



Using Digitized Specimen Data in Research: Applications for Ecology, Phylogenetics, and Biogeography



Pam Soltis
University of Florida



iDigBio
Integrated Digitized Biocollections



BiotaPhy



Topics

- Intro to data
- Downloading data
- Cleaning data
- Georeferencing
 - GEOLocate
- Ecological Niche Modeling
 - Maxent
- ENM Analysis & Interpretation
- Scaling Up – BiotaPhy
- Additional uses of collections data
- Discussion/Q&A
- Manual and R-based options

Workshop Leaders

- Pam Soltis
- Makenzie Mabry
- Lauren Gillett
- Shelly Gaynor
- JT Miller
- Elizabeth White
- Maria Cortez
- Doug Soltis
- Malu Ore Rengifo
- Contributions from many previous lab members/colleagues:
 - Blaine Marchant, Charlotte Germain-Aubrey, Andre Naranjo, Anthony Melton, Tal Kinser, Mike Belitz, Rhett Rautsaw, João Vidal, Jr.



BiotaPhy



Logistics

- Dropbox folder
 - Word doc with schedule and instructions
 - Presentations folder
 - Demos folder
- Workshop evaluation survey
 - by email
- Pace/breaks



Herbaria Worldwide

NYBG STEERE HERBARIUM

Home Collections Discover Index Herbariorum Virtual Herbarium Loans Digitization



Index Herbariorum

3,400 herbaria
400,000,000 specimens!



Natural History Collections



1-2 billion specimens
in the US
3-4 billion specimens
worldwide

Smithsonian

~1600 natural history
collections in the US,
thousands in the world



Systematics & Taxonomy



Carl Linné, aka Carolus Linnaeus

The Linnean Collections

LINNEAN
SOCIETY
of London



[Linnean Society Home](#)

Linnaea borealis, 1732

Uses of Natural History Collections



...Species interactions
Phenology
Biogeography
More!

Genetics
Genomics
Chemistry...

Smithsonian



Uses of Natural History Collections



Most specimens locked away in cabinets, unavailable for general use.

Smithsonian



Uses of Natural History Collections



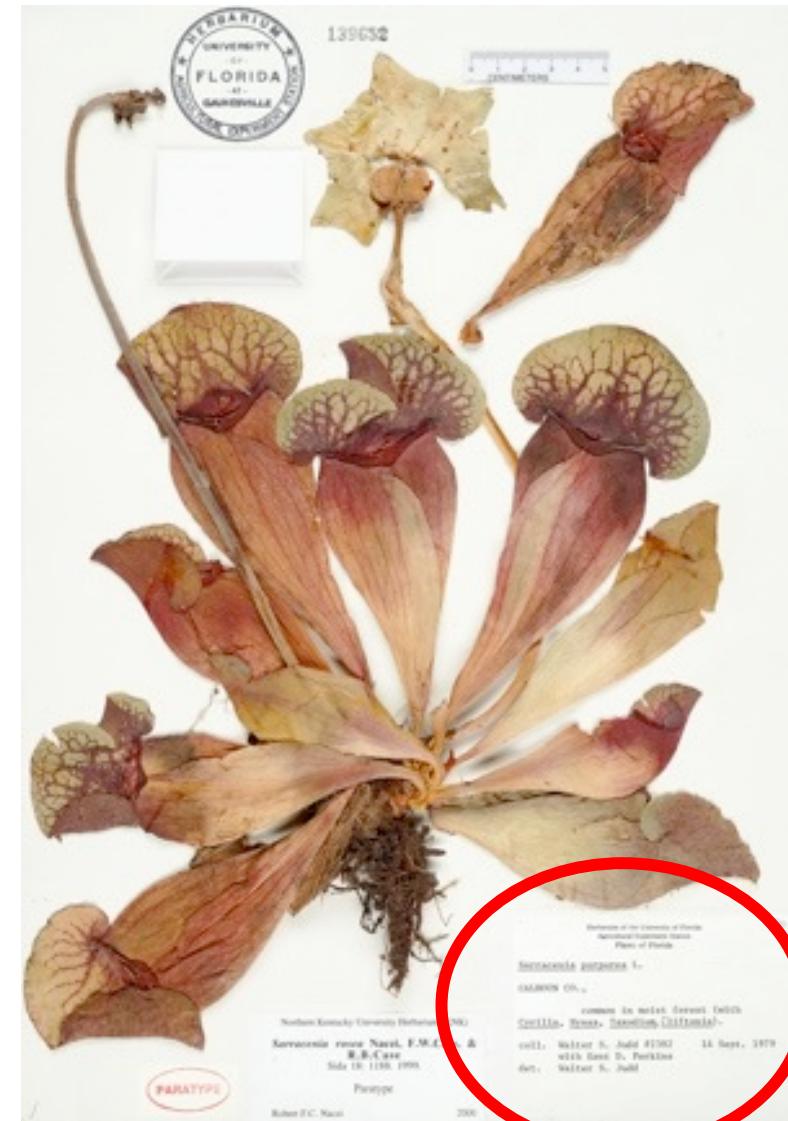
DIGITIZATION



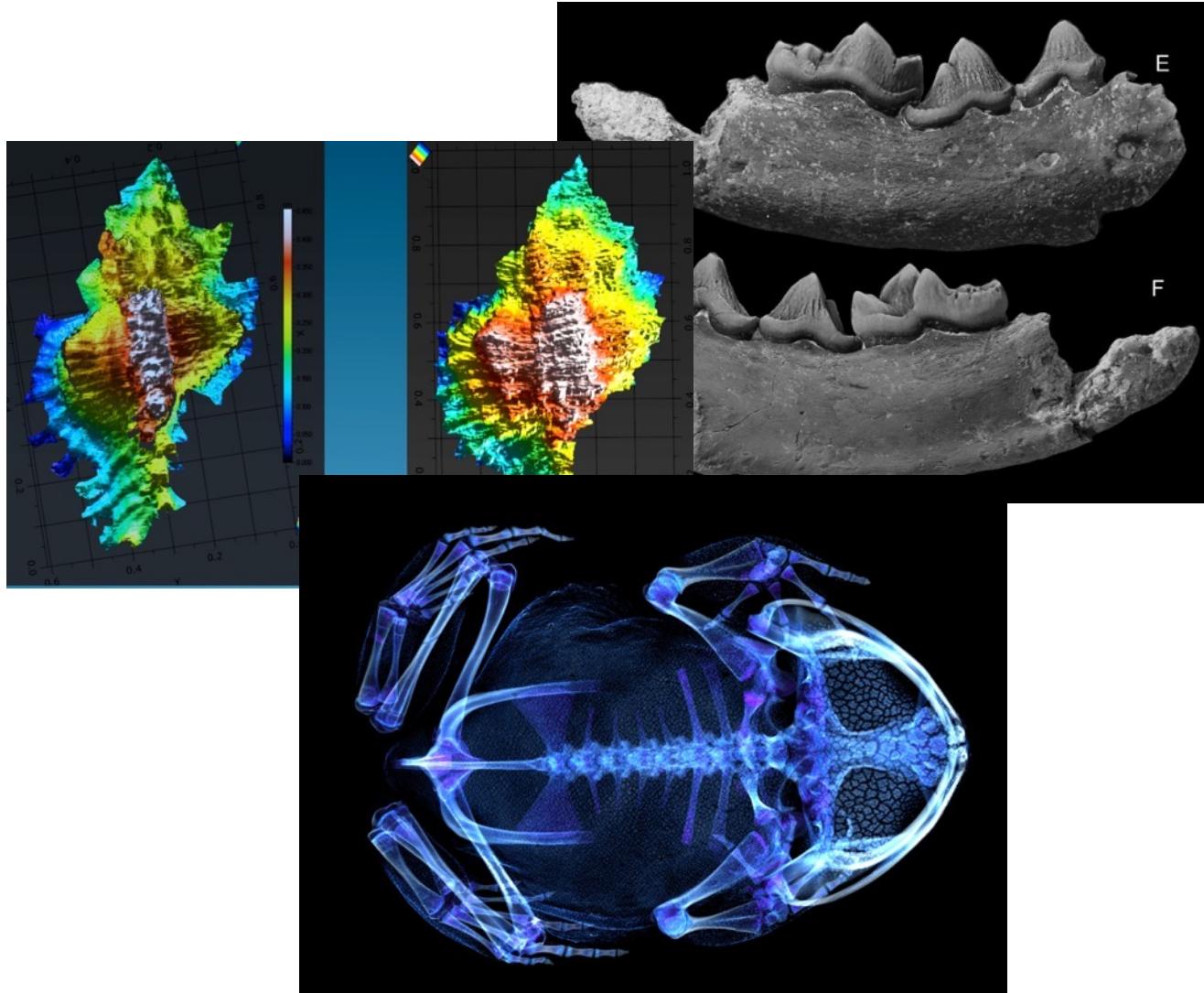
Smithsonian

Digitization: Data and Images

- Scientific name – including authority
- Date
- Collector
- Location – state, county, specific site, GPS coordinates
- Associated species
- Notes



Digitization: Data and Images



iDigBio: www.idigbio.org



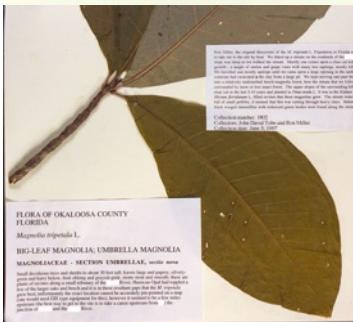
National Coordinating Center For Digitization of Biodiversity Collections

Ingest, serve, integrate data:

Localities

Dates

Images



iDigBio: A Source for Specimen Records

The screenshot shows the iDigBio website homepage. At the top, there's a navigation bar with links for "About iDigBio", "Research", "Technical Information", "Education", and "Log In". Below the navigation is a search bar with the placeholder "ENHANCED BY iDigBio" and a magnifying glass icon. The main content area features a large image of a biological specimen (a porous, yellowish-brown structure) on the left. Overlaid on this image is the text: "Making data and images of millions of biological specimens available on the web". To the right of the image are three statistics: "137,571,430 Specimen Records", "52,529,865 Media Records", and "1,797 Recordsets". Below these stats is a green button labeled "Search the Portal". Further to the right is a yellow section titled "WHY DIGITIZE?" featuring a video thumbnail and the text "Why digitization matters" and "More about what we do and why". At the bottom, there are five colored boxes representing different services: "Digitization" (green, camera icon), "Sharing Collections" (light green, double arrows icon), "Working Groups" (light blue, people icon), "Proposals" (light teal, lightbulb icon), and "Citizen Scientists" (light blue, microscope icon).

About iDigBio | **Research** | **Technical Information** | **Education**

ENHANCED BY iDigBio

Log In

137,571,430
Specimen Records

52,529,865
Media Records

1,797
Recordsets

Search the Portal

WHY DIGITIZE?

Why digitization matters
More about what we do and why

Digitization
Learn, share and develop best practices

Sharing Collections
Documentation on data ingestion

Working Groups
Join in, contribute, be part of the community

Proposals
New tool and workshop ideas

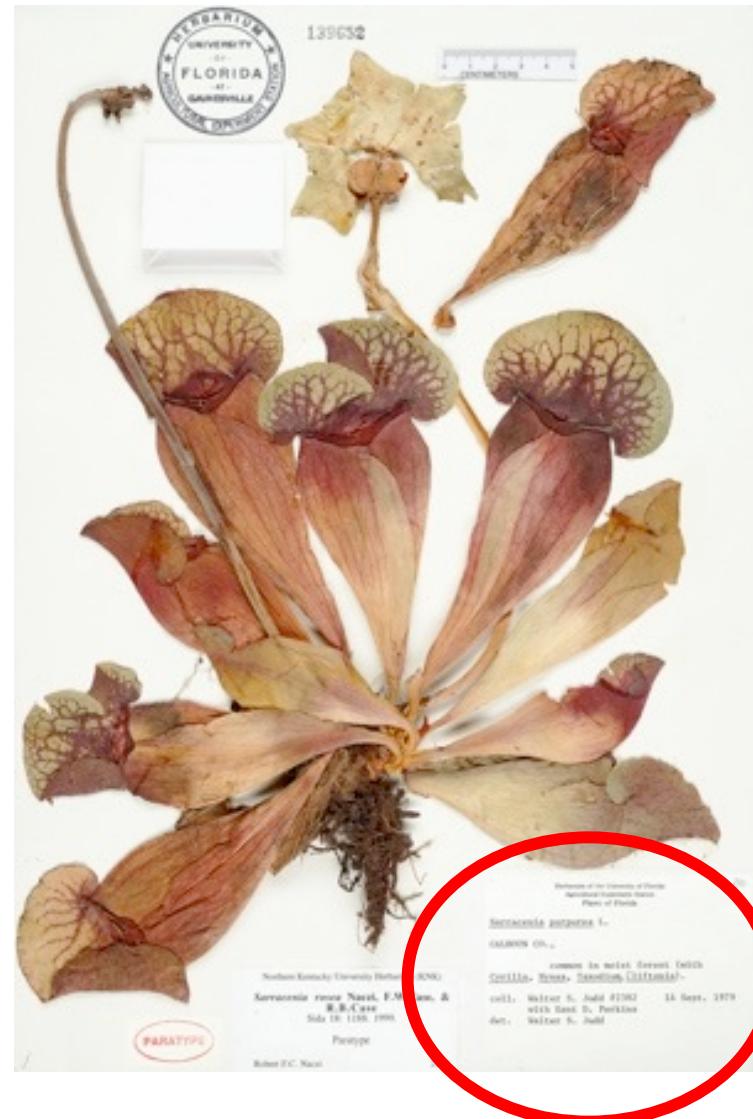
Citizen Scientists
How can you help biological collections?

Specimen Occurrences in iDigBio

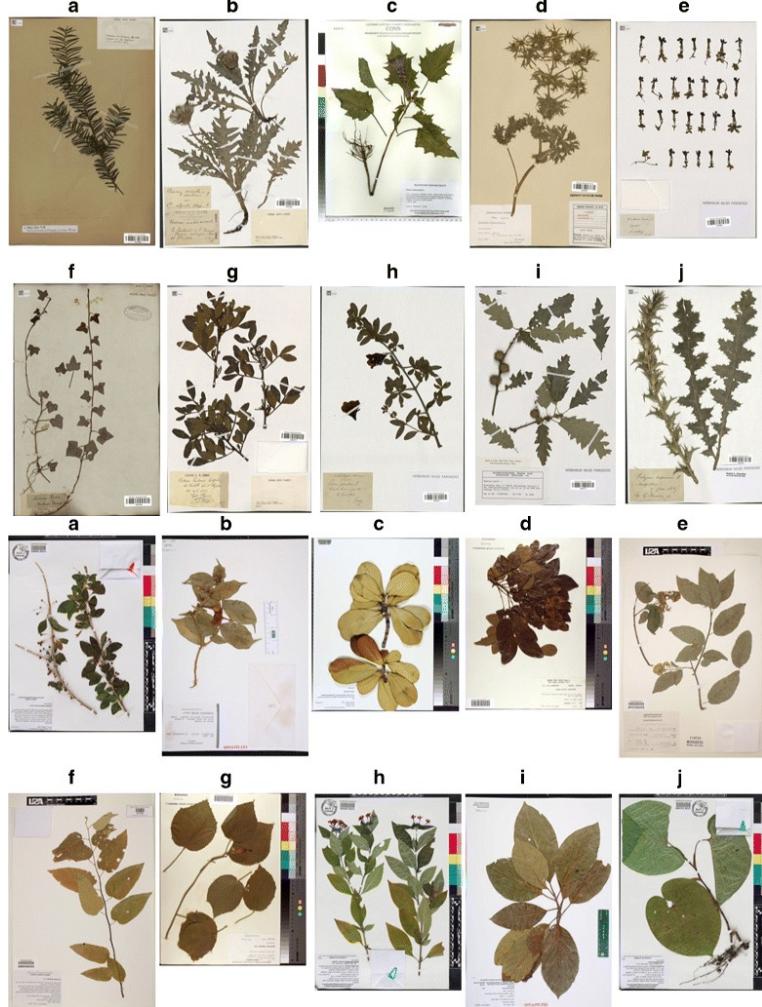


Label Data from Herbarium Specimens

- Scientific name – including authority
- Date
- Collector
- Location – state, county, specific site, GPS coordinates
- Associated species
- Notes



Machine Learning & Biodiversity Research

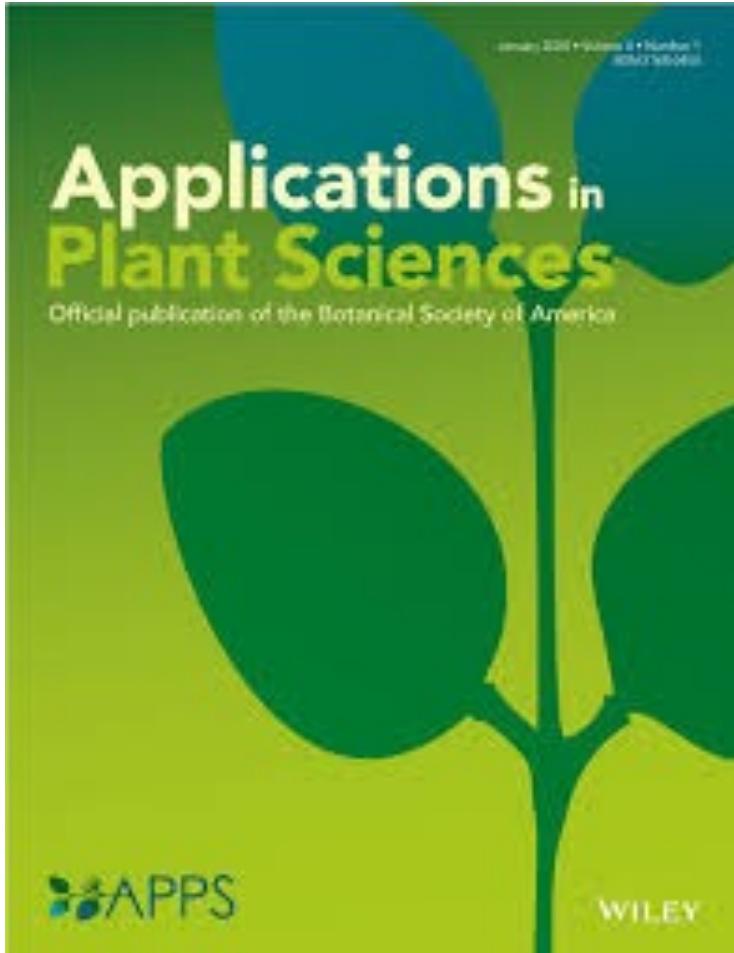


Carranza-Rojas et al. 2017



Pearson et al. 2020

Machine Learning & Plant Biology: Special Issue of APPS (2020)



June: Herbarium Specimen Images

- Phenology (2)
- Morphology
- Herbivory
- Species Identification (2)
- Software: Leaf Detection
- Software: Leaf Recognition and Measurement

Edited by P. Soltis, G. Nelson, A. Zare, E. Meineke

Other Data Aggregators

The image is a collage of logos for various biodiversity data aggregators, arranged in a grid-like structure. The logos include:

- iDigBio**: Integrated Digitized Biocollections. Logo features four stylized leaf and shell icons in green, blue, yellow, and orange.
- VerNe**: A logo featuring the letters 'Ver' in black and 'Ne' in green, with small plant icons integrated into the 'e's.
- Global Biodiversity Information Facility**: Free and Open Access to Biodiversity Data. The background is a green world map.
- NSII**: National Specimen Information Infrastructure. Logo features large blue letters 'NSII' with a green circular icon, and Chinese text '国家标本平台' below it.
- Canadensys**: A red banner with white text and a globe icon.
- Atlas of Living Australia**: ala.org.au. Logo features a red stylized 'X' or 'ala' shape.
- specieslink**: The word 'species' in orange and 'link' in grey.

'Big Data' Research in Biodiversity Science

- Monitoring shifts in biodiversity
- Tracking invasive species
- Ecological Niche Modeling, climate change
- Past movements and climate change
- Tracking phenological shifts
- Landscape genetics
- Integration of ENM with phylogeny
- Community phylogenetics/assembly
- Biogeography



Overview of Day's Activities

- | | |
|--------------|---|
| 8:00 | Welcome and Overview of the Workshop – Pam |
| 8:15 | Darwin Core and Data Fields – Pam |
| 8:30 | Data Downloading Overview – Makenzie |
| 9:00 | <i>Activity:</i> Manual Data Downloads, iDigBio Portal – Lauren |
| 9:30 | <i>Activity:</i> R-based Data Downloads – Shelly |
| 9:45 | Data Cleaning Overview – JT |
| 10:00 | Break |
| 10:30 | <i>Activity:</i> Data Cleaning (A) Manual (JT)
(B) R-based (Shelly) |
| 11:00 | Georeferencing Overview – Lauren |
| 11:15 | <i>Activity:</i> Georeferencing (A) Manual (Lauren)
(B) R-based (Makenzie) |

Overview of Day's Activities

12:15	Lunch
1:00	Climatic Processing Overview – Shelly
1:20	<i>Activity:</i> Climatic Processing – (A) Manual (Elizabeth) (B) R-based (Shelly)
2:00	Ecological Niche Overview – Makenzie
2:15	ENM Settings Overview – JT
2:30	<i>Activity:</i> Ecological Niche Modeling (A) Manual (Elizabeth) (B) R-based (Shelly)
3:00	Break
3:30	Interpreting ENMs Overview – Shelly
3:40	<i>Activity:</i> Interpreting ENMs (A) Manual (Elizabeth/Lauren/Makenzie) (B) R-based (Shelly)
4:00	Post-ENM analysis – Shelly
4:15	BiotaPhy: An Overview – Doug/Maria
4:30	<i>Wrap-up and Q&A</i>
5:00	End