID should not change when a specimen is moved/donated/gifted to another collection.	Validate uniqueness. Validate prefix?	http://rs.tdwg.org/dwc/terms/occurrenc https://www.idigbio.org/content/idigbio-guid-statement	Specimen
4	BarcodeValue			Machine readable alpha-numeric identifiers given to the collection object. Usually unique within a collection.	If different than AccessionID		Barcode
5	AccessionID		Occurrence:catalogNumber	Historical alpha-numerical identifiers given to collection objects.			Accessior Number
6	CollectionNumber		Occurrence:recordNumber	Collector's number, the identifier given by the collector to a specimen or sample in the field and which is likely to have been written in associated field notes. The CollectionNumber isn't the same as the AccessionID, which is usually only applied once the specimen gets accessioned into a collection.			Collector'
7	OtherCatalogNumber	Y		Previous or alternate fully qualified catalog numbers or other human-used identifiers for the same Occurrence, whether in the current or any other data set or collection	Differs from Occurrence:otherCatalogNumbers, in that this should not be a concatenated list, but a "many" term.		Other Cat
8	Preparation	Y		How the specimen has been prepared or presented.	Differs from Occurrence:preparations, in that this should not be a concatenated list, but a "many" term.		Preparati
	CollectionDateIntervalVerbatim		event:verbatimEventDate	Verbatim date and time when the object was collected,	For the interpreted date, consult		Verbatim

Sheet1 Sheet2 Sheet3 Dwc

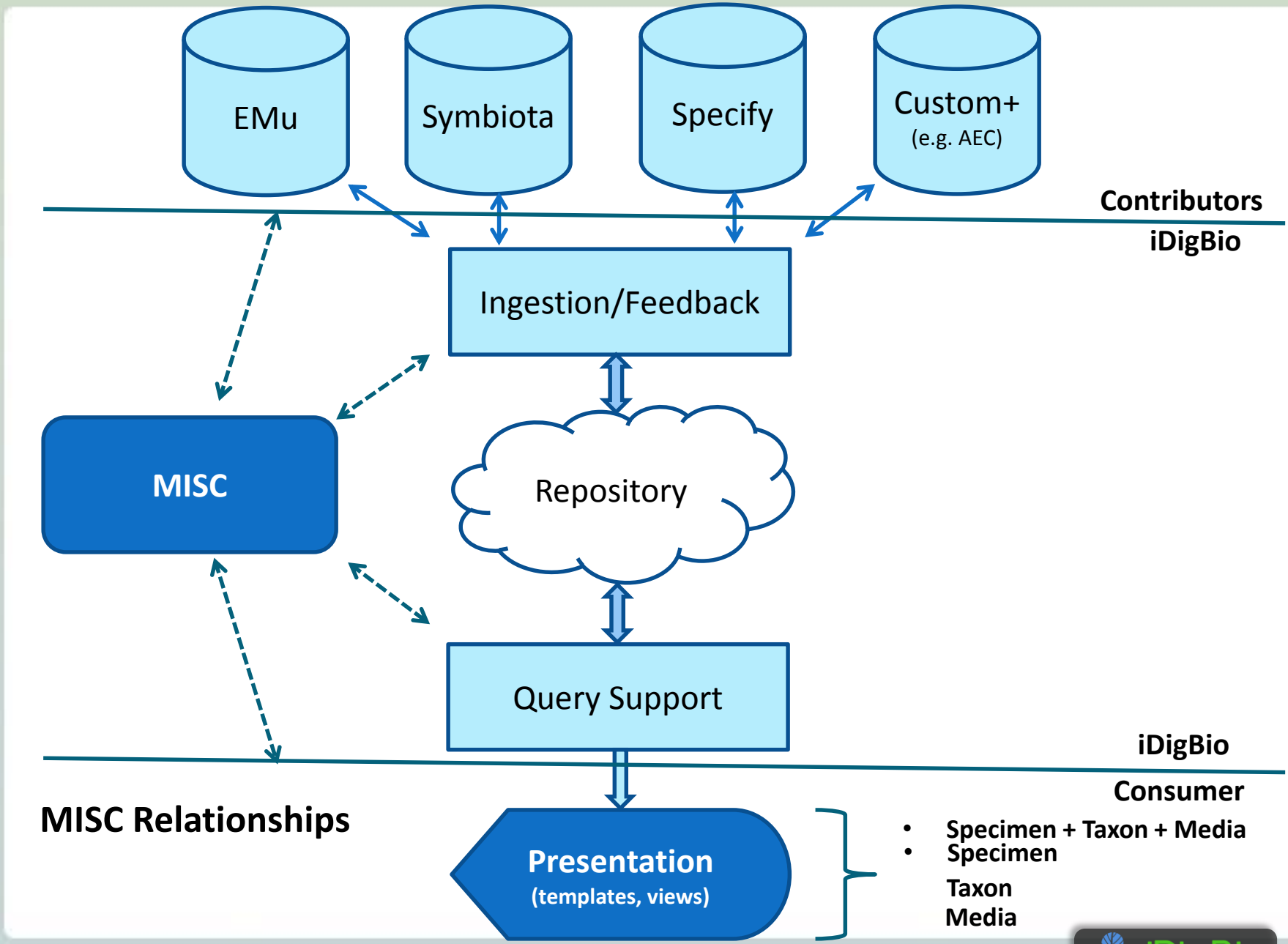


Media

File Edit View Insert Format Data Tools Help All changes saved

	A	B	C	D	E	F	G	H	I
1	Media								
2	iDigBio	Many	Dwc	TDWG/LSID	Definition	Validation/Notes	References	User Interface Label	Examples
3	MediaGUID				GUID for a digital multimedia object.	User ingesting owns the GUID prefix. iDigBio GUIDs unlike Audubon core v.1.8 dcterms:identifier, need to be globally unique.	http://purl.org/dc/terms/identifier		
4	MediaUrl	YG1			A URL providing access to the multimedia object; media files should be recorded at the highest quality possible.		Similar to: http://rs.tdwg.org/ac/terms/accessURI		
5	MediaType	YG1			Technical type (format) of digital multimedia object.	MIME types used as a controlled vocabulary.	http://purl.org/dc/terms/format	Media Type	"image/jpeg","image/tiff","video/mp4","audio/ earth.kml+xml","application/gml+xml"
6	MediaMetadata	YG1			Blob of text containing metadata about the media (e.g., from EXIF, IPTC)			EXIF Information	
7	SemanticType	Y			Semantic type of digital multimedia object.	Should indicate "Label" or "FieldBook" for a photograph of a label or field book, "PhysicalSpecimen" or "LivingSpecimen" for a photograph of a specimen, "Illustration" for a man-made drawing, and so on. This can be more fine grained if a controlled vocabulary can be created.	http://purl.org/dc/terms/type http://rs.tdwg.org/ac/terms/subtype http://rs.tdwg.org/ac/terms/tag	Tag	"Label","Fieldbook","PhysicalSpecimen","Livin "SlideSpecimen","BirdSong","MultipleBirdSong
8	Description	YG2			Description of the individual resource, containing the Who, What, When, Where and Why as free-form text.		http://purl.org/dc/terms/description	Description	"Scanned herbarium sheet with specimen colle of Plant City 4 miles from Mango Jct., on Hwy 92
9	LanguageCode	YG2			Code for the language used in the title and description.	Must be in ISO 639-1 format.	http://purl.org/dc/terms/language	Language Code	en, "es", "pt"
10	Title	YG2			Concise title, name, or brief descriptive label of individual resource. This field should include the complete title with all the subtitles, if any.		http://purl.org/dc/terms/title	Title	"Ilex glabra from FSU"
11	DigitizationDevice				Free form text describing the device or devices used to create the resource.		http://rs.tdwg.org/ac/terms/captureDevice	Capture Device	"Canon Supershot 2000","Makroscan Scanner 2000","Zeiss Axioscope with Camera IIIu","SEM Electron Microscope)"
12	MediaSizeInBytes				Size in bytes of the multimedia resource accessible through the MediaUrl.	Positive, numeric.	Similar to: http://purl.org/dc/terms/extent	Media Size	"12345"
13	OcrOutput	YG3			Output of the process of applying OCR to the multimedia object.			OCR Output	
14	OcrTechnology	YG3			Free form text describing the software utilized for OCR as well as any additional technique (cropping, color alteration applied, controlled vocabulary).				"Tesseract version 3.01 on Windows, latin cha
15									
16									
17	Relationships to Other Concepts								
18	SpecimenGUID	Y			Relates to one or more specimens (CollectionObject concept).				
19									
20	Terms Common to All Concepts								
21	ProviderCreatedByGUID				Relates the media entry to the original author of the record according to the provider (Person concept).				
22	ProviderCreatedTimestamp				Date and time when the record was originally created on the provider data management system.	Must be in ISO 8601 format with precision up to seconds.	http://rs.tdwg.org/ac/terms/digitizationDate	Date and Time Digitized	"2012-08-15T00:14:17.880Z"
23	ProviderModifiedByGUID				Relates the media entry to the original author of the record modification according to the provider (Person concept).				

BagOfTerms Media

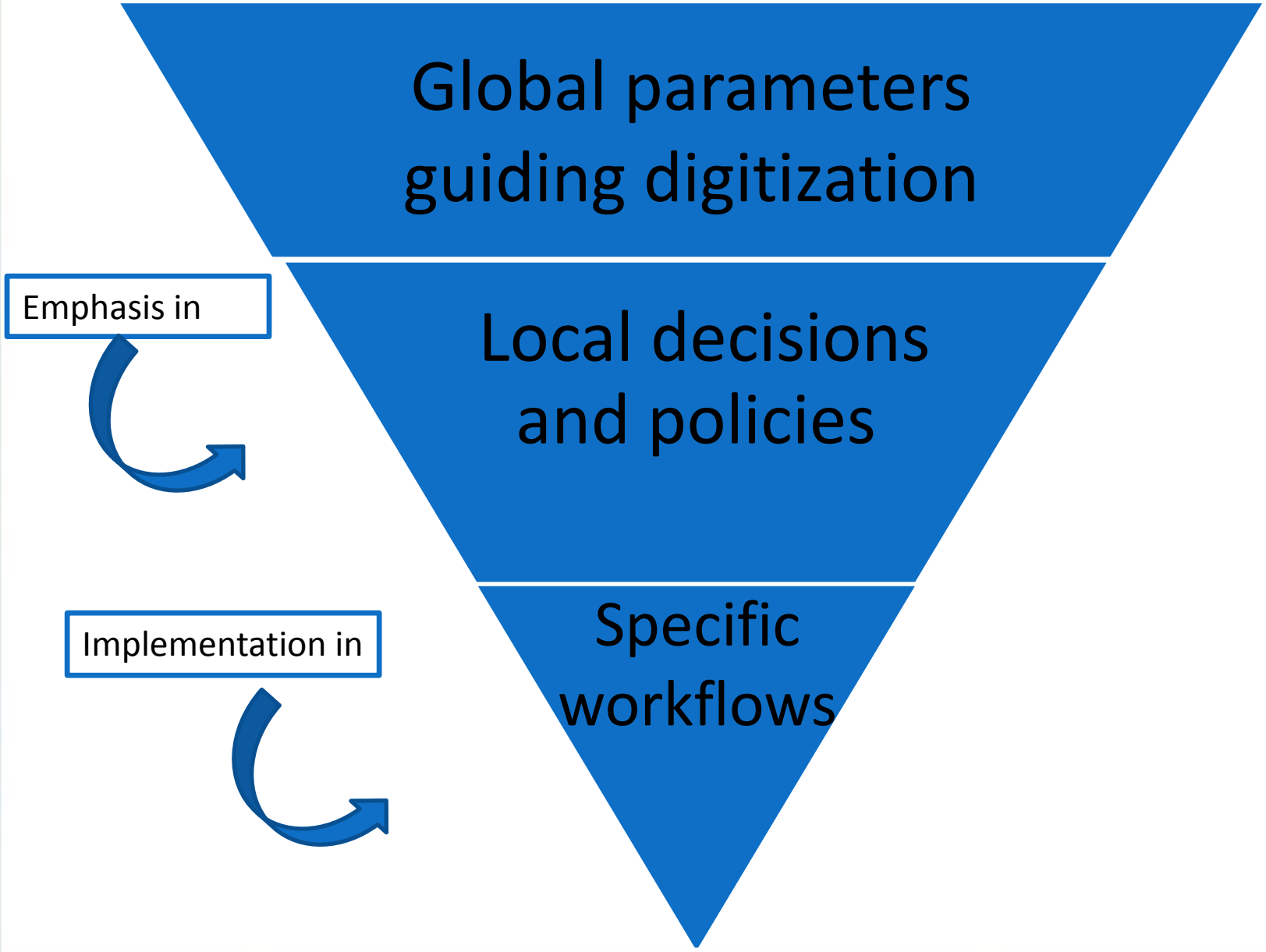

EMu
Symbiota
Specify
Custom+
(e.g. AEC)
Contributors
iDigBio
Ingestion/Feedback
MISC
Repository
Query Support
iDigBio
Consumer
MISC Relationships
Presentation
(templates, views)
• Specimen + Taxon + Media
• Specimen
Taxon
Media

Biodiversity Digitization: Ultimate Goals

Output level: An abundance of scientifically useful and accessible data.

Constituency level: High quality exposure of the content and value of scientific collections.

Improvement level: Collaboration and workflow sharing across the collections community.


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

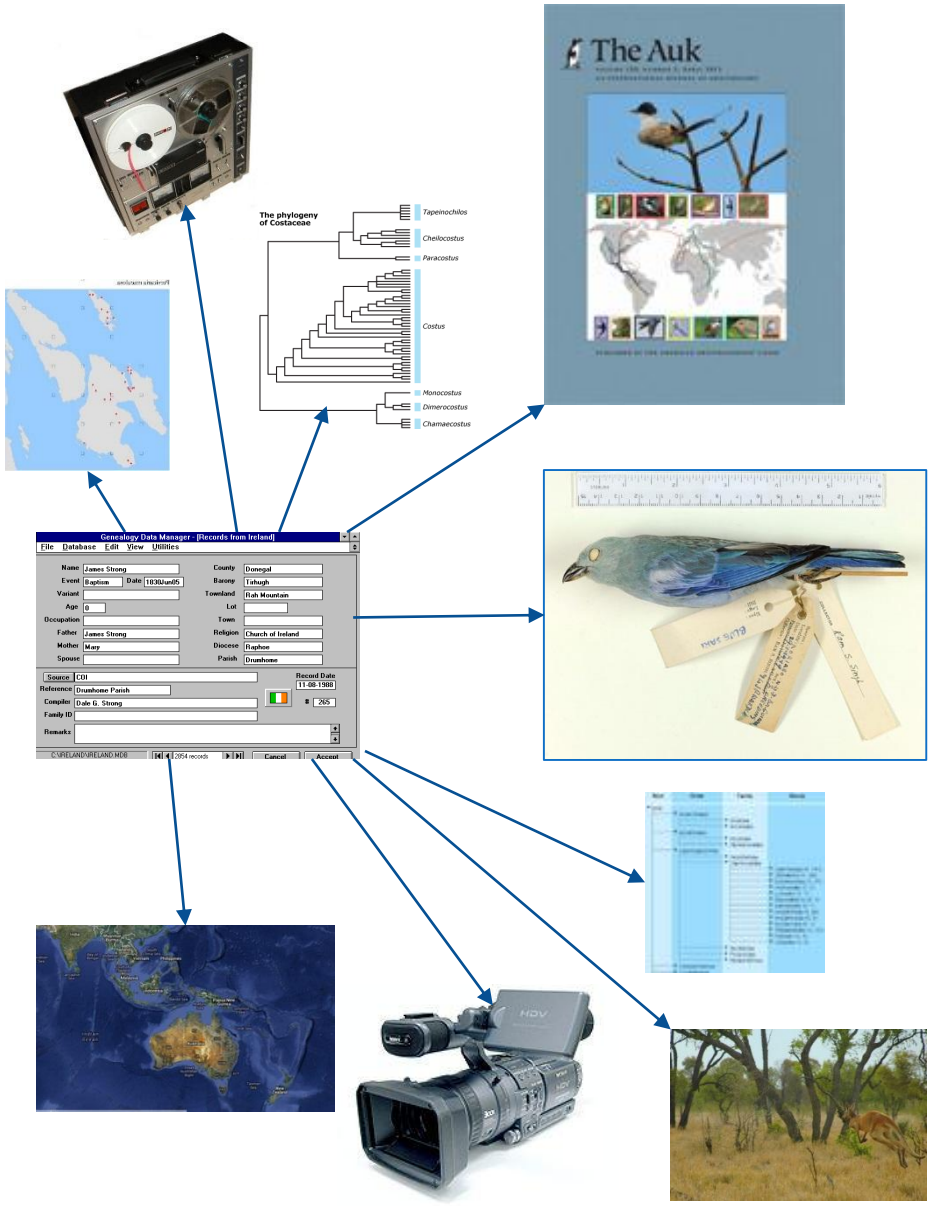
Tracks to Digitization

• **Taking the inside track [short view]** 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 [long view]** 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.

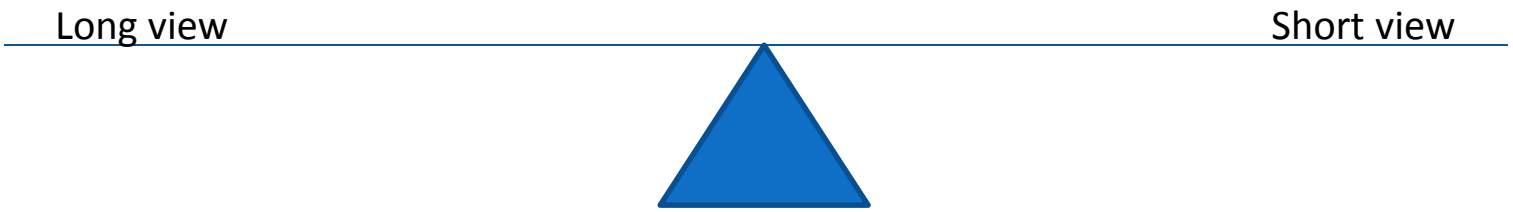
Scan and Deliver: Managing User-initiated Digitization in Special Collections and Archives, 2011
J. Schaffner, F. Snyder. S. Supple

The Long View

Fully populated collection object records searchable across all embedded label data and linked to specimen images, field images, related audio and video recordings, duplicate collections, white papers, grey literature, published works, collecting localities, georeferences, nomenclatural histories, collector information, generic habitat descriptions, taxonomic trees, phylogenies…and anything else related to the specimen that might help scientists and others better understand the collection object in question.



Balancing the long view with the short view: The local decision

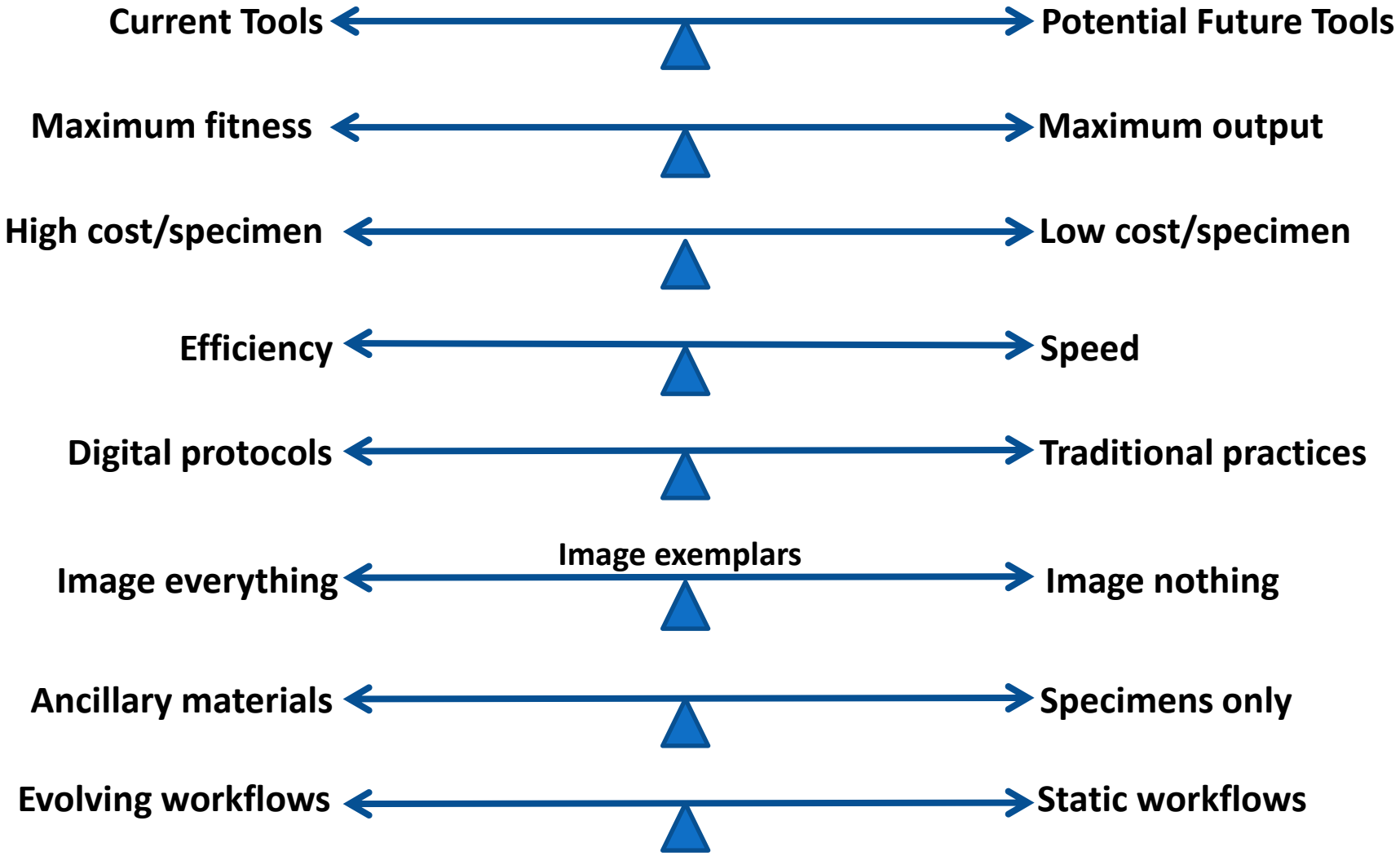


How does an institution develop doable, effective, and sustainable strategies for balancing long term goals with short term constraints, including a commitment to implementing future enhancements?

Pressures mitigating the long view

So much data, so little time.
Collections are not getting smaller (proactive vs. legacy).
Funding agencies have high output/low cost expectations.
We only have 3 years to get this done (sustainable models?).
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 (inside track).

Digitization Continua/Decision Points



Future Tools Favoring the Inside/Middle Tracks

OCR, NLP, and ICR (handwriting analysis) improvements.
Automated image analysis for data extraction.
Data mining of labels.
Robotic technologies, conveyor belts, etc.
Improvements in discovery/capture/use of duplicates.
Improvements in voice recognition and other data entry technologies.
Post-digitization tools for curation and quality control (FilteredPush).
Field data capture.

Long view ⟷ Short View

Facilitators

- Emphasize fitness for use
- Robust datasets
- Data validation/cleaning
- Integrated quality control
- Integrated georeferencing
- Intensive curation
- Record historical annotations
- Staff specialization
- Small collection
- Emphasize images
- High quality images

Facilitators

- Emphasize output
- Spartan datasets
- Defer validation/cleaning
- Deferred/minimum quality control
- Deferred georeferencing
- Deferred or cursory curation
- Record current determination
- Staff generalization
- Large collection
- Emphasize data
- Low quality images

Metrics

Issues in Measuring Productivity

Comparability: Just what is being measured?

- What is included in the output?
- Are all steps in the process accounted for?
- Are all expenditures of time accounted for?
- How do we arrive at a true per specimen cost?

Measuring productivity (comparability across collections):

- Unit (output per unit time vs. expenditure/project totals)
- Data fitness (should data robustness be factored in the calculus?)

Measuring cost:

- Should digitization be a competitive event?
- Output per hour at given fitness?
- $$ per specimen at given fitness?
- Accounting for variances in prep type, regional pay rates, robustness, etc.?

Metrics

Issues in Measuring Use

Measuring use:

- **Number of virtual visitors to the collection?**
- **Types of visitors?**
- **Average time per visitor?**
- **Type of data accessed?**

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

ZooKeys 209: 19–45 (2012)
doi: 10.3897/zookeys.209.3135
www.zookeys.org

RESEARCH ARTICLE

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

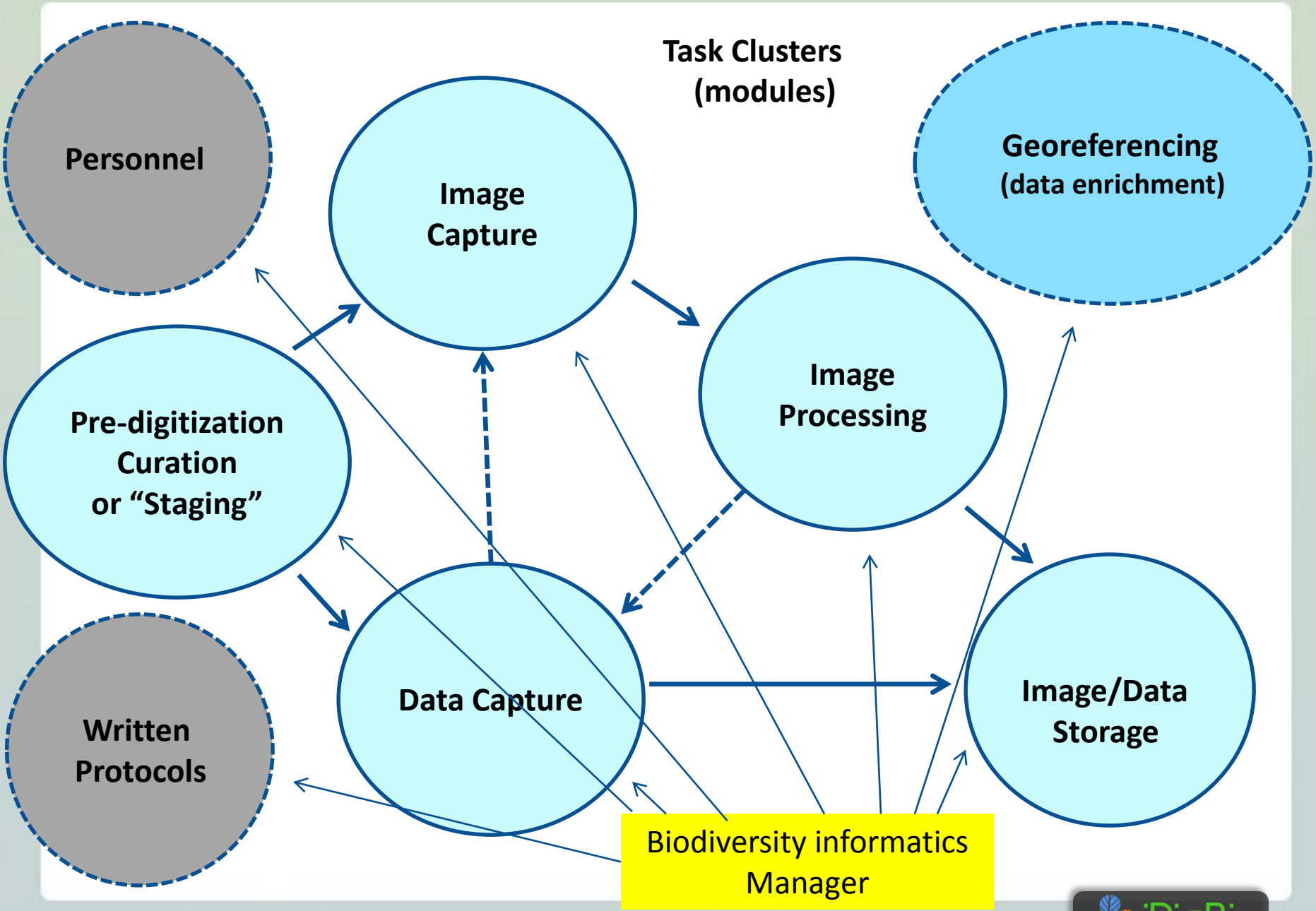
Five task clusters that enable efficient and effective digitization of biological collections

Gil Nelson[1], Deborah Paul[1], Gregory Riccardi[1], Austin R. Mast[2]

1 *Institute for Digital Information, Florida State University, Tallahassee, FL 32306-2100, United States* **2** *Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, United States*

Corresponding author: *Gil Nelson* (gnelson@bio.fsu.edu)

iDigBio
Integrated Digitized Biocollections


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

Personnel Dimensions

- Equipped for taxonomic judgment
- Personnel management skills
- Expertise with standard references
- Keen observational skills
- Specimen handling skills
- Select specimens to image

❖ Many parts of this work do not require trained staff

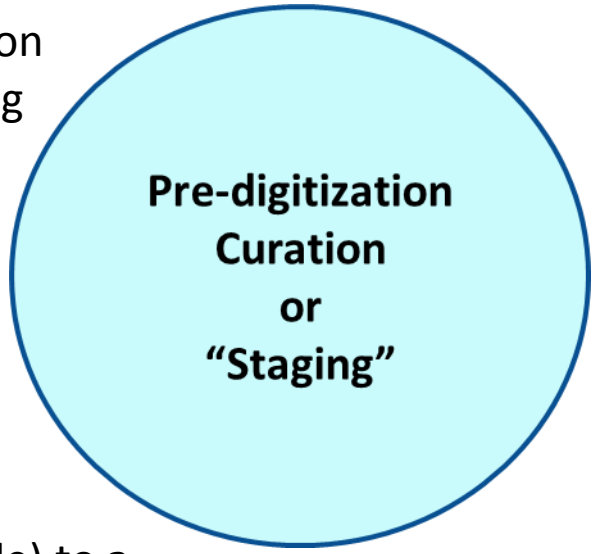
Pre-digitization Curation or "Staging"

Potential workflow bottleneck

Activities

Determination/annotation (a professional)
Data verification (a professional)
Drawer/cabinet organization (trained techs)
Re-pinning (trained techs)
Barcode application (trained techs)

Collections managers reported that pre-digitization curation and staging also provides **unanticipated benefits**, including opportunities to:



- inspect for and repair specimen damage and evaluate collection health (route to conservation),
- inventory the collection,
- re-pin or remount specimens and replenish or replace preservatives in containers,
- treat specimens for pests,
- attach a unique identifier (most often a 1- or 2-D barcode) to a specimen, container, or cabinet,
- discover important but previously unknown, lost, or dislocated holdings (e.g. those owned by other institutions or the federal government),
- update nomenclature and taxonomic interpretation,
- reorganize the contents of cabinets, cases, trays, and containers, especially when these are the units of digitization,
- vet type specimens, and
- select exemplars for digitization, when that approach is appropriate.

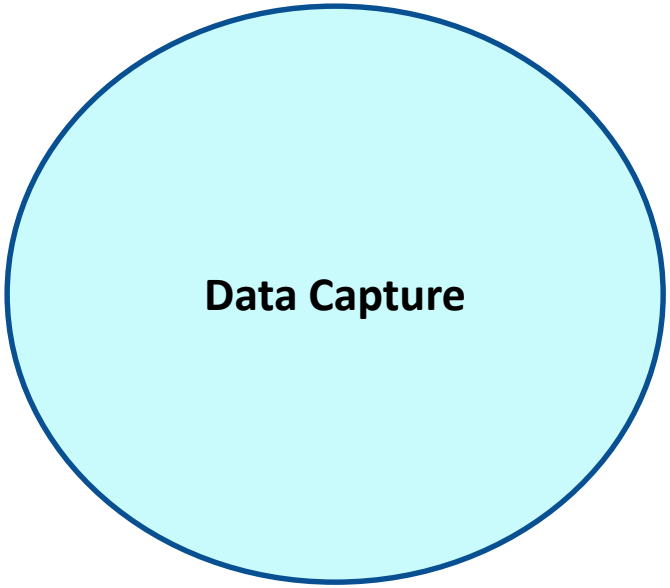
Personnel Dimensions

- Accurate
- Efficient
- Focused
- Tolerant of tedium
- Productive
- Speedy
- Oriented toward improving process
- ❖ Discipline-specific training or interest is often not particularly important

Source Documents

- Specimens/labels
- Images
- Ledgers/catalogs
- Field notebooks
- Monographs

Potential workflow bottleneck



Process dimensions

- Keyboarding
- Voice capture
- OCR
  - QC
  - Data extraction
  - Barcode value extraction
- Data import

Data Import Issues

- Source
  - Internal (legacy)
  - External
- Data quality/trust
- Data format
- Transformation/ field mapping
- Post-import cleanup and quality control

Imaging Configuration

- Camera/lens
- Microscope/camera
- Flatbed scanner
- Herbscan
- Copy stand/lighting
- Lightbox
- Scale bar
- Color standard
- Barcode scanner
- Associated instruments
  - pinning blocks
  - forceps
  - dissecting needles

Stacking software

- Helicon Focus
- Zerene Stacker
- Automontage
- CombineZ

Image Capture

Organisms being imaged

Plants	Mammals
Insects	Paleo
Bryophytes	Herps
Birds	Fish

iDigBio Integrated Digitized Biocollections

Recommendations for the Acquisition, Processing, Storage, and Distribution of Digital Images

iDigBio is pleased to offer the following recommendations for capturing, processing, archiving, and distributing digital images based on work of the Developing Robust Object-to-Image-to-Data Workflows Working Group (DROID), a review of industry standards*, input from the broader collections community, and feedback from iDigBio staff. These recommendations are divided into four clusters: image acquisition, image archiving, image derivatives, and image distribution.

Image Acquisition Recommendations

- Camera RAW is the preferred format for recording camera images.
- Scanner images should be recorded in RAW, if available, or 24 bit or higher TIFF.
- Record at the highest native resolution available for the imaging device.
- Adjust white balance settings to match the light source.
- A visible color checker and scale are recommended, as appropriate.
- Adobe RGB or sRGB are the preferred color spaces and should be stated in the metadata.
- CMYK color space should be avoided.
- Image file names should be restricted to alpha numeric characters, without spaces or other special characters, underscore (_) and hyphen (–) excepted.

Image Archiving Recommendations

- Archiving images is strongly recommended and is viewed as an institutional responsibility, governed by institutional policy.
- Archived image files should be permanently retained in a secure, redundant environment on institutional infrastructure or commercial back-up services.
- Uncompressed DNG is the preferred archival format.
- Images captured in proprietary camera RAW format (e.g., CR2, NEF, PEF, etc.) should be converted to Digital Negative format (DNG).
- Conversion to DNG can be accomplished via Adobe DNG Converter, available free from Adobe.com, or via Adobe Photoshop or Adobe Lightroom.
- Uncompressed or lossless compressed, unmodified TIFF at 24 bits or higher is an alternative archival format.
- Uncompressed or lossless compressed TIFF at 24 bits or higher is preferred for archiving processed images.

Image Derivative Recommendations

- JPEG and JPEG 2000 images should be saved at their native dimensions (spatial resolution).
- Manually increasing resolution (e.g., alterations to image size or ppi or dpi settings) beyond that recorded by the camera may insert derived data or unwanted visual artifacts into an image file and is not recommended.
- EXIF, IPTC, and other image metadata (including JPEG 2000 XML data), including that added at time of processing, should be preserved in perpetuity and distributed with all derivative and distributed copies of images.
- Image processing should be carried out only on DNG or other RAW files to prevent accumulated data losses and potential image degradation by repeated processing of JPEG and TIFF image files.
- Excessive image manipulation or compositing, including sharpening, enhancing color saturation or balance, and adjusting contrast should be avoided.

Image Distribution Recommendations

- Images uploaded to aggregators and portals should be fit for display.
- Lower quality images should be distributed to aggregators and portals only when there is no higher quality image available.
- When a low quality image is uploaded or distributed, the metadata for that image should indicate that it is the only image available for that specimen.
- JPEG format at native resolution and minimal compression is preferred for distribution via the internet.
- Lossless compression is preferred.
- When lossless compression is not an option, lossy compression at the lowest level is preferred.

***Industry References**

- Universal Photographic Digital Imaging Guidelines
- The AIC Guide to Digital Photography and Conservation Documentation
- American Institute for Conservation of Historic and Artistic Works
- OpenRAW.org
- Adobe Digital Negative (DNG) Specification

Biodiversity Informatics Managers

Enjoy broad institutional support
Essential to successful digitization
Personable and positive
Oriented toward service to staff
Strong background in the biological sciences
Strong technology skill set
Often “home-grown”
Provide rapid response to requests and suggestions
Open to new and/or enhanced protocols and techniques
Actively seek ways to increase efficiency and productivity
Ability to customize or create software that contributes to digitization efficiency

Biodiversity Informatics Manager Working Group

Mission

The mission of this working group is to circumscribe *biodiversity informatics manager* as a recognized career path within the natural history museum and biological collections communities, and to raise awareness of this career path to collections managers, museum directors, academicians, and potential biodiversity informatics managers and professionals.

In-scope activities

- Define a universal *biodiversity informatics manager* job description.
- Delineate specific and detailed job dimensions drawn from that job description.
- Determine the technical and personal skill set (knowledge, skills, and attitudes) required for effective and successful performance as a biodiversity informatics manager.
- Describe the career paths followed by current successful biodiversity informatics managers.
- Define the role of biodiversity informatics managers within the larger museum or organization.
- Justify this role and recommend its placement within the organizational hierarchy.
- Make concrete recommendations regarding the desired academic preparation of biodiversity informatics managers, to include strategic guidelines for recruiting new professionals into this career path.

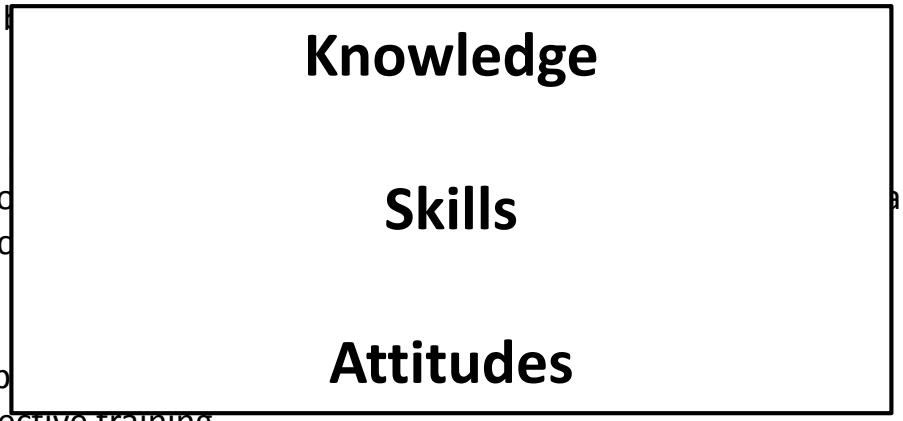
Knowledge

Biology/Collections Management
Working knowledge of collections management, including processing loan requests, coordinating with collectors, field collection methods, tissue sampling, specimen handling, annotation protocols, ancillary documents and data, etc.
Strong background in the l

Engineering
Understand the existence
Complete understanding c and use.
Current and forthcoming c

Institutional Context
Understand the range of b
Design and delivery of effective training.
Understanding institution and community wide biodiversity informatics tools and resources.

External/Community
Expert knowledge of existing and developing data and image aggregators.
Understand the institutional milieu/data policies
Understanding of cloud-based technology.
Understanding of Open Source issues.
Volunteerism, public participation, and citizen science.
Working knowledge of biodiversity data standards (e.g., Darwin Core, Dublin Core, ABCD, Audubon Core, etc.).
Working knowledge of image metadata standards (e.g. EXIF, IPTC).
Be conversant with the uses of biodiversity data for research and discovery across multiple domains and the grand challenges that these data can address.



Knowledge

Biology/Collections Management
Working knowledge of collections management, including processing loan requests, coordinating with collectors, field collection methods, tissue sampling, specimen handling, annotation protocols, ancillary documents and data, etc.
Strong background in the biological sciences.

Engineering
Understand the existence and use of web services.
Complete understanding of enterprise-level database design, structure, philosophy, and use.
Current and forthcoming computer, databasing, and imaging technology.

Institutional Context
Understand the range of biodiversity and related data to be digitally tracked.
Design and delivery of effective training.
Understanding institution and community wide biodiversity informatics tools and resources.

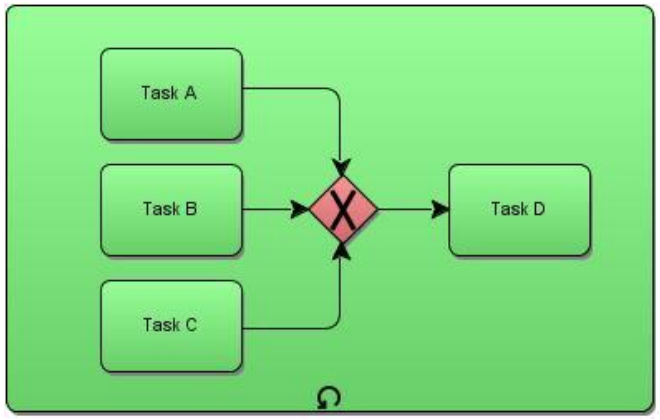
External/Community
Expert knowledge of existing and developing data and image aggregators.
Understand the institutional milieu/data policies
Understanding of cloud-based technology.
Understanding of Open Source issues.
Volunteerism, public participation, and citizen science.
Working knowledge of biodiversity data standards (e.g., Darwin Core, Dublin Core, ABCD, Audubon Core, etc.).
Working knowledge of image metadata standards (e.g. EXIF, IPTC).
Be conversant with the uses of biodiversity data for research and discovery across multiple domains and the grand challenges that these data can address.

Three-part community survey

- **Biodiversity informatics managers**
- **Collections managers**
- **Museum administrators**

Workflows

Values of defined workflows



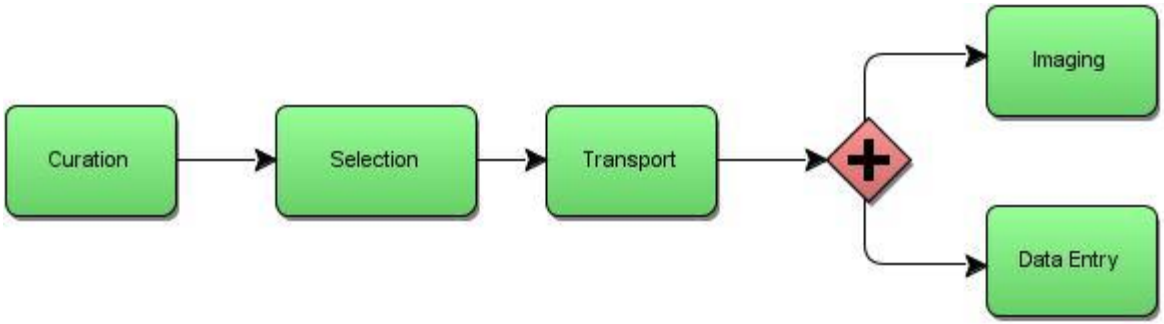
- Promote efficiency and automation of processes
- Facilitate routing and scheduling of activities
- Provide for balancing workloads
- Ensure that processes are visible and predictable
- Allow for escalations and notifications
- Enhance tracking of tasks
- Foster collaboration of all parties involved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

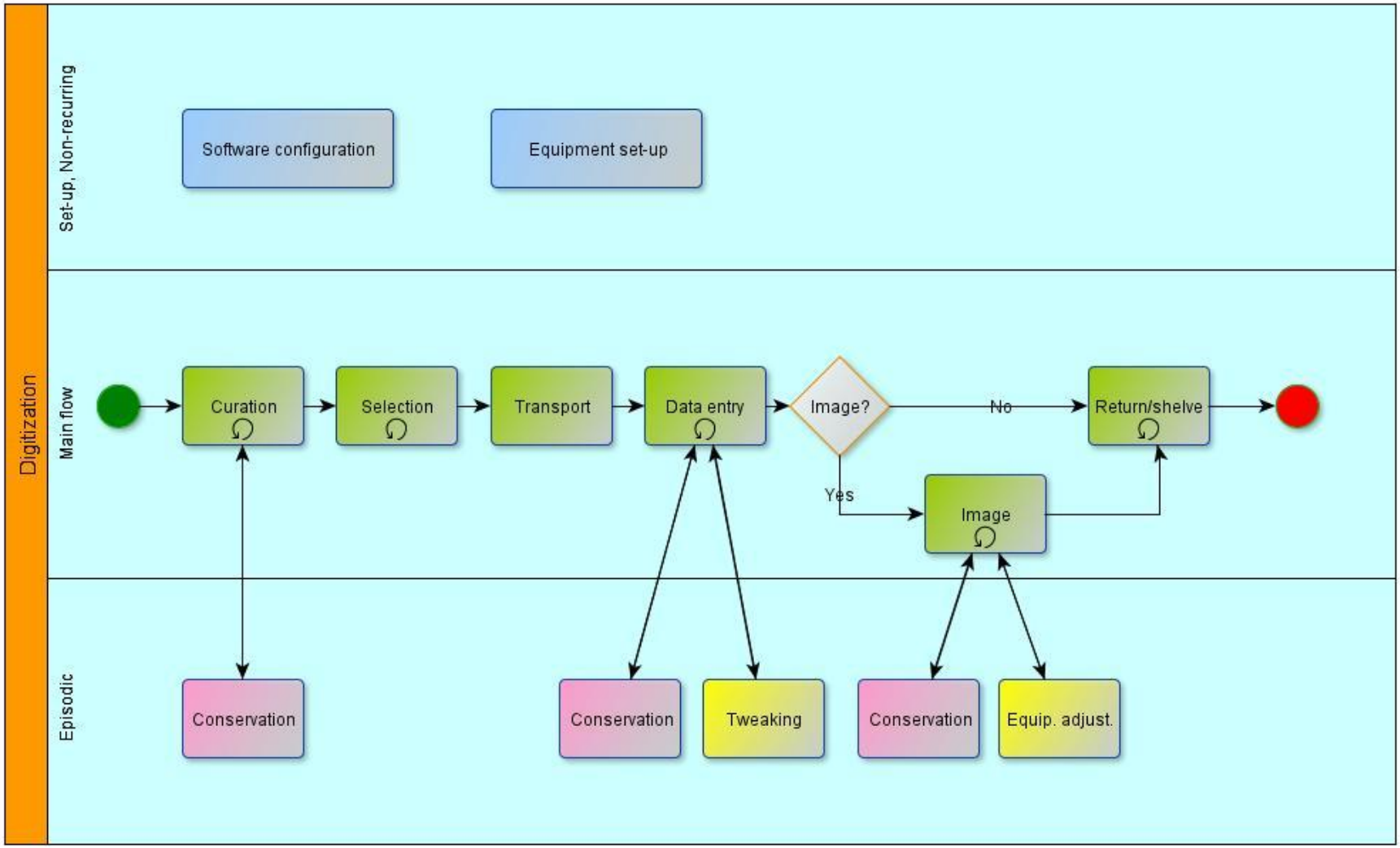
Pre-planning a Workflow Process

- Identify the database management system and imaging equipment to be used.
- Identify the process/module for which the workflow will be designed.
- Identify (*in excruciating detail*!) the tasks (or task clusters) that constitute the process/module.
- Identify the specific actions to be taken to accomplish the task.
- Identify roles (and only secondarily the people who will fill them).
- Identify points/processes/parameters for notifications and escalations.
- Identify dependencies, transitions, and iterations.
- Determine minimal data requirements for defining a complete record.
- Determine how records and objects will be uniquely identified in a global environment.
- Determine how identifiers will be assigned.
- Determine if/how identifiers will be affixed to the specimen/lot/collection object.
- Determine a consistent file naming strategy for images, attachments, other related materials.
- Determine file storage needs and location for data, images, and ancillary materials.
- Define and diagram flow.

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	Data entry, imaging
Specimen return/shelving	Recurring	Imaging or data entry




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.

A sample, detailed task list.

1. Open Capture NX2 and View NX2.
2. Open Camera Control Pro 2.
3. Open default.ncc as settings file:
   Settings->Load Control Settings
   ->My Documents->CameraSettings->default.ncc.
4. Create a folder in X:\SpecimenImages\NEF, using the current date as the folder name, as 2013-02-27.
5. Retrieve next specimens to image from cabinet.
6. Insert Image “From Here” tag to proper place in cabinet.
7. Set image number in Camera Control 2 to next bar code:
   tools->download options
   Edit
   Start numbering at: <Enter next bar code number; no leading zeros>.
8. In Download Options, set the default folder to the one you created in step 4.
9. Position specimen in frame, ensuring complete specimen is visible.
10. Open Live View, position the focus square on specimen.
11. Click AF to test.
12. Click AF and Shoot.
13. Once the first image loads, navigate to it in Capture NX2 or View NX2.
14. Open the image, zoom in and check margins to ensure all of the specimen is visible.
15. Repeat 8-11 until satisfied, resetting image number each time.
16. Close Live View.
17. Load next specimen in frame.
18. Use remote release on camera and record the images.
19. As you shoot, check each image bar code to ensure it is in sequence with the one preceding it and matches the next one in the series.
20. For out-of-sequence bar codes, change the number in the download options.
21. Repeat 17-20 until all specimens are imaged.

Rules for Defining Roles

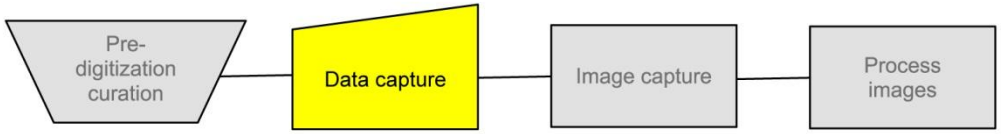
Tasks are assigned to roles, not to individuals.

Generic roles require less maintenance than highly specific ones. Modifying roles that resolve to an individual must be done much more frequently than roles that resolve to a person group.

If possible, associate each module with a single role, with the contents of that module encompassing the tasks that a person in that role must be able to perform.

Modules that require more than one role should be divided into separate modules, or at least into submodules.

Role definitions should include tasks required for managing transitions between modules.



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.

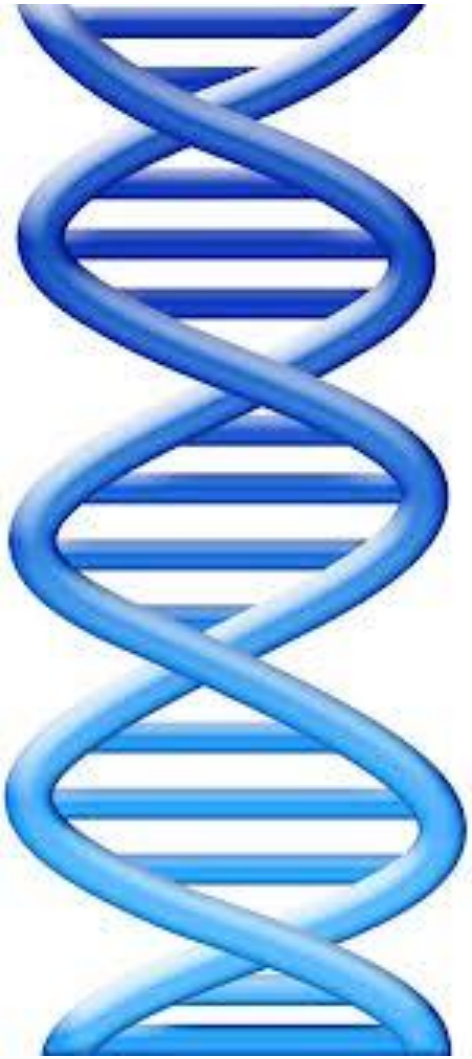
Create task lists

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

Guiding Principles

- **Provide for efficiency**
  - Contrasted with speed.
  - Reduce technician fatigue.
  - Maintain technician focus.
  - Optimize task execution times (time/motion)
    - Record statistics
    - Make adjustments.

- **Ensure conservation of movement**
  - Positioning and compactness of work station components.
  - Left to right vs. right to left.
  - Starting and ending locations.
  - Proximity of equipment (including mouse).



Guiding Principles

- **Multi-tasking**
  - Making the most of down time (regardless how long).
  - Nesting shorter tasks (start one task, start another)
    - Overcoming distractedness.

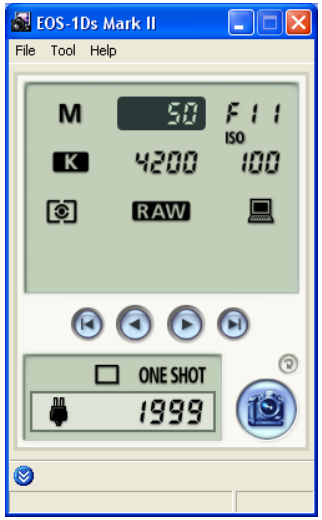
- **Workflow simulation and modeling**
  - Analyzing temporal juxtaposition of workflow task clusters.
  - Analyzing spatial juxtaposition of workflow task clusters.

- **Task list simulation**

Guiding Principles

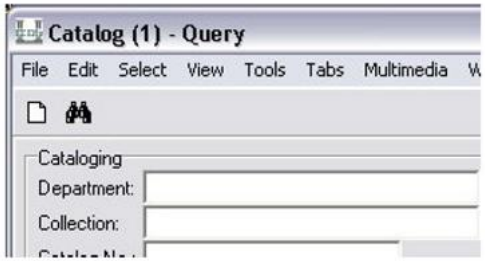
- **Segmenting clusters and subroutines**
  - Standalone repetitive processes.
    - Barcoding.
    - Imaging.
    - Image processing.
    - Re-shelving.
  - Conservation and repair.
  - Georeferencing.
  - OCR.

Documentation and Instructions



- **Written Protocols**
  - Essential!
  - Include 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


iDigBio
Integrated Digitized Biocollections