



Welcome to ADBC!

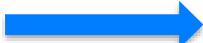
The iDigBio Team
ADBC Summit 2017

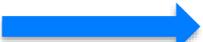


2010. Strategic Plan for the Network Integrated Biocollections Alliance (NIBA)

“The plan outlines the elements required for an inclusive, vibrant, partnership of US biological collections that collectively will **document the nation’s biodiversity resources and create a dynamic electronic resource** that will serve the country’s needs in answering critical questions about the environment, human health, biosecurity, commerce and the biological sciences.”

Collections community

 **NSF's Advancing Digitization of Biodiversity Collections (ADBC) Program
(1st funding in 2011)**

 **2013. Implementation Plan for the
Network Integrated
Biocollections Alliance (NIBA)**

U.S. National Science Foundation

In response to the scientific community

Advancing Digitization of Biodiversity Collections Program (ADBC)

- The goal of **ADBC** is to remove the inaccessibility of data through digitization and making the data available online--for researchers, educators, policymakers, etc.



\$100 million over 10 years
non-federal collections



Coordinating Center Advancing Digitization of Biodiversity Collections (ADBC)



Principal Investigators:

Greg Riccardi, Digitization & Training (FSU)
Jose Fortes, Computational Activities (UF/ACIS)
Pam Soltis, Research Collaborations (FLMNH)
Bruce MacFadden, E&O Collaborations (FLMNH)
Larry Page, Director (FLMNH)

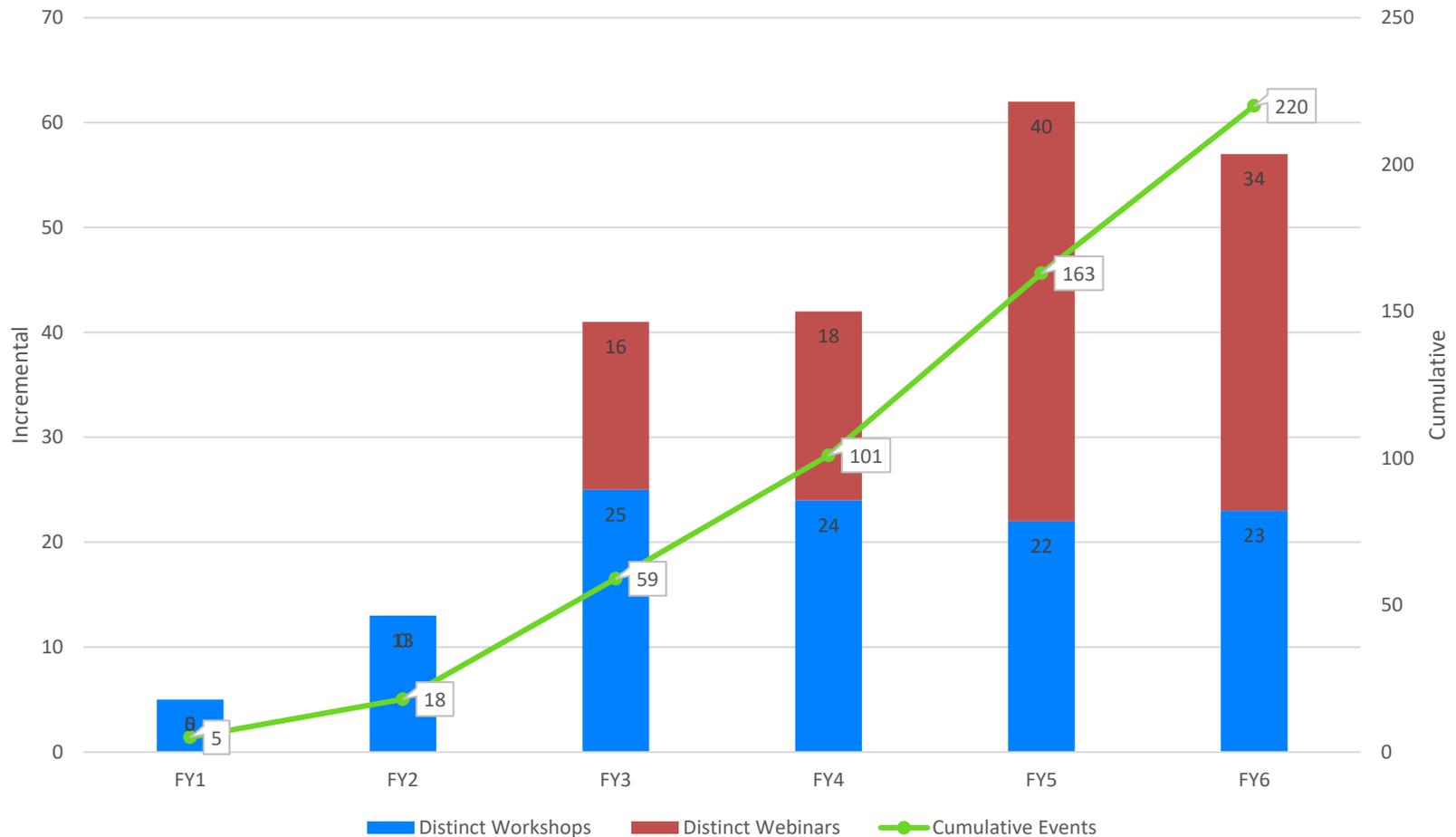
What does iDigBio do?

- **Engage the collections community** – find the specimens
- **Enable digitization of biodiversity collections data**
 - Develop efficient & effective standards & workflows
 - Workforce education & training via workshops/webinars
- **Provide portal access to biodiversity data in a cloud computing environment**
 - Enable data access & discoverability
 - Respond to cyberinfrastructure needs
- **Promote use of biodiversity data to address key environmental and economic challenges**
 - Researchers, educators, general public, policy-makers, etc.
- **Assist in planning for long-term sustainability of the national digitization network & effort**



Number of Workshops and Webinars

Number of iDigBio Workshops and Webinars
7/1/2011 thru 7/11/2017



20 Thematic Collections Networks (TCNs)

- **InvertNet:** An Integrative Platform for Research on Environmental Change, Species Discovery and Identification (*Illinois Natural History Survey, University of Illinois*)
- **Plants, Herbivores, and Parasitoids:** A Model System for the Study of Tri-Trophic Associations (*American Museum of Natural History*)
- **North American Lichens and Bryophytes:** Sensitive Indicators of Environmental Quality and Change (*University of Wisconsin Madison*)
- **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 Arthropods Network (SCAN):** A Model for Collections Digitization to Promote Taxonomic and Ecological Research (*Northern Arizona University*)
- **The Macroalgal Herbarium Consortium:** Accessing 150 Years of Specimen Data to Understand Changes in the Marine/Aquatic Environment (*University of New Hampshire*)
- Developing a Centralized Digital Archive of **Vouchered Animal Communication Signals** (*Cornell University*)
- **Fossil Insect Collaborative:** A Deep-Time Approach to Studying Diversification and Response to Environmental Change (*University of Colorado at Boulder*)
- **Great Lakes Invasives:** Documenting the Occurrence through Space and Time of Aquatic Non-indigenous Fish, Mollusks, Algae, and Plants Threatening North America's Great Lakes (*University of Wisconsin Madison*)
- **InvertEBase:** Reaching Back to See the Future: Species-rich Invertebrate Faunas Document Causes and Consequences of Biodiversity Shifts (*Field Museum of Natural History*)
- **The Key to the Cabinets:** Building and Sustaining a Research Database for a Global Biodiversity Hotspot (*Appalachian State University*)
- **The Microfungi Collections Consortium:** A Networked Approach to Digitizing Small Fungi with Large Impacts on the Function and Health of Ecosystems (*INHS, University of Illinois*)
- **Documenting Fossil Marine Invertebrate Communities of the Eastern Pacific:** Faunal Responses to Environmental Change over the last 66 million years (*University of California-Berkeley*)
- **Cretaceous World:** The Cretaceous World: Digitizing Fossils to Reconstruct Evolving Ecosystems in the Western Interior Seaway (*University of Kansas*)
- **LepNet:** Lepidoptera of North America Network: Documenting Diversity in the Largest Clade of Herbivores (*Northern Arizona University*)
- **MAM:** The Mid-Atlantic Megalopolis: Achieving a greater scientific understanding of our urban world (*University of Pennsylvania*)
- **SoRo:** Using Herbarium Data to Document Plant Niches in the High Peaks and High Plains of the Southern Rockies (*University of Colorado*)
- **oVert:** Open Exploration of Vertebrate Diversity in 3D (*University of Florida*)

279 institutions

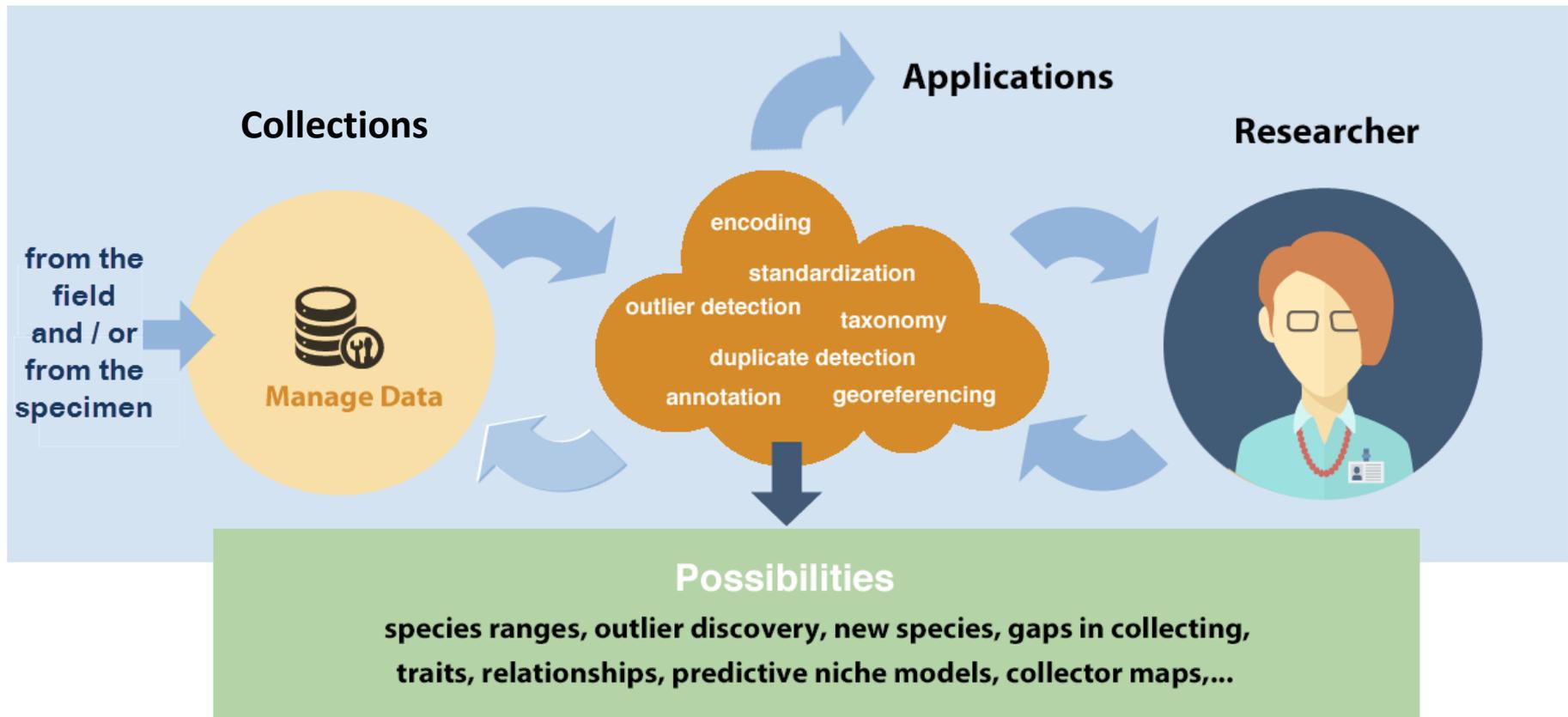
**iDigBio is working with 561
collections in 336 institutions**

**iDigBio Portal has
105M records for
~315M specimens
with 22M
associated media
records**



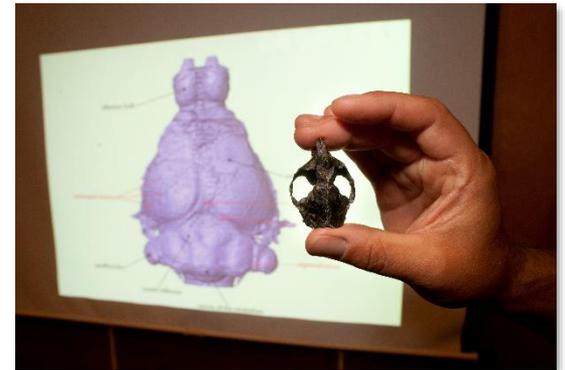
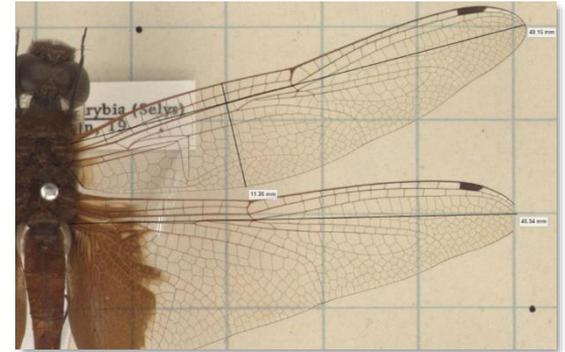


So Many Acronyms!



An acronym overview

- ADBC, TCNs and PENs
- Digitization, Data mobilization, Community Building
- Aggregators +
- Education, Outreach, Inreach
- Publishing
- Got more? (Of course you do!)
- Full glossary on the Wiki
 - <http://bit.ly/adbcterm>



Got More?

<http://bit.ly/adbcterm>

ACIS		ry	GEOLocate	LBCC	NSF	SCAN
ADBC				Net	OCR	SEINET
				aries of Life	Open Refine	SEPASAL
AIM-UP!	DataCite		GUID, UUID	Lifemapper	ORCiD	SERNEC
ALA	DataONE		GUODA	MaCC	oVert	SoRo
API	Digitarium		ICBN	MaM	PaleoCore	Spark
Arctos	DiSSCo		ICZN	MaNIS	PaleoDb	Specify
ARK	DOI		iDigBio	MapsData	Paleoniches	SPNHC
ARPHA	Drupal		IGSN	Mendeley	PEN	SQL
Audubon Core	DataDryad		iNaturalist	MiCC	Pensoft	SQLite
Axiell EMu	DwC		InverteBase	Morphbank	PhyloLink	SYMBIOTA
BCoN	EarthCube		InvertNet	Morphobank	Picturae	SYNTHESYS
BHL	EoL		iPlant	MorphoSource	Plazi	TCN
BISON	ePandda		IPNI	MySQL	PlutoF	TDWG
Canadensys	EPICC		IPT	NEON	Python	TTD
CartoDB	figShare		iSamples	NIBA	QGIS	USVH
CONABIO	FOSSIL		ITIS	Notes from	R, ridigbio,	VACS
CRIA	GBIF		IUCN	Nature	rgbif, rvertnet	VertNet
CSIRO	GenBank		Red List	NEVP	ROpenSci	WeDigBio
DAMS	GitHub		Kurator	NSCA	RStudio	WoRMS

From NIBA to BCoN & the ADBC

- Networked Integrated Biocollections Alliance (NIBA)
 - Resulted in Advancing the Digitization of Biological Collections (ADBC)
- Biodiversity Collections Network (BCoN)
 - AIBS (American Institute for Biological Sciences), iDigBio & NSC Alliance (Natural Science Collections Alliance)

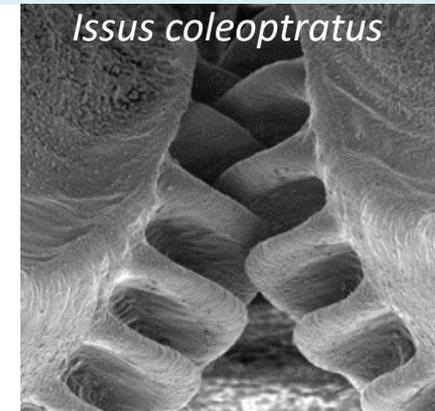


**BIODIVERSITY
COLLECTIONS NETWORK**

<http://bcon.aibs.org/>

Advancing the Digitization of Biological Collections (ADBC)

- iDigBio, that's
 - Integrated Digitized Biocollections
- Thematic Collection Networks (TCNs)
 - network of institutions digitizing around a research theme
- Partners to Existing Networks (PENs)
 - digitizing specimens for existing TCN to fill gaps



Credit: Malcolm Burrows

Thematic Collections Networks (TCNs) and Partners to Existing Networks (PENs)

TCN: network of institutions strategically digitizing information for a particular research theme, such as impacts of climate change or biota of a region.



iDigBio – related projects

- Libraries of Life (a free app)
 - Collections in 3-D, developed in collaboration with TCNs, Arizona State University
- WeDigBio (Worldwide Engagement for Digitized Collections)
 - 4-day worldwide transcription event
- The FOSSIL Project
 - Engaging amateur fossil clubs and societies as downstream users of digitized data



Digitization

Collection Management Software

- Arctos 
- Symbiota 
- Specify 
- Axiell EMu 
- collectionspace 

Community Building

- SPNHC 
- GRBio 
- SCNet 
- ECN 
- Darwin Core Hour 
- Working Groups: DROID, GWG, SWG, PaleoDigi, NANSH,...

Data Mobilization

- TDWG - Biodiversity Information Standards
- DwC - Darwin Core
- AC - Audubon Core
-  **GBIF** INTEGRATED PUBLISHING TOOLKIT (IPT)
free and open access to biodiversity data
- Kurator
-  **GEOLocate**
- DAMS – Digital Asset Management
- ABBYY, Tesseract
- *Unique identifiers*
 - GUID, UUID, ARK, IGSN

Data aggregation and Data use

Biodiversity Data Mobilization and Use

- **EoL** - Encyclopedia of Life
- **BHL** – Biodiversity Heritage Library
- **NEON** – National Ecological Observatory Network
- **CyVerse** (iPlant) - facilitating scientific research in the cloud
- **GBIF** – Global Biodiversity Information Facility
- **iDigBio**
- **VertNet**
- **BISON** – Biodiversity Serving our Nation
 - GBIF North American Node
- **USVH** – United States Virtual Herbarium
- **ALA, DigiVol** – Atlas of Living Australia
- **Canadensys**
- **SiBBR** – Brazilian Biodiversity Information System
- **CONABIO** (Mexico)
- **CRIA** (Brazil) and **SpeciesLink**



Education, Outreach, Inreach



- AIM-UP!
 - Advancing Integration of Museums into Undergraduate Programs; using collections data in undergraduate education



- BLUE
 - Biodiversity Literacy in Undergraduate Education



- Notes from Nature
 - Crowdsourcing collections transcription, creating communities



- iNaturalist
 - observation app, connecting scientists and the general public



- ePANDDA
 - Enhancing Paleontological and Neontological Data Discovery API, connecting scientific literature with specimens



- The Carpentries - Data Carpentry & Software Carpentry, Reproducible Science Curriculum,...

- **Biodiversity informatics skills** for those in the biodiversity community. Focus on tidy data, fit for reproducible research.



Publishing



• Pensoft

- ARPHA Writing Tool (AWT), BISS



• DataCite, DOI (Digital Object Identifier)

- Find, share and reuse, cite data, connect and get credit



• Mendeley

- Free reference manager; organize papers, read & annotate your PDFs



• ORCID

- Open Researcher and Contributor ID



• DataONE

- Data Observation Network for Earth (DataONE) – data repository and data management best practices



• Data Dryad

- Curated resource making the data underlying scientific publications discoverable, freely reusable, and citable general-purpose home for a wide diversity of data types



• Figshare

- online repository where researchers can preserve and share their research outputs, including figures, datasets, images, and videos. #openData

AZiA
ADW
AIM-L
ALA
Amph
APC
API
APNI
Arbor
ArcGIS
Arctos
ARK
ARPHA
AudioNote
Audubon Core
AutoMontage
Axiell
bccvl
BCoN
BoL
BerkeleyMapper
BHL
BHL - China
BHL - Europe
Bio2RDF
BioCAsE
biocode
BioGeoMancer
BioGUID
BioNames
BioOffice
BioSharing
BIOTA
BioVeL
BiSciCol
BISON
BITC
BOLDSystems
BONAP
BSA
BugGuide
CABIN
CalPhotos
Canadensys
CartoDB
Catalogue of Life
CBIF
CCH
CF21

COOL
CETAF

Exploring Genomics
Data
EXIF
EZID
Falling rain
Fauna Europaea
FIC

herbaria@home
HERBIS
HerpNET
HOLOS
HUBzero
Hymenoptera Online
ICRN

MaM
MaNIS
MANTIS
MapWindow GIS
MapReduce
MapsData
MBB
Mendeley
MIBBI
MICC
MOBOT
ModestR
MOL
Morphbank
Morphobank
MorphoSource
Morphster
MHC
mx
MySQL
NameBank
ClassificationBank
National Biodiversity
Data Center
NatureServe
NCSA
NEOMAP
NEON
NeotomaDB
NESCent
NEVP
NIBA
NIMBUS
NLP
Notes from Nature
NSCA
NSF
NSII
OBIS
OBO Foundry
OCR
OCROPUS
Open Atrium
Atrium
openModeller
Open Refine
OpenStack
Open Tree of Life
OpenUp!
OpenZoom
Oracle
ORCID

ORNIS
OWL
PaleoCore
PaleoDb
PaleoNICHES
PaleoPortal
PEN
Pensoft
PeriodO
PhenoBlast
Phenoscape
Photoshop
Photosimile
Phylocode
Phylojive
PhyloLink
Picturae
Planting Science
PLANTS Database
Plazi
PlutoF
PNW Herbaria
PostGIS
PostgreSQL
prefixcommons
pro-iBiosphere
PROJ.4
Psql'list
Python
QGIS
R
Raintree
RDA
RDF
Recorder 6
Re:discovery
Redmine
ReFindit
REST
ridigbio
RMCA
ROpenSci
RStudio
SCAN
Scratchpads
SEINET
SEPASAL
SERNEC
SiB
SilverCollection
SilverLining
SOAP
Spark
SPARQL
Species 2000

speciesLink
Specify
SPNHC
SQL
SQLite
STEM
SYMBIOTA
SYNTHESES
SysTax
TACC
TAPIR
Taxonworks
TCN
TDWG
Tesseract
The Field Book Project
TNRS
TOLKIN
TORCH
tranScriptorium
Transkribus
TreeBASE
Trifacta
TROPICOS
TRY
TTD
Tri-Trophic TCN
TurboScan
uBio
UFBI
UNITE
Universal Chalcidoidea
Database
USVH
VACS
VASCAN
Vernon Systems
VertNet
VIBRANT
VIVO
VMWare
WCSP
Windows Azure
WeDigBio
WISFlora
WoRMS
WSDL
Xen
XML
XSEDE
YAMZ
Zenodo
Zoomify
Zooniverse

Got More?

<http://bit.ly/adbcterm>

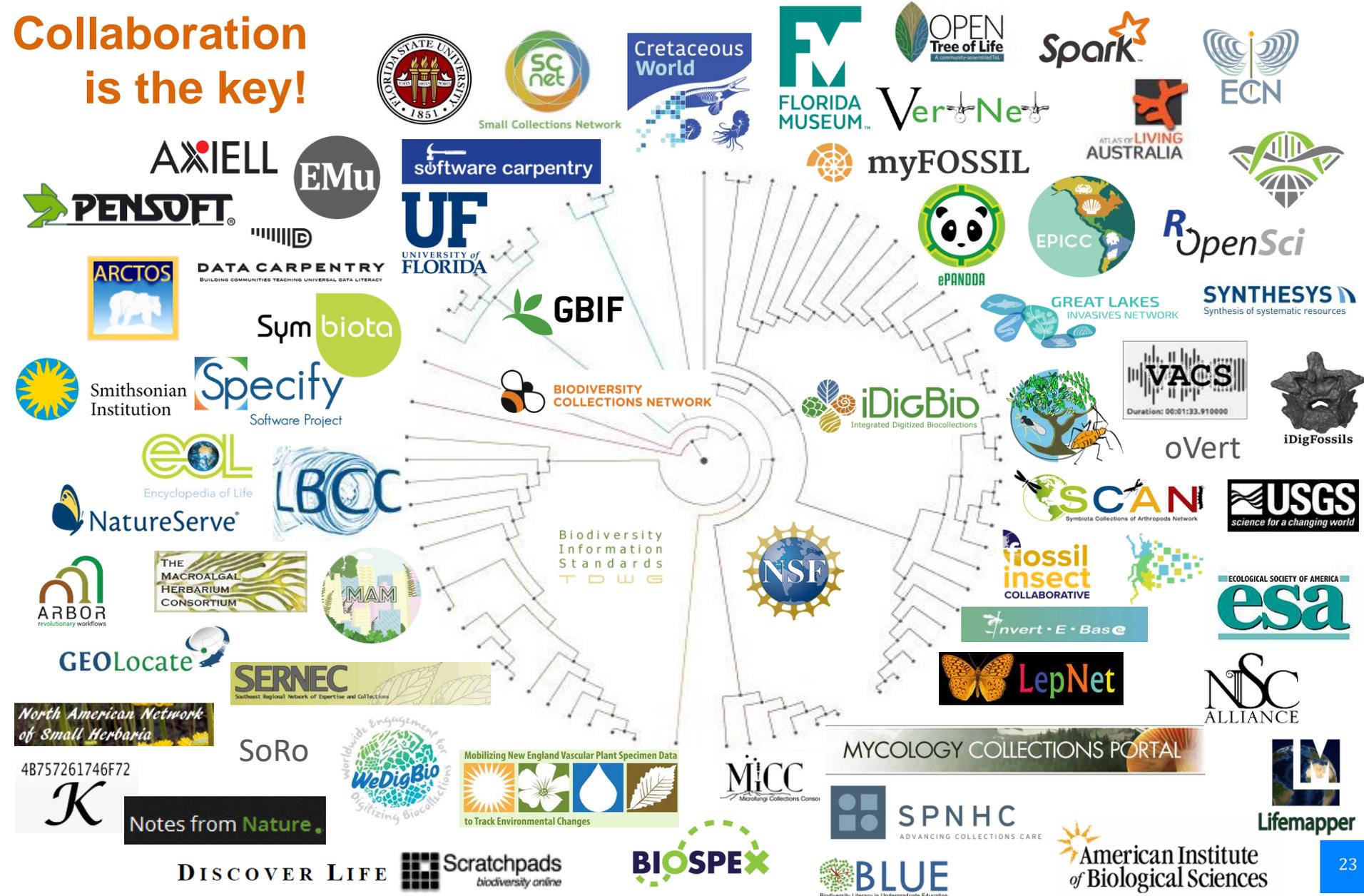
FOSSIL
From the Page
Fuzzy Gazetteer
GBIF
GCI
Grey Card Index
GDAL
GenBank
geneious
Gene Ontology
Geopaparazzi
Geotag Photos Pro
Thesaurus of
Geographic Names
GitHub
GIMP
GLANSIS
GloBI
GLOBIS
GNIS
GEOlocate
Georeferencing
Georeferencing
Calculator
GEOS
Geotools
GLI
GigaPan
CaptureStudio
Global Plants
GLOBIS
GLOBIS-B
GNA
Google Analytics
Google Earth
Google Maps
GPlates
GPS Visualizer
Grass GIS
GRBio
GUID
UUID
Hadoop

CRIA
CrossRef
CSIRO
CyVerse
DAMS
DataCarpentry
Software Carpentry
DataCite
DataONE
DataONE Dash
DataTurbine
DBTNT
DiGIR
Digital Florida
Digitarium
DINA
Discover Life
DIVA-GIS
djatoka
DOI
DroidDB
Drupal
Dryad
DSLr
DwC
DwC Terms
EarthCube
eBiodiversity
eBird
EcoInforma
ECOS
Effechecka
EMU
KE EMU
EnviroAtlas
EOL
EpiCollect
ePanda
ESFRI
EPICC
EU BON
EUROPEANA

FOSSIL
From the Page
Fuzzy Gazetteer
GBIF
GCI
Grey Card Index
GDAL
GenBank
geneious
Gene Ontology
Geopaparazzi
Geotag Photos Pro
Thesaurus of
Geographic Names
GitHub
GIMP
GLANSIS
GloBI
GLOBIS
GNIS
GEOlocate
Georeferencing
Georeferencing
Calculator
GEOS
Geotools
GLI
GigaPan
CaptureStudio
Global Plants
GLOBIS
GLOBIS-B
GNA
Google Analytics
Google Earth
Google Maps
GPlates
GPS Visualizer
Grass GIS
GRBio
GUID
UUID
Hadoop

IDQ
ISBER
IGSN
IH
IK
ILDIS
ImageMagick
iNaturalist
INBio
Index Fungorum
InverteBase
InvertNet
iPlant
IPNI
IPT
IPT 2
IrfanView
iSamples
ITIS
IUCN
Red List
IWGSC
JAI
Jetstream
JPEG 2000
JSON
JSTOR
JSTOR Plant Science
JTS
KML
Kurator
LBCC
LepNet
LibXML
Libraries of Life
LifeDesks
Lifemapper
LifeWatch
LINCAOCNET
Lucid
Macaulay Library
MaCC
Magnolia grandifLORA

Collaboration is the key!





iDigBio Resources

Digitization Process

rev. 7/3/2009
rev. 5/8/2012

START

END

PROOFREAD

PROVIDE
FEEDBACK
ON DATA ENTRY

REVIEW AND
RESOLVE
QUESTIONS

PROOFREAD
AND CORRECT
DATA

Researchers

Browse our specimen portal →

Collections Staff

Learn how your collection can benefit from our work →

Teachers & Students

Learning resources & opportunities to engage →

Undergraduate Resources

Online Resources for Undergraduate Students and Educators



WeDigBio collaborated with collectionseducation.org to create an undergraduate classroom exercise about citizen science. Find this resource (others coming soon) on their [education exercises](#) page.

Other Resources

AIM-UP!: Advancing Integration of Museums into Undergraduate Programs



"AIM-UP! is an NSF-funded Research Coordination Network exploring the use of natural history collections in undergraduate education. AIM-UP! is refining existing efforts and developing new integrated approaches to collections-based training in large-scale questions using the expertise of educators, curators, collection managers, database managers, and scientists whose work spans disciplines and relates topics covering a spectrum of time and space."

The AIM-UP! project includes [educational modules](#) and [tutorials](#) to be used in college-level courses.

Learn more about AIM-UP! through a recent presentation ([pdf](#) and [recording link](#)) given at SPNHC 2015 or the recent publication "[Natural History Collections as Emerging Resources for Innovative Education.](#)"

SERC: the Science Education Resource Center at Carleton College

"Whether you're an undergraduate faculty member or a K-12 science teacher our suite of partner projects have materials you can use today. Connect with a community of peers as they share what really works in their classrooms."

SERC the Science Education Resource Center at Carleton College

SERC includes resources for both undergraduate education and K-12 education with an

APPLY

PHOTOGRAPH SPECIMEN

DISCARD IMAGE

CROP, SHARPEN, ADJUST BRIGHTNESS, CONTRAST, ETC.; NAME AND SAVE CORRECTED IMAGE FILE

CHECK FOR RARITY STATUS

YES

BLUR LOCALITY DATA ON IMAGE; RARE-STAMP IMAGE; SAVE CORRECTED IMAGE FILE

iDigBio Website Resources <https://www.idigbio.org>

General

- iDigBio and TCN info
- TCN Resources page
- Collaborators map
- ADBC proposal tips
- Staff Directory
- Calendar of upcoming events (workshops, webinars...)
- News
- Event recaps
- Press releases
- Community announcements

Research

- Monthly Research Spotlights
- Recordings of conference presentations
- List of genetic repositories
- Tutorials
- API information
- iDigBio R package
- Research tools
- Collaborators
- Links to GitHub

iDigBio Website Resources <https://www.idigbio.org>

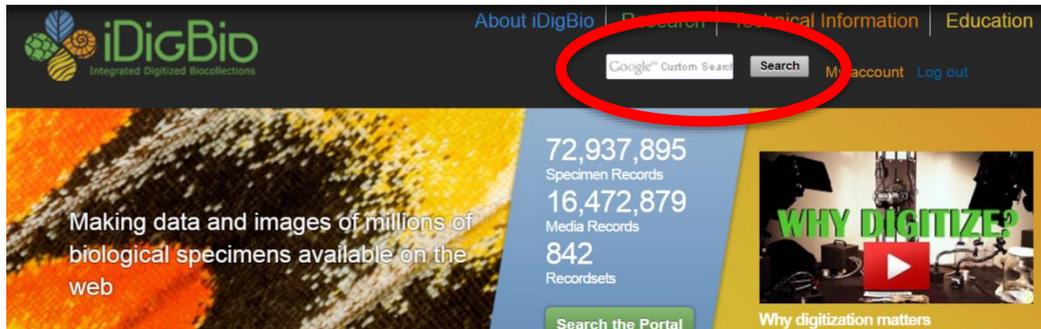
Technical Information

- Working group information
- Bibliography of relevant presentations and publications
- Digitization workflows
- Equipment recommendations
- Workshop summaries
- Workshop and webinar recordings
- Data ingestion guidance

Education

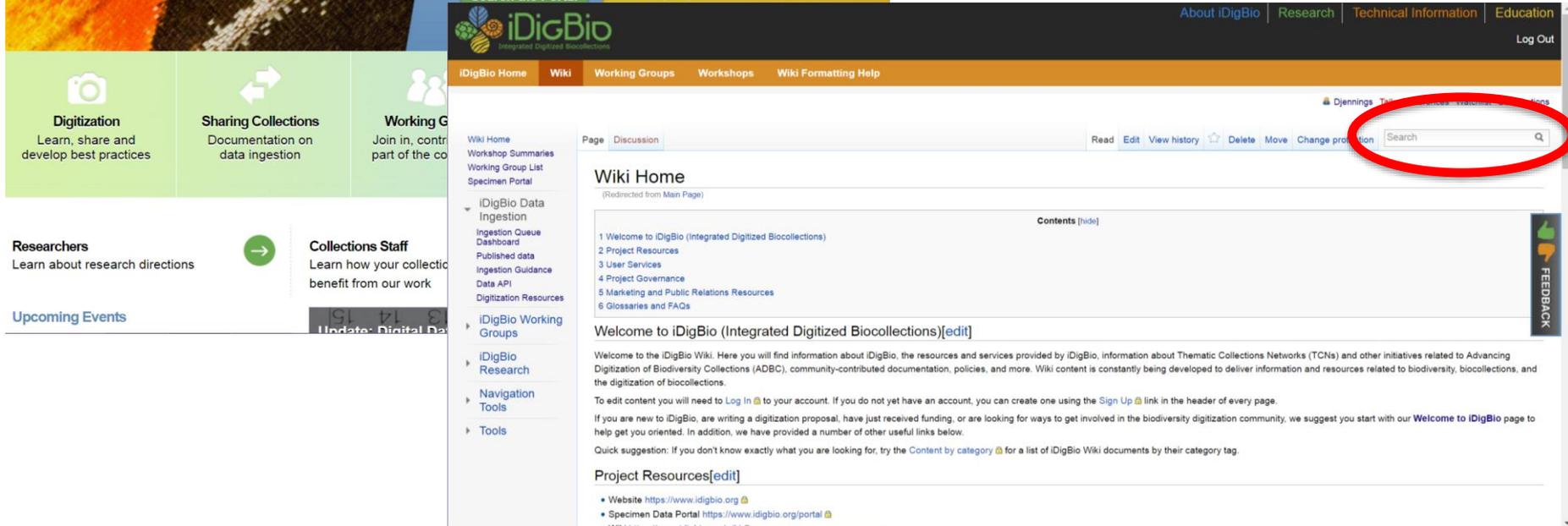
- Links to ADBC educational products
- Monthly Biodiversity Spotlights
- Portal Curiosities
- Coding Corner
- K-12 lesson plans
- Undergraduate modules
- Information about Citizen Science
- Educational collaborators

I can't find X, or I want to know about Y



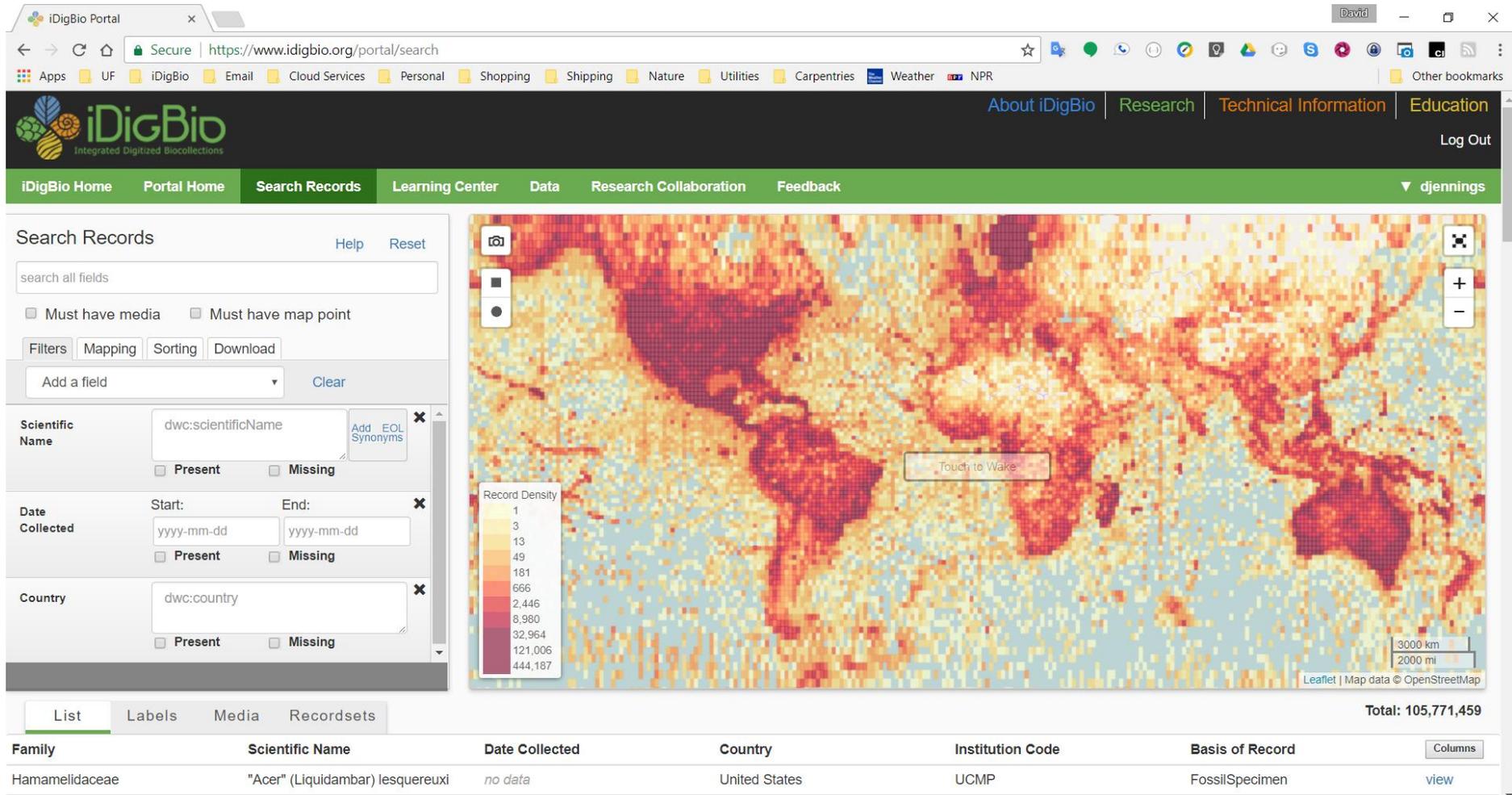
The screenshot shows the top navigation bar of the iDigBio website. A search box labeled "Google™ Custom Search" with a "Search" button is circled in red. To the right of the search box are links for "My account" and "Log out". Below the navigation bar, the main content area features a large image on the left with the text "Making data and images of millions of biological specimens available on the web". On the right, there are statistics: "72,937,895 Specimen Records", "16,472,879 Media Records", and "842 Recordsets". A video thumbnail titled "WHY DIGITIZE?" is also visible.

Use the
search boxes!



The screenshot shows the iDigBio Wiki page. The top navigation bar includes "About iDigBio", "Research", "Technical Information", and "Education", along with a "Log Out" link. Below this is a secondary navigation bar with "iDigBio Home", "Wiki", "Working Groups", "Workshops", and "Wiki Formatting Help". A search box is circled in red in the top right corner. The main content area is titled "Wiki Home" and includes a list of contents: "1 Welcome to iDigBio (Integrated Digitized Biocollections)", "2 Project Resources", "3 User Services", "4 Project Governance", "5 Marketing and Public Relations Resources", and "6 Glossaries and FAQs". A "Project Resources" section is also visible at the bottom, listing links for the website and specimen data portal.

Flexible search across all data, indexed fields, media, geolocation, map boundary, auto-completion, synonyms, ...



The screenshot displays the iDigBio Portal search interface. The top navigation bar includes links for About iDigBio, Research, Technical Information, and Education. The main search area features a search bar, filter options (Must have media, Must have map point), and a list of search filters (Scientific Name, Date Collected, Country) with dropdown menus and checkboxes for 'Present' and 'Missing'. A heatmap map shows record density across a geographical area, with a legend indicating density values from 1 to 444,187. Below the map, a table of search results is shown, with the first row displaying data for the family Hamamelidaceae, scientific name "Acer" (Liquidambar) lesquereuxi, and date collected as no data.

Search Records [Help](#) [Reset](#)

search all fields

Must have media Must have map point

Filters Mapping Sorting Download

Add a field [Clear](#)

Scientific Name [Add EOL Synonyms](#)

Present Missing

Date Collected Start: End:

Present Missing

Country

Present Missing

Record Density

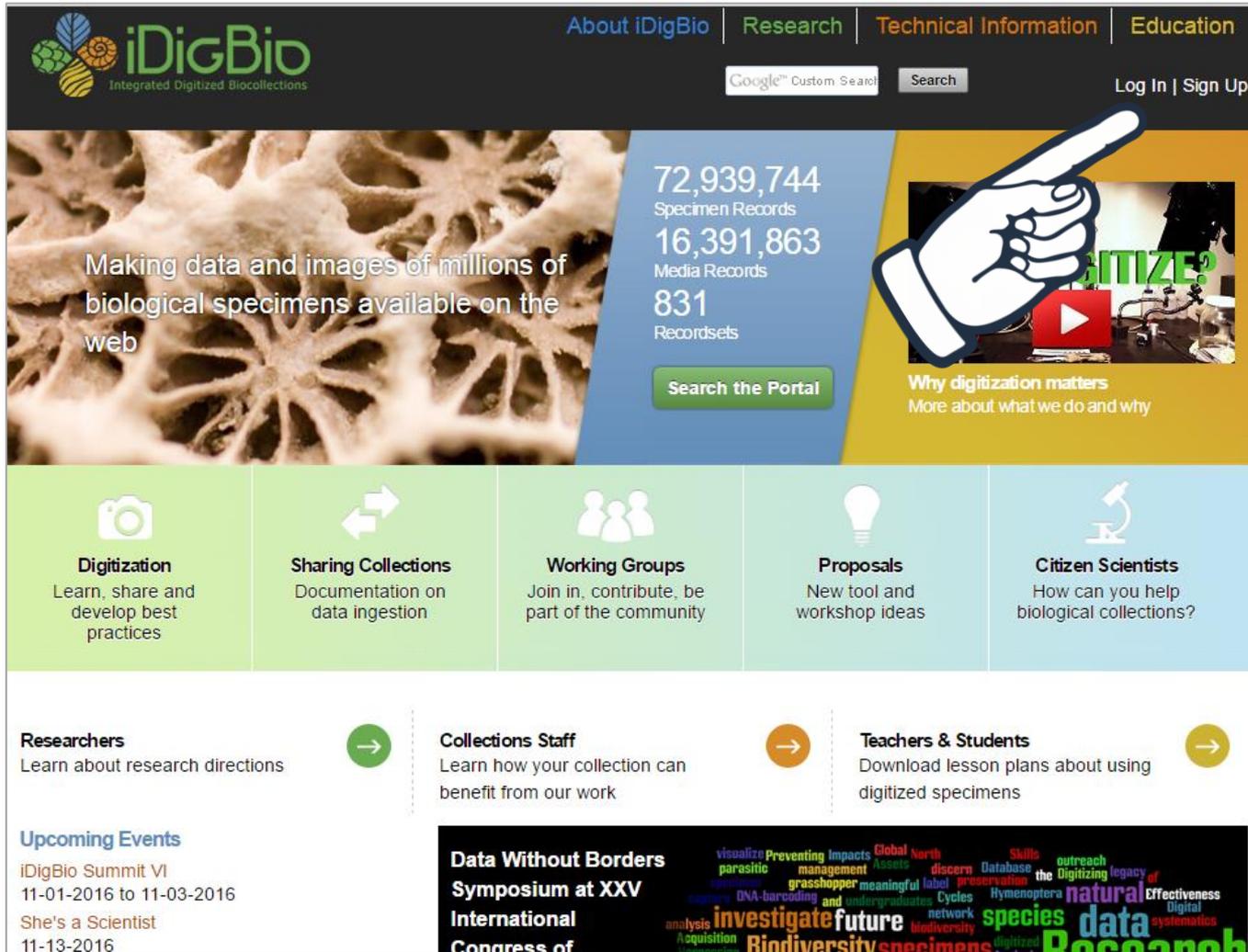
- 1
- 3
- 13
- 49
- 181
- 666
- 2,446
- 8,980
- 32,964
- 121,006
- 444,187

Total: 105,771,459

Family	Scientific Name	Date Collected	Country	Institution Code	Basis of Record	Columns
Hamamelidaceae	"Acer" (Liquidambar) lesquereuxi	no data	United States	UCMP	FossilSpecimen	view

Get Involved!

Step 1: Sign up for an account!



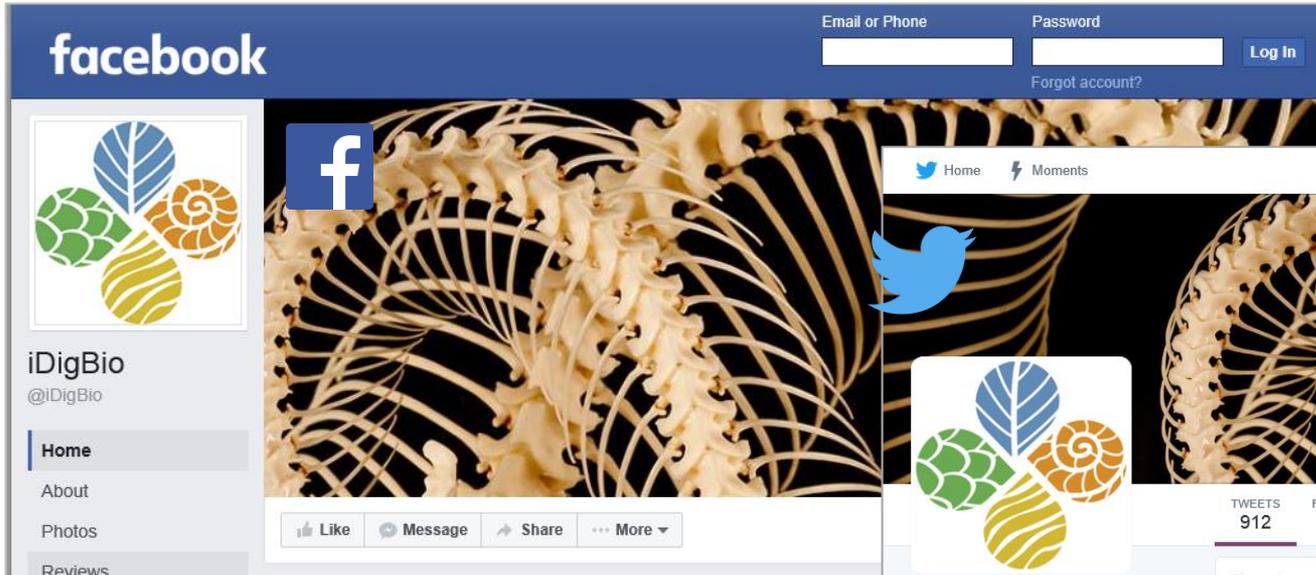
The screenshot shows the iDigBio website homepage. At the top, there is a navigation bar with links for 'About iDigBio', 'Research', 'Technical Information', and 'Education'. Below this is a search bar with 'Google™ Custom Search' and a 'Search' button, and a 'Log In | Sign Up' link. The main banner features a background image of a biological specimen and text stating 'Making data and images of millions of biological specimens available on the web'. To the right of this text, statistics are displayed: 72,939,744 Specimen Records, 16,391,863 Media Records, and 831 Recordsets. A 'Search the Portal' button is located below these statistics. A large hand cursor icon points to the 'Sign Up' link in the top right corner. Below the banner, there are five main service tiles: '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), and 'Citizen Scientists' (How can you help biological collections?). At the bottom, there are three sections: 'Researchers' (Learn about research directions), 'Collections Staff' (Learn how your collection can benefit from our work), and 'Teachers & Students' (Download lesson plans about using digitized specimens). There is also an 'Upcoming Events' section listing 'iDigBio Summit VI' (11-01-2016 to 11-03-2016) and 'She's a Scientist' (11-13-2016). A 'Data Without Borders Symposium at XXV International Congress of...' is partially visible. A word cloud at the bottom right contains terms like 'investigate future', 'species data', 'Biodiversity', 'specimens', 'Research', 'DNA-barcoding', 'network', 'Hymenoptera', 'natural', 'Effectiveness', 'Digital', 'systematics', 'analysis', 'Acquisition', 'visualize', 'Preventing Impacts', 'Global North', 'Assets', 'discern', 'Database', 'outreach', 'legacy', 'the', 'Digitizing', 'management', 'grasshopper', 'meaningful', 'label', 'preservation', 'Cycles', 'Hymenoptera', 'natural', 'Effectiveness', 'Digital', 'systematics', 'analysis', 'Acquisition', 'investigate future', 'Biodiversity', 'specimens', 'Research'.

Step 2: Sign up for the iDigBio Newsletter

- TCN and digitization news
- Upcoming workshops and webinars
- Event recaps
- Articles featuring innovative collections-based research
Biodiversity Spotlights



Step 3: Social media



vimeo.com/idigbio



idigbio.org/rss-feed.xml

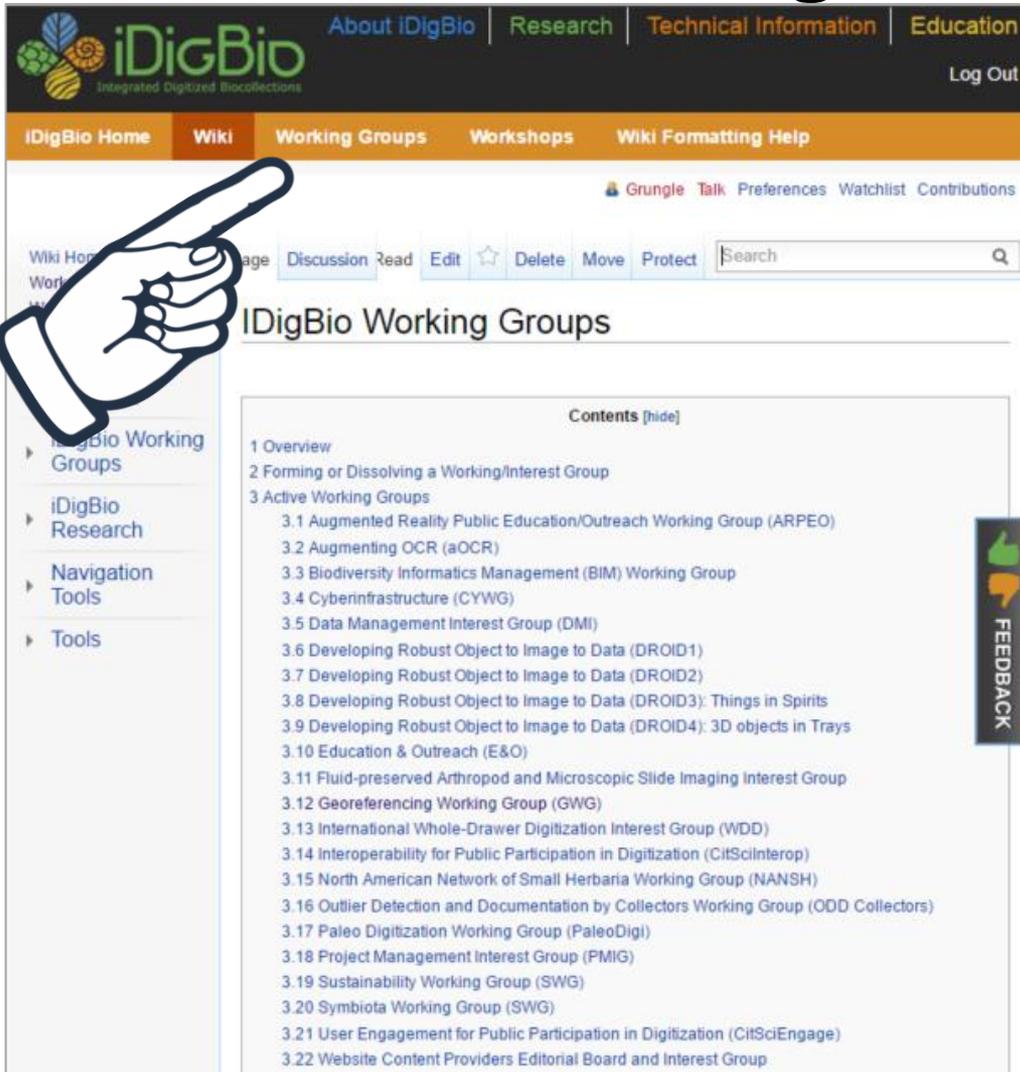


idigbio.org/events-calendar/export.ics



www.idigbio.org/wiki

Step 4: Get involved with a Community Working Group



iDigBio About iDigBio Research Technical Information Education Log Out

iDigBio Home Wiki Working Groups Workshops Wiki Formatting Help

Grungle Talk Preferences Watchlist Contributions

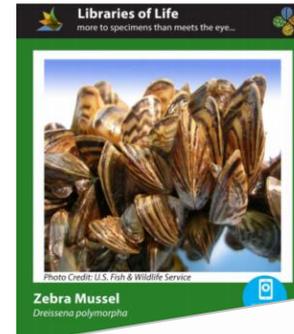
Discussion Read Edit Delete Move Protect Search

iDigBio Working Groups

Contents [hide]

- 1 Overview
- 2 Forming or Dissolving a Working/Interest Group
- 3 Active Working Groups
 - 3.1 Augmented Reality Public Education/Outreach Working Group (ARPEO)
 - 3.2 Augmenting OCR (aOCR)
 - 3.3 Biodiversity Informatics Management (BIM) Working Group
 - 3.4 Cyberinfrastructure (CYWG)
 - 3.5 Data Management Interest Group (DMI)
 - 3.6 Developing Robust Object to Image to Data (DROID1)
 - 3.7 Developing Robust Object to Image to Data (DROID2)
 - 3.8 Developing Robust Object to Image to Data (DROID3): Things in Spirits
 - 3.9 Developing Robust Object to Image to Data (DROID4): 3D objects in Trays
 - 3.10 Education & Outreach (E&O)
 - 3.11 Fluid-preserved Arthropod and Microscopic Slide Imaging Interest Group
 - 3.12 Georeferencing Working Group (GWG)
 - 3.13 International Whole-Drawer Digitization Interest Group (WDD)
 - 3.14 Interoperability for Public Participation in Digitization (CitSciInterop)
 - 3.15 North American Network of Small Herbaria Working Group (NANSW)
 - 3.16 Outlier Detection and Documentation by Collectors Working Group (ODD Collectors)
 - 3.17 Paleo Digitization Working Group (PaleoDigi)
 - 3.18 Project Management Interest Group (PMIG)
 - 3.19 Sustainability Working Group (SWG)
 - 3.20 Symbiota Working Group (SWG)
 - 3.21 User Engagement for Public Participation in Digitization (CitSciEngage)
 - 3.22 Website Content Providers Editorial Board and Interest Group

FEEDBACK



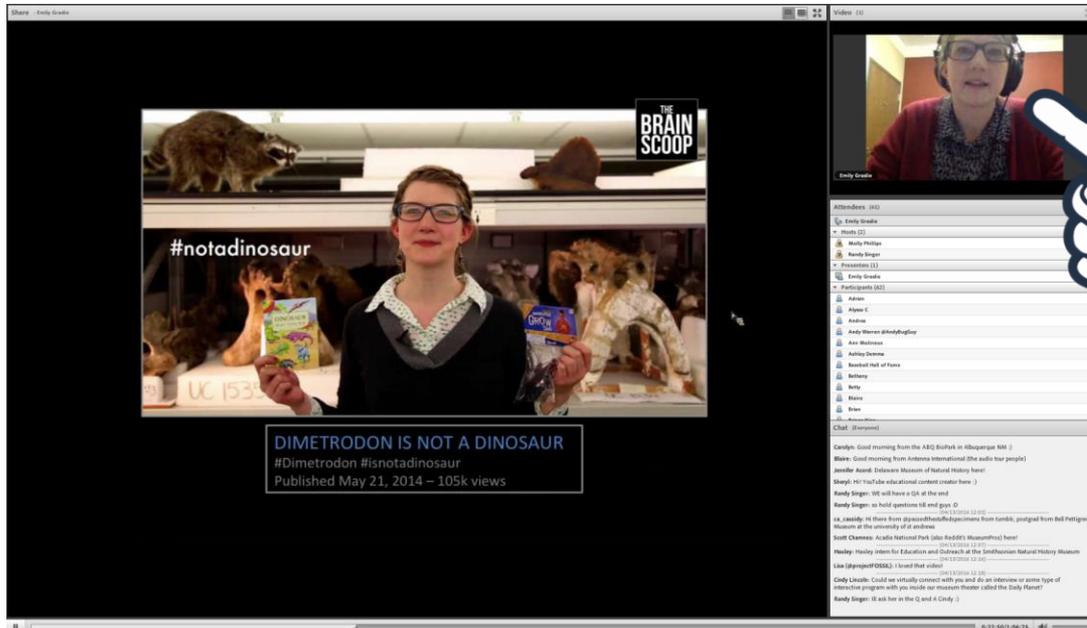
Documentation
API development
Workflows
Standards
Best practices
Hackathons
Workshops, Webinars



Step 5: Watch a webinar...or star in one!

<https://www.idigbio.org/tags/webinar>

https://www.idigbio.org/wiki/index.php/Web_Conferencing



Workshops
Symposia
Webinars



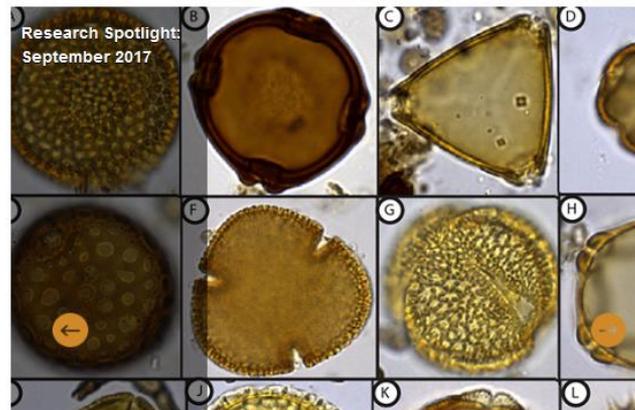
Be like Emily: get a headset preferably with a microphone!

Step 6: Contribute to the iDigBio website

- Submit an article for the **Research Spotlight**
- Write an article about **your project**
- Contribute your **workflows**
- **Update** your individual TCN wiki pages
- Write about your **iDigBio experience**
- **Post an event**

Bering Land Bridge and the MyCoPortal

Contributed by: Teresa Iturriaga, Rhianna Baldree, Alex Kuhn, Andrew Miller

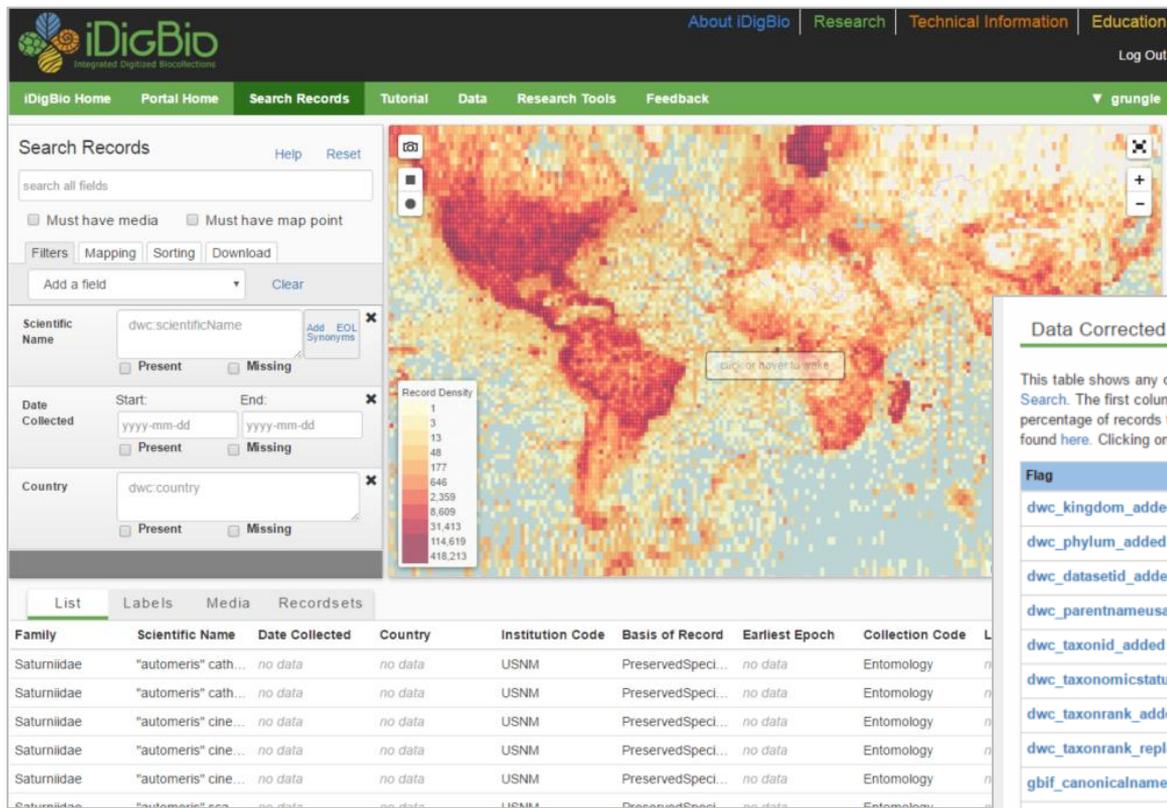


Mycologists long to collect areas remote to most men where fungi today may thrive keeping plants, trees, and cycles alive.

Bridges are to their liking since one can go underneath connecting with what lies beneath. About fungi this is most striking.

In summer some may float if the bridge is over a moat. Fungi are versatile and persistent to new niches they aren't resistant.

Step 7: Use the portal for research and data cleaning – feedback!



Search Records

search all fields

Must have media Must have map point

Filters Mapping Sorting Download

Add a field Clear

Scientific Name: Add EOL Synonyms

Present Missing

Date Collected: Start: End:
 Present Missing

Country:
 Present Missing

Record Density

1
3
13
48
177
646
2,359
8,609
31,413
114,619
418,213

Family	Scientific Name	Date Collected	Country	Institution Code	Basis of Record	Earliest Epoch	Collection Code
Saturniidae	"automeris" cath...	no data	no data	USNM	PreservedSpeci...	no data	Entomology
Saturniidae	"automeris" cath...	no data	no data	USNM	PreservedSpeci...	no data	Entomology
Saturniidae	"automeris" cine...	no data	no data	USNM	PreservedSpeci...	no data	Entomology
Saturniidae	"automeris" cine...	no data	no data	USNM	PreservedSpeci...	no data	Entomology
Saturniidae	"automeris" cine...	no data	no data	USNM	PreservedSpeci...	no data	Entomology

Data Corrected Data Use Raw Recordset

This table shows any data corrections that were performed on this recordset to improve the capabilities of iDigBio Search. The first column represents the correction performed. The last two columns represent the number and percentage of records that were corrected. A complete list of the data quality flags and their descriptions can be found here. Clicking on a data flag name will take you to a search for all records with this flag in this recordset.

Flag	Records With This Flag	(%) Percent With This Flag
dwc_kingdom_added	224195	99.94
dwc_phylum_added	223860	99.79
dwc_datasetid_added	221486	98.732
dwc_parentnameusageid_added	221486	98.732
dwc_taxonid_added	221486	98.732
dwc_taxonomicstatus_added	221486	98.732
dwc_taxonrank_added	221486	98.732
dwc_taxonrank_replaced	221486	98.732
gbif_canonicalname_added	221486	98.732
gbif_taxon_corrected	221486	98.732
geopoint_datum_missing	218751	97.513
gbif_reference_added	217460	96.938

Step 8: Collaborate!



iDigBio Collaborations Enabling Research

To facilitate the study of biodiversity, a number of software products are being collaboratively developed with researchers and projects. These websites, tools, and workflows take advantage of the data being digitized at US and global institutions and made available by iDigBio through our [data services](#). Many other tools and services can be found through the [Biodiversity Catalogue](#). If you have a great idea for using iDigBio data and web services, [get in touch with us](#) or [submit a proposal!](#)

Researchers

Browse our specimen portal



Collections Staff

Learn how your collection can benefit from our work



Teachers & Students

Learning resources & opportunities to engage



4B757261746F72

K

riDigBio: an R interface to the iDigBio Data API

Familiar with R? The [riDigBio](#) package is a great way to interface with the data shared by iDigBio. Contributions to this R package can be made via [GitHub](#).

Biodiversity
Information
Standards
T D W G

Symbiota

FreshData and Effechecka





TCN Responsibilities

TCN Responsibilities

- **Maintain a TCN wiki page**
- Submit requested info for Summit resources
- Participate in TCN meetings
- Submit bi-monthly reports to iDigBio
- Provide feedback via annual community survey and other solicitations
- Prepare annual report for NSF

Using Herbarium Data to Document Plant Niches in the High Peaks and High Plains of the Southern Rockies - Past, Present, and Future

Contents [\[hide\]](#)

- 1 [Digitization TCN: Using Herbarium Data to Document Plant Niches in the High Peaks and High Plains of the Southern Rockies - Past, Present, and Future \(SoRo\)](#)
 - 1.1 [Project Summary](#)
 - 1.2 [Current Research](#)
 - 1.3 [Project Websites & Social Media](#)
 - 1.4 [Citizen Science & Outreach Projects](#)
 - 1.5 [Project Leadership](#)
 - 1.6 [Project Collaborators](#)
 - 1.7 [Protocols & Workflows](#)
 - 1.8 [Publications](#)
 - 1.9 [Professional Presentations](#)
 - 1.10 [Other project documentation](#)

Digitization TCN: Using Herbarium Data to Document Plant Niches in the High Peaks and High Plains of the Southern Rockies - Past, Present, and Future (SoRo)[\[edit\]](#)

Project Summary[\[edit\]](#)

The rugged and expansive terrain of the Southern Rocky Mountains (SoRo) yields the most crucial resource for human existence in western North America: Water. From upper reaches of the high peaks of Colorado, Wyoming, New Mexico, and surrounding states, the headwaters of numerous major rivers of the West originate and give rise to the highest outflow and freshwater runoff west of the Mississippi River: The Platte, Arkansas, Rio Grande, Colorado, and Green Rivers and portions of the Snake and Missouri Rivers. The cleanliness and reliability of these water resources are in large part attributable to the plant life that forms the basis for all SoRo ecosystems. Plant species from the high peaks and adjacent high plains of the SoRo derive from different geographic origins, evolutionary histories, and ecological affinities. They grow in varied habitats and represent one of the most narrowly adapted floras in the world. This unique and fragile flora is widely documented in natural history collections (i.e., herbaria), but specimens themselves and the scientific data that accompany them remain poorly 'visible' owing to a lack of data in digitized format. The SoRo Herbarium Consortium brings together 38 collaborating institutions (universities, botanical gardens, national parks, Native American Nations) to digitize more than 1.7 million botanical specimens from the study area. The overarching goal of this project is to make these data available to a broad community of users (scientists, educators, government officials, land managers, the general public) via openly accessible, free data portals.

The Southern Rocky Mountains (SoRo) support a diverse and highly adapted flora of species with varied ecologies, ranging from alpine to sagebrush plains to shortgrass prairies. The plant biota of the SoRo shares important evolutionary and geological histories with the adjacent plains and prairies, and together these ecosystems are among the most endangered landscapes in North America. Human demands on these systems are escalating, and risk factors such as fire, development, and environmental change are predicted to grow. Thus, building digital resources to document plant niches in the SoRo is met with a sense of urgency. The proposed work will address a major gap in accessible information among North American natural history collections by digitizing more than 1.7 million botanical collections from the SoRo. Specifically, 38 collaborative institutions (universities, botanical gardens, national parks, Native American Nations) will collaborate to generate these new resources and make them available via data portals, including iDigBio (www.idigbio.org). Additionally, new tools will be developed within the open source Symbiota platform that will allow users to quickly search and compile vetted, herbarium data records that can then be used for analysis in external software applications such as niche modeling packages, or to update source data repositories for subsequent in-house curation.

Current Research[\[edit\]](#)

Proposed research uses of data generated through oVert project include:

Project Websites & Social Media[\[edit\]](#)

Citizen Science & Outreach Projects[\[edit\]](#)

Project Leadership[\[edit\]](#)

Project sponsor: University of Colorado, Boulder (NSF Award) [@](#)

Principal Investigator (PI): Erin Tripp

Erin Tripp (Curator), w/ Tim Hogan (Collections Manager), University of Colorado, Boulder

Dina Clark (Collections Manager), University of Colorado, Boulder

SoRo TCN
Project Summary
Current Research
Project Websites
Network Map @
Publications

TCN Responsibilities

- Maintain a TCN wiki page
- Submit requested info for Summit resources
- Participate in TCN meetings
- Provide feedback via annual community survey and other solicitations
- **Submit bi-monthly reports to iDigBio**
- **Prepare annual report for NSF**

Helpful Links for Reporting

- idigbio.org/content/tcn-bi-monthly-progress-report-idigbio
- idigbio.org/wiki/images/3/34/ADBC_AnnualReportInfoSheet.pdf



How To Get Your Data To iDigBio, Data Standards and IPT

Joanna McCaffrey, iDigBio Biodiversity Informatics Manager
Summit 2017
1 November, 2017

https://www.idigbio.org/wiki/index.php/Data_Ingestion_Guidance



What's In This For You?

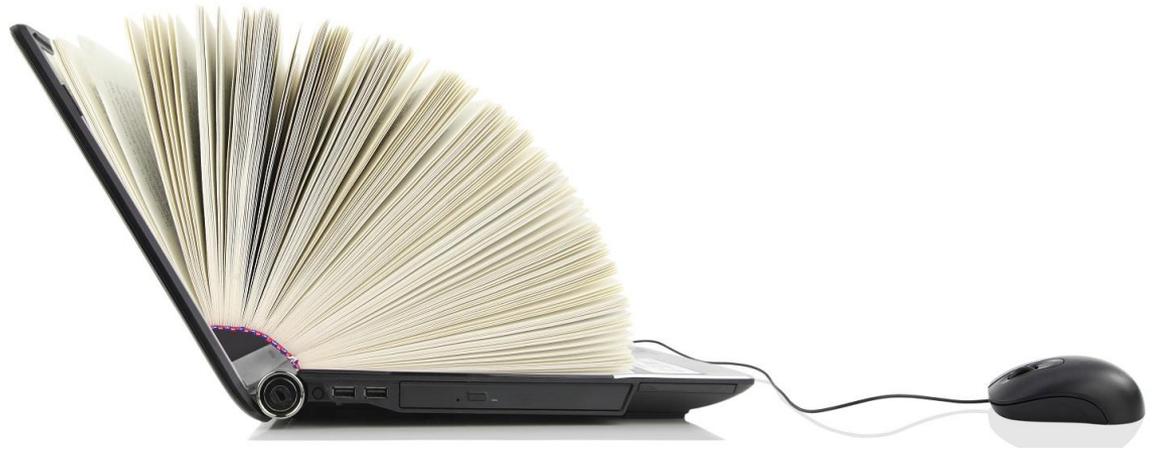
Learn how to get your data *published*

- correctly,
- fully,
- with an eye to quality,
- optimized for the aggregate

What do we mean by **publishing** data?

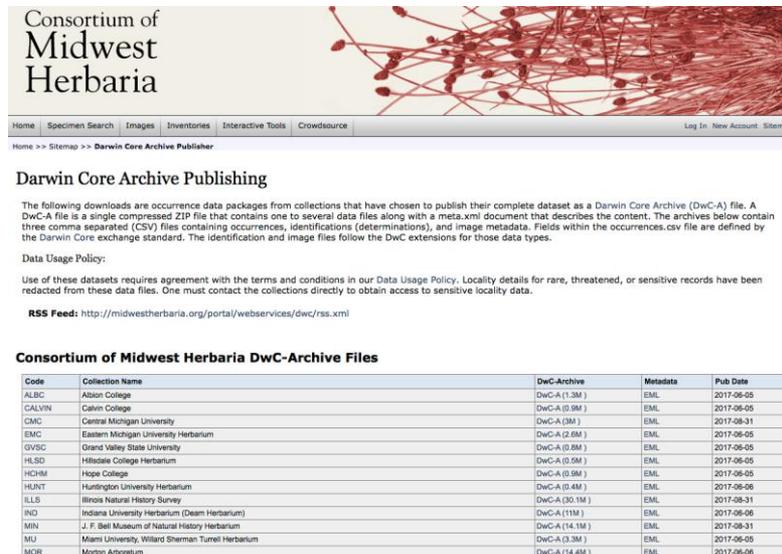
making biodiversity data publicly accessible & discoverable, in a standardized form, via a URL.

** that is reproducible and automated*



Data publishing – where to begin

- Email data@idigbio.org - "I'm ready"
- The way to mobilize data -> IPT:
 - 1) Your IPT: IPT, VertNet, Symbiota
 - 2) Our IPT: <http://ipt.idigbio.org>



Consortium of Midwest Herbaria

Home | Specimen Search | Images | Inventories | Interactive Tools | Crowdsourcing | Log In | New Account | Sitemap

Darwin Core Archive Publishing

The following downloads are occurrence data packages from collections that have chosen to publish their complete dataset as a Darwin Core Archive (DwC-A) file. A DwC-A file is a single compressed ZIP file that contains one to several data files along with a meta.xml document that describes the content. The archives below contain three comma separated (CSV) files containing occurrences, identifications (determinations), and image metadata. Fields within the occurrences.csv file are defined by the Darwin Core exchange standard. The identification and image files follow the DwC extensions for those data types.

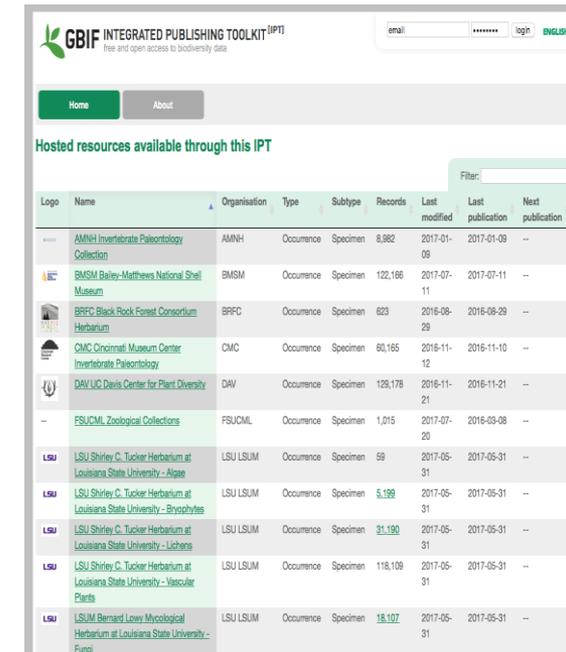
Data Usage Policy:

Use of these datasets requires agreement with the terms and conditions in our Data Usage Policy. Locality details for rare, threatened, or sensitive records have been redacted from these data files. One must contact the collections directly to obtain access to sensitive locality data.

RSS Feed: <http://midwestherbaria.org/portal/webservices/dwc/rss.xml>

Consortium of Midwest Herbaria DwC-Archive Files

Code	Collection Name	DwC-Archive	Metadata	Pub Date
ALBC	Albion College	DwC-A (1.3M)	EML	2017-06-05
CALVIN	Calvin College	DwC-A (0.9M)	EML	2017-05-05
CMC	Central Michigan University	DwC-A (8M)	EML	2017-08-31
EMC	Eastern Michigan University Herbarium	DwC-A (2.6M)	EML	2017-06-05
GVSC	Grand Valley State University	DwC-A (0.8M)	EML	2017-06-05
HLSD	Hillsdale College Herbarium	DwC-A (0.5M)	EML	2017-06-05
HCRM	Hope College	DwC-A (0.9M)	EML	2017-06-05
HUNT	Huntington University Herbarium	DwC-A (0.4M)	EML	2017-06-06
ILLS	Illinois Natural History Survey	DwC-A (20.1M)	EML	2017-08-31
IND	Indiana University Herbarium (Deam Herbarium)	DwC-A (11M)	EML	2017-06-06
MIN	J. F. Bell Museum of Natural History Herbarium	DwC-A (14.1M)	EML	2017-08-31
MU	Miami University, Willard Sherman Tunell Herbarium	DwC-A (3.3M)	EML	2017-06-05
MOR	Morton Arboretum	DwC-A (14.4M)	EML	2017-06-06



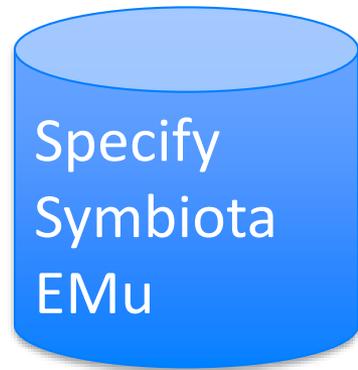
GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)
Free and open access to biodiversity data

Home | About

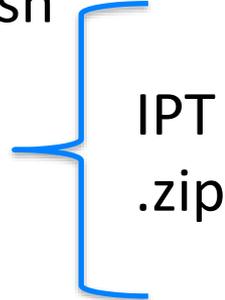
Hosted resources available through this IPT

Logo	Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication
	AMNH Invertebrate Paleontology Collection	AMNH	Occurrence	Specimen	8,982	2017-01-09	2017-01-09	--
	BMSM Bailey-Matthews National Shell Museum	BMSM	Occurrence	Specimen	122,166	2017-07-11	2017-07-11	--
	BRPC Black Rock Forest Consortium Herbarium	BRPC	Occurrence	Specimen	623	2016-08-29	2016-08-29	--
	CMC Cincinnati Museum Center Invertebrate Paleontology	CMC	Occurrence	Specimen	60,165	2016-11-12	2016-11-10	--
	DAV UC Davis Center for Plant Diversity	DAV	Occurrence	Specimen	128,178	2016-11-21	2016-11-21	--
	FSUCM Zoological Collections	FSUCM	Occurrence	Specimen	1,015	2017-07-29	2016-03-08	--
LSU	LSU Shirley C. Tucker Herbarium at Louisiana State University - Algae	LSU LSUM	Occurrence	Specimen	59	2017-05-31	2017-05-31	--
LSU	LSU Shirley C. Tucker Herbarium at Louisiana State University - Bryophytes	LSU LSUM	Occurrence	Specimen	5,199	2017-05-31	2017-05-31	--
LSU	LSU Shirley C. Tucker Herbarium at Louisiana State University - Lichens	LSU LSUM	Occurrence	Specimen	91,190	2017-05-31	2017-05-31	--
LSU	LSU Shirley C. Tucker Herbarium at Louisiana State University - Vascular Plants	LSU LSUM	Occurrence	Specimen	118,109	2017-05-31	2017-05-31	--
LSU	LSUM Bernard Lowy Mycological Herbarium at Louisiana State University - Fungi	LSU LSUM	Occurrence	Specimen	18,107	2017-05-31	2017-05-31	--

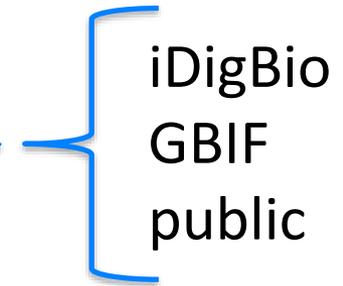
Where do I start?



(re)publish



register



publish -> register -> { republish cycle }

IPT

Definition: The Integrated Publishing Toolkit (IPT) is a free open source software tool written in Java that is used to publish and share biodiversity datasets through the GBIF network. (and other aggregators)

Produces a Darwin Core Archive (DwC-A) [COLO]

Symbiota IPT

(<https://botanydb.colorado.edu/collections/datasets/datapublisher.php>)

Name	^	Date Modified	Size	Kind
 eml.xml		May 18, 2017, 10:01 AM	8 KB	XML File
 identifications.csv		May 18, 2017, 10:01 AM	2 KB	comma-separated values
 images.csv		May 18, 2017, 10:01 AM	109.9 MB	comma-separated values
 meta.xml		May 18, 2017, 10:01 AM	9 KB	XML File
 occurrences.csv		May 18, 2017, 10:01 AM	128.6 MB	comma-separated values

The go-to guide for data ingestion

Everything you wanted to know about preparing data for ingestion:

- https://www.idigbio.org/wiki/index.php/Data_Ingestion_Guidance
 - Identifiers
 - **Darwin Core** – occurrence data (specimen records)
 - **Audubon Core** - media

DATA WAY #1: IPT benefits

- Make data available to GBIF, iDigBio, public
- Setup assistance and ongoing service
- Self-serve if that is your preference
- **GBIF -> Once endorsed by our country node manager – USGS – data are registered for ingestion by GBIF**



DATA Way #2 - Symbiota

- When you click your data to publish, all the necessary parts of the Darwin Core archive package (.zip) are generated.
 - Custom Darwin Core Archive (DwC-A) on an RSS feed produced by Symbiota
 - And almost automatic media
 - <http://symbiota.org>



Symbiota

Promoting
Bio-Collaboration



Publishing – more info to remember

- Searching is not the same as publishing on a Symbiota node. The data can be in the portal, data are searchable, but not necessarily published.

<http://symbiota.org/docs/darwin-core-archive-data-publishing/>



You have to re-publish the DwC-A for us to pick up your updates [RSS feed]

3 ways to get media into iDigBio and linked to the specimen record:

1. Import Audubon Core data into the IPT extension
2. If using Symbiota – built in support
3. Media appliance

Media Metadata – the necessities

- id (coreid of the specimen record it links to)
- identifier (media has its own GUID)
- format (image/jpeg)
- accessURI (public path to your best quality jpg)

Can include richer data – EXIF, IPTC

DATASET INFO: info about the provider (metadata)

Document your dataset **metadata** with your provider information (eml.xml):

- responsible parties (name, address, email, role)
- institution name, institution code, collection code, logo
- URL to the collection at your institution
- descriptive paragraph about the institution, collection, and the dataset

DATASET INFO: rights

- Use Creative Commons standards:

– CC0 for data (not copyrightable)



– CC BY for media (at least)



DATASET INFO: update collections lists

1. iDigBio Collections

Catalog <https://www.idigbio.org/portal/collections>

2. Index Herbariorum

<http://sweetgum.nybg.org/ih/>

3. GRBio.org

Repositories: <http://grbio.org/find-biorepositories>

Please use a collectionCode

Do you know what your **institutionCode** is?

IDENTIFIERS

- Every specimen and media record needs an identifier. [Robust and persistent]

- We like UUIDs with a prefix:

urn:uuid:2d5d3a8f-7a18-4825-a129-4a32b4ae58b8

Data Quality: Consider searchability in the aggregate

Dates – dwc:eventDate, dwc:day, dwc:month, dwc:year:

- this is not a month: Spring
- this is not a day: 10-18
- this is not a year: 1989? Or [1989]

Taxonomy – fill in dwc:scientificName, parse out the elements, fill in higher taxonomy

- this is not a species: shrimp, daisy

Tics: * [] {} ?, 0 fillers, NA place holders

- Use the verbatim and remarks fields for things that do not fit the definitions.

Data Quality: Grooming and tics – top 10

Your dataset **is no longer just for making labels**, there are other considerations for being digital, and out in the wild:

- 1) Put dates in ISO 8601 format, i.e., YYYY-MM-DD, e.g., 2015-09-17
- 2) Parse apart scientific name
- 3) Conversely, put the piece parts into a scientific name
- 4) Provide as much higher taxonomy as you feel comfortable with, fill in tribe, sub+super family, kingdom, division, class, order), get out of ‘family’ land.
- 5) Make sure lat and lon coordinates are in decimal, and no N, S, E, W
- 6) Do not export '0' “n/a” in fields to represent no value, e.g., lat or lon, height
- 7) put elevation in METERS units in the elevation field without the units (e.g., the fields dwc:minimumElevationInMeters and dwc:maximumElevationInMeters already assume the numeric values are in meters, so there no need to include the units with the data)
- 8) Leave out ‘County’, ‘Co.’ in county field
- 9) And not to get too esoteric, do not use un-escaped newline characters or embedded tabs
- 10) Watch out for diacritics, save in UTF-8

à á â ã ä å

When is my work done?

- Digitization is never done
 - Label data
 - Georeferenced
 - Image
 - + **AND** +

For TCNs/PENs: Not until your data are in iDigBio.

- It is not enough to get to it to Symbiota
 - Publish, re-publish with updates



Symbiota Users / IPT users

- Your dataset naming
 - Give it a complete name, institution, collection/herbarium
 - Description of the collection – what is in THIS data
 - Good contacts - the person who will respond to requests
- Join the Symbiota working group – community, webinars

Example of good naming (2)

University of Vermont, Pringle Herbarium, North American bryophytes

Specimen Records: 17,697

Media Records: 16,941

iDigBio Last Ingested Date: 2016-07-28



The Pringle Herbarium (VT) contains 300,000 specimens, including vascular plants, bryophytes, lichens, algae and fungi. This portal contains our North American bryophyte specimens, numbering about 18,000. Other digitization projects cover type specimens, vascular plant specimens, North American lichens, macroalgae and macrofungi. These images and data are available through various portals. The herbarium does not maintain its own online database.

Contacts

Name *none*
Role *none*
Email CNABHadmin@asu.edu

Name Dorothy Allard, Virtual Herbarium
Coordinator
Role *none*
Email djallard@uvm.edu





Get involved!



idigbio.org/wiki



facebook.com/iDigBio



twitter.com/iDigBio



vimeo.com/iDigBio



idigbio.org/rss-feed.xml



idigbio.org/events-calendar/export.ics





Contact us!



Cathy Bester
Event Coordinator
cbester@flmnh.ufl.edu



Joanna McCaffrey
Biodiversity
Informatics Manager
jmccaffrey@flmnh.ufl.edu



David Jennings
Project Manager
djennings@flmnh.ufl.edu



Gil Nelson
Digitization Process
Specialist
gnelson@bio.fsu.edu



Deb Paul
Informatics Analyst
dpaul@fsu.edu



Alex Thompson
IT Expert
godfoder@acis.ufl.edu



Shari Ellis
Project Evaluator
sellis@ufl.edu



Jillian Goodwin
Project Assistant
jvgoodwin@fsu.edu



Molly Phillips
Education & Outreach
Coordinator
mphillips@flmnh.ufl.edu



Kevin Love
IT Expert
klove@flmnh.ufl.edu