iDigBio
Integrated Digitized Biocollections

IDIGBIO ORIENTATION WEBINAR SERIES

Project Management and Data Mobilization

David Jennings, iDigBio Project Manager

Cat Chapman, iDigBio Biodiversity Informatics Coordinator

November 3, 2021

iDigBio is funded by grants from the National Science Foundation [DBI-1115210 (2011-2018), DBI-1547229 (2016-2022), & DBI-2027654 (2021-2026)]. 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.

iDigBio Orientation Webinar Series

October 6 - Welcome to iDigBio Phase 3

October 20 - Ongoing iDigBio collaborations and events

November 3 - Project management and data mobilization considerations for digitization networks

November 17 - Resources for digitizing biodiversity collections and managing the digitized data

December 1 - Increasing diversity and inclusion in biodiversity collections

December 15 - Accessing and using digitized data from biodiversity collections

https://www.idigbio.org/content/idigbio-orientation-2021

Today's agenda

- This orientation session doubles as our regular quarterly TCN meeting

- We are discussing timely topics relevant to all TCNs, PENs, and other digitization projects:
  - Digitization project management
  - TCN responsibilities and reporting
  - Importance of your Data Management Plan
  - How to get your data to iDigBio

- Questions & Answers

Digitization Project Management

David Jennings
iDigBio Project Manager
djennings@flmnh.ufl.edu

iDigBio is funded by grants from the National Science Foundation [DBI-1115210 (2011-2018), DBI-1547229 (2016-2022), & DBI-2027654 (2021-2026)]. 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.

Project management is important for success!

- We generally present something at the annual orientation
- We presented a webinar in our *Adapting to COVID-19* series that specifically focused on project management:
  - Highly recommended viewing!
  - Includes a broader overview of project management (presented by me)
  - Includes a series of **tips and tricks** on managing projects (presented by me, Diego Barroso, and Jen Zaspel)
  - https://www.idigbio.org/wiki/index.php/Webinar_Series:_Adapting_to_COVID-19:_Resources_for_Natural_History_Collections_in_a_New_Virtual_World

Why do we keep bringing up project management?

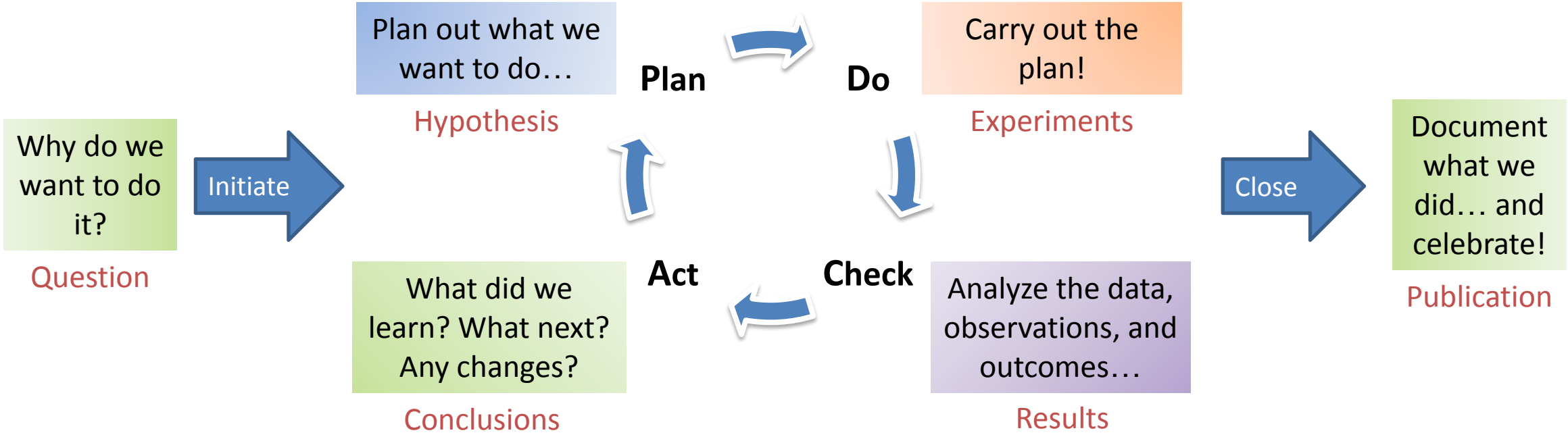
- Common challenges with scientific and research projects:
  - Scientists want to do research, not administration
  - Project management and administration can be time consuming
  - Administration distracts scientists from their main interests
  - Scientists often lack training or experience in project management
  - Funding agencies desire accurate cost estimates and predictable outcomes

- There must be a better way:
  - Educate emerging scientists in the basics of project management and administration
  - Transfer knowledge and lessons from experienced scientists to those in emerging projects

https://www.nsf.gov/awardsearch/showAward?AWD_ID=0121190

Project management is not as abstract as you think…

"Plan-do-check-act (PDCA) represents an intersection between the scientific method and everyday operations. Given an objective, whatever process is defined to achieve that objective can be—and often unknowingly is—subject to PDCA."



Nelson T. Dan (2017). The Gravity of PDCA. *Quality Progress*, *50*(1): 64.
https://search.proquest.com/openview/1d159396d94f73d3d12b23ead675e43e/1?pq-origsite=gscholar&cbl=34671

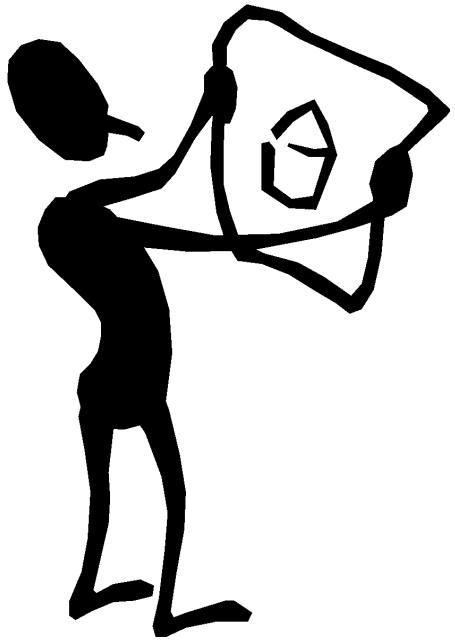
Project Management **is often intertwined with** ***Leadership***

- **Management** is about *coping with complexity* to bring order and predictability to a situation
- **Leadership** is about *coping with change* to survive or compete effectively in a new environment

Management		Leadership
Planning and budgeting	⟺	Setting direction
Organizing and Staffing	⟺	Aligning people
Control and problem-solving	⟺	Motivation

Kotter J. P. (1990). What leaders really do. *Harvard Business Review*, *68*(3), 103–111.
https://hbr.org/2001/12/what-leaders-really-do

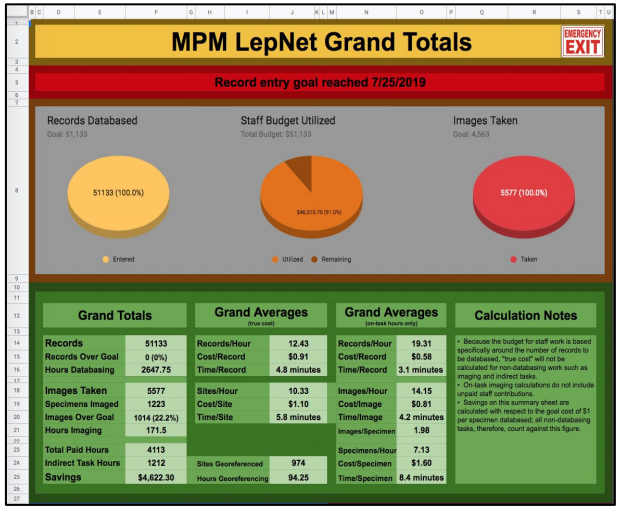
Project management starts with a *Workplan*

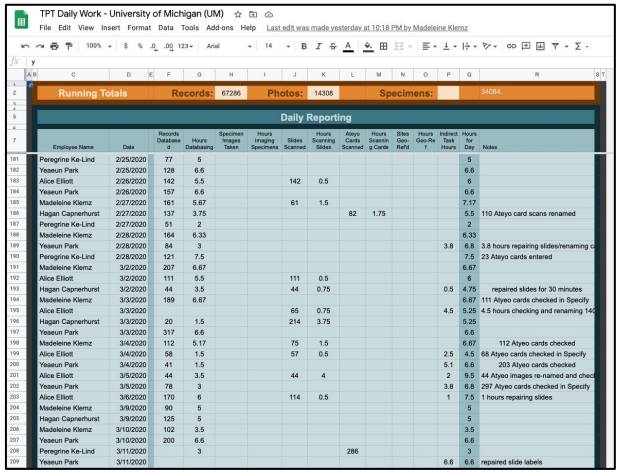


- A workplan serves as a map and guide for your team:
  - **What** needs to be done?
  - **Who** is doing the work?
  - How much will the work cost (**budget**)?
  - When will the work be done (**schedule**)?
  - How the work will be done (**workflows**)?
  - How will you manage **communications**?
  - How will you manage **risk**?
  - How will you manage **change**?
  - What **metrics** will you use to track and measure **progress**, **quality**, and **scope**?
  - What are your internal and external **dependencies**?
- Remember that a workplan is a living document → adaptable
- The planning process is vitally important for your team

A few tips and tricks from our collective experience...

- Manage your risks *early* and *often*
- Be flexible → you can act on *incomplete* data, but *insufficient* data does not compute
- Seek to build relationships → know who is who and what makes them unique
- Create a series of templates/forms for emails, reports, advertisements, etc.
- Use a centralized repository to facilitate access → documentation, training resources
- Create reports as you go (not all at the end)
- Prioritize and keep costs in mind
- Hold regular meetings
- Create a workplan (and a status dashboard)
- Keep your stakeholders in the loop → newsletter





https://www.idigbio.org/wiki/index.php/Webinar_Series:_Adapting_to_COVID-19:_Resources_for_Natural_History_Collections_in_a_New_Virtual_World

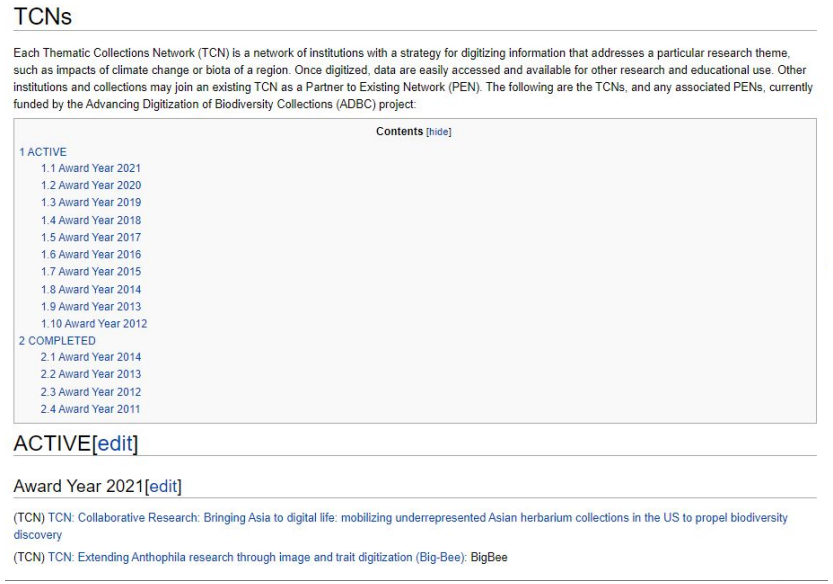
TCN Responsibilities and Reporting

David Jennings
iDigBio Project Manager
djennings@flmnh.ufl.edu

iDigBio is funded by grants from the National Science Foundation [DBI-1115210 (2011-2018), DBI-1547229 (2016-2022), & DBI-2027654 (2021-2026)]. 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.

TCN Responsibilities

1. **Maintain your TCN wiki page**
   - https://www.idigbio.org/wiki/index.php/TCNs
   - Have your Project Manager email us to get an account to allow editing
2. Submit requested info for the annual Summits
3. Provide feedback via community survey and other solicitations



TCNs

Each Thematic Collections Network (TCN) is a network of institutions with a strategy for digitizing information that addresses a particular research theme, such as impacts of climate change or biota of a region. Once digitized, data are easily accessed and available for other research and educational use. Other institutions and collections may join an existing TCN as a Partner to Existing Network (PEN). The following are the TCNs, and any associated PENs, currently funded by the Advancing Digitization of Biodiversity Collections (ADBC) project:

Contents [hide]

1 ACTIVE
- 1.1 Award Year 2021
- 1.2 Award Year 2020
- 1.3 Award Year 2019
- 1.4 Award Year 2018
- 1.5 Award Year 2017
- 1.6 Award Year 2016
- 1.7 Award Year 2015
- 1.8 Award Year 2014
- 1.9 Award Year 2013
- 1.10 Award Year 2012

2 COMPLETED
- 2.1 Award Year 2014
- 2.2 Award Year 2013
- 2.3 Award Year 2012
- 2.4 Award Year 2011

ACTIVE[edit]

Award Year 2021[edit]

(TCN) TCN: Collaborative Research: Bringing Asia to digital life: mobilizing underrepresented Asian herbarium collections in the US to propel biodiversity discovery

(TCN) TCN: Extending Anthophila research through image and trait digitization (Big-Bee): BigBee

TCN: Extending Anthophila research through image and trait digitization (Big-Bee)

Contents [hide]

1 Extending Anthophila research through image and trait digitization (Big-Bee)
- 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
  - 1.6.1 Known Collaborators
- 1.7 Protocols & Workflows
- 1.8 Publications
- 1.9 Professional Presentations
- 1.10 Other project documentation

Extending Anthophila research through image and trait digitization (Big-Bee)[edit]

Project Summary[edit]

Declining populations of bees impact plant-pollinator interactions in both natural and agricultural systems. While bees and other insects pollinate most wild plants, and are critical to sustain a large proportion of global food production, they are decreasing in both numbers and diversity. Our understanding of the factors driving these declines is limited because we lack sufficient data on the distribution of bee species, and on the behavioral and anatomical traits that may make them either vulnerable or resilient to human-induced environmental changes, such as habitat loss and climate change. Fortunately, wild bees have been collected by researchers and deposited in natural history collections for over 100 years, retaining a wealth of associated attributes that can be extracted from specimen images. This project will digitally capture data and images from these historic specimens, develop tools to measure bee traits from these images, and generate a comprehensive bee trait and image dataset to measure changes through time. This will increase our understanding of specific traits that put bee species at risk of decline - a critical need for both sustaining our agricultural economy and the conservation of our natural resources. In addition, the large image datasets created by this project can be used for new artificial intelligence identification tools that will help

BigBee TCN

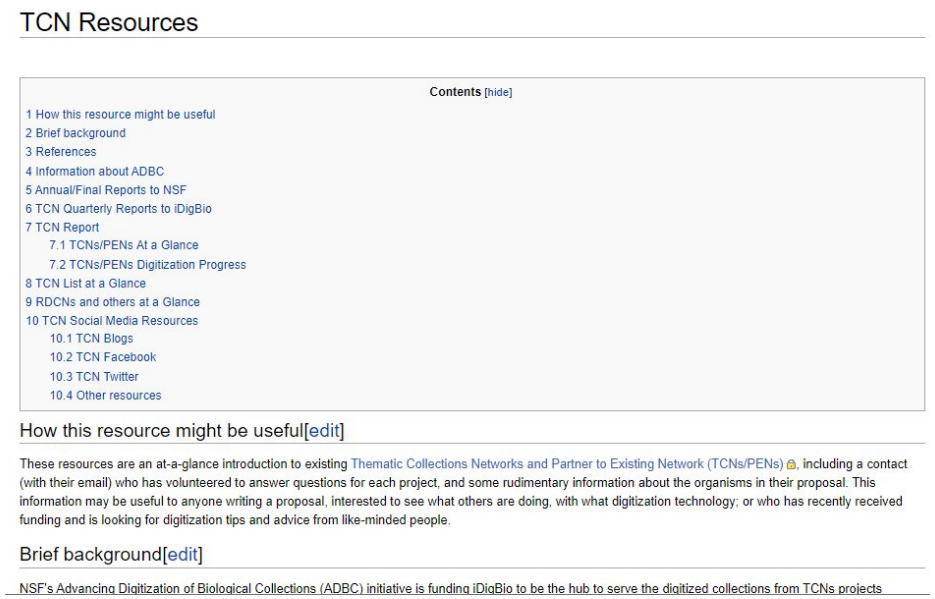
Quick Links

Project Summary

TCN Responsibilities (cont.)

4. Prepare your annual report for NSF

- Are you the lead institution or not? R.T.F.D.
- NSF guidance document for TCN reports: https://www.idigbio.org/wiki/images/3/34/ADBC_AnnualReportInfoSheet.pdf
- You can find the above document linked on the TCN Resources page: https://www.idigbio.org/wiki/index.php/TCN_Resources
- If in doubt, email Reed Beaman (rsbeaman@nsf.gov) or your fellow TCN PIs (TCNPIS-L@lists.ufl.edu)



TCN Resources

Contents [hide]

1 How this resource might be useful
2 Brief background
3 References
4 Information about ADBC
5 Annual/Final Reports to NSF
6 TCN Quarterly Reports to iDigBio
7 TCN Report
7.1 TCNs/PENs At a Glance
7.2 TCNs/PENs Digitization Progress
8 TCN List at a Glance
9 RDCNs and others at a Glance
10 TCN Social Media Resources
10.1 TCN Blogs
10.2 TCN Facebook
10.3 TCN Twitter
10.4 Other resources

How this resource might be useful[edit]

These resources are an at-a-glance introduction to existing Thematic Collections Networks and Partner to Existing Network (TCNs/PENs), including a contact (with their email) who has volunteered to answer questions for each project, and some rudimentary information about the organisms in their proposal. This information may be useful to anyone writing a proposal, interested to see what others are doing, with what digitization technology; or who has recently received funding and is looking for digitization tips and advice from like-minded people.

Brief background[edit]

NSF's Advancing Digitization of Biological Collections (ADBC) initiative is funding iDigBio to be the hub to serve the digitized collections from TCNs projects

Advancing Digitization of Biological Collections

Division of Biological Infrastructure
Directorate for Biological Sciences
National Science Foundation

Guidelines for Annual and Final Reports

Lead PI's Annual/Final Report

1) Request information from all the collaborating PIs in your TCN who have awards that were active in this reporting cycle. Compile what has been accomplished (i.e., digitization activities, other activities), who did the work, and the outcomes for the overall project. The Lead PI must combine what has been accomplished by the TCN as a whole into an Integrated Project Report. Make sure this Integrated Project Report addresses:

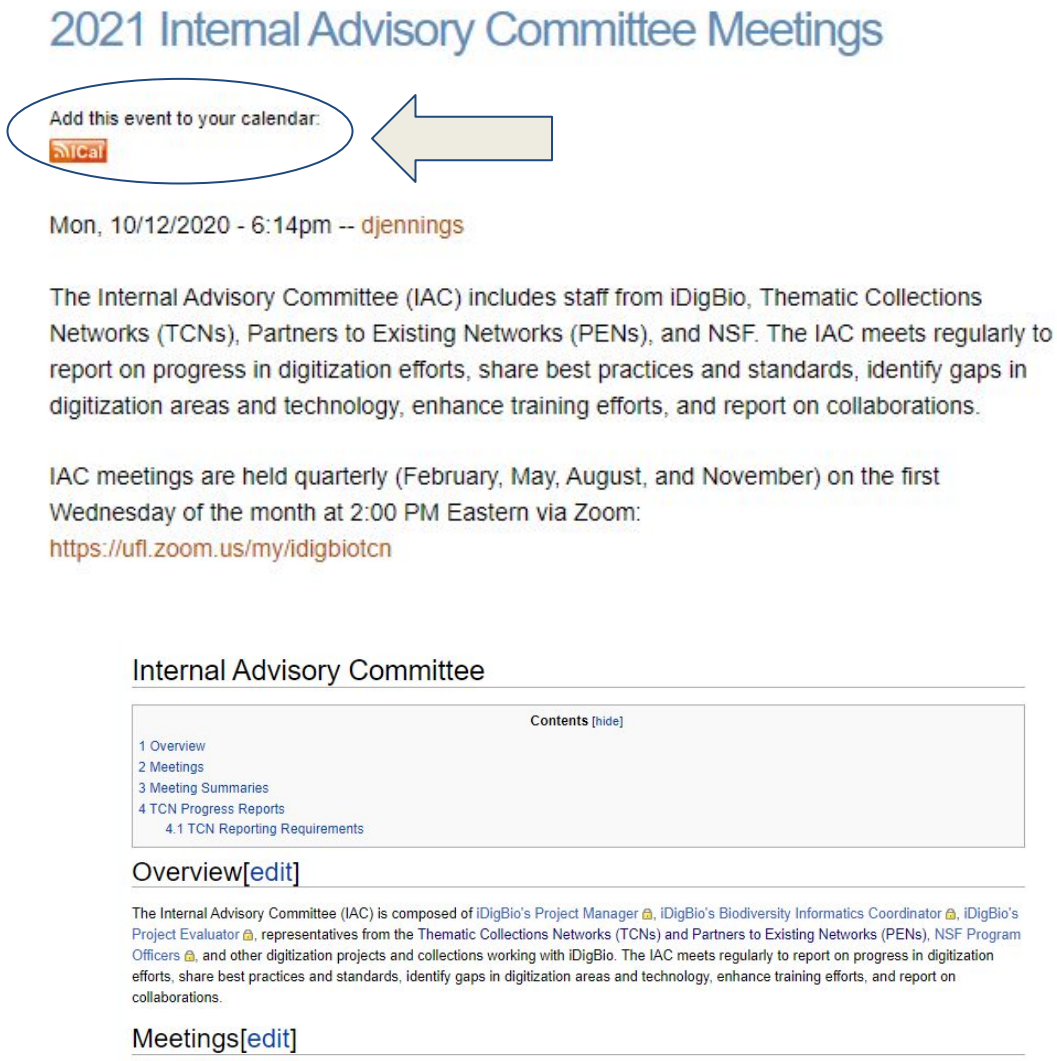
Participants:

- All participants involved in the entire project (important: only those who have worked >160 hours on the project, including undergraduate participants).

TCN Responsibilities (cont.)

5. Participate in the quarterly meetings

- First Wednesday of every February, May, August, and November @ 2:00 PM Eastern
- Schedule including links to subscribe to the meeting events:
  - https://www.idigbio.org/content/2021-internal-advisory-committee-meetings
  - https://www.idigbio.org/content/2022-internal-advisory-committee-meetings
- Minutes are posted on the wiki: https://www.idigbio.org/wiki/index.php/Internal_Advisory_Committee



2021 Internal Advisory Committee Meetings

Add this event to your calendar: iCal

Mon, 10/12/2020 - 6:14pm -- djennings

The Internal Advisory Committee (IAC) includes staff from iDigBio, Thematic Collections Networks (TCNs), Partners to Existing Networks (PENs), and NSF. The IAC meets regularly to report on progress in digitization efforts, share best practices and standards, identify gaps in digitization areas and technology, enhance training efforts, and report on collaborations.

IAC meetings are held quarterly (February, May, August, and November) on the first Wednesday of the month at 2:00 PM Eastern via Zoom: https://ufl.zoom.us/my/idigbiotcn

Internal Advisory Committee

Contents [hide]

1 Overview
2 Meetings
3 Meeting Summaries
4 TCN Progress Reports
4.1 TCN Reporting Requirements

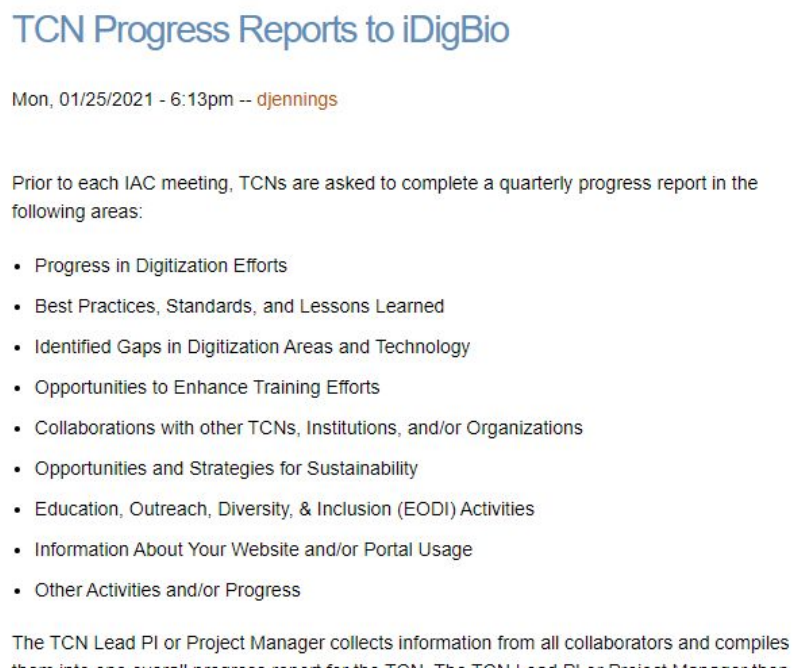
Overview[edit]

The Internal Advisory Committee (IAC) is composed of iDigBio's Project Manager, iDigBio's Biodiversity Informatics Coordinator, iDigBio's Project Evaluator, representatives from the Thematic Collections Networks (TCNs) and Partners to Existing Networks (PENs), NSF Program Officers, and other digitization projects and collections working with iDigBio. The IAC meets regularly to report on progress in digitization efforts, share best practices and standards, identify gaps in digitization areas and technology, enhance training efforts, and report on collaborations.

Meetings[edit]

TCN Responsibilities (cont.)

6. **Submit quarterly reports to iDigBio**
   - ***Due by the quarterly meeting***
   - Instructions for completing and submitting your report are here: https://www.idigbio.org/content/tcn-progress-reports-idigbio
   - Compilations are posted on the wiki: https://www.idigbio.org/wiki/index.php/Internal_Advisory_Committee



TCN Progress Reports to iDigBio

Mon, 01/25/2021 - 6:13pm -- djennings

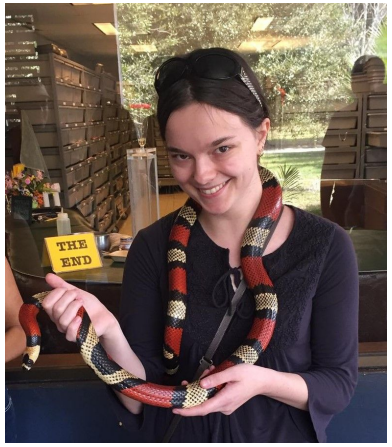
Prior to each IAC meeting, TCNs are asked to complete a quarterly progress report in the following areas:

- Progress in Digitization Efforts
- Best Practices, Standards, and Lessons Learned
- Identified Gaps in Digitization Areas and Technology
- Opportunities to Enhance Training Efforts
- Collaborations with other TCNs, Institutions, and/or Organizations
- Opportunities and Strategies for Sustainability
- Education, Outreach, Diversity, & Inclusion (EODI) Activities
- Information About Your Website and/or Portal Usage
- Other Activities and/or Progress

The TCN Lead PI or Project Manager collects information from all collaborators and compiles

Importance of Your Data Management Plan

Cat Chapman
iDigBio Biodiversity Informatics Coordinator
cchapman@floridamuseum.ufl.edu

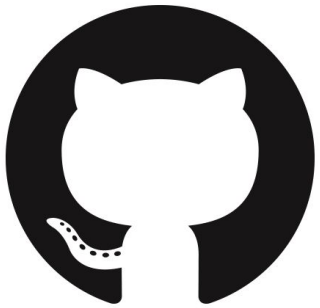


iDigBio is funded by grants from the National Science Foundation [DBI-1115210 (2011-2018), DBI-1547229 (2016-2022), & DBI-2027654 (2021-2026)]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

What nobody wants to think about… 🙈🙉🙊

- "What happens after the $$$ runs out?"
  - What will happen to your data?
  - Where will your data live?
  - Where will your documents live?
    - Protocols? Workflows?
- iDigBio can provide training and guidance
  - We cannot provide funding
- Your proposal has a data management plan
  - Implementation is your responsibility
  - But we CAN help you

Some ideas

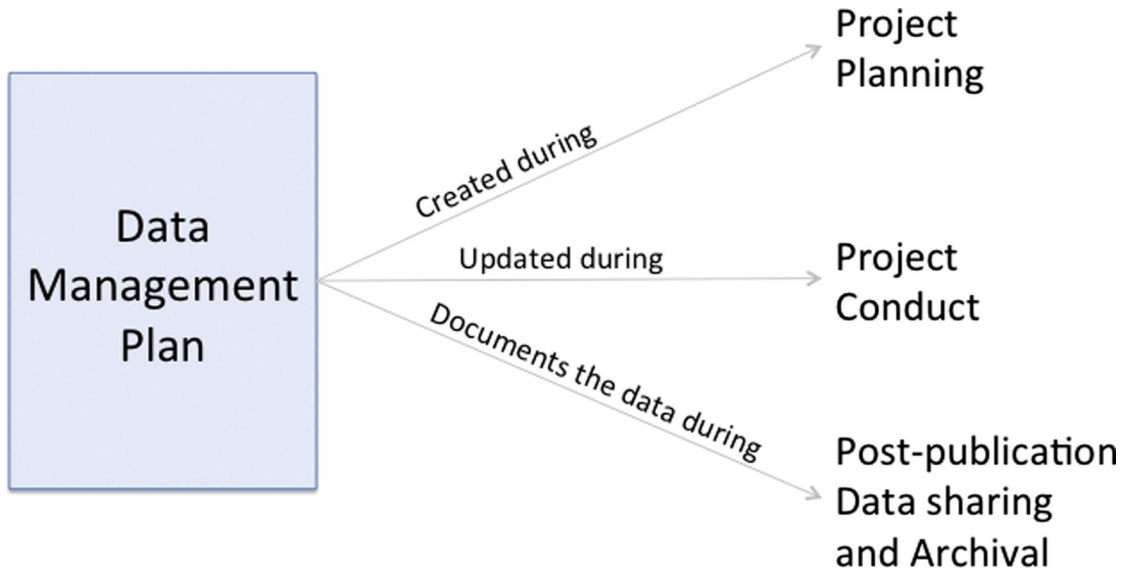


- Internet Archive (archive.org)
- Your institution
- If all else fails… GitHub
  - Third-party private entities probably aren't the best, but better than nothing

Any other ideas? 👀

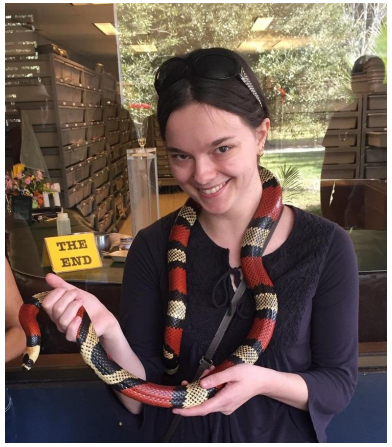
TL;DR

- Data Management Plans = very important!
- Your DMP = your responsibility to fulfill!
- We CAN help you out!
  - Other, established TCNs are also a valuable knowledge resource
- Think of the long-term
- Data loss makes the biodiversity informatics world sad
- Data persistence makes the biodiversity informatics world happy



How to Get Your Data to iDigBio

Cat Chapman
iDigBio Biodiversity Informatics Coordinator
cchapman@floridamuseum.ufl.edu



iDigBio is funded by grants from the National Science Foundation [DBI-1115210 (2011-2018), DBI-1547229 (2016-2022), & DBI-2027654 (2021-2026)]. 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.

What's In This For You?

Meet the iDigBio Staff

Overview of the ingestion process

Learn how to get your data ***published***

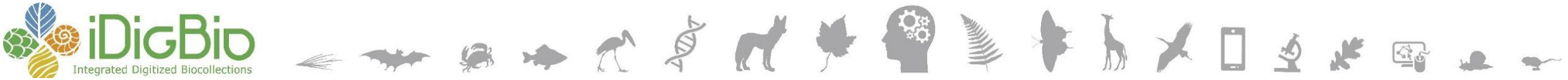
iDigBio Data Mobilization Staff

Caitlin “Cat” Chapman
cchapman@floridamuseum.ufl.edu

Biodiversity Informatics Coordinator

Dan Stoner
dstoner@acis.ufl.edu

Data Integration Expert



data@idigbio.org

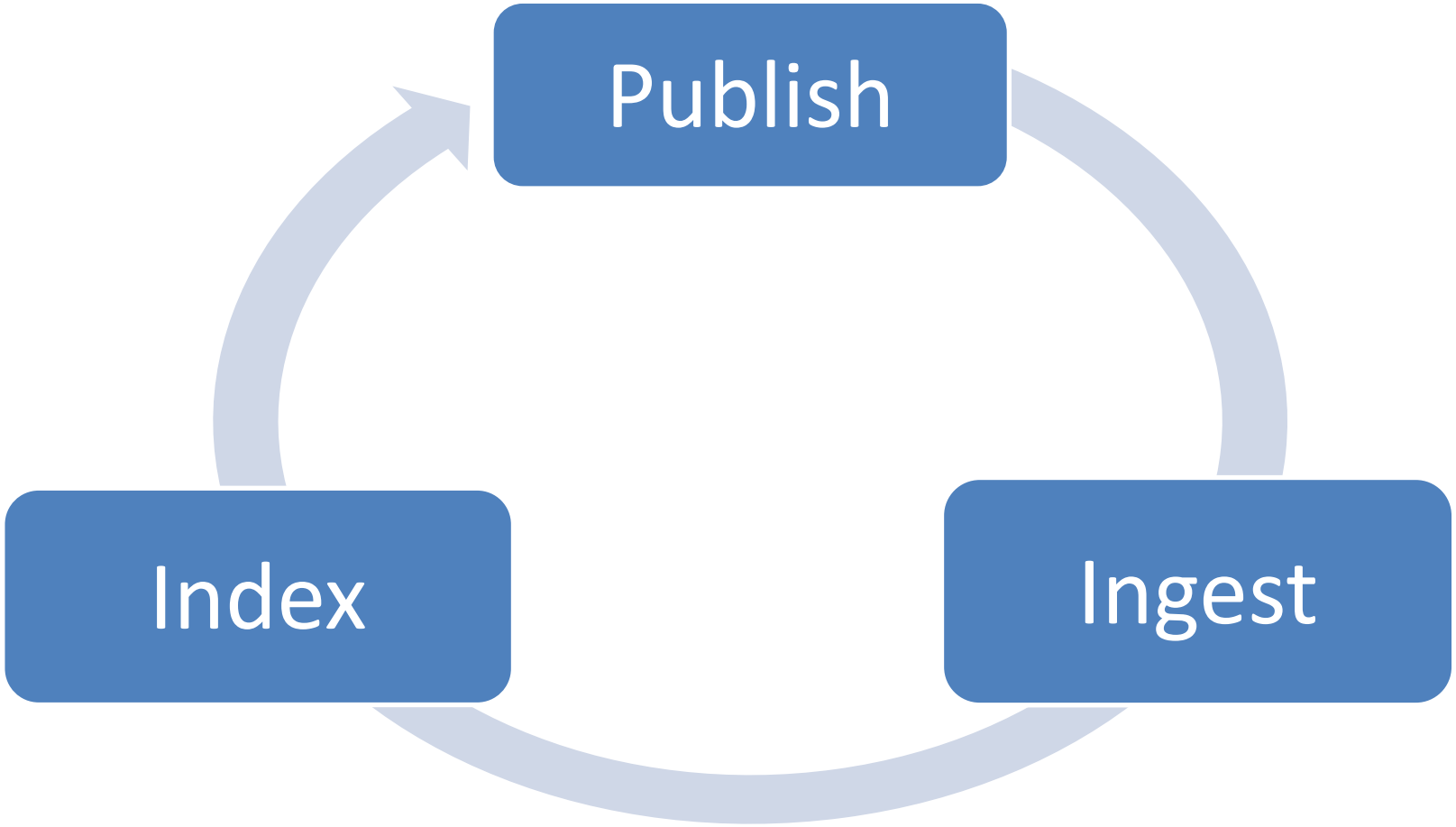
The go-to guide for data ingestion

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

Ingestion Process



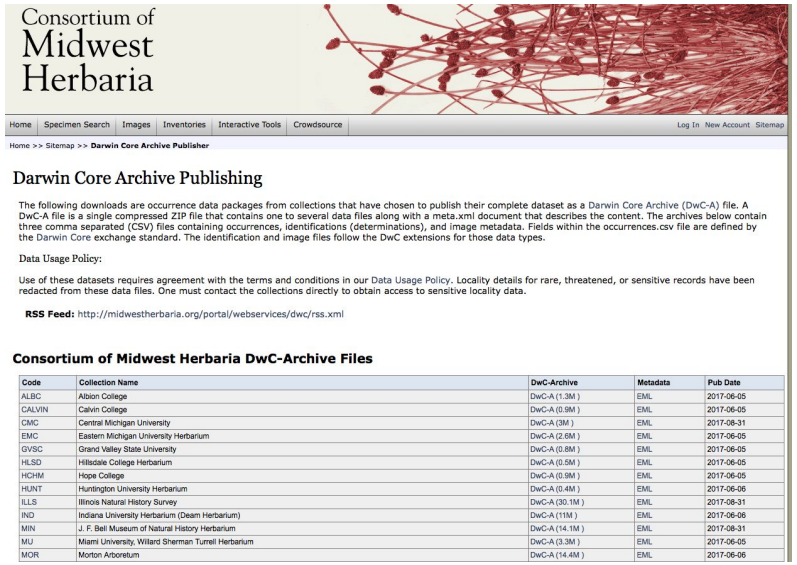
What do we mean by publishing data?

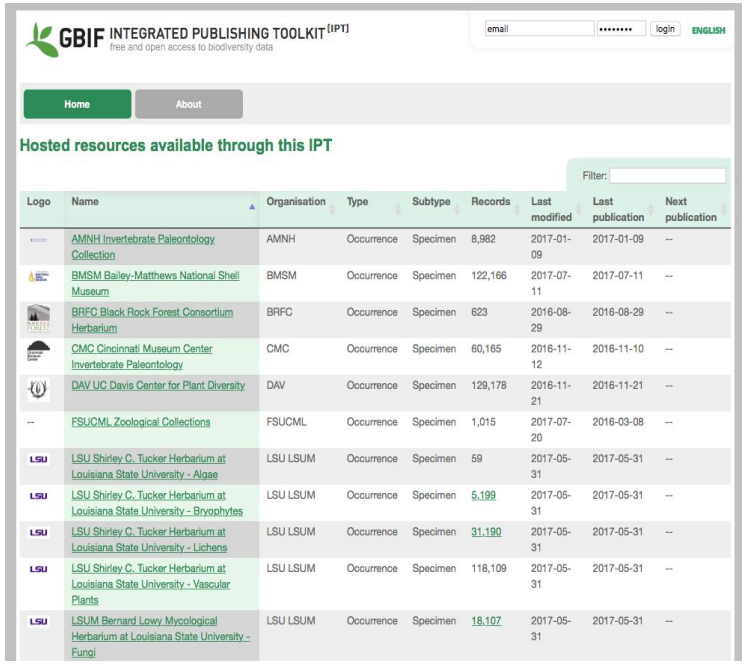
making biodiversity data publicly accessible & discoverable, in a standardized form, via a URL.

** that is reproducible and automated*

Data publishing – where to begin

- Email data@idigbio.org
  "*I'm ready*"
- Review your data and publishing options together



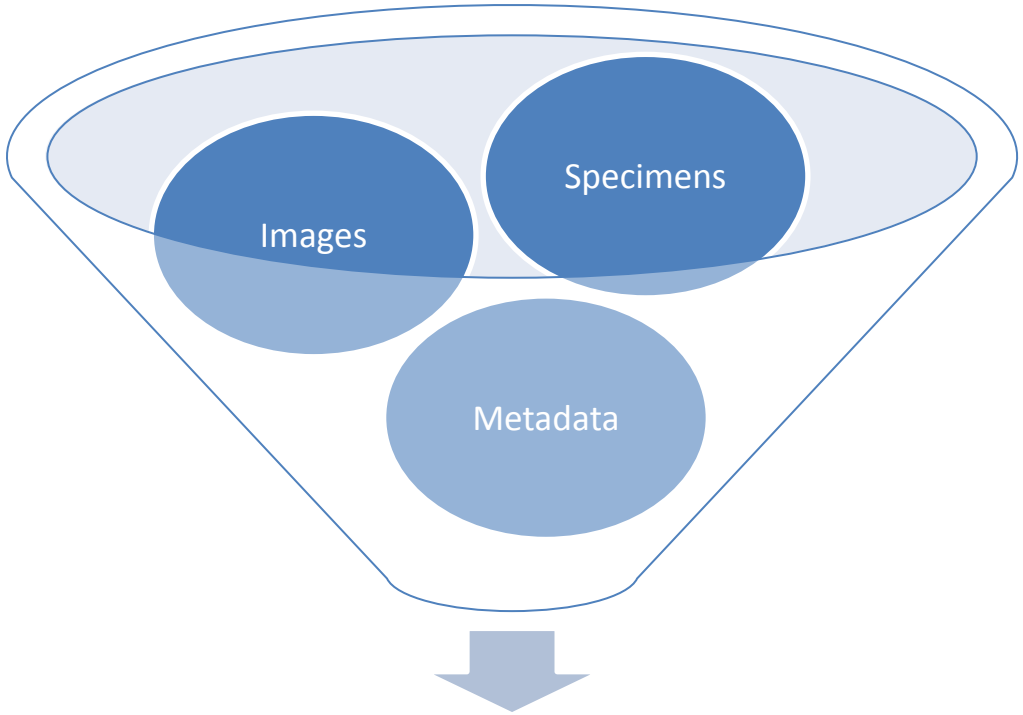


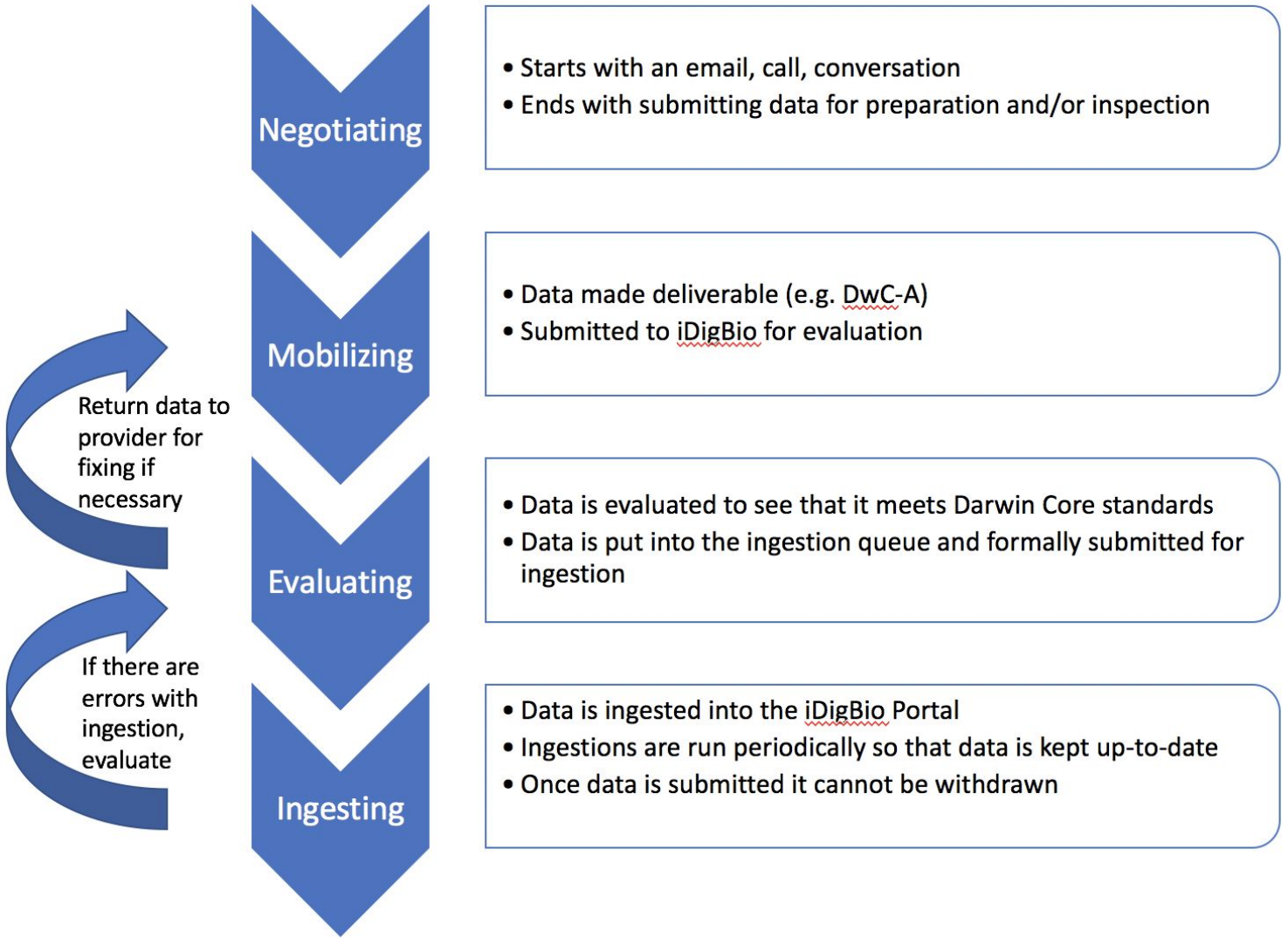
Ingestion Queue

- https://www.idigbio.org/wiki/index.php/Data_Ingestion_Report

Milestones:

- Negotiating
- Mobilizing
- Evaluating
- Ingesting





- Starts with an email, call, conversation
- Ends with submitting data for preparation and/or inspection

- Data made deliverable (e.g. DwC-A)
- Submitted to iDigBio for evaluation

- Data is evaluated to see that it meets Darwin Core standards
- Data is put into the ingestion queue and formally submitted for ingestion

- Data is ingested into the iDigBio Portal
- Ingestions are run periodically so that data is kept up-to-date
- Once data is submitted it cannot be withdrawn

DATASET INFO: info about the provider (metadata)

Document your dataset metadata with your provider information:

- 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)

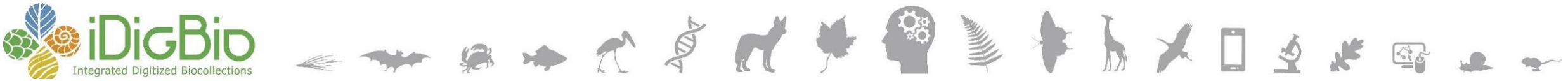
IDENTIFIERS

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

- Think of a unique identifier as a Social Security Number for your catalogued object!
- If it changes, it needs a “paper trail”!

Example UUID:

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



Remember, when you're ready:

data@idigbio.org

Thank You!

www.idigbio.org

facebook.com/iDigBio

twitter.com/iDigBio

vimeo.com/idigbio

idigbio.org/rss-feed.xml

webcal://www.idigbio.org/events-calendar/export.ics

iDigBio is funded by grants from the National Science Foundation [DBI-1115210 (2011-2018), DBI-1547229 (2016-2022), & DBI-2027654 (2021-2026)]. 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.

Join us for the next episodes in the Orientation series...

October 6 - Welcome to iDigBio Phase 3

October 20 - Ongoing iDigBio collaborations and events

November 3 - Project management and data mobilization considerations for digitization networks

November 17 - Resources for digitizing biodiversity collections and managing the digitized data

December 1 - Increasing diversity and inclusion in biodiversity collections

December 15 - Accessing and using digitized data from biodiversity collections

https://www.idigbio.org/content/idigbio-orientation-2021

Orientation Webinar Series Survey:

https://ufl.qualtrics.com/jfe/form/SV_8dj1FtMT6Sp45T0