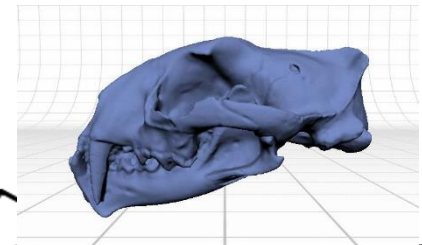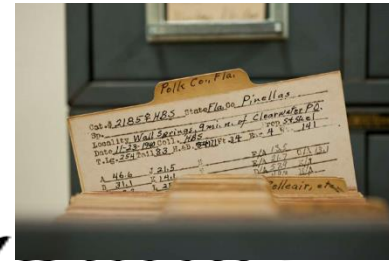# Digitizing the Past and Present for the Future

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

Gil Nelson, iDigBio Digitization Process Expert – gnelson@bio.fsu.edu

OBFS Annual Meeting, 21 Sep 2018, Schoodic Institute at Acadia National Park
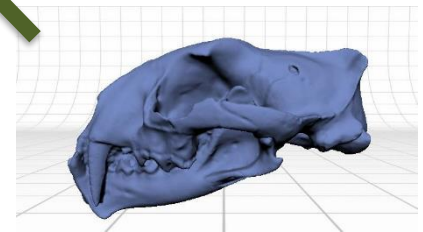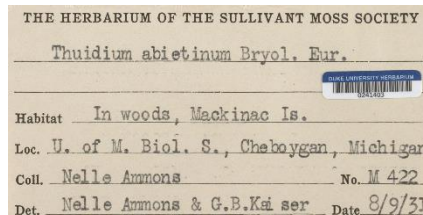
Why digitize?

Estimates suggest there are between
500 million and 1 billion
biological and paleobiological specimens
in the United States and potentially
3-4 billion worldwide
Many are digitized, but <u>most are not</u>.
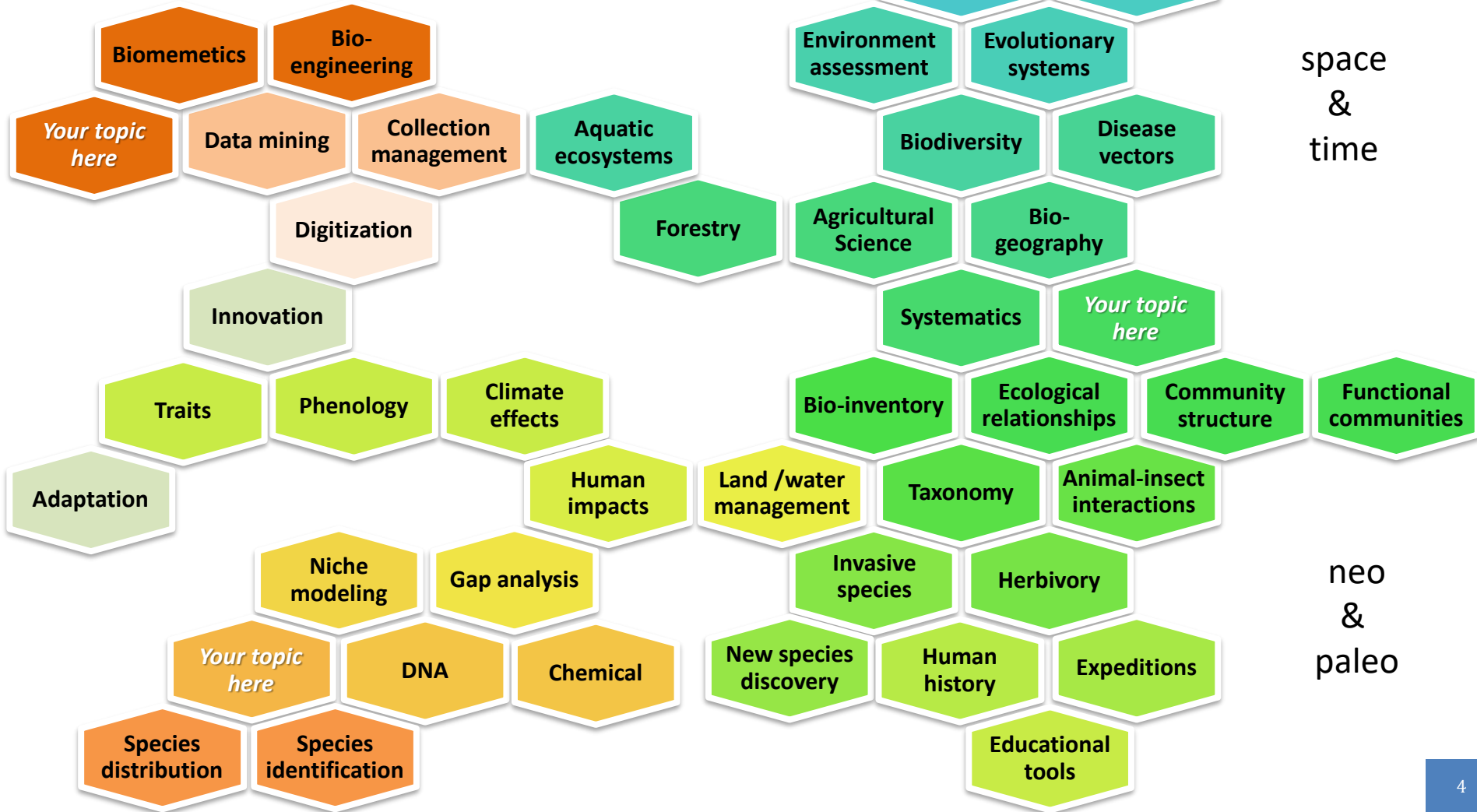***A vast untapped source of information!***

Why digitize?

- Digitization = Conversion of analog information (e.g., text, photos, sound, etc.) into digital information

- Principal benefits:
  - Generates **accessibility**
  - Enables **discovery**
  - Enables **sharing**
  - Facilitates **research**
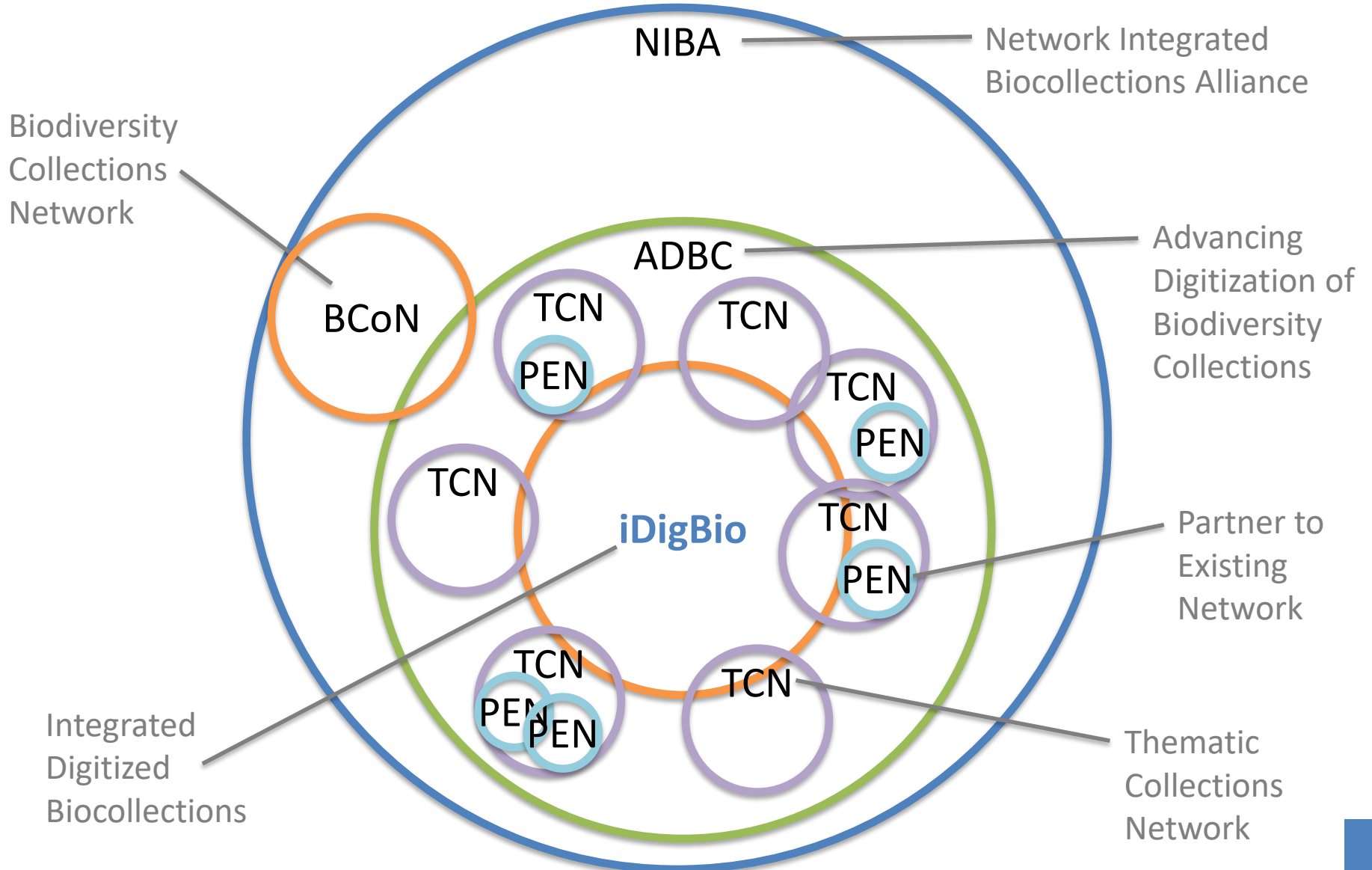 - Informs **policy**

THE HERBARIUM OF THE SULLIVANT MOSS SOCIETY

Thuidium abietinum Bryol. Eur.

Habitat   In woods, Mackinac Is.
Loc. U. of M. Biol. S., Cheboygan, Michigan
Coll. Nelle Ammons   No. M 422
Det. Nelle Ammons & G.B.Keiser   Date 8/9/31

Myriad of potential uses for biodiversity specimen data

Plant-animal relationship

Conservation

Restoration

Diet analysis

Parasites & microbes

Biomemetics

Bio-engineering

Environment assessment

Evolutionary systems

space & time

Your topic here

Data mining

Collection management

Aquatic ecosystems

Biodiversity

Disease vectors

Digitization

Forestry

Agricultural Science

Bio-geography

Innovation

Systematics

Your topic here

Traits

Phenology

Climate effects

Bio-inventory

Ecological relationships

Community structure

Functional communities

Adaptation

Human impacts

Land /water management

Taxonomy

Animal-insect interactions

Niche modeling

Gap analysis

Invasive species

Herbivory

neo & paleo

Your topic here

DNA

Chemical

New species discovery

Human history

Expeditions

Species distribution

Species identification

Educational tools

4

Improving data accessibility through digitization
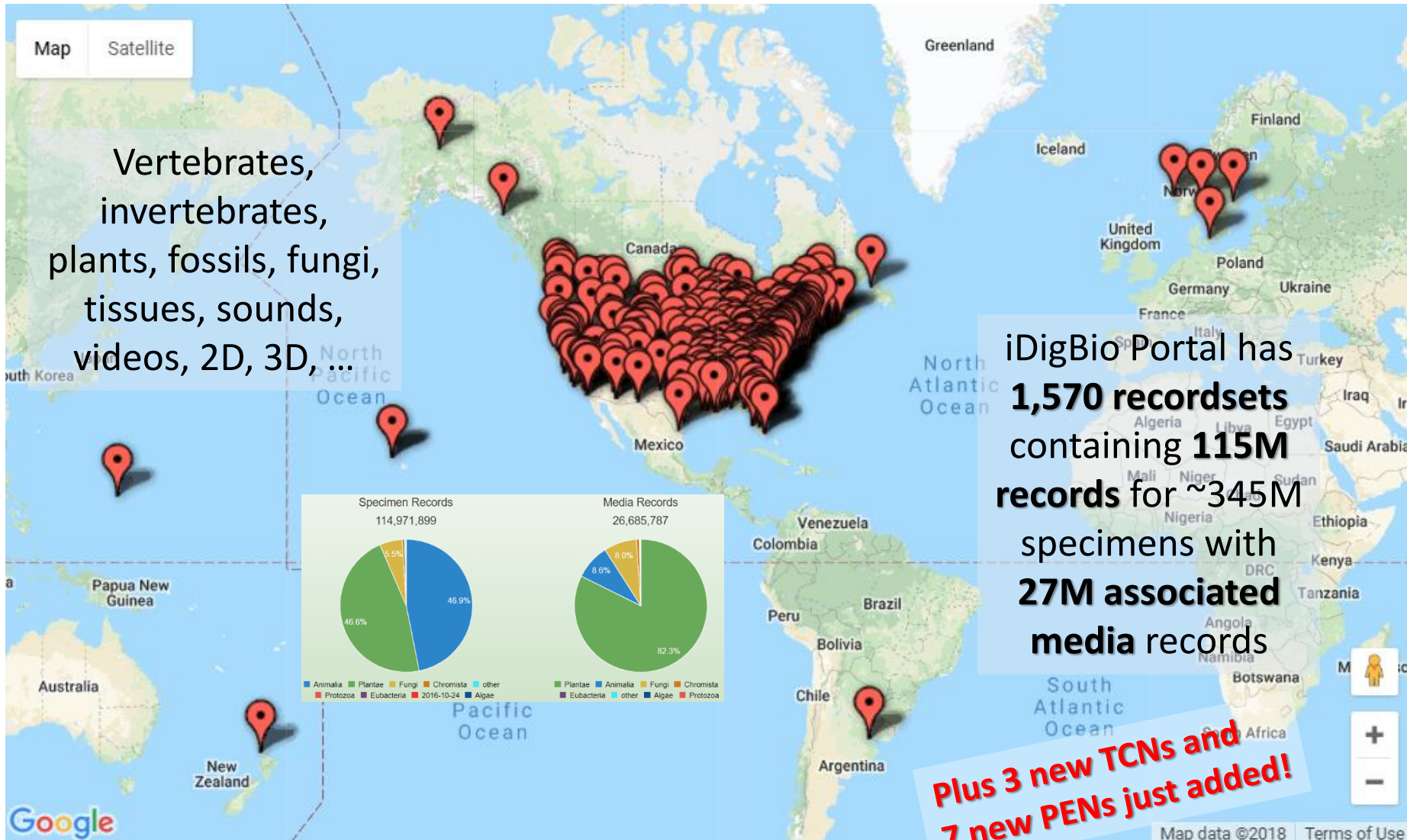
What is iDigBio?

- ❖ iDigBio is the **coordinating center** for the national effort to **national effort to digitize** non-federal U.S. natural history collections.

- ❖ NSF's **Advancing Digitization of Biodiversity Collections** (ADBC) program
  - 10-year, $100 million national effort

- ❖ **Networks of institutions organized by theme** to focus research, drive digitization efforts, and build communities
  - **Thematic Collection Networks** (TCNs)
  - **Partners to Existing Networks** (PENs)

"To advance scientific knowledge by improving access to digitized information in vouchered scientific collections across the U.S."

ADBC National Digitization Network

Vertebrates, invertebrates, plants, fossils, fungi, tissues, sounds, videos, 2D, 3D, …

iDigBio Portal has **1,570 recordsets** containing **115M records** for ~345M specimens with **27M associated media** records

Plus 3 new TCNs and 7 new PENs just added!

709 participating collections in 397 institutions (18 TCNs + 17 PENs)

What does iDigBio do?

1. **Facilitate digitization** of biodiversity collections data
   - Standards, workflows, protocols, task clusters
   - Workforce education, training, capacity building

2. Provide **access** to and **discoverability** of biodiversity data in a cloud computing environment

3. **Facilitate use** of biodiversity data to address key research, environmental, educational, and economic challenges

4. Envision the long-term **sustainability** of the **national digitization** effort
   - Expand participation
   - Proliferate & broaden uses of biodiversity data

What does iDigBio do?

Data Liberation
Flexible search
Indexed fields
Media
Geolocation
Synonyms
Data use
Collections
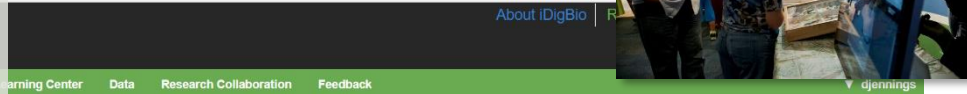
Research Use
Tool collaboration
Portal development
Data quality feedback
Research Spotlight
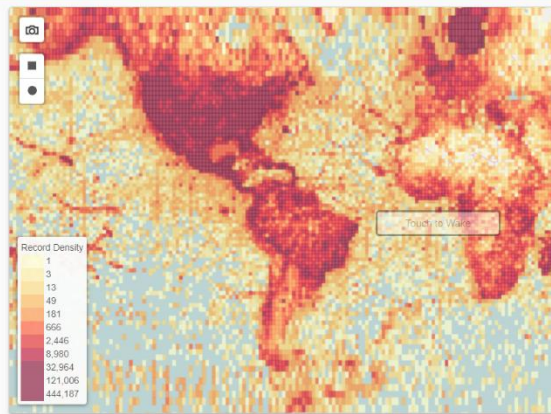
Training
Data skills
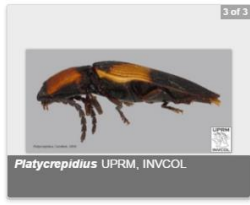Data literacy
Collections software
Imaging
Project management

EO&D
Citizen Science
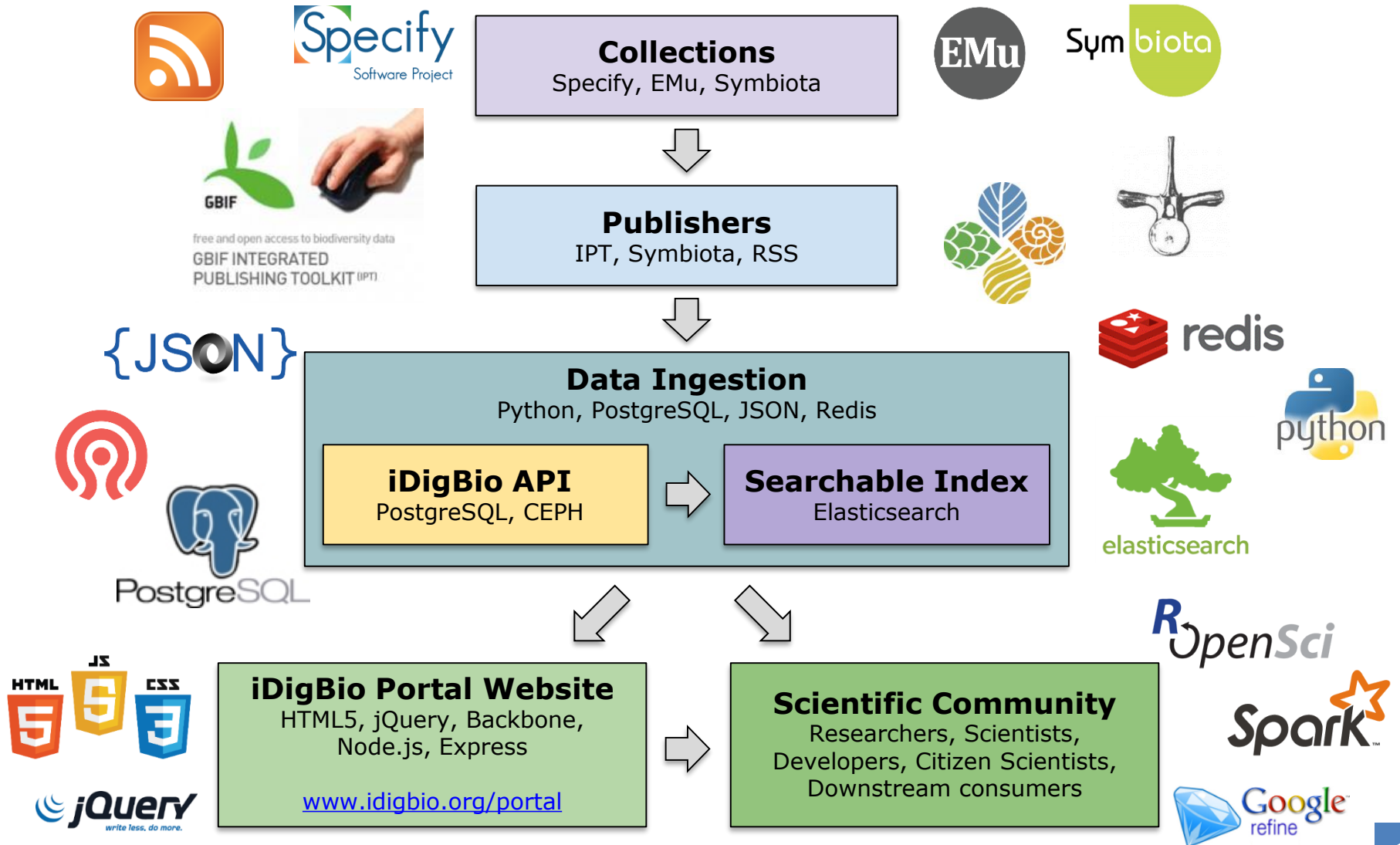K-12 materials
Undergraduate
Mentor teachers
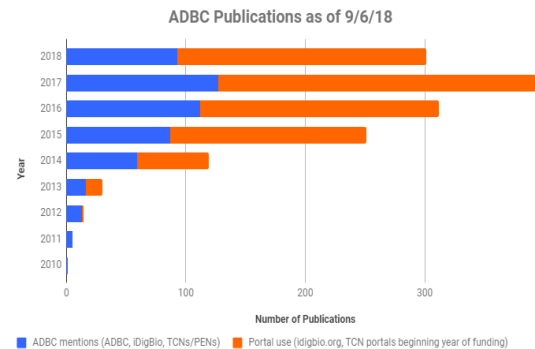
Digitization
Workflows
Protocols
Task clusters
Data standards

9

iDigBio Data Flow

Collections
Specify, EMu, Symbiota

Publishers
IPT, Symbiota, RSS

Data Ingestion
Python, PostgreSQL, JSON, Redis

iDigBio API
PostgreSQL, CEPH

Searchable Index
Elasticsearch

iDigBio Portal Website
HTML5, jQuery, Backbone, Node.js, Express

www.idigbio.org/portal

Scientific Community
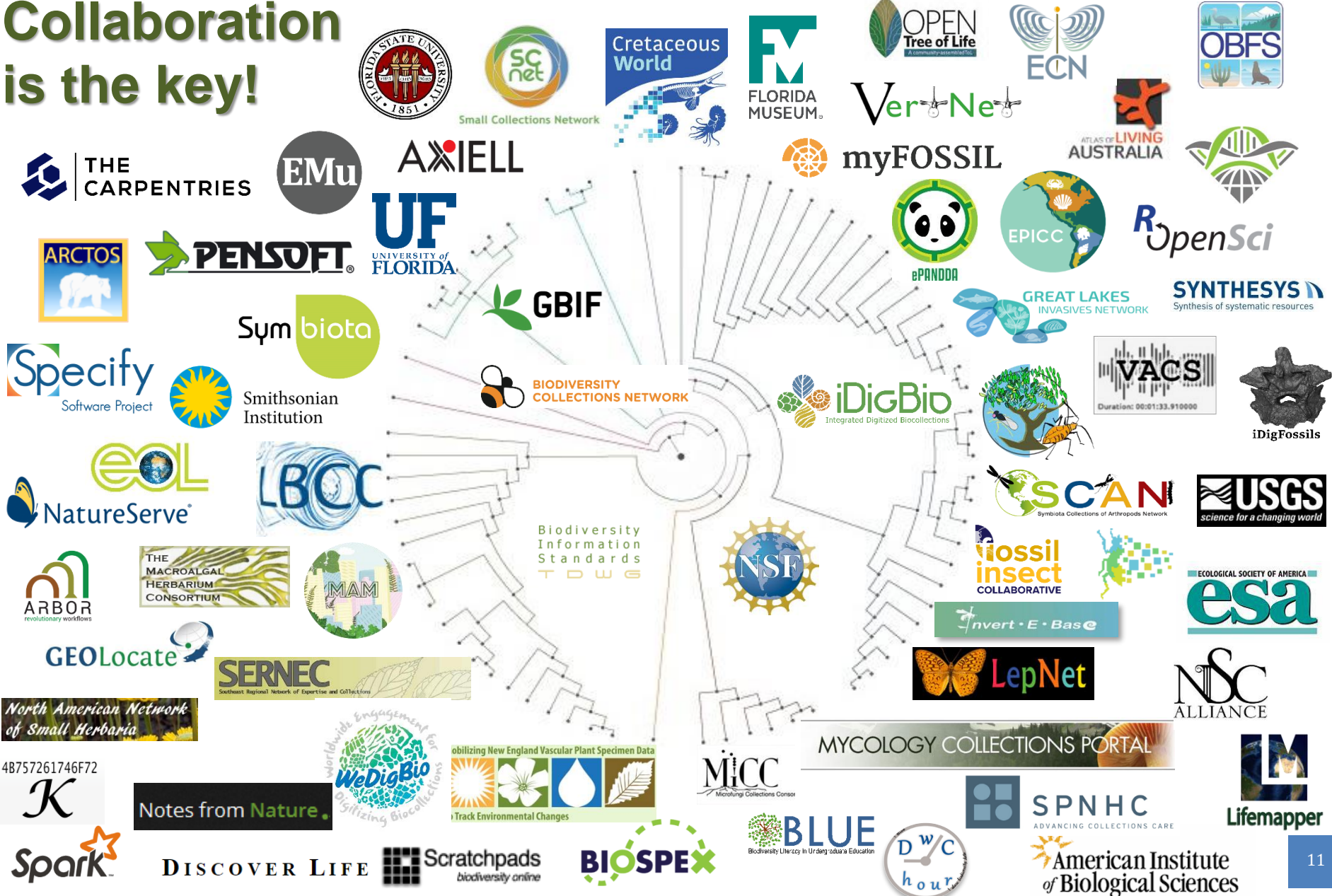Researchers, Scientists, Developers, Citizen Scientists, Downstream consumers

Collaboration is the key!

We want to engage with you!

- iDigBio provides **access to data** with the means to **answer research questions**

- Many opportunities for **collaboration** and potential **funding**:
  - **Public participation** in digitization
    - e.g., host a WeDigBio event, host a DwC hour
  - **Research** using the data already in the portal
    - e.g., niche modeling, conservation, etc.
  - **Data mining** the portal for new discoveries
    - e.g., extract measurements/characteristics
  - Gathering all of the "**dark data**"
    - Help us get "all" collections digitized and online!
    - e.g., proposals for TCN, PEN, CSBR, IMLS, etc.
  - **Enriching** the data
    - e.g., data linking, field notes, etc.
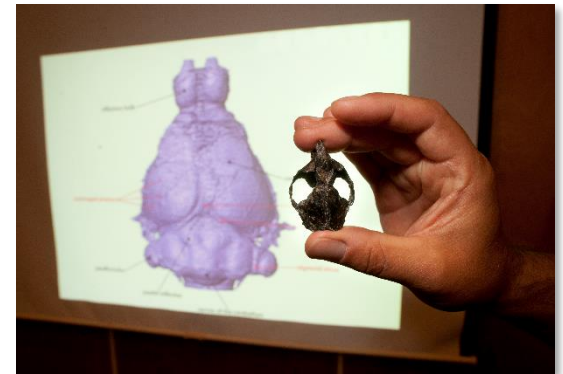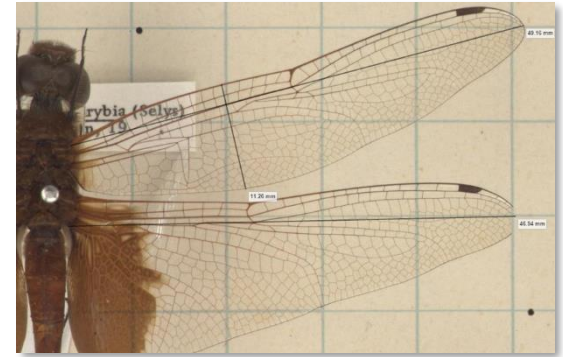
We want everyone involved!

www.idigbio.org

facebook.com/iDigBio

twitter.com/iDigBio

vimeo.com/idigbio

idigbio.org/rss-feed.xml

idigbio.org/events-calendar/export.ics

Join us 18-21 October for WeDigBio 2018! Sign up at wedigbio.org !