



THE OHIO STATE
UNIVERSITY



AI for Nature: Exploring our Natural Heritage

Elizabeth G. Campolongo

Senior Data Scientist

Imageomics Institute and AI & Biodiversity Change (ABC) Global Center



This material is based upon work supported by the National Science Foundation under Awards No. 2118240 and 233042, NSERC Award No. 58513, and draws on research supported by the Social Sciences and Humanities Research Council.





HDR Institute: NSF

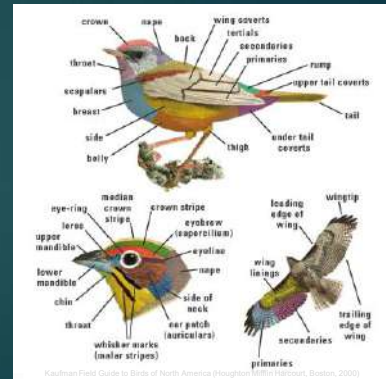
imageomics.org

Imageomics: A New Frontier of Biological Information

Powered by Knowledge-Guided Machine Learning



Making biological traits computable from images, video, 3D, ...

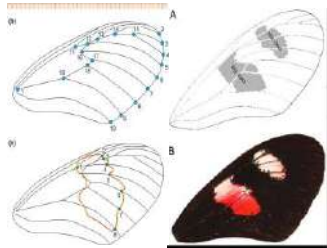


The scope of Imageomics-enabled tasks

Micro

Global

Segmentation &
landmarking



Slide Credit: Dr. Harry Chao

Automating Carabidae Morphometric Measurements



DINOv3
Foundation
Model



Heatmap
Head



Frozen

Trainable

The scope of Imageomics-enabled tasks

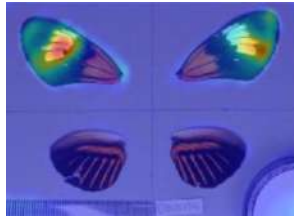
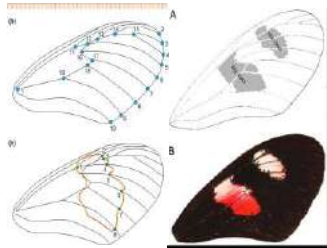
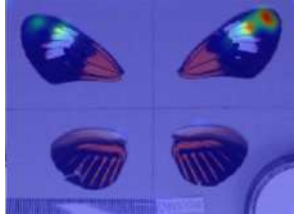
Micro

Global

Segmentation &
landmarking



Trait
extraction &



Slide Credit: Dr. Harry Chao

Sparse Autoencoders (SAEs) for Interpretability and Scientific Exploration



Training objective

Reconstruct original embedding with minimal active features per image (*sparsity*).

Practical Goal

Ability to explore and understand the model embedding space.



saev repo

Scientific Aspiration

Trait discovery: Identify *potential or candidate traits*—features that strongly activate on a particular, repeated, visual trait (e.g., wing pattern or beak structure).

Requires **domain expert** interpretation and analysis.



Known Dichotomous Keys

H. erato

H. melpomene

SAE
Activations

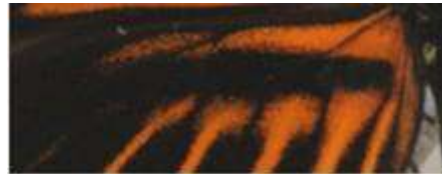
H. erato
ssp. *cyrbia*

H. melpomene
ssp. *cythera*



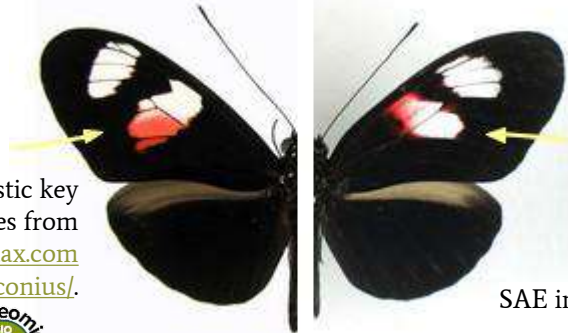
ssp. *lativitta*

ssp. *malleti*



ssp. *plesseni*

ssp. *notabilis*



Diagnostic key
images from
[cliniquevetodax.com](http://cliniquevetodax.com/Heliconius/)
[/Heliconius/](http://Heliconius/).



SAE images from [Heliconius Collection \(Cambridge Butterfly\)](http://HeliconiusCollection.org), a compilation of images taken by the Butterfly Genetics Group (University of Cambridge): Chris Jiggins, Gabriela Montejo-Kovacevic, Ian Warren, and Patricio Salazar. Image DOI: [10.5281/zenodo.3477891](https://doi.org/10.5281/zenodo.3477891).

SAEs: Evaluating Potential Traits

Original



Activation



Increased



Decreased



The scope of Imageomics-enabled tasks

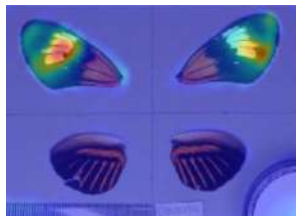
Micro

Global

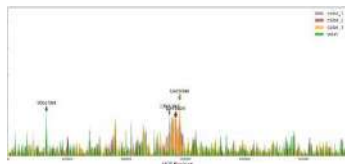
Segmentation & landmarking



Trait extraction &



Trait & genotype alignment



Wildlife monitoring



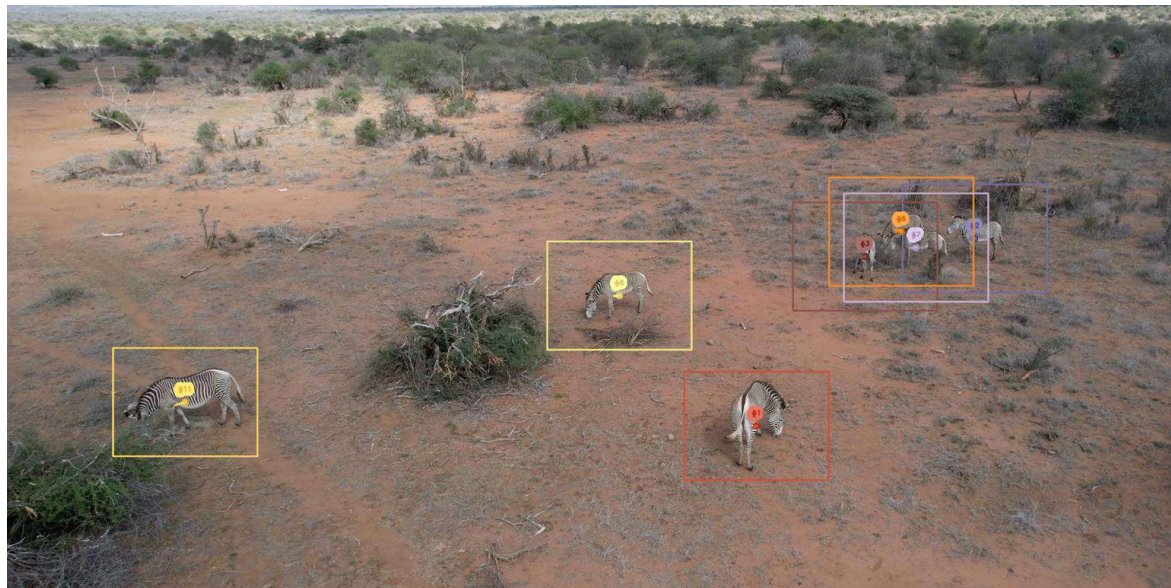
Walk

Graze



Slide Credit: Dr. Harry Chao

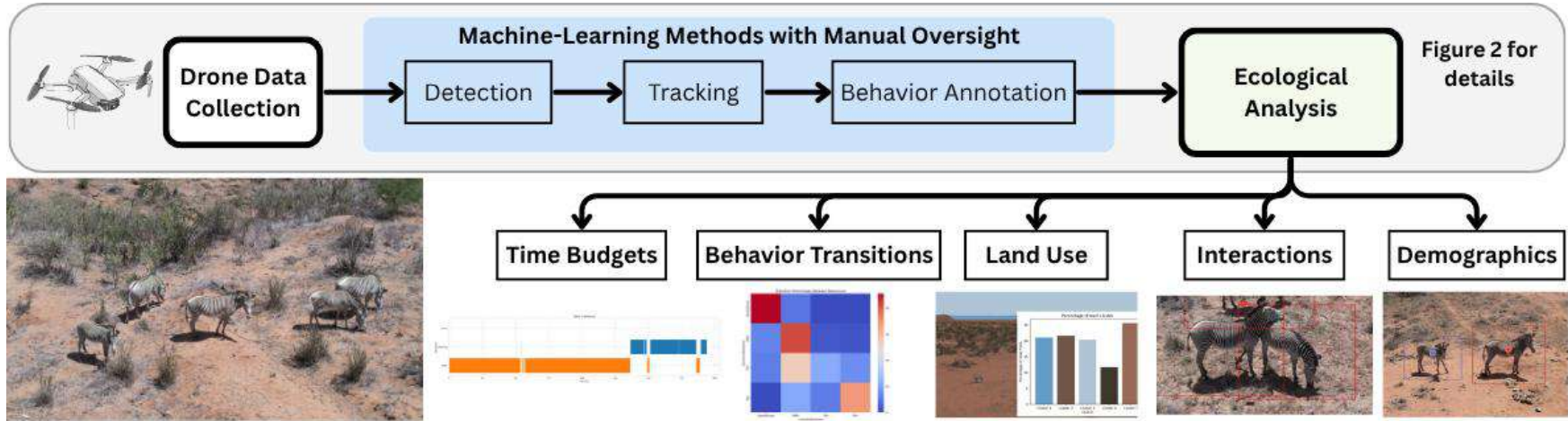
Kenyan Animal Behavior Recognition (KABR)



Learn more about
the KABR Project



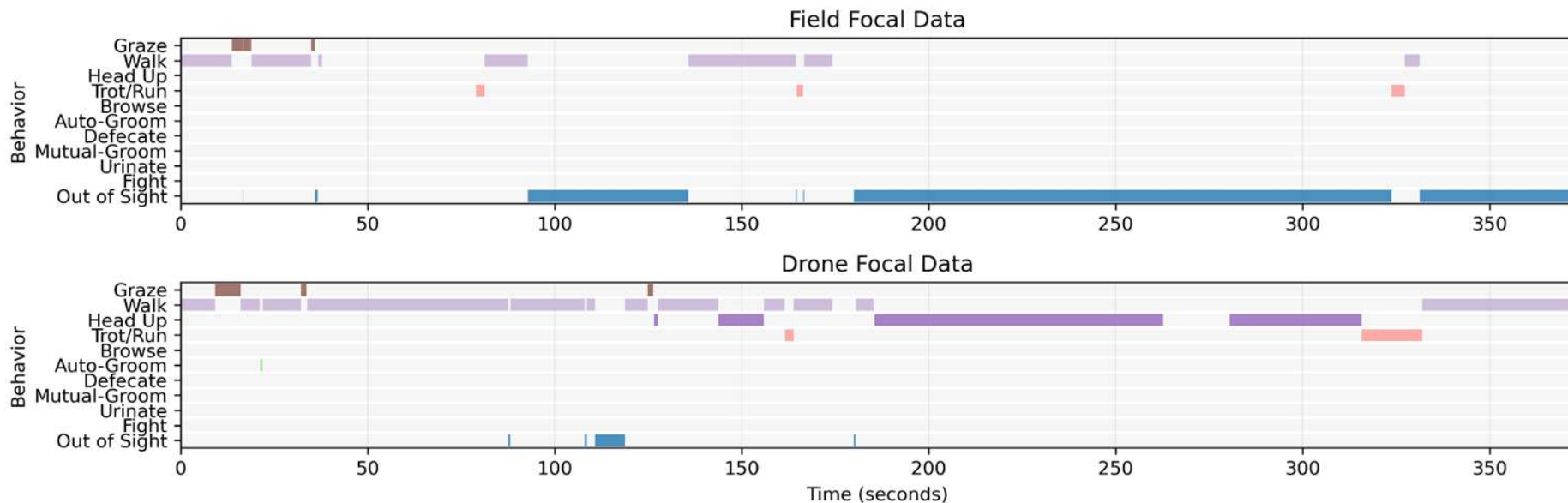
Automating Animal Behavior Monitoring



kabr-tools package



Automating Animal Behavior Monitoring



kabr-tools package



The scope of Imageomics-enabled tasks

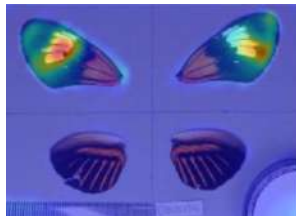
Micro

Global

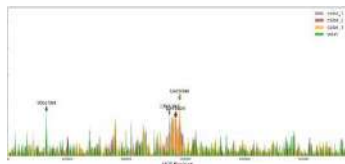
Segmentation & landmarking



Trait extraction &



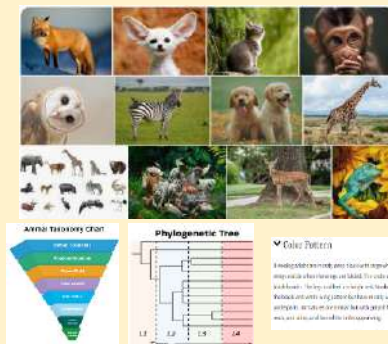
Trait & genotype alignment



Wildlife monitoring



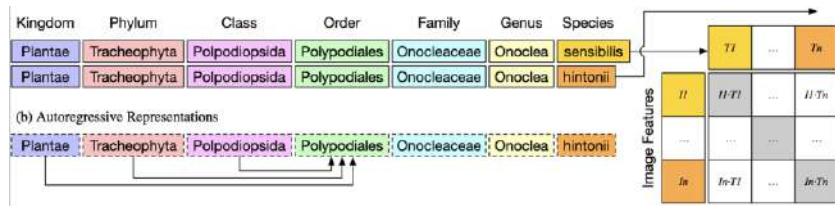
Imageomics foundation model



Multi-task enabling Foundation Model



BioCLIP: A Vision Foundation Model for the Tree of Life



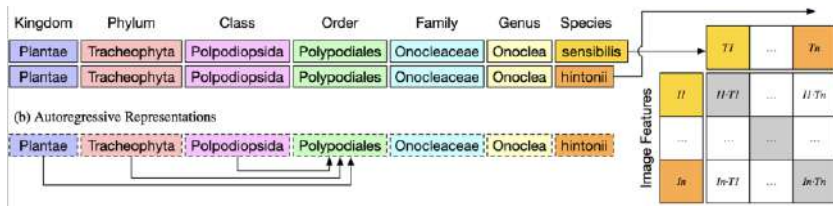
Samuel Stevens, Jiaman Wu, Matthew J Thompson, Elizabeth G Campolongo, Chan Hee Song, David Edward Carlyn, Li Dong, Wasila M Dahdul, Charles Stewart, Tanya Berger-Wolf, Wei-Lun (Harry) Chao, Yu Su

imageomics.github.io/bioclip-ecosystem

10M images, ~450K taxa



BioCLIP 2: A Vision Foundation Model for the Tree of Life



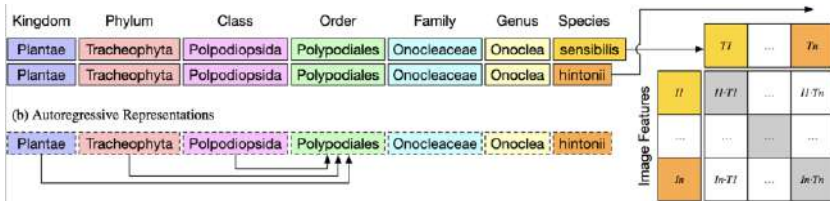
Jianyang Gu, Sam Stevens, Elizabeth Campolongo, Matt Thompson, Net Zhang, Jiaman (Lisa) Wu, Andrei Kopanev, Zheda Mai, Alex White, James Balhoff, Wasila Dahdul, Daniel Rubenstein, Hilmar Lapp, Tanya Berger-Wolf, Wei-Lun (Harry) Chao, Yu Su

imageomics.github.io/bioclip-ecosystem

214M images, ~1M taxons (868K species)

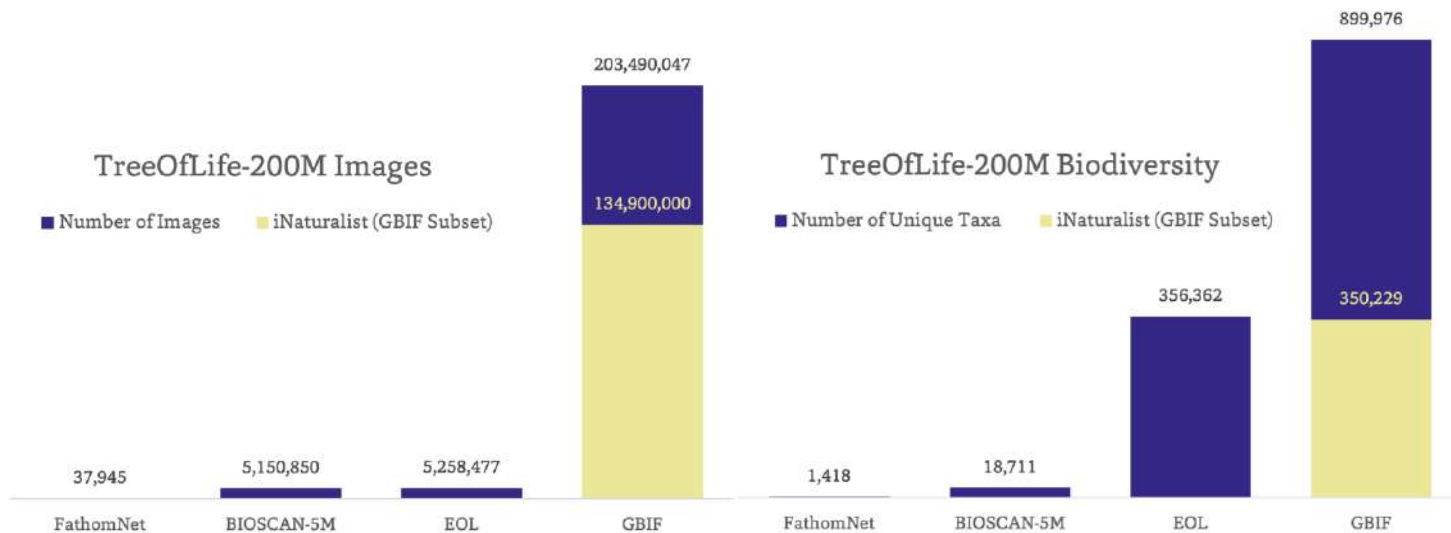


BioCLIP 2: A Vision Foundation Model for the Tree of Life



Jianyang Gu, Sam Stevens, Elizabeth Campolongo, Matt Thompson, Net Zhang, Jiaman (Lisa) Wu, Andrei Kopanev, Zheda Mai, Alex White, James Balhoff, Wasila Dahdul, Daniel Rubenstein, Hilmar Lapp, Tanya Berger-Wolf, Wei-Lun (Harry) Chao, Yu Su

imageomics.github.io/bioclclip-ecosystem



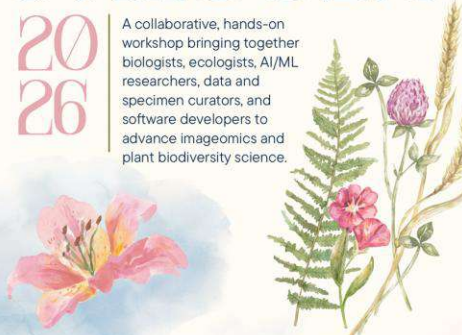
BioCLIP in Action: Community

 THE IMAGEOMICS INSTITUTE PRESENTS

FLORA PALOOZA

2026

A collaborative, hands-on workshop bringing together biologists, ecologists, AI/ML researchers, data and specimen curators, and software developers to advance imageomics and plant biodiversity science.



 AUGUST 17-20, 2026

 HYATT HOUSE COLUMBUS/OSU COLUMBUS, OH

 SCAN TO LEARN MORE AND APPLY!

 **INTERDISCIPLINARY COLLABORATION**
Work and collaborate with other people of different expertise.

 **BUILD TOOLS, DATA, AND WORKFLOWS**
Co-create open resources that empower the imageomics community.

 **ADVANCE RESEARCH**
Develop innovative approaches to understand life on our changing planet.

GET YOUR FUNCTIONALITY ON.

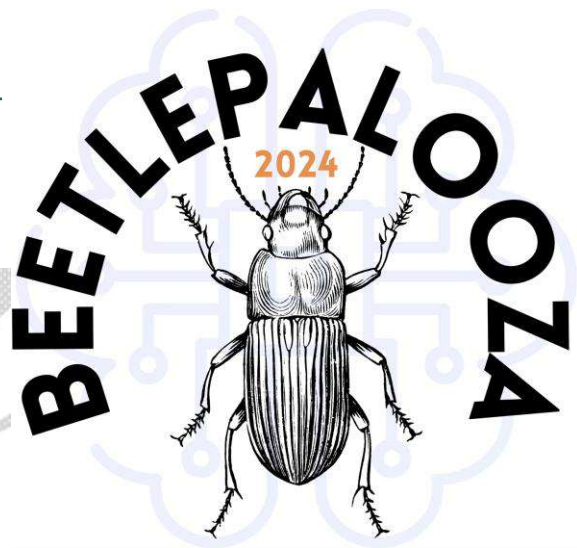
FUNCA PALOOZA



AUGUST 18-21, 2025

A COLLABORATIVE HANDS-ON MACHINE LEARNING EVENT FOR EXTRACTING FUNCTIONAL TRAITS FROM IMAGES.





Real-world Impact: BioCLIP in Action



Challenge:

- Manual annotation is time-consuming and expensive.

Solution:

- Integrate BioCLIP into Mothbox pipeline with pybioclip.
- Utilizes FiftOne for UI.

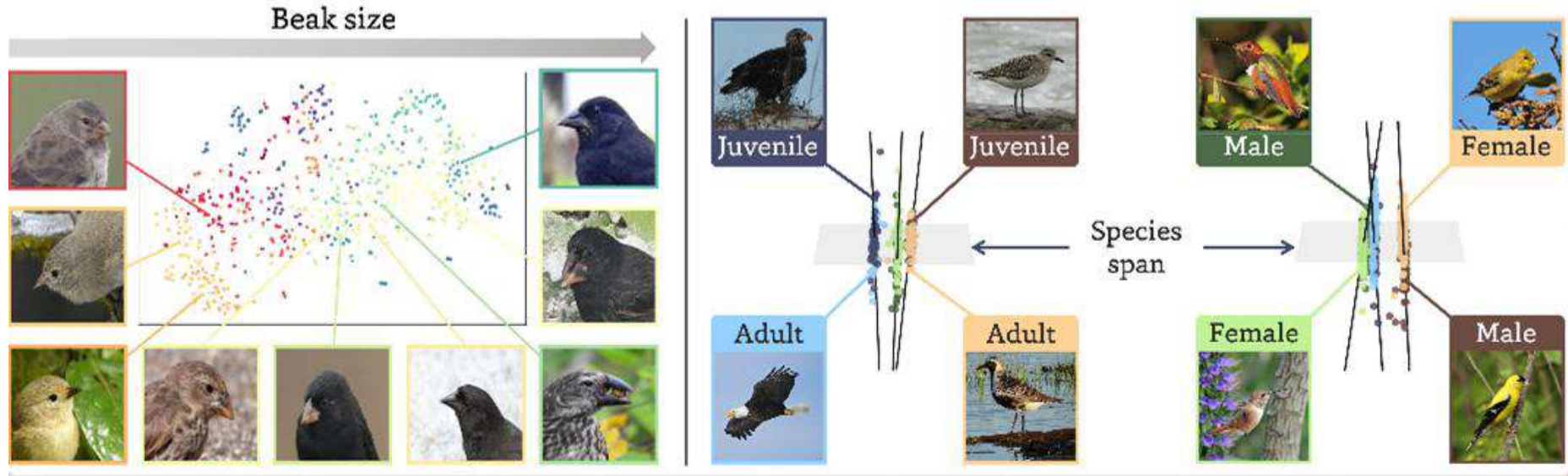
Result:

- Expert can review a full night's images and label to next rank in 15-20 minutes.

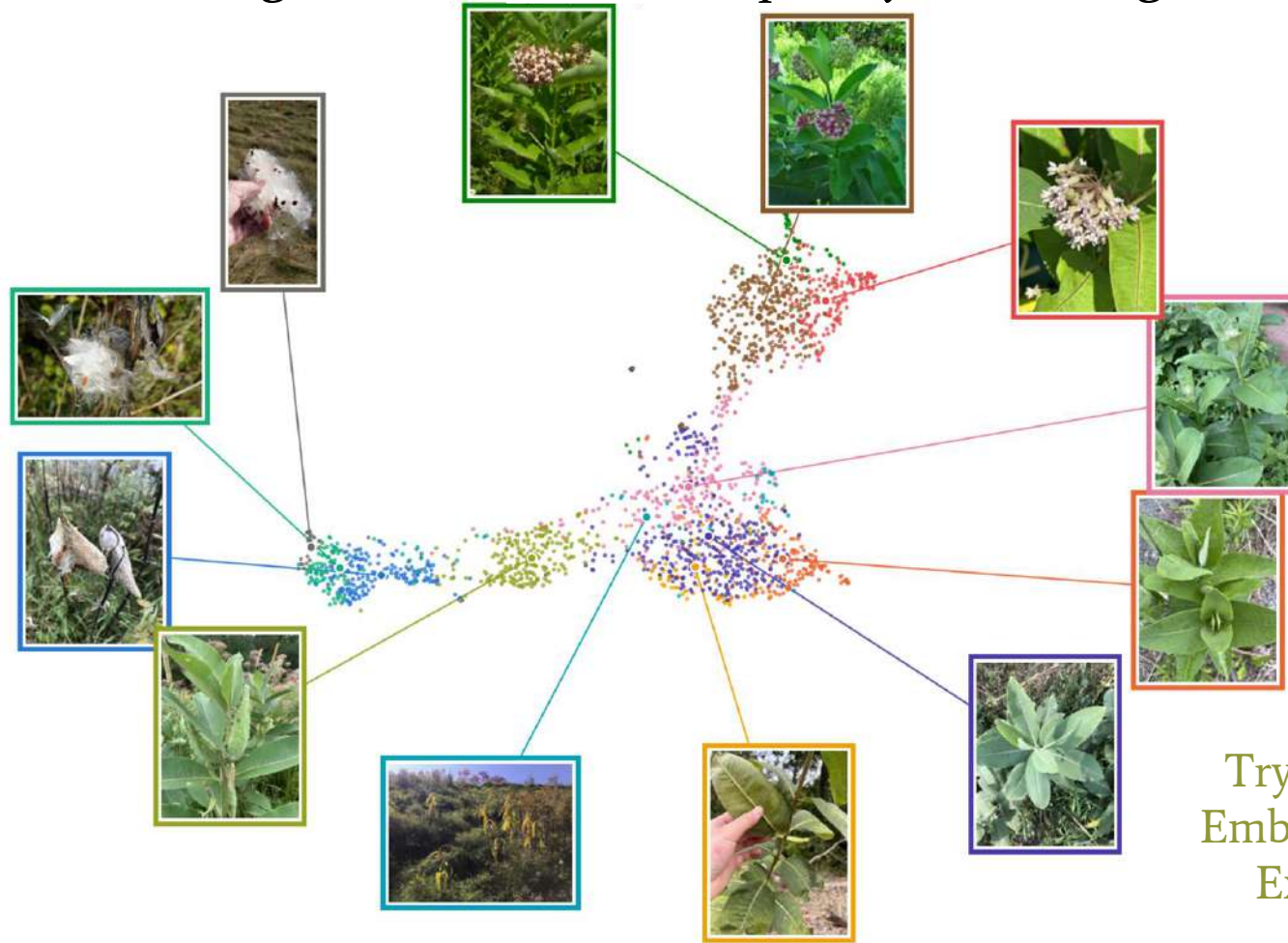


Images courtesy of [Mothbox](#).

BioCLIP 2: Beyond Classification



BioCLIP 2 Embeddings of milkweed (*Asclepias syriaca*) images



Try Image
Embedding
Explorer



This is a ...



BioCLIP 2:

*Lamprotornis
chalybaeus* (Greater
Blue-eared Starling)

Sometimes we don't get the animal we expect

Solutions

Isolate individuals to identify all species.

Mask those to exclude from the image for prediction using the model patches.

bioclip patch zebra-and-bird.png (1/1)

SELECT PATCHES TO MASK

PREVIEW (MASKED)

MODE

- ☒ Grid (14x14)
- ☐ Freeform

MASK FILL

- ☒ Mean fill
- ☐ Zero fill (black)
- ☐ Gaussian blur

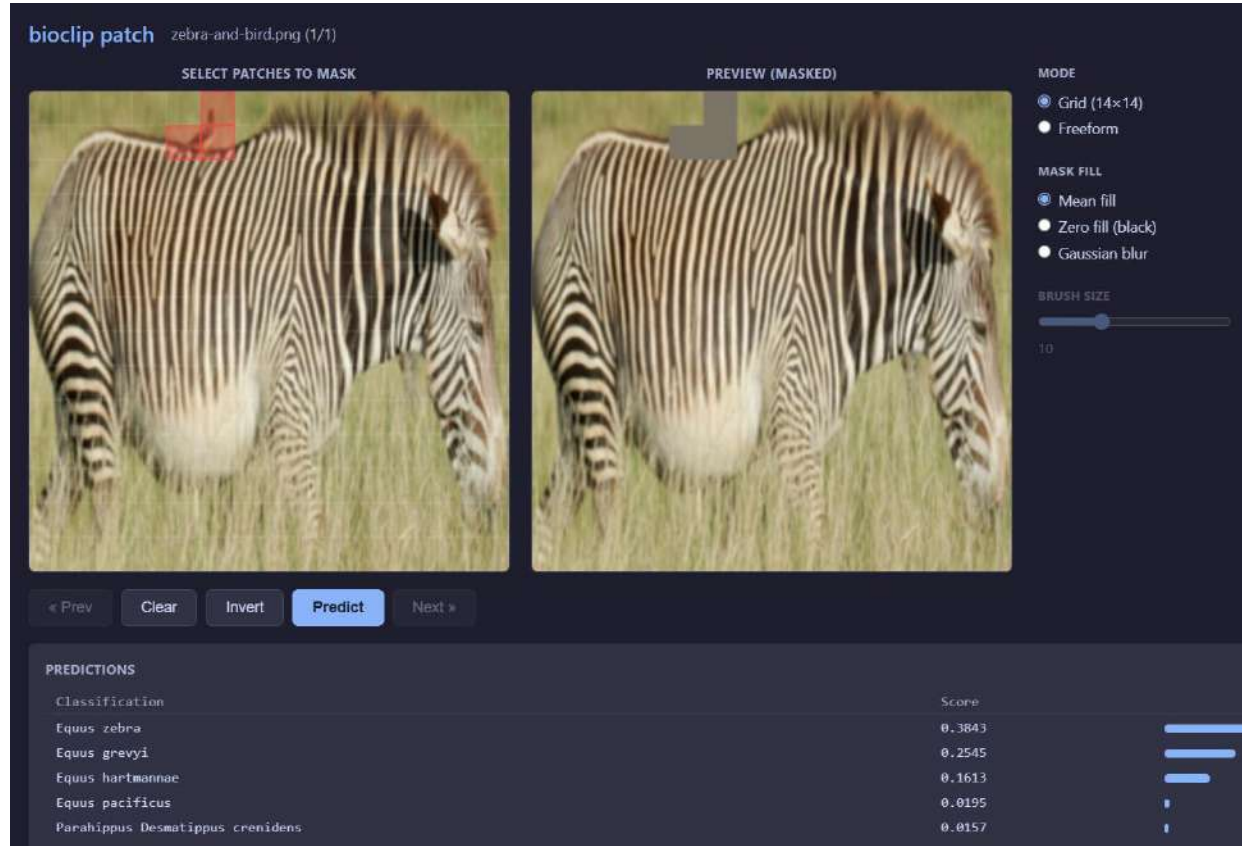
BRUSH SIZE

10

« Prev Clear Invert Predict Next »

PREDICTIONS

Classification	Score
Equus zebra	0.3843
Equus grevyi	0.2545
Equus hartmannae	0.1613
Equus pacificus	0.0195
Parahippus Desmatippus crenidens	0.0157



We can find species interactions!



Search
TreeOfLife-200M
yourself!

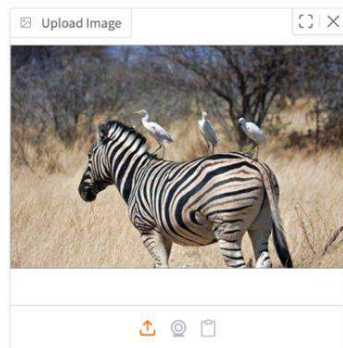


Image Source Scope
Filter results by image source availability.

URL-Available Only

Search Depth (nprobe)

27

IVF partitions to search. Higher = more accurate, slower.

1

128

Top N Results

40

1

128

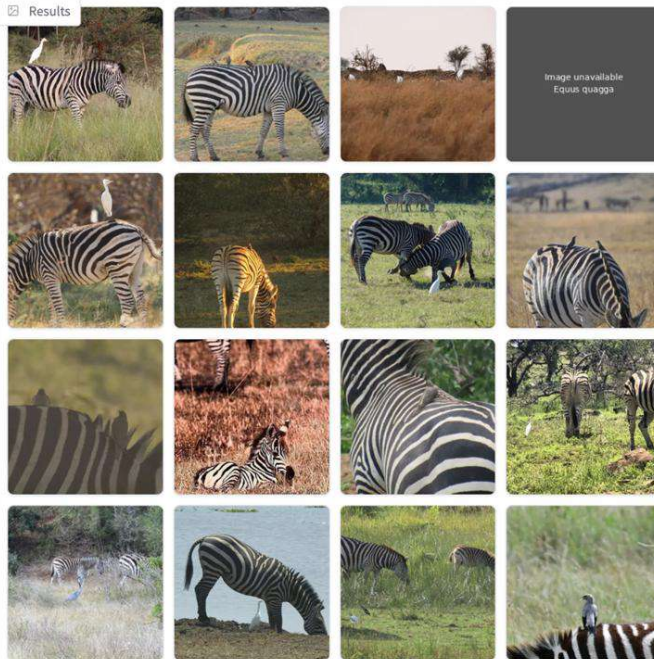
Search

Export Result Metadata

Search Results

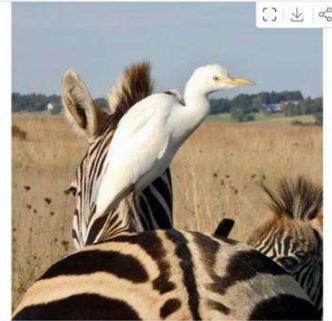
Prediction

Results



Selected

Summary



#25 Cattle Egret

Bubulcus ibis

Distance: 0.1809

Image: ok_cached

Taxonomy

Source

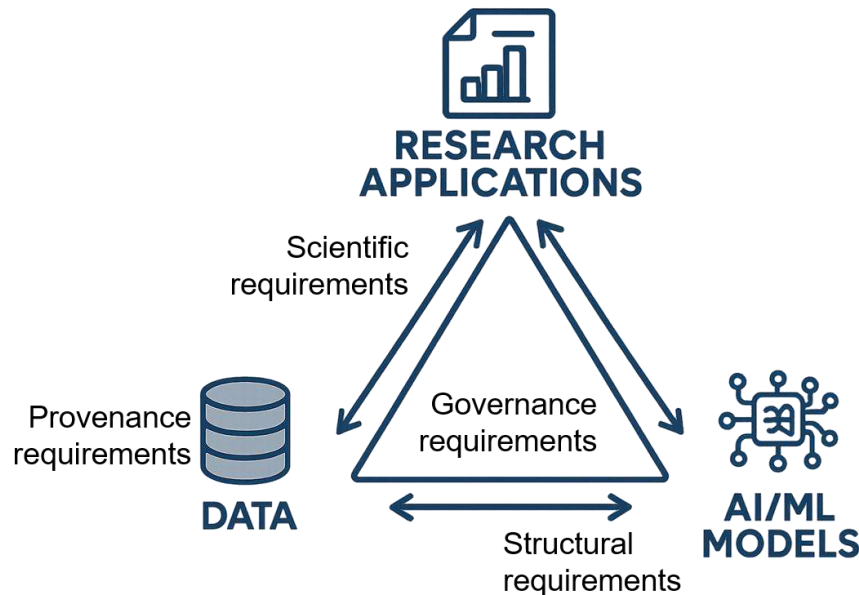
Rank	Name
Kingdom	Animalia
Phylum	Chordata
Class	Aves
Order	Pelecaniformes

FAIR4AI-Bio: AI-enabled Biodiversity Connections



AI-FAIR Criteria

- ✓ Structural
- ✓ Scientific
- ✓ Provenance
- ✓ Governance

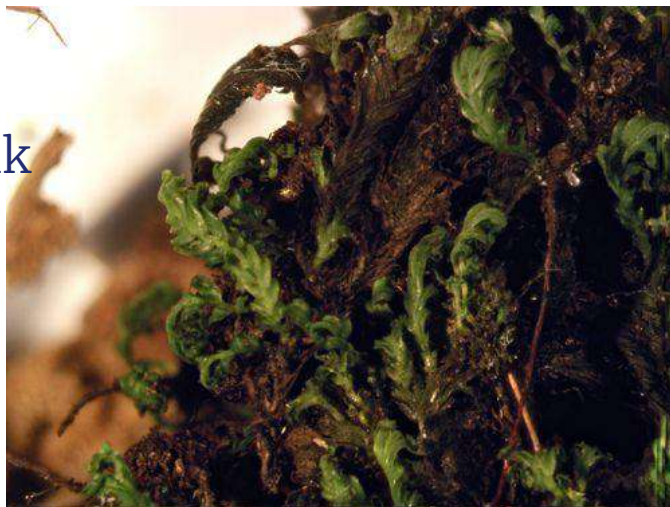


The FAIR4AI-Bio initiative is funded by the US National Science Foundation's Findable Accessible Interoperable Reusable Open Science (FAIROS) program under Awards [2531924](#), [2531922](#), [2531925](#), [2600200](#) (Collaborative Research: Disciplinary Improvements: AI-ready Ecology and Biodiversity Data Infrastructure for Open Science).

From FAIR to FAIR4AI

Beyond FAIR criteria

- Data/metadata can be queried without downloading large files or specialized file types
 - Subsetting for only desired data:
 - Image type
 - Taxonomic rank



Fissidens santa-clarensis Thér.

Collected in United States of America
by University of Florida (licensed
under

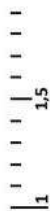
<http://creativecommons.org/licenses/by-nc/4.0/>)



Florida Museum of Natural History (2025).
University of Florida Herbarium - Bryophytes.
Occurrence dataset <https://doi.org/10.15468/cuvh8t>
accessed via GBIF.org on 2025-08-08.
<https://www.gbif.org/occurrence/5138270268>

Motivation

Beyond FAIR criteria:
Data point context and content



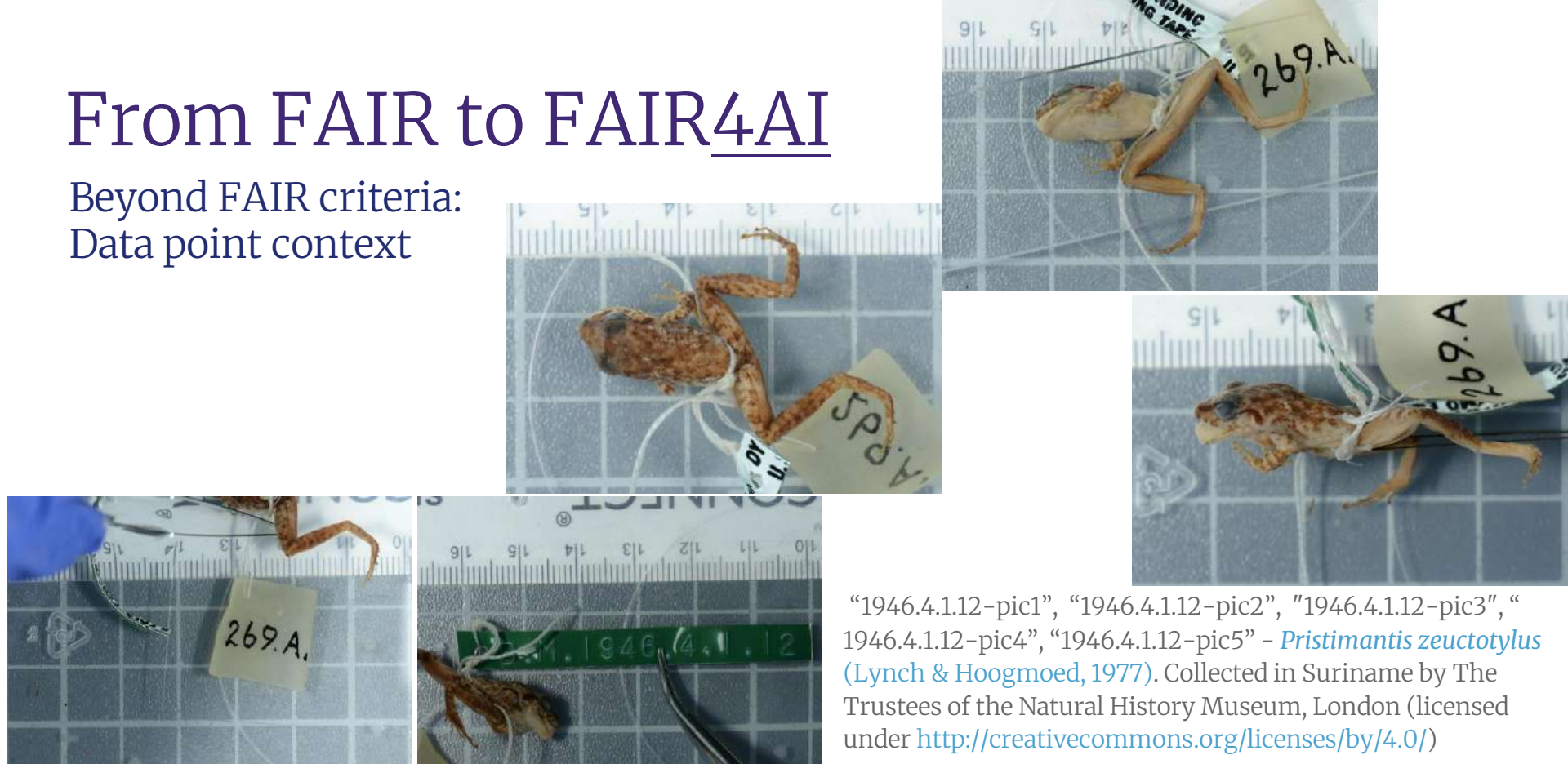
"CANL10002094_u.jpg" (left) and "CANL10002094_t.jpg" (right) – *Caloplaca cerina* (Hedw.) Th.Fr. Collected in Canada by Credit Line: © Canadian Museum of Nature (licensed under <http://creativecommons.org/licenses/by-nc/4.0/>)

Corynoneura eocenica Doitteau & Nel, 2007. Collected in France. (licensed under <http://creativecommons.org/licenses/by/4.0/>)



From FAIR to FAIR₄AI

Beyond FAIR criteria:
Data point context



“1946.4.1.12-pic1”, “1946.4.1.12-pic2”, “1946.4.1.12-pic3”, “1946.4.1.12-pic4”, “1946.4.1.12-pic5” – *Pristimantis zeuctotylus* (Lynch & Hoogmoed, 1977). Collected in Suriname by The Trustees of the Natural History Museum, London (licensed under <http://creativecommons.org/licenses/by/4.0/>)

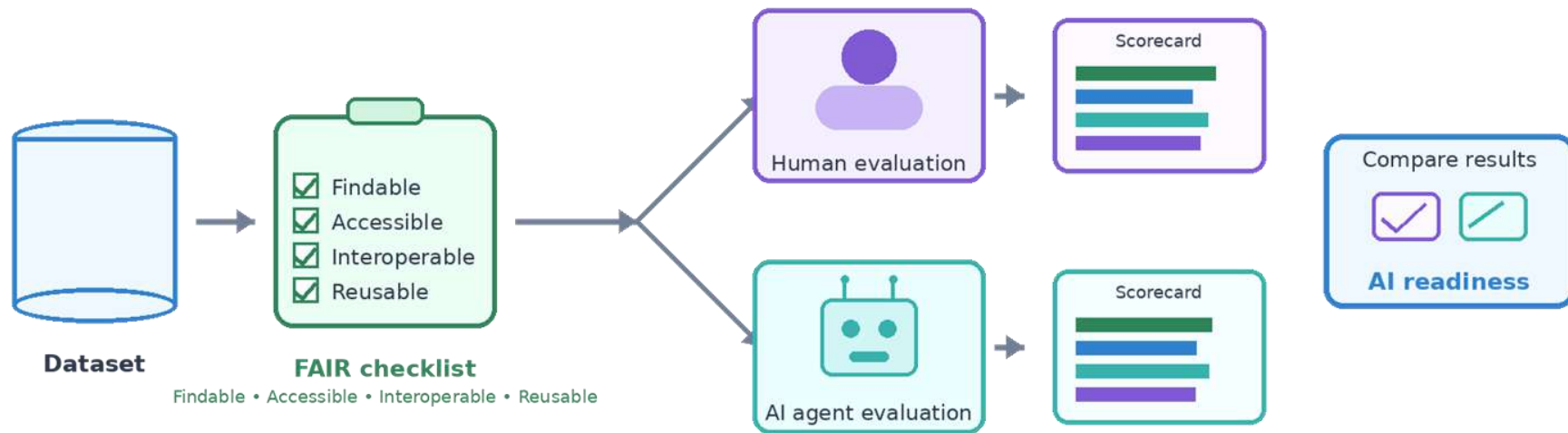
Natural History Museum (2025). Natural History Museum (London) Collection Specimens. Occurrence dataset <https://doi.org/10.5519/qd.eqf5xtwc> accessed via GBIF.org on 2025-08-08.
<https://www.gbif.org/occurrence/1056329684>



FAIR4AI Checklist Evaluation

Source Checklists:

- ☐ ESIP AI ready data checklist (v1.0)
- ☐ Imageomics data checklist
- ☐ Imageomics metadata checklist



Objectives:

1. Can a machine (e.g., AI agent) read metadata from a dataset landing page?
2. How FAIR4AI does the agent perceive the dataset to be?
3. How does agent-based evaluation compare to human-based evaluation?

<https://github.com/FAIR4AI-Bio/FAIR4AI-checklists>





Thank you!

Questions?

catalog.imageomics.org



Join the FAIR4AI Initiative:
campolongo.4@osu.edu

Join the Imageomics
mailing list



Hugging Face

Follow Imageomics
on Hugging Face
and GitHub!



GitHub