Existing AI tools and services: A Galaxy Platform focus

27 Janvier 2026

Using Galaxy-Ecology as a sustainable Biodiversity common virtual research environment: Focus on integration, access and automation of AI algorithms

Yvan Le Bras
MNHN / PNDB-Data Terra / UAR DoHNÉE
French BON
France

Coline Royaux
MNHN / PNDB – DATA TERRA / UAR DoHNÉE / French BON
France

Marie Jossé
CNRS / – DATA TERRA / French BON
France

Pauline Seguineau
MNHN / PNDB – DATA TERRA / UAR DoHNÉE / French BON
France

Arthur Barreau
SU / PNDB – DATA TERRA / UMR ISYEB / French BON
France

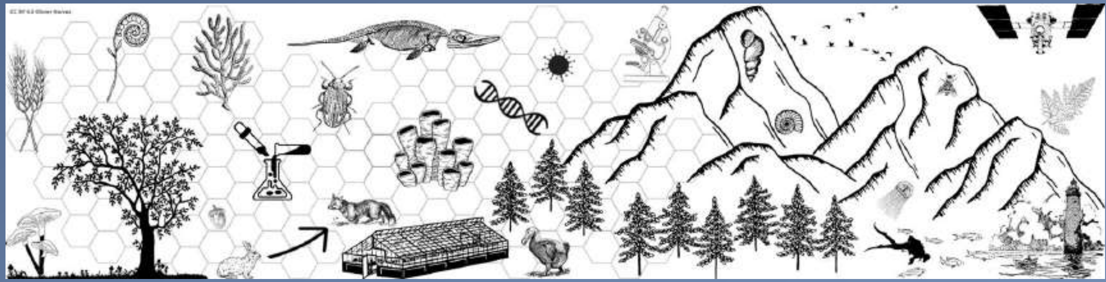
Nadine Le Bris
SU / PNDB – DATA TERRA / UMR ISYEB / French BON
France

Olivier NORVEZ
CNRS / PNDB & Theia – DATA TERRA / French BON
France

Francis Clément
MNHN / PNDB – DATA TERRA / French BON
France

Nicolas Buisine
MNHN / PNDB – DATA TERRA / UAR DoHNÉE / French BON
France

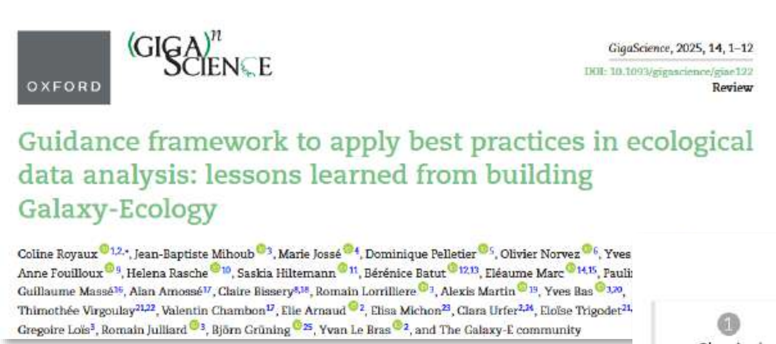
Jian-Sheng Sun
MNHN / PNDB – DATA TERRA / UAR DoHNÉE / French BON
France



Galaxy Ecology: From a platform to best practices in Biodiversity data analysis

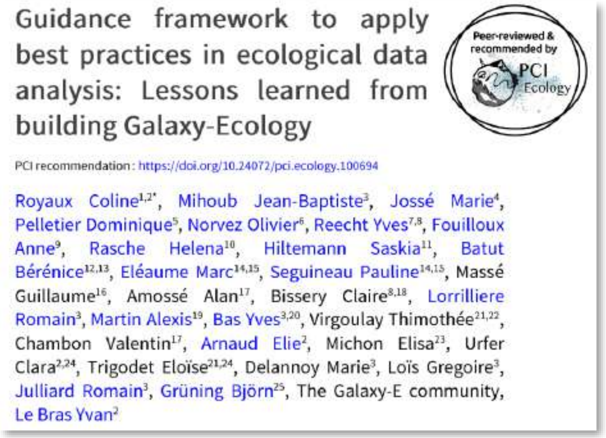
Applying “Galaxy-E” best practices to reuse!

Atomization & Generalization to reuse & transpose


OXFORD
(GIGA)^{n} SCIENCE
GigaScience, 2025, 14, 1–12
DOI: 10.1093/gigascience/giae122
Review

Guidance framework to apply best practices in ecological data analysis: lessons learned from building Galaxy-Ecology

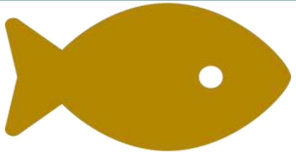
Coline Royaux, Jean-Baptiste Mihoub, Marie Jossé, Dominique Pelletier, Olivier Norvez, Yves Anne Fouilloux, Helena Rasche, Saskia Hiltemann, Bérénice Batut, Eléaume Marc, Pauli… Guillaume Massé, Alan Amossé, Claire Bissery, Romain Lorrillere, Alexis Martin, Yves Bas, Thimothée Virgoulay, Valentin Chambon, Elie Arnaud, Elisa Michon, Clara Urfer, Eloïse Trigodet, Gregoire Loïs, Romain Julliard, Björn Grüning, Yvan Le Bras, and The Galaxy-E community


Guidance framework to apply best practices in ecological data analysis: Lessons learned from building Galaxy-Ecology

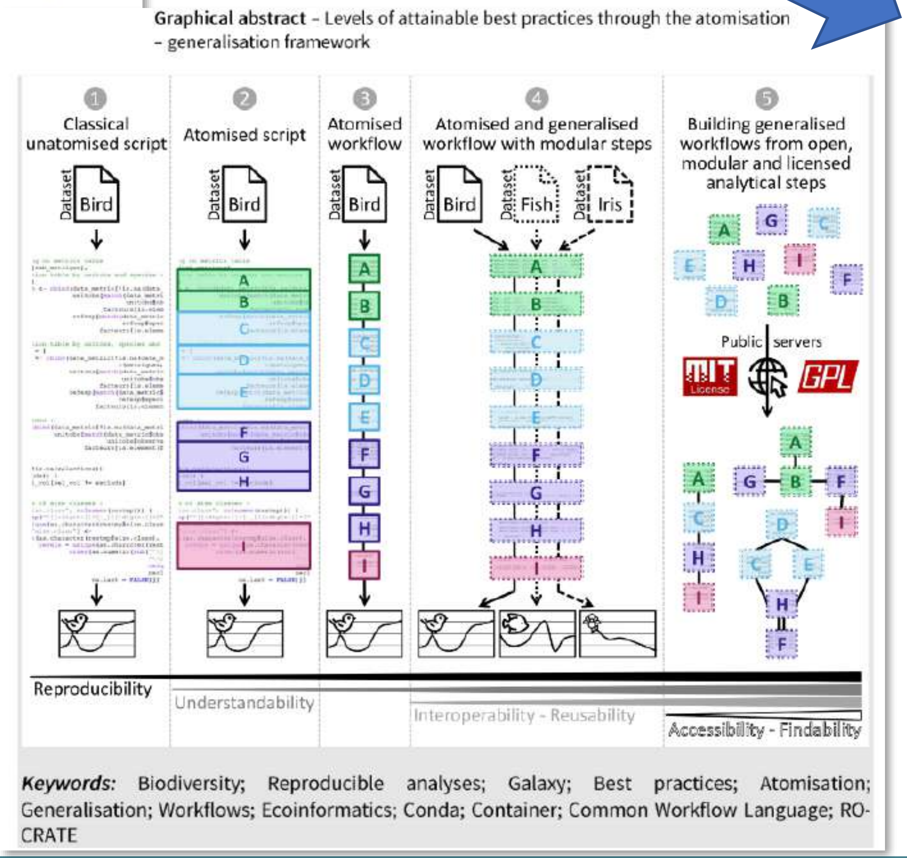
Peer-reviewed & recommended by PCI Ecology

PCI recommendation: https://doi.org/10.24072/pci.ecology.100694

Royaux Coline[1,2*], Mihoub Jean-Baptiste[3], Jossé Marie[4], Pelletier Dominique[5], Norvez Olivier[6], Reecht Yves[7,8], Fouilloux Anne[9], Rasche Helena[10], Hiltemann Saskia[11], Batut Bérénice[12,13], Eléaume Marc[14,15], Seguineau Pauline[14,15], Massé Guillaume[16], Amossé Alan[17], Bissery Claire[8,18], Lorrilliere Romain[3], Martin Alexis[19], Bas Yves[3,20], Virgoulay Thimothée[21,22], Chambon Valentin[17], Arnaud Elie[2], Michon Elisa[23], Urfer Clara[2,24], Trigodet Eloïse[21,24], Delannoy Marie[3], Loïs Gregoire[3], Julliard Romain[3], Grüning Björn[25], The Galaxy-E community, Le Bras Yvan[2]

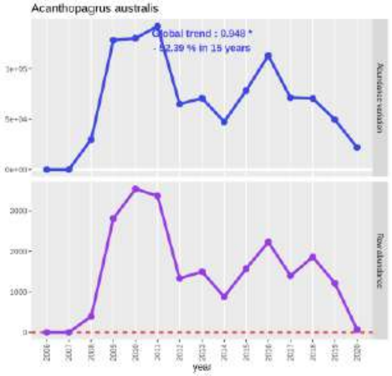


PAMPA

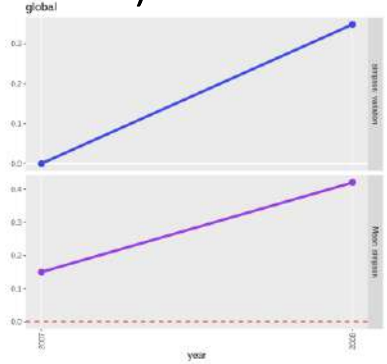

Graphical abstract – Levels of attainable best practices through the atomisation – generalisation framework

Keywords: Biodiversity; Reproducible analyses; Galaxy; Best practices; Atomisation; Generalisation; Workflows; Ecoinformatics; Conda; Container; Common Workflow Language; RO-CRATE

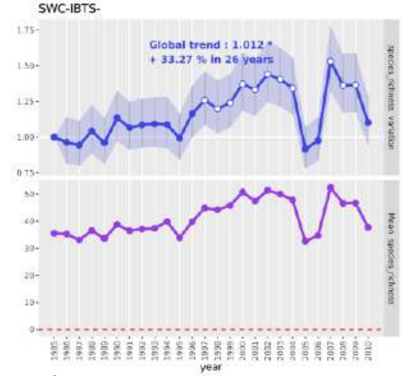
REEF LIFE SURVEY fishes data



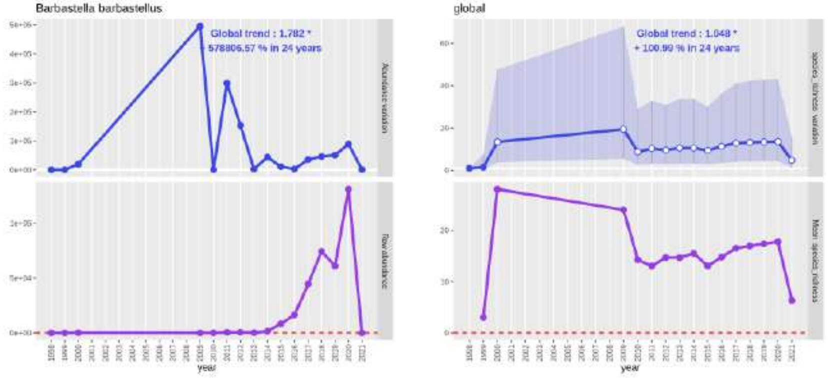
Marine Protecterd Areas (PAMPA) fishes data



DATRAS fishing data



VIGIE CHIRO bats data



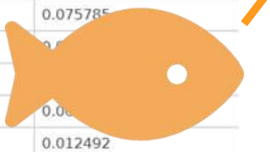
Applying “Galaxy-E” best practices to reuse!

Different data, same primary variables, same workflow

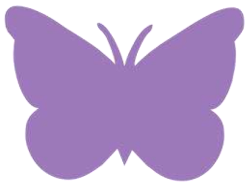
Location **Year** **Species** **Occurrence**

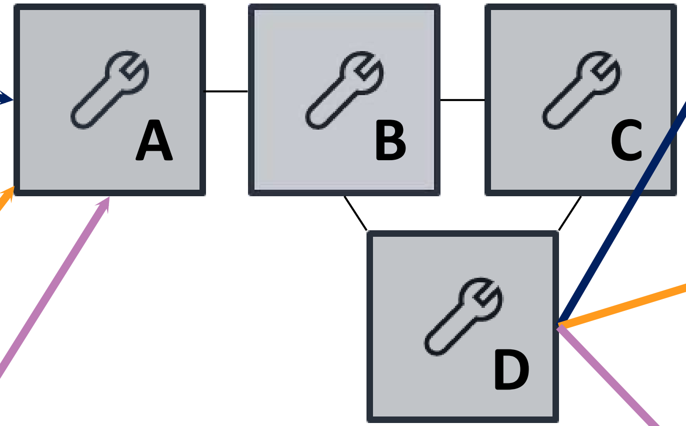
	carre	annee	espece	abond
1	2	2016	ACCGEN	0
2	2	2017	ACCGEN	0
3	2	2018	ACCGEN	0
4	2	2019	ACCGEN	0
5	183	2016	ACCGEN	0
6	183	2017	ACCGEN	0

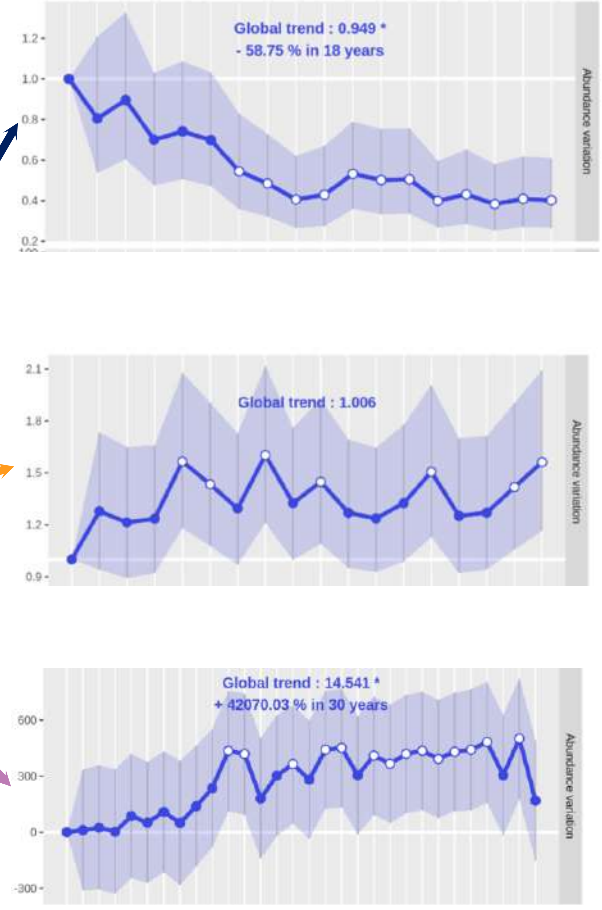
	Survey	Year	Quarter	Area	AphiaID	Species	LngtClass	CPUE_number_per_hour
1	BITS	1991	1	22	126281	Anguilla anguilla	0	0.000000
2	BITS	1991	1	22	126281	Anguilla anguilla	720	0.009160
3	BITS	1991	1	22	126417	Clupea harengus	0	0.000000
4	BITS	1991	1	22	126417	Clupea harengus	80	0.075785
5	BITS	1991	1	22	126417	Clupea harengus	85	[illegible]
6	BITS	1991	1	22	126417	Clupea harengus	95	[illegible]
7	BITS	1991	1	22	126417	Clupea harengus	100	0.0[illegible]
8	BITS	1991	1	22	126417	Clupea harengus	105	0.012492
9	BITS	1991	1	22	126417	Clupea harengus	110	0.618357



	UnitObs	rotation	codeSp	sexe	taille	classe_taille	poids	nb_ind
1	AS140155	3	Hemifasc	-999	-999	P	-999	1
2	AS140159	1	Nasosp.	-999	-999	P	-999	3
3	AS140159	3	Gompvari	-999	-999	P	-999	1
4	AS140160	3	Gompvari	-999	-999	P	-999	1







Applying “Galaxy-E” best practices to reuse!

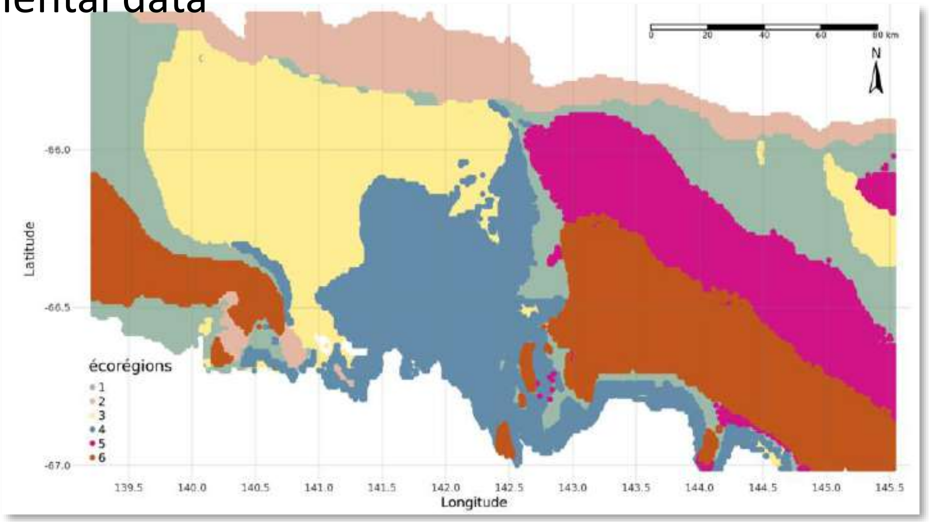
Different data, same primary variables, same workflow

Transpose to other taxonomic and/or geographical and/or temporal context

From Antarctica to North East Atlantic

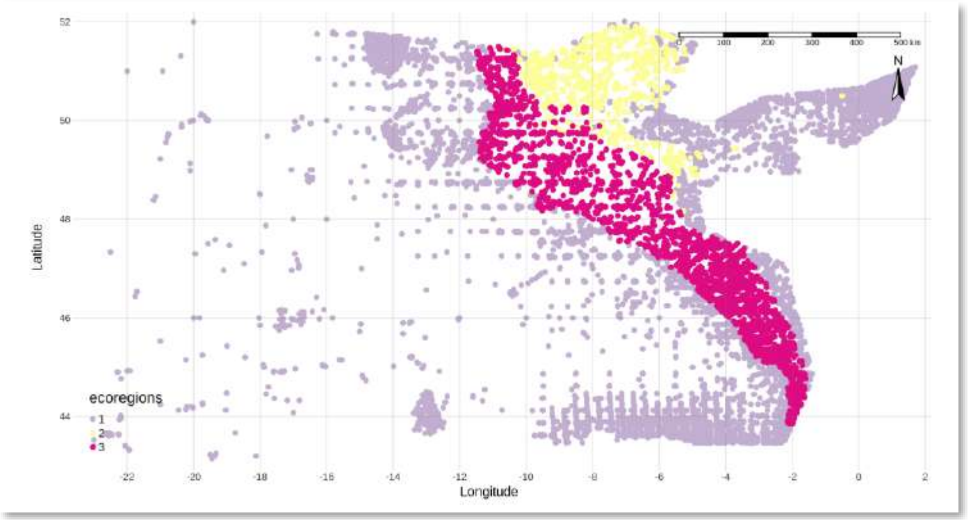
From Invertebrates to Fishes

Project data (invertebrates)
Project environmental data



North Atlantic
OBIS marine occurrence data (fishes)
OBIS marine environmental data (all)

c1: decimallongitude × c2: decimallatitude × c5: shoredistance × c6: bathymetry × c7: sst × c8: sss ×



Applying “Galaxy-E” best practices to reuse!

Different data, same primary variables, same workflow

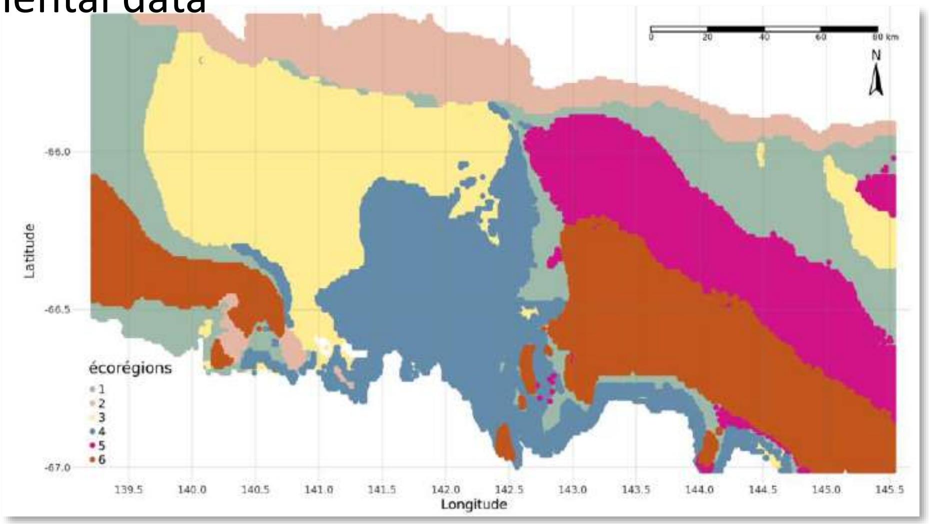
Test several scenarios

From Antarctica to North East Atlantic

From Invertebrates to Fishes

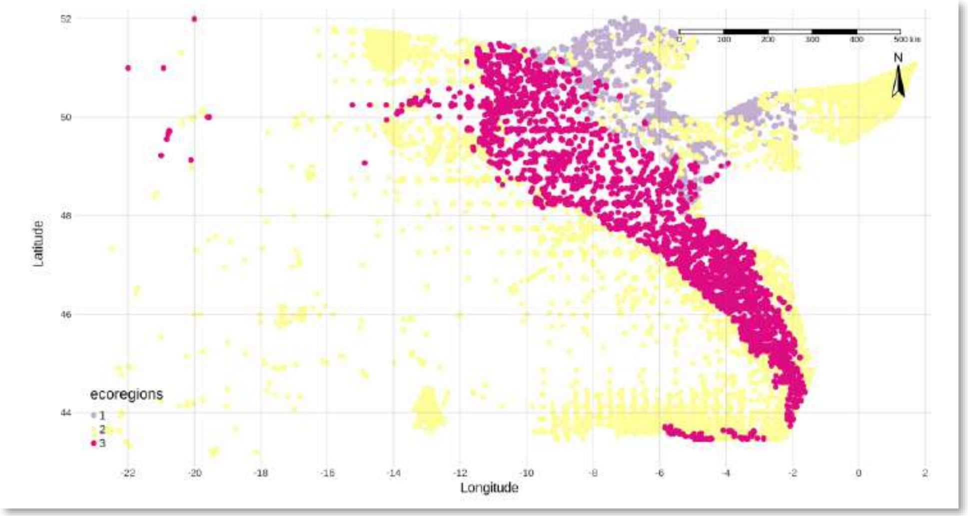
Modifying modelling parameters

Project data (invertebrates)
Project environmental data



North Atlantic
OBIS marine occurrence data (fishes)
OBIS marine environmental data (sst / bath

c1: long | c2: lat | c4: bathymetry | c5: sst | c6: sss



Training tutorials for Ecology

Galaxy Training Network

Data and Metadata Management

These tutorials are focusing on data and metadata management in Ecology.

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Cleaning GBIF data using OpenRefine biodiversity					
Creating FAIR Quality assessment reports and draft of Data Papers from EML metadata with MetaShRIMPS Metadata · EML · FAIR · Data Paper · biodiversity					
Creating metadata using Ecological Metadata Language (EML) standard with EML Assembly Line functionalities Metadata · EML · Ecology · Biodiversity · FAIR · Data Paper					
Data submission using ENA upload Tool					

Data access

These lessons focus on ways to access data classicaly used in Ecology

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
QGIS Web Feature Services earth-system · GIS · Geographical Information System · WFS · Spatial data · Maps · OGC · biodiversity					

Data preprocessing

These lessons focus on manners to preprocess data used in Ecology

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Biodiversity data exploration taxonomic data · data quality · biodiversity					
Checking expected species and contamination in bacterial isolate illumina · bacteria · microgalaxy					
Cleaning GBIF data for the use in Ecology gbif · data management · data cleaning · biodiversity					

Data visualization

These tutorials showcase data visualization in Ecology

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Visualization of Climate Data using NetCDF xarray Map Plotting pangeo					
Visualize EBV cube data with Panoply netCDF viewer interactive-tools · EBV cube · Data visualization					

Data analysis

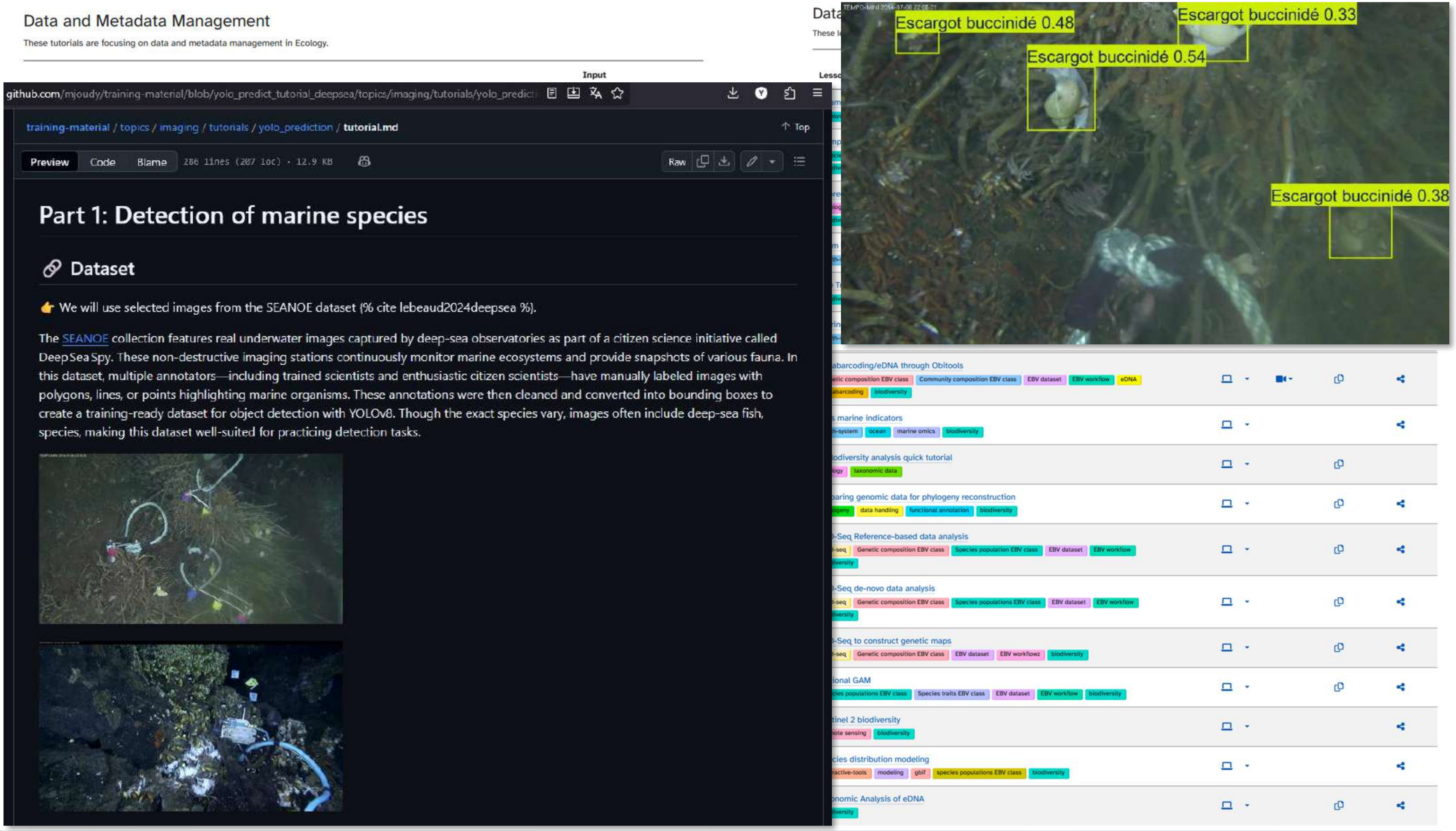
These lessons focus on ways to analyse data in Ecology

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Champs blocs indicators Ecosystem EBV class · EBV dataset · EBV workflow · Marine ecosystems · biodiversity					
Compute and analyze biodiversity metrics with PAMPA toolsuite Species population EBV class · Community composition EBV class · EBV dataset · EBV workflow · modeling · biodiversity					
Ecoregionalization workflow tutorial ecology · taxonomic data · EBV workflow · modeling · gbif · ocean · earth-system · interactive-tools · biodiversity					
From NDVI data with OpenEO to time series visualisation with Holoviews earth-system · land degradation · NDVI · copernicus · holoviews					
Life Traits Ecoregionalization workflow biodiversity					
Marine Omics identifying biosynthetic gene clusters earth-system · ocean · marine omics · biodiversity					
Metabarcoding/eDNA through Obitools Genetic composition EBV class · Community composition EBV class · EBV dataset · EBV workflow · eDNA · Metabarcoding · biodiversity					
Obis marine indicators earth-system · ocean · marine omics · biodiversity					
Phylodiversity analysis quick tutorial ecology · taxonomic data					
Preparing genomic data for phylogeny reconstruction phylogeny · data handling · functional annotation · biodiversity					
RAD-Seq Reference-based data analysis RAD-seq · Genetic composition EBV class · Species population EBV class · EBV dataset · EBV workflow · biodiversity					
RAD-Seq de-novo data analysis RAD-seq · Genetic composition EBV class · Species populations EBV class · EBV dataset · EBV workflow · biodiversity					
RAD-Seq to construct genetic maps RAD-seq · Genetic composition EBV class · EBV dataset · EBV workflowz · biodiversity					
Regional GAM Species populations EBV class · Species traits EBV class · EBV dataset · EBV workflow · biodiversity					
Sentinel 2 biodiversity Remote sensing · biodiversity					
Species distribution modeling interactive-tools · modeling · gbif · species populations EBV class · biodiversity					
Taxonomic Analysis of eDNA biodiversity					

Training tutorials for Ecology

New tutorial using AI model (*in review*)

[Link to the PR]



Galaxy Ecology: From citizen science to AI

Citizen science & crowdsourcing

HORIZON 2020

SPIPOLL

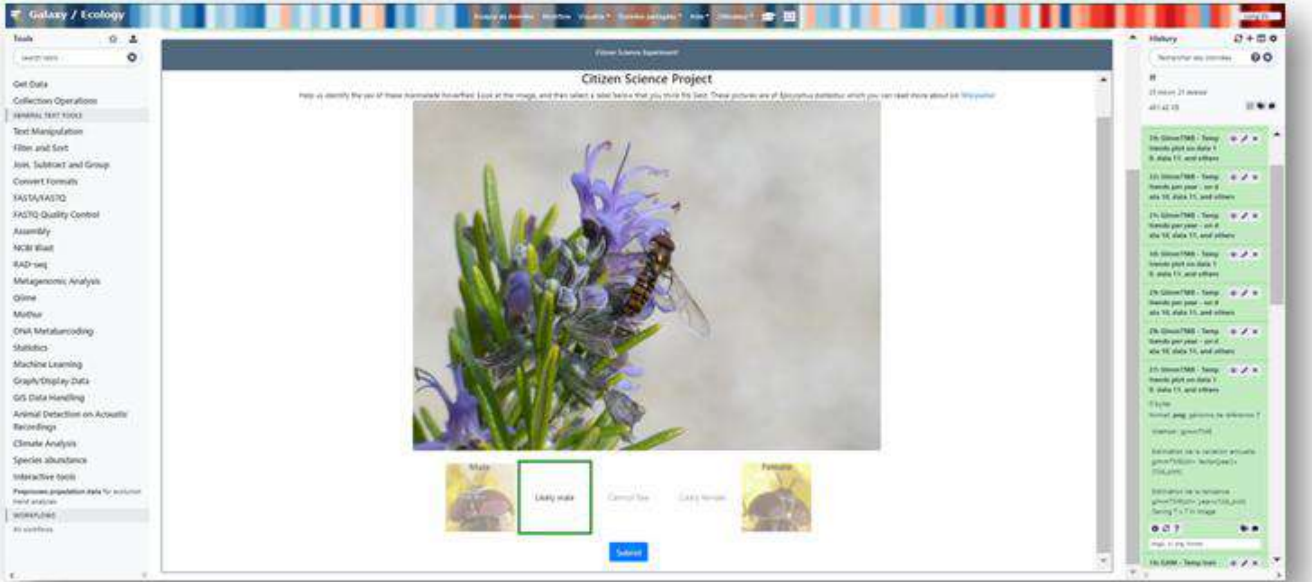
VIGIENATURE

Un réseau de citoyens qui fait avancer la science

HORIZON 2020

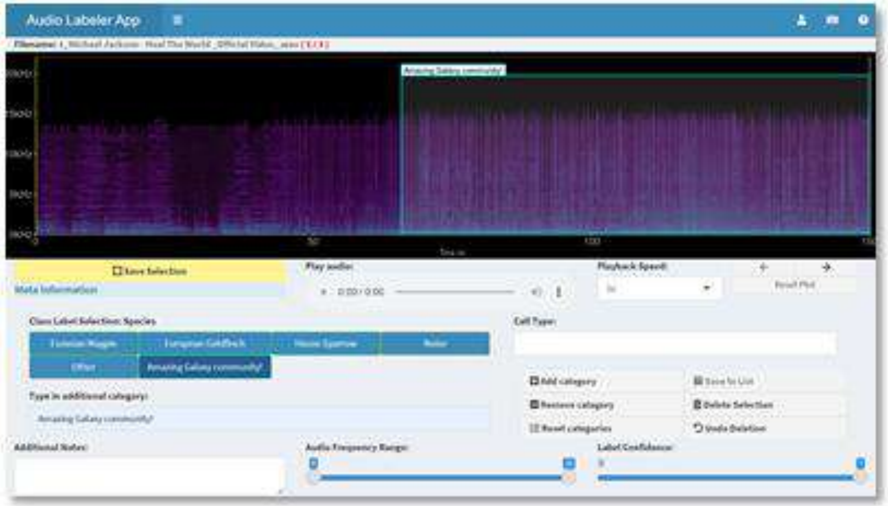
Gamification of participatory science for training and education purposes

Crowdsourcing image



>3400 classifications per month / >110 per day

Annotation of sounds



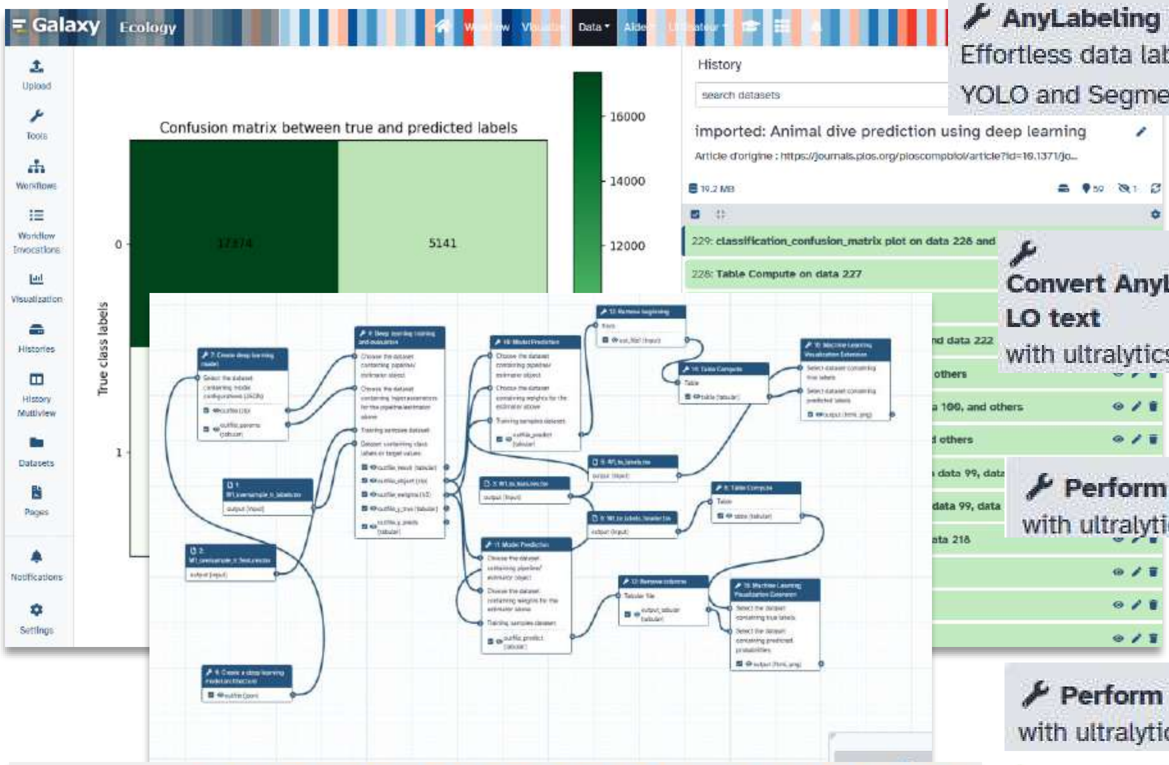
Annotation of images

MOOREV

SCIENCES SORBONNE UNIVERSITÉ

Annotations & AI

Reproduce analyses from research article ***Deep learning from creating the model to training and evaluation***




PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

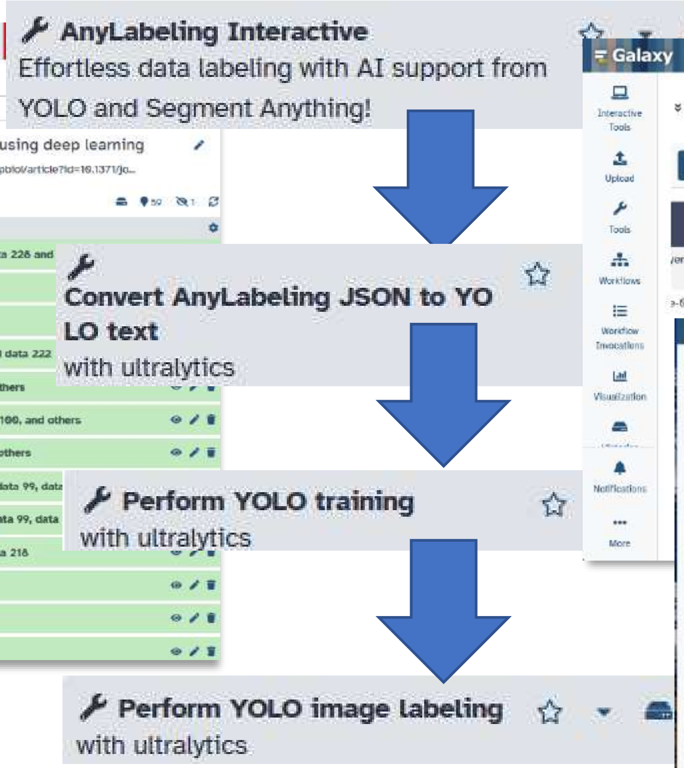
Deep inference of seabird dives from GPS-only records: Performance and generalization properties

Amédée Roy, Sophie Lanco Bertrand, Ronan Fablet

Version 2 Published: March 11, 2022 • https://doi.org/10.1371/journal.pcbi.1009890

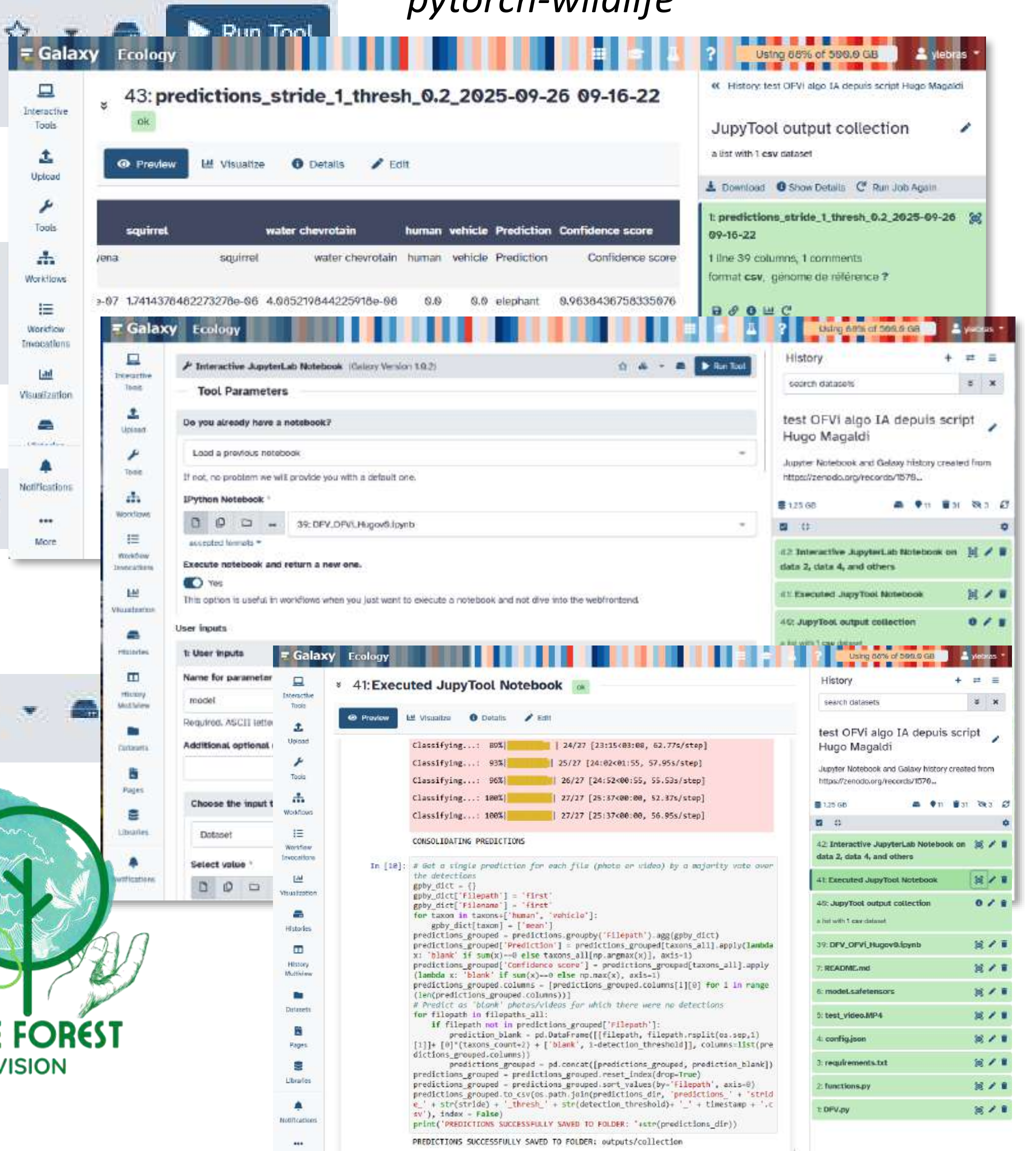
Adapt existing tools to use cases

From Interactive annotation to training to labeling



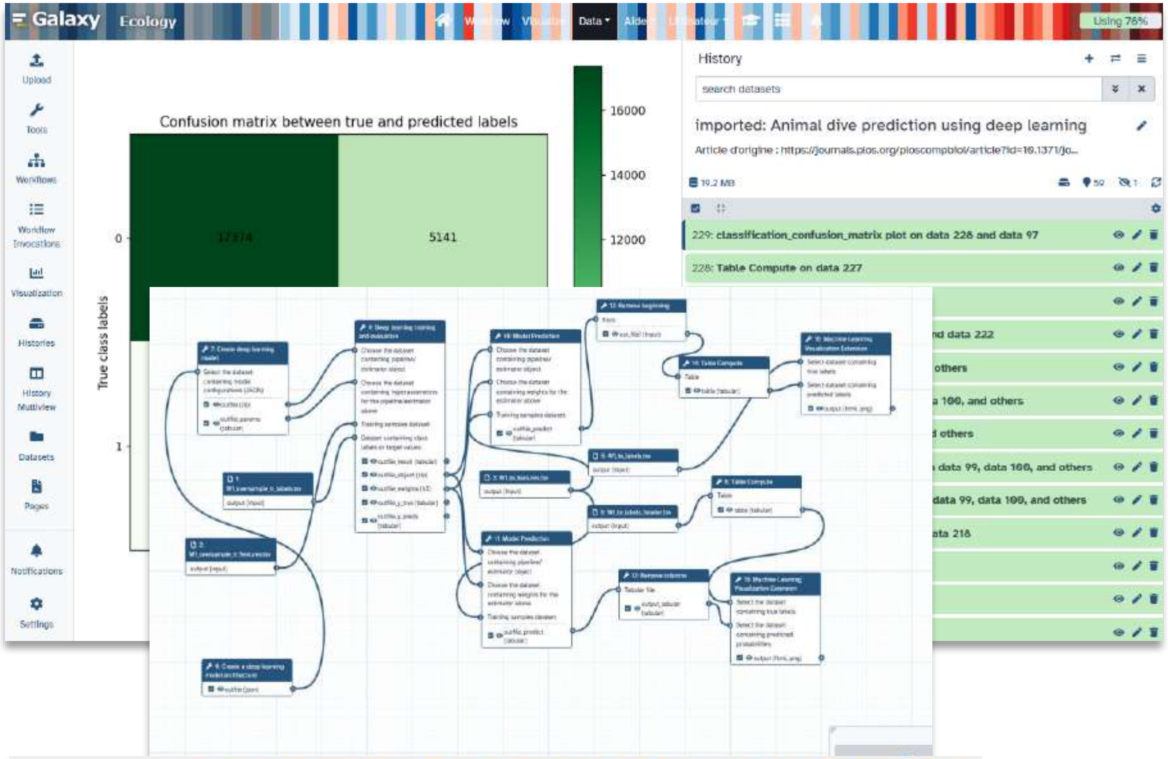
Using Jupyter notebook

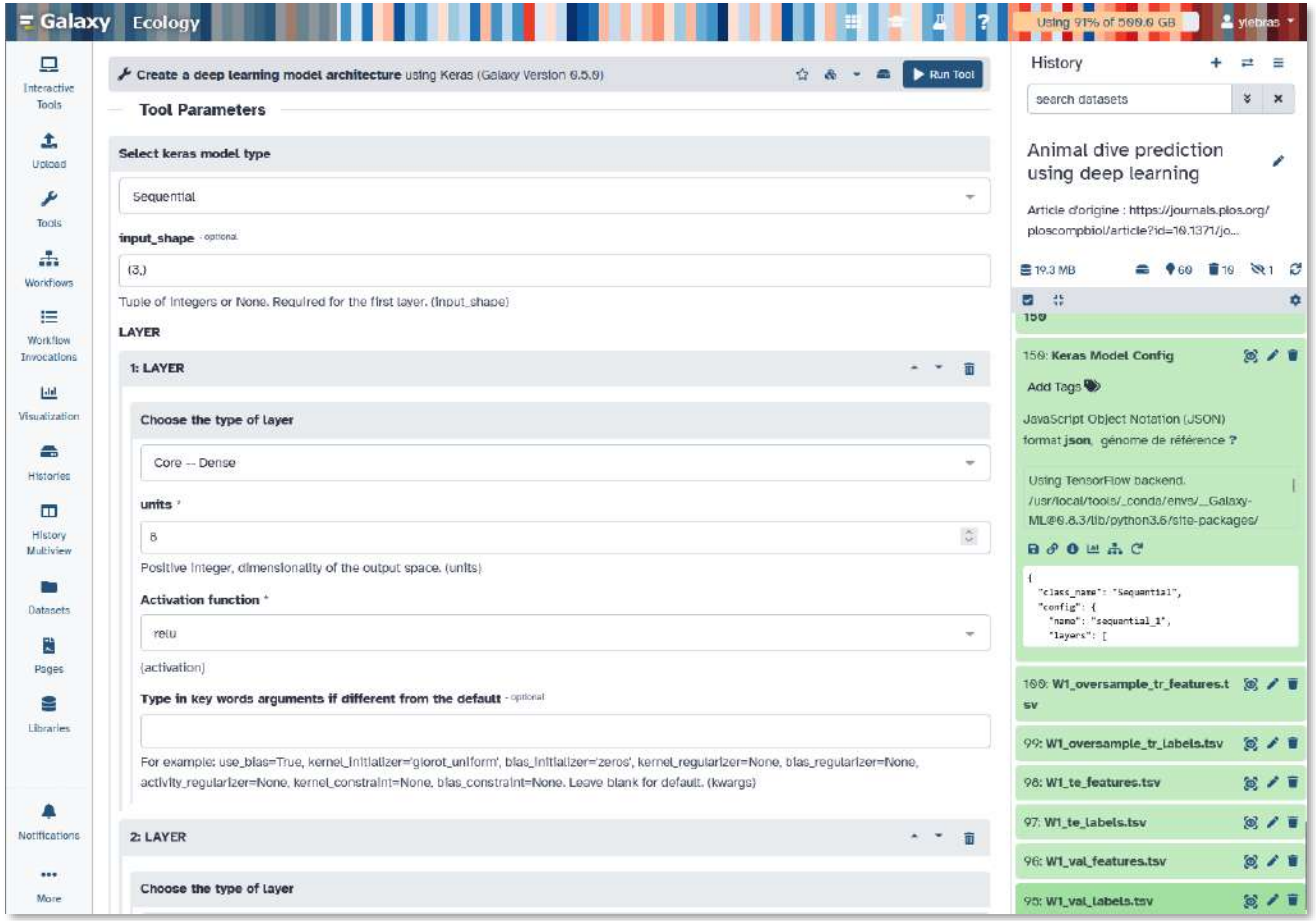
From camera trap video to species detection with megadetector & pytorch-wildlife



Annotations & AI

Reproduce analyses from research article *Deep learning from creating the model to training and evaluation*





PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

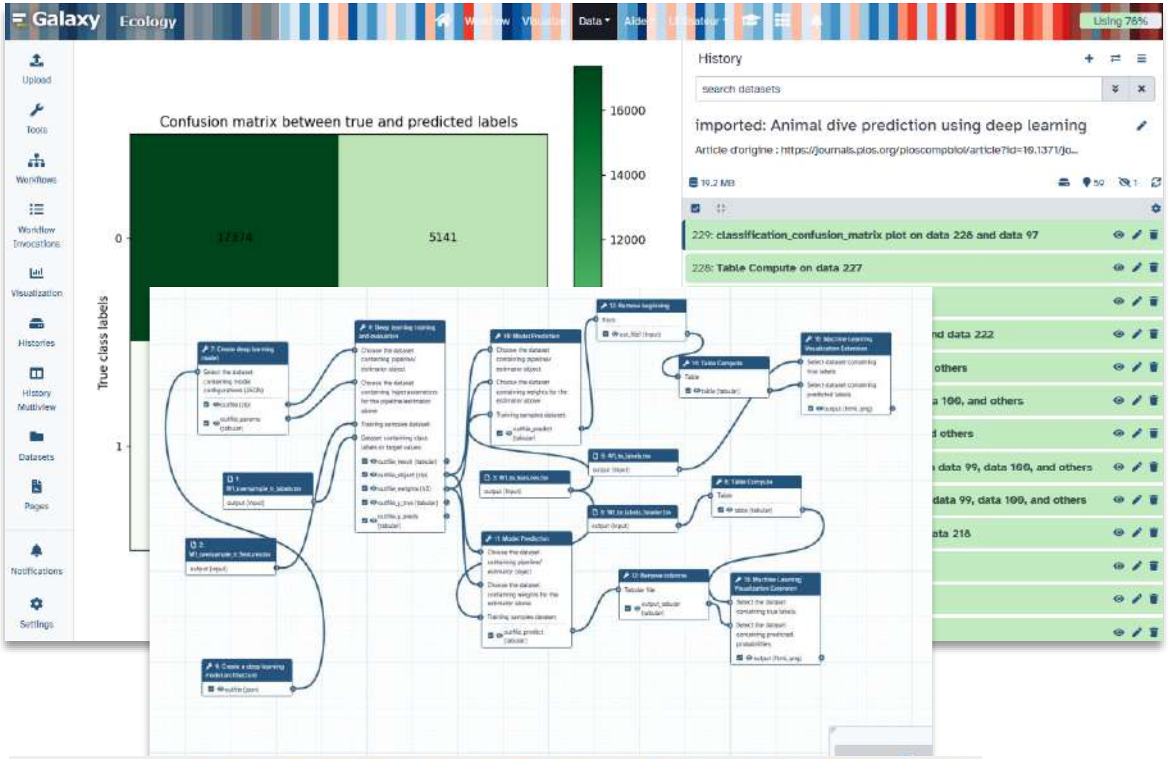
Deep inference of seabird dives from GPS-only records: Performance and generalization properties

Amédée Roy, Sophie Lanco Bertrand, Ronan Fablet

Version 2 Published: March 11, 2022 • https://doi.org/10.1371/journal.pcbi.1009890

Annotations & AI

Reproduce analyses from research article *Deep learning from creating the model to training and evaluation*




PLOS COMPUTATIONAL BIOLOGY

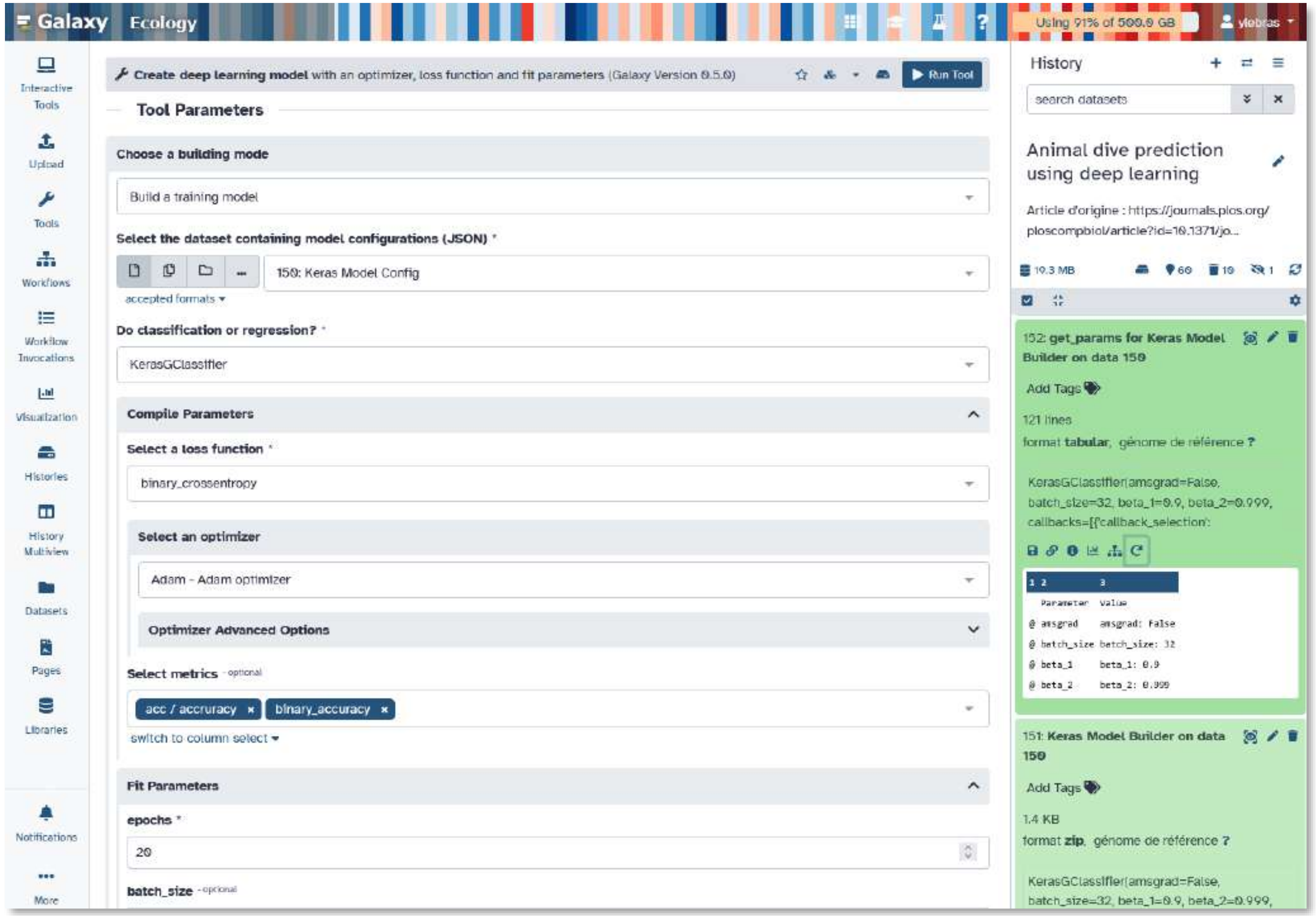
OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Deep inference of seabird dives from GPS-only records: Performance and generalization properties

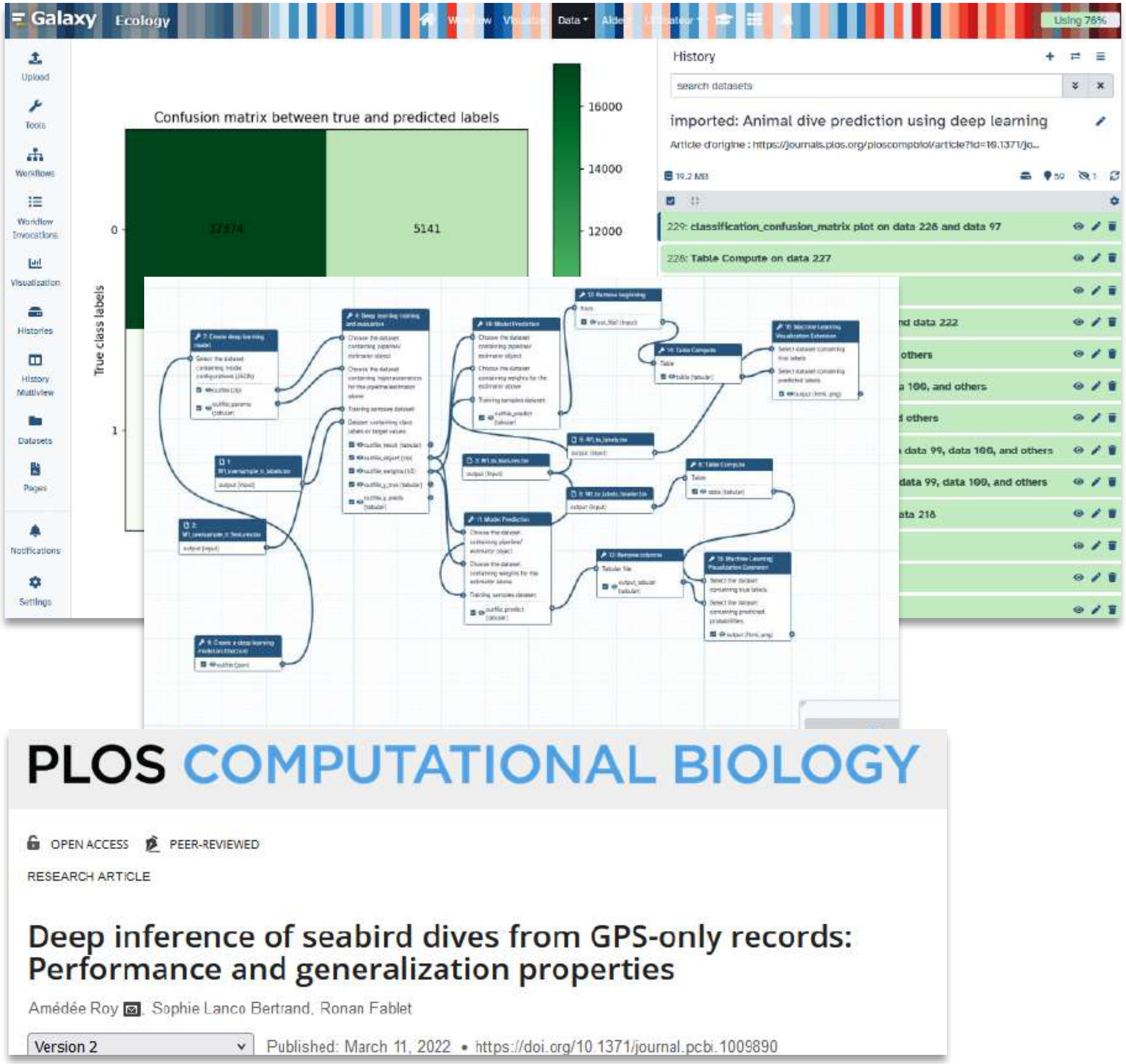
Amédée Roy, Sophie Lanco Bertrand, Ronan Fablet

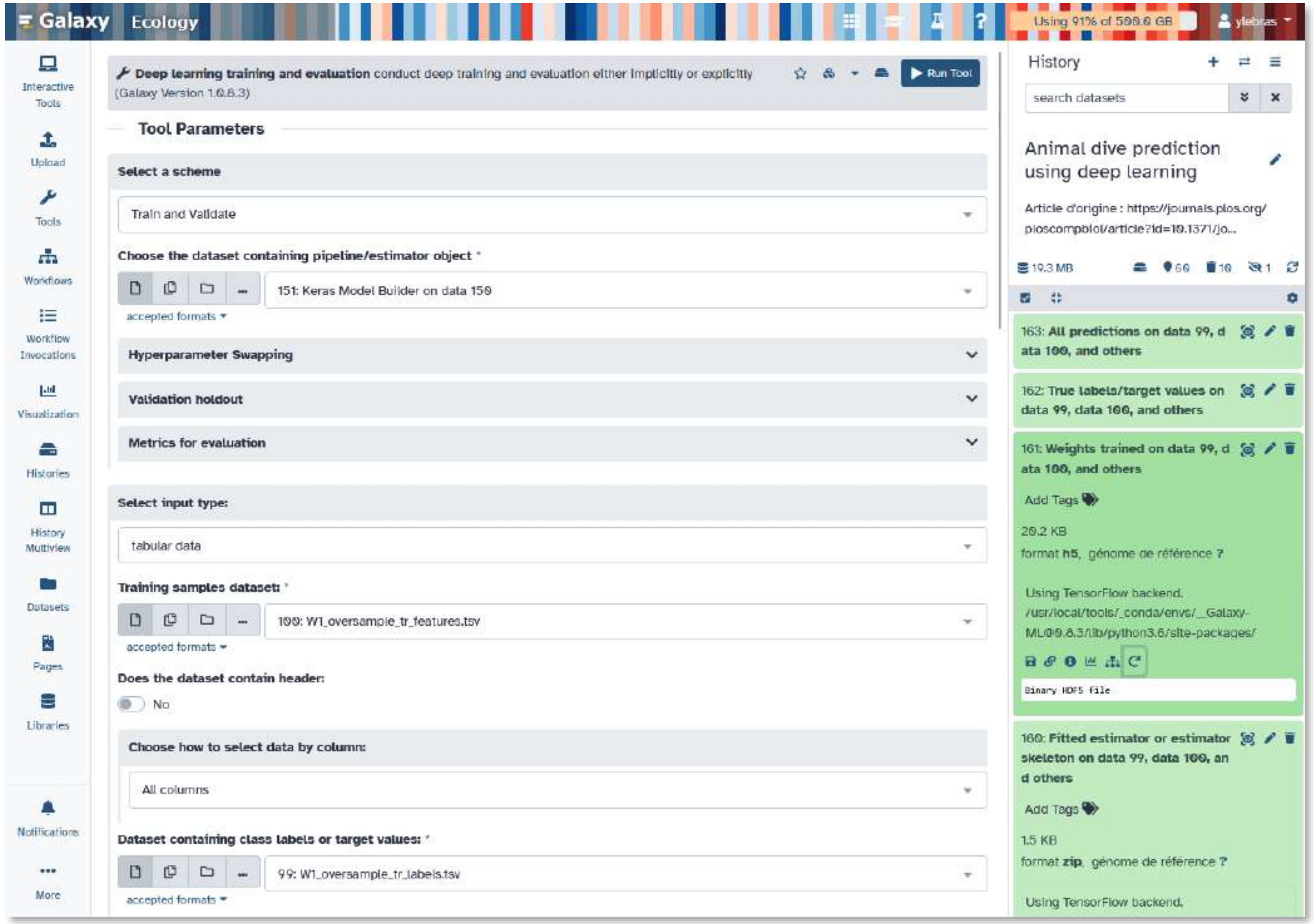
Version 2 | Published: March 11, 2022 • https://doi.org/10.1371/journal.pcbi.1009890



Annotations & AI

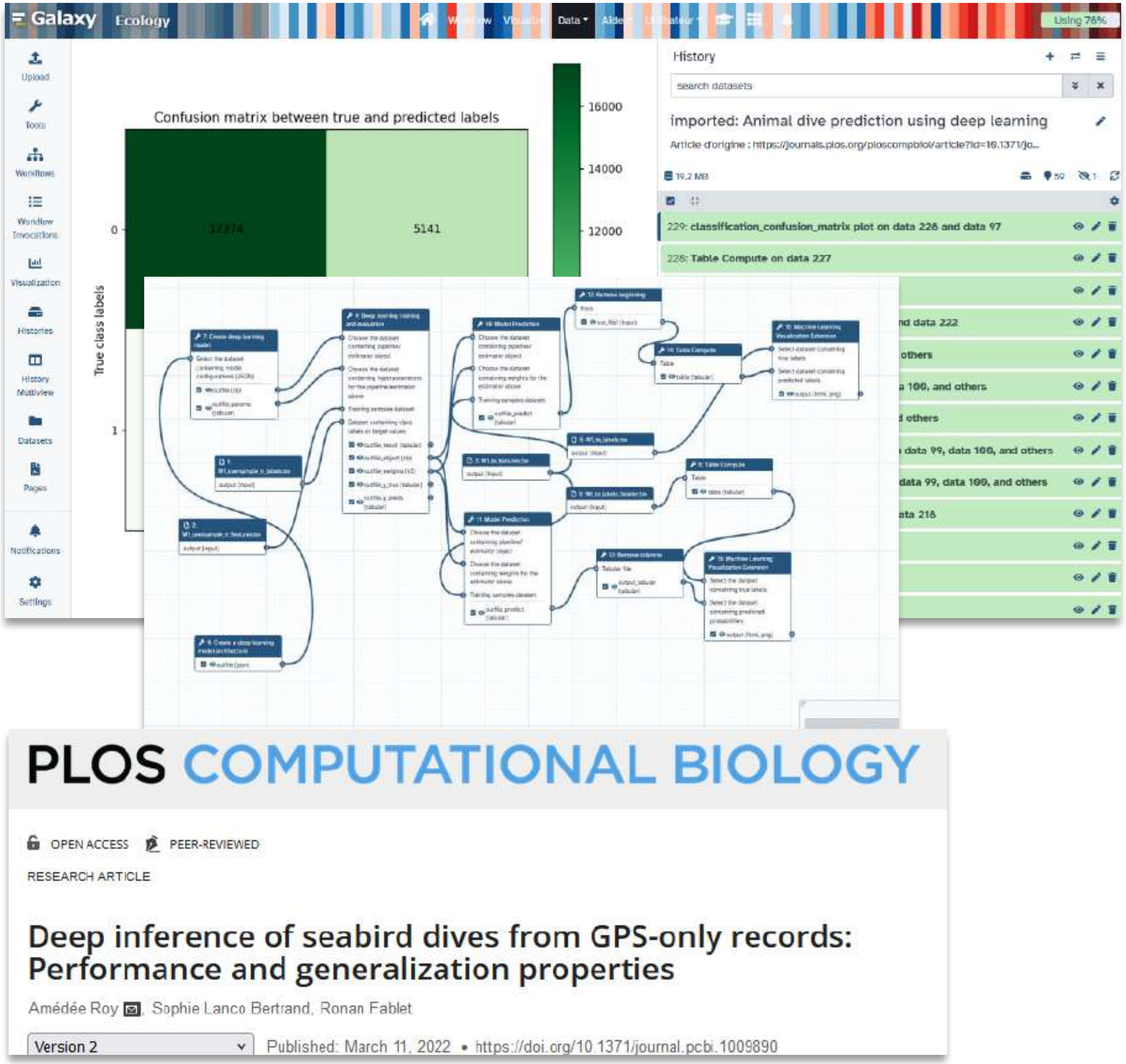
Reproduce analyses from research article ***Deep learning from creating the model to training and evaluation***

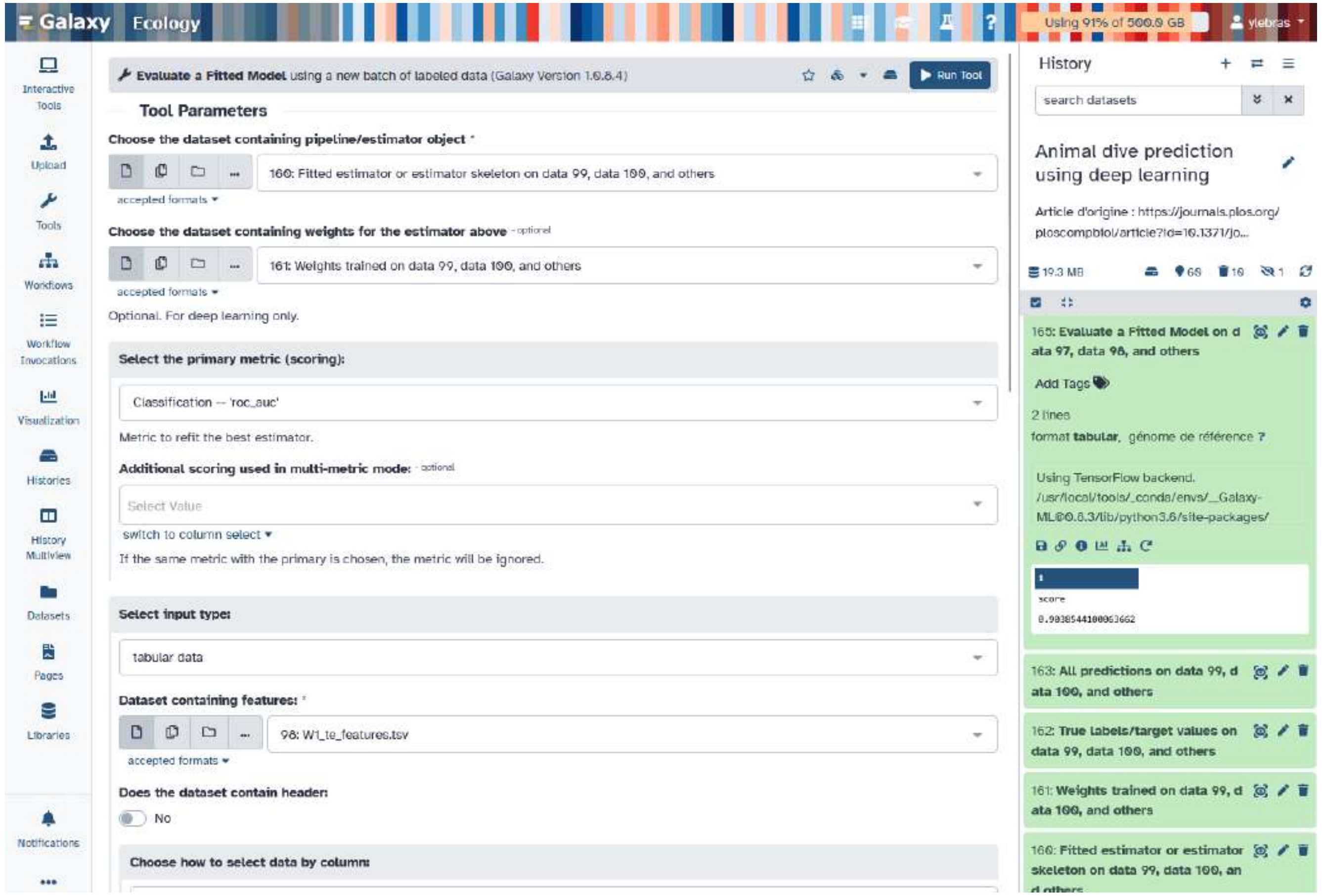




Annotations & AI

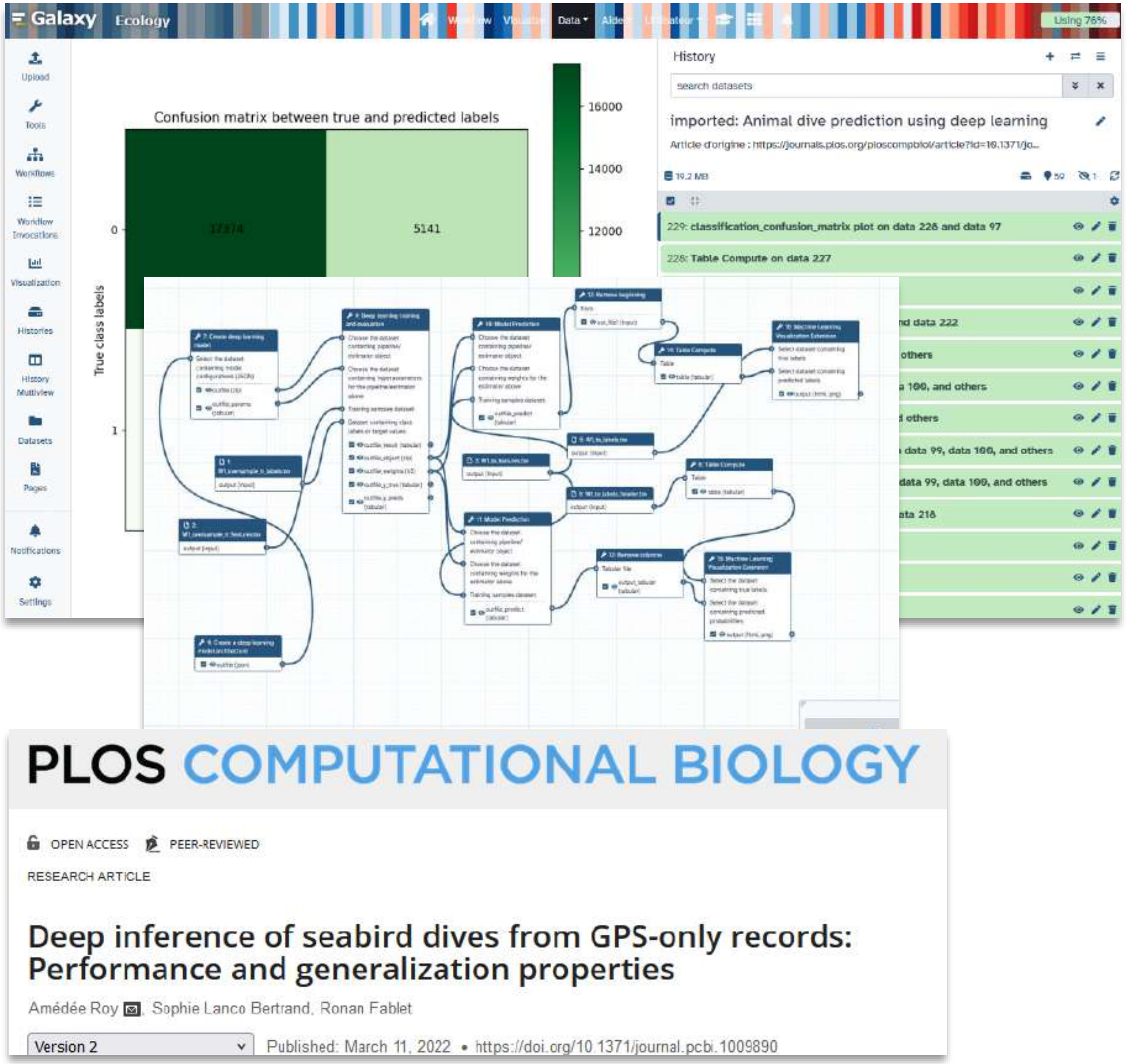
Reproduce analyses from research article *Deep learning from creating the model to training and evaluation*

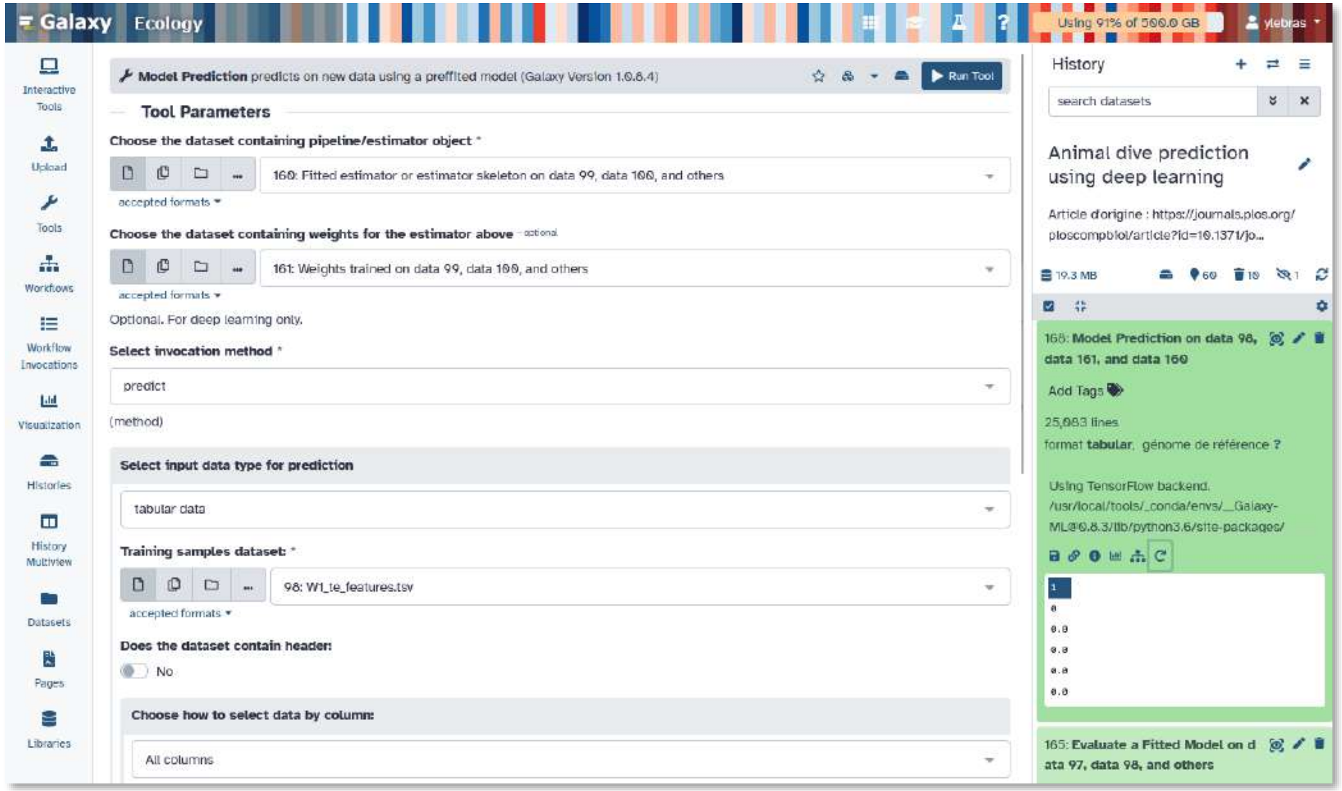




Annotations & AI

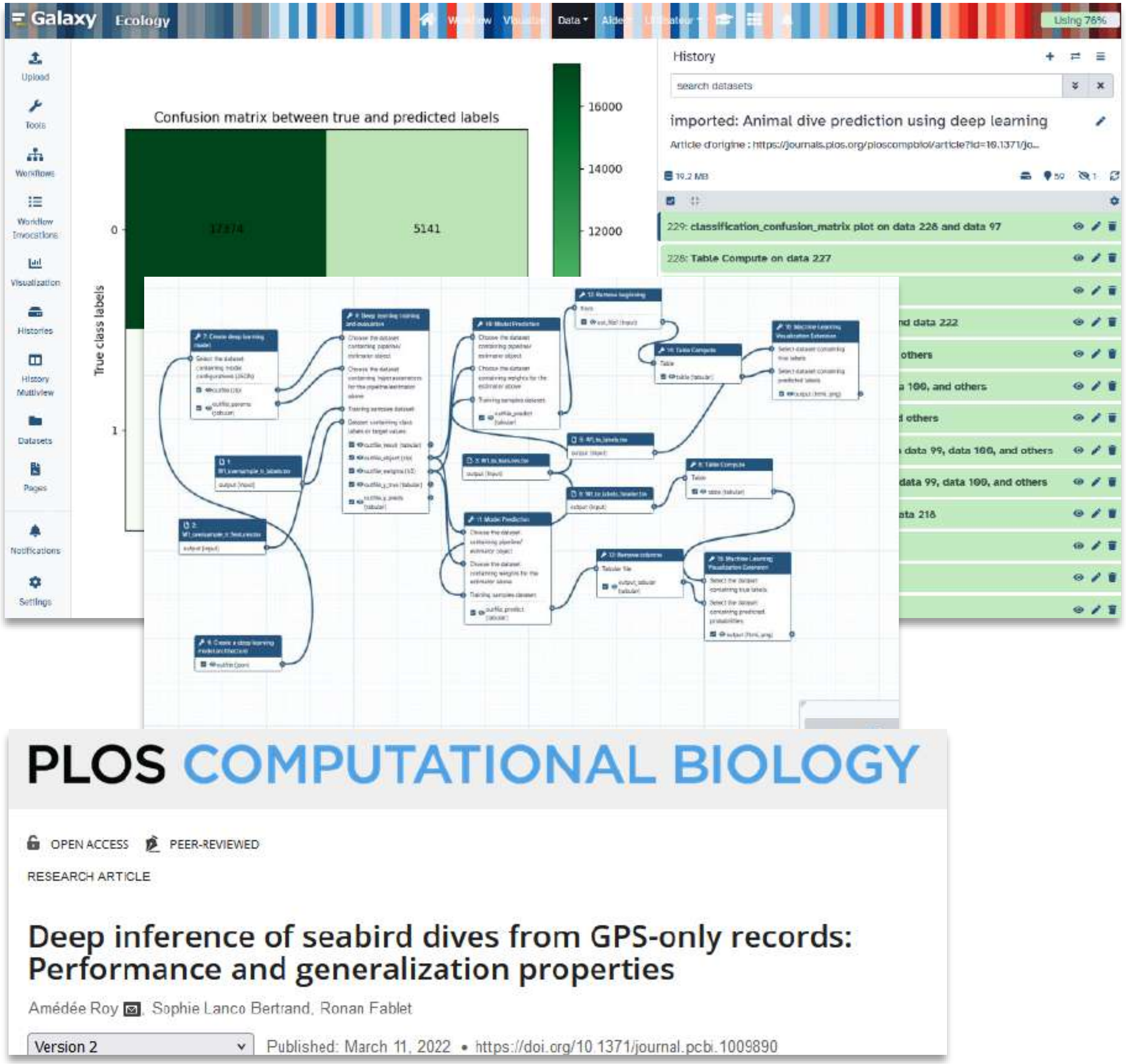
Reproduce analyses from research article *Deep learning from creating the model to training and evaluation*

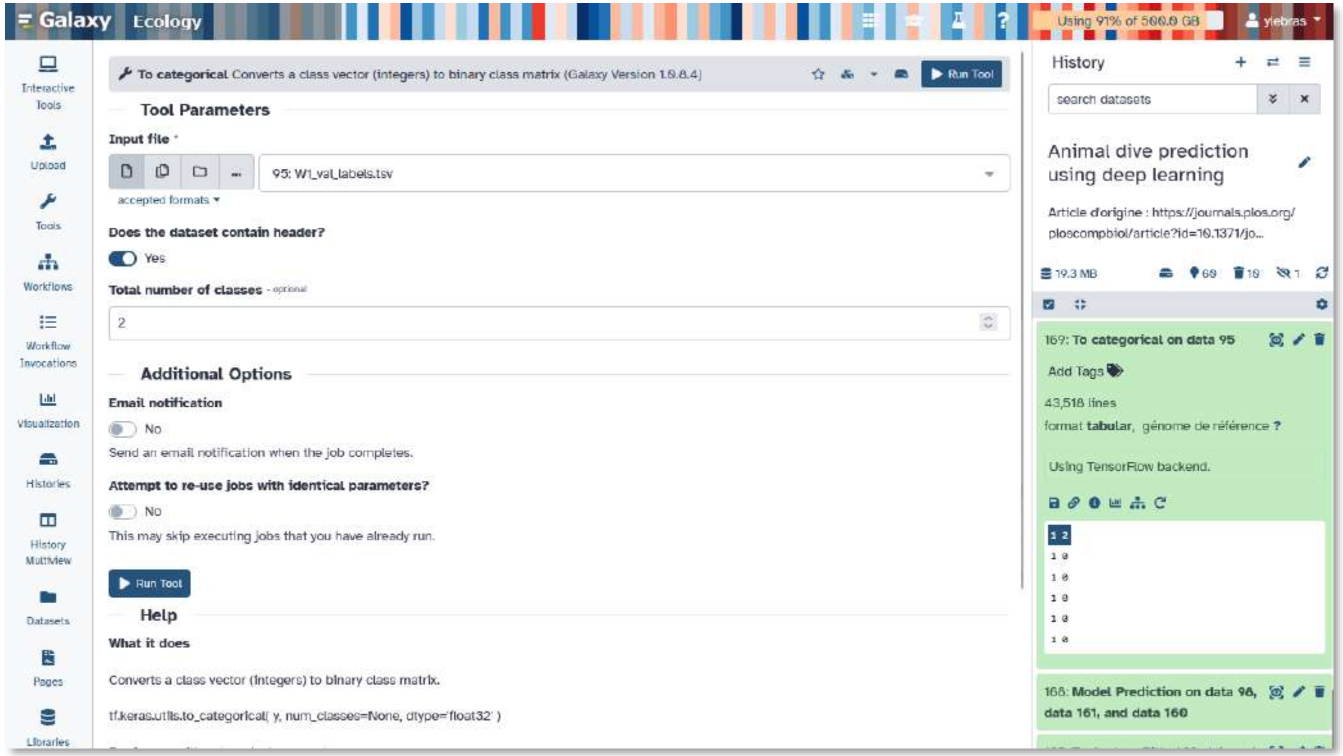




Annotations & AI

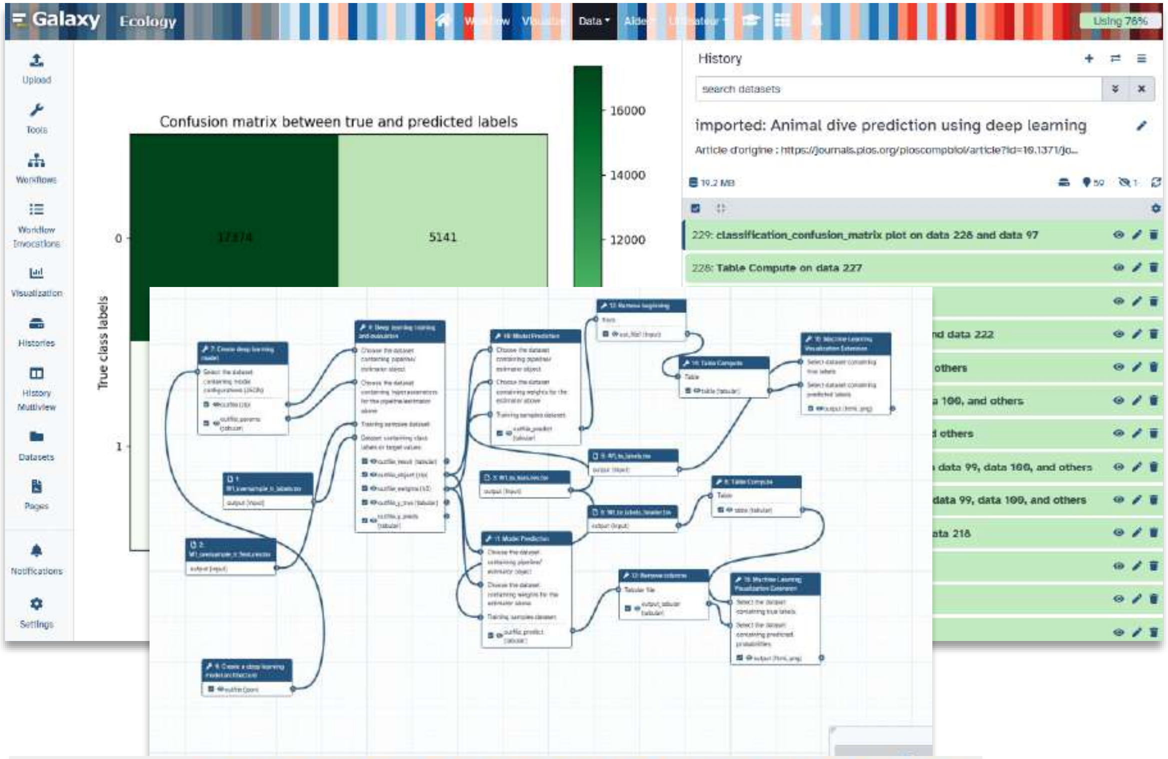
Reproduce analyses from research article *Deep learning from creating the model to training and evaluation*





Annotations & AI

Reproduce analyses from research article *Deep learning from creating the model to training and evaluation*




PLOS COMPUTATIONAL BIOLOGY

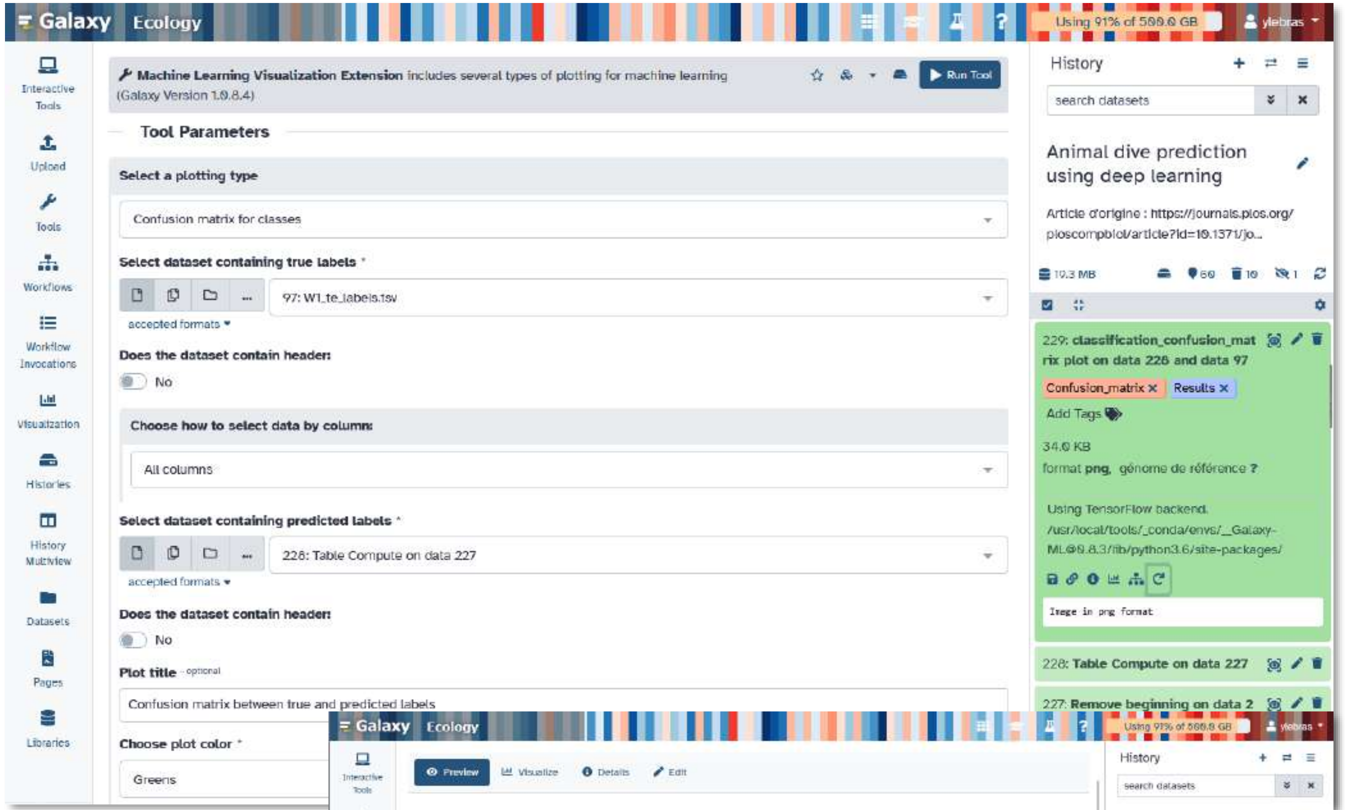
OPEN ACCESS PEER-REVIEWED

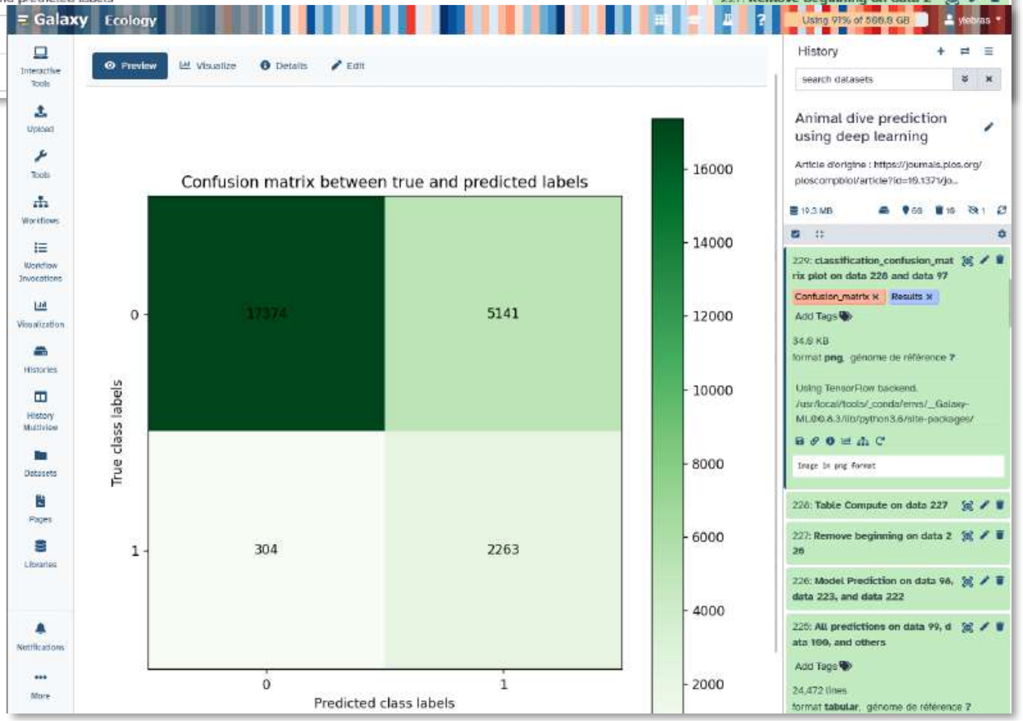
RESEARCH ARTICLE

Deep inference of seabird dives from GPS-only records: Performance and generalization properties

Amédée Roy, Sophie Lanco Bertrand, Ronan Fablet

Version 2 Published: March 11, 2022 • https://doi.org/10.1371/journal.pcbi.1009890

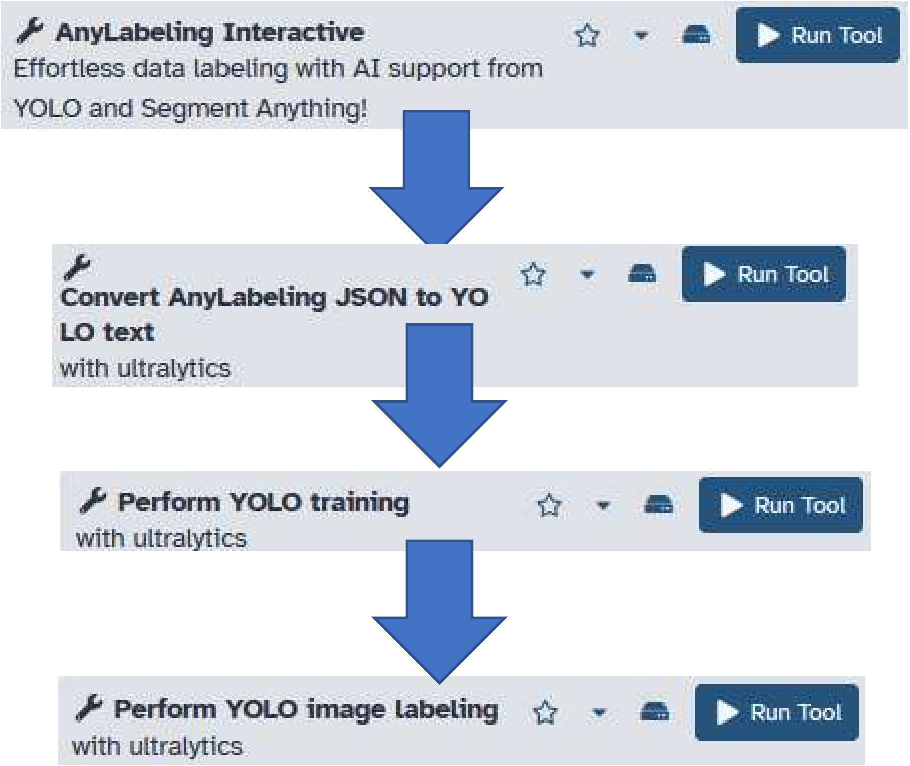




Annotations & AI

Adapt existing tools to use cases

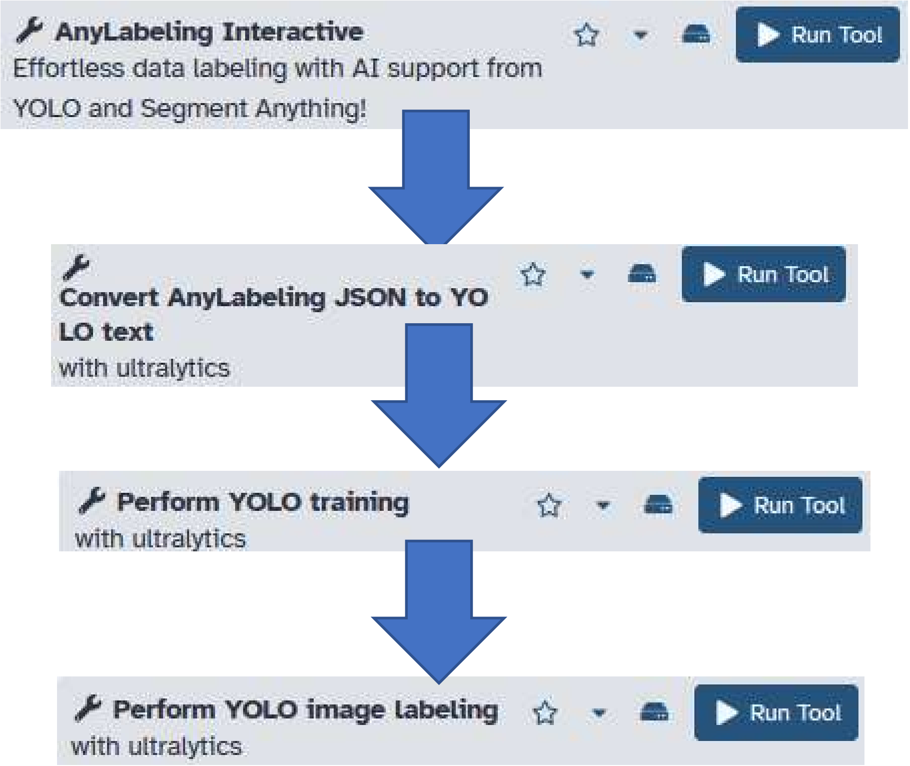
From Interactive annotation to training to labeling



Annotations & AI

Adapt existing tools to use cases

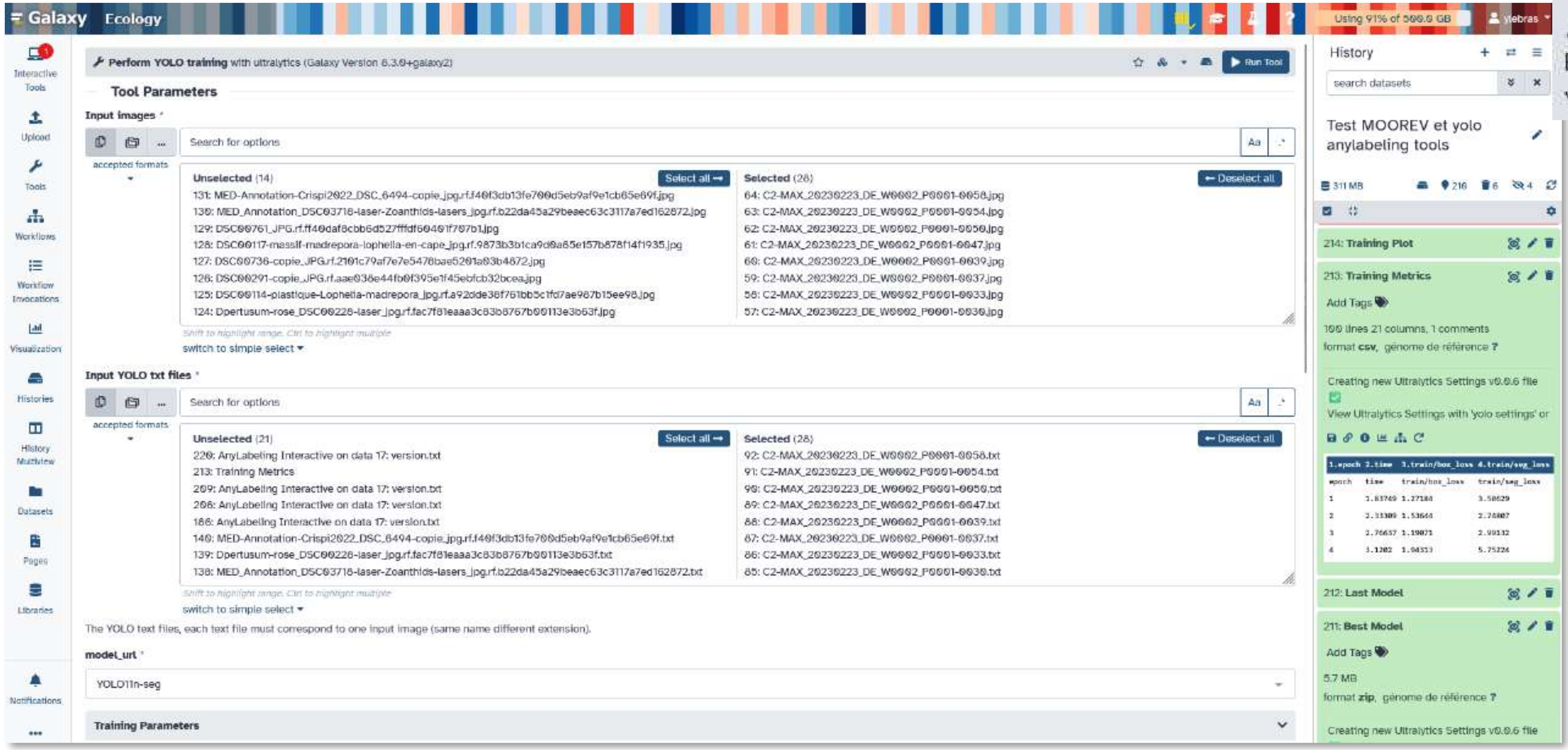
From Interactive annotation to training to labeling



Annotations & AI

Adapt existing tools to use cases

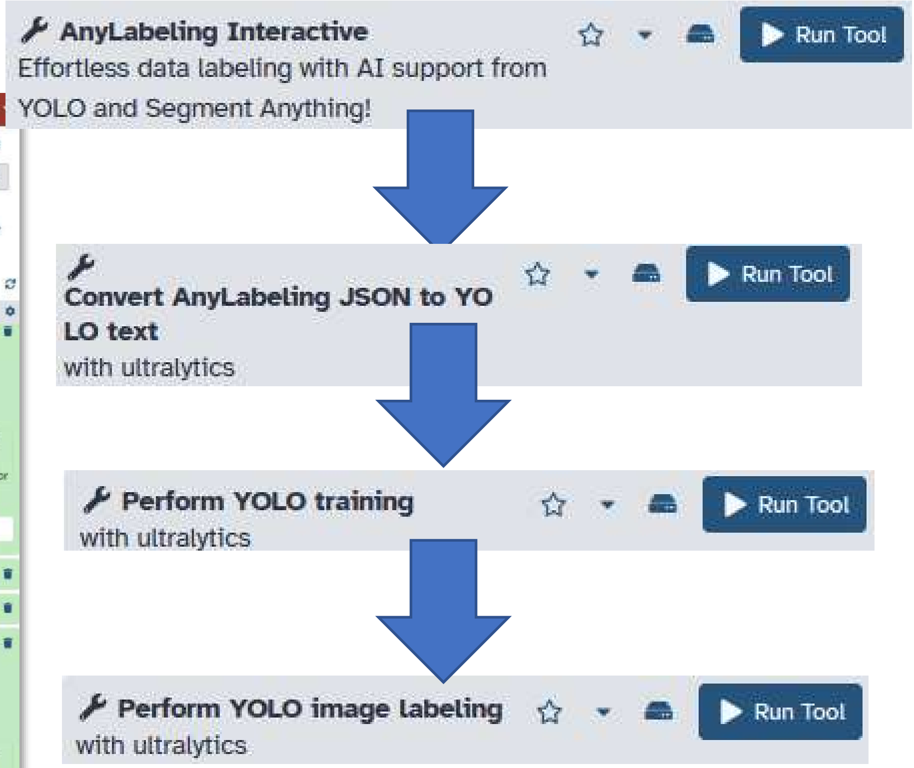
From Interactive annotation to training to labeling

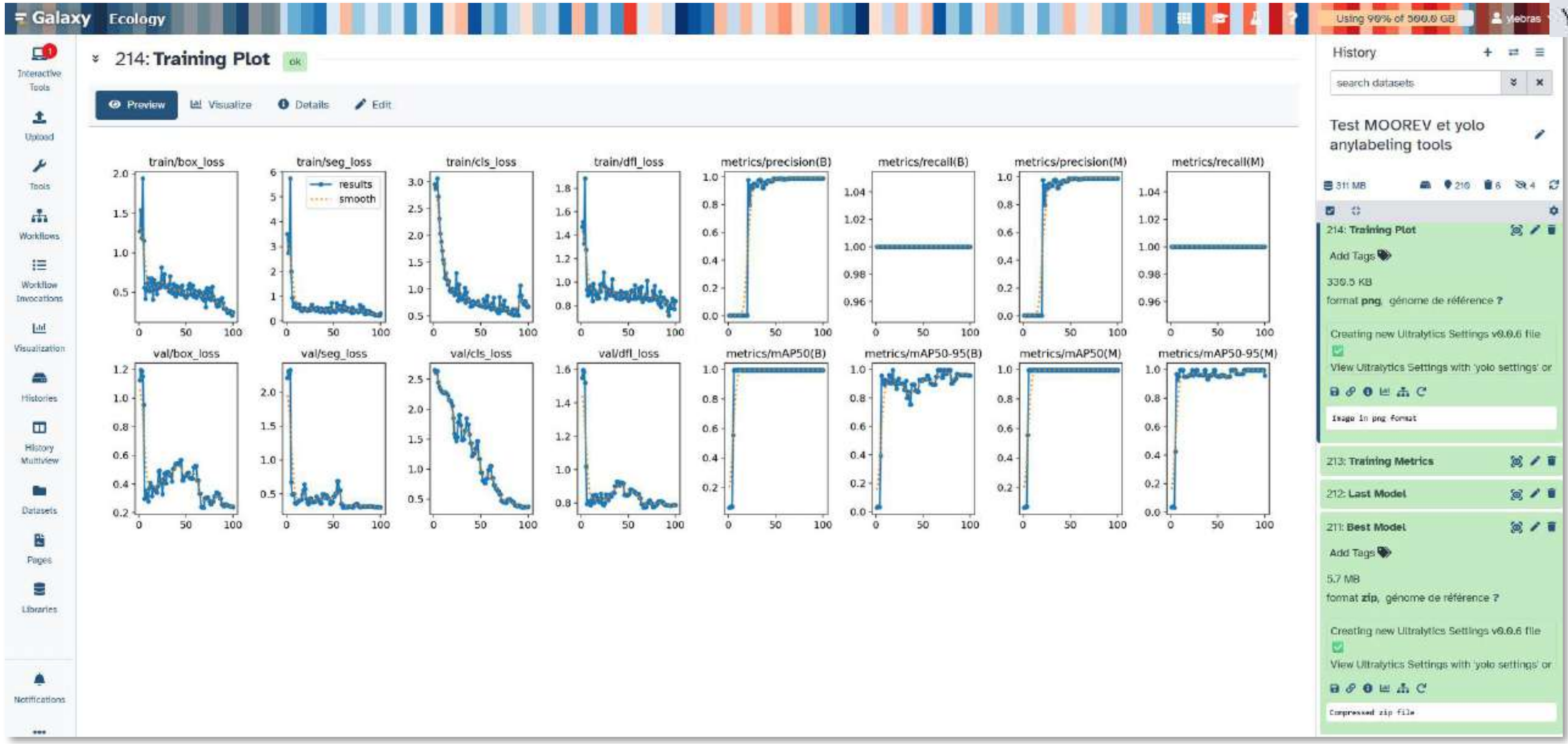


Annotations & AI

Adapt existing tools to use cases

From Interactive annotation to training to labeling





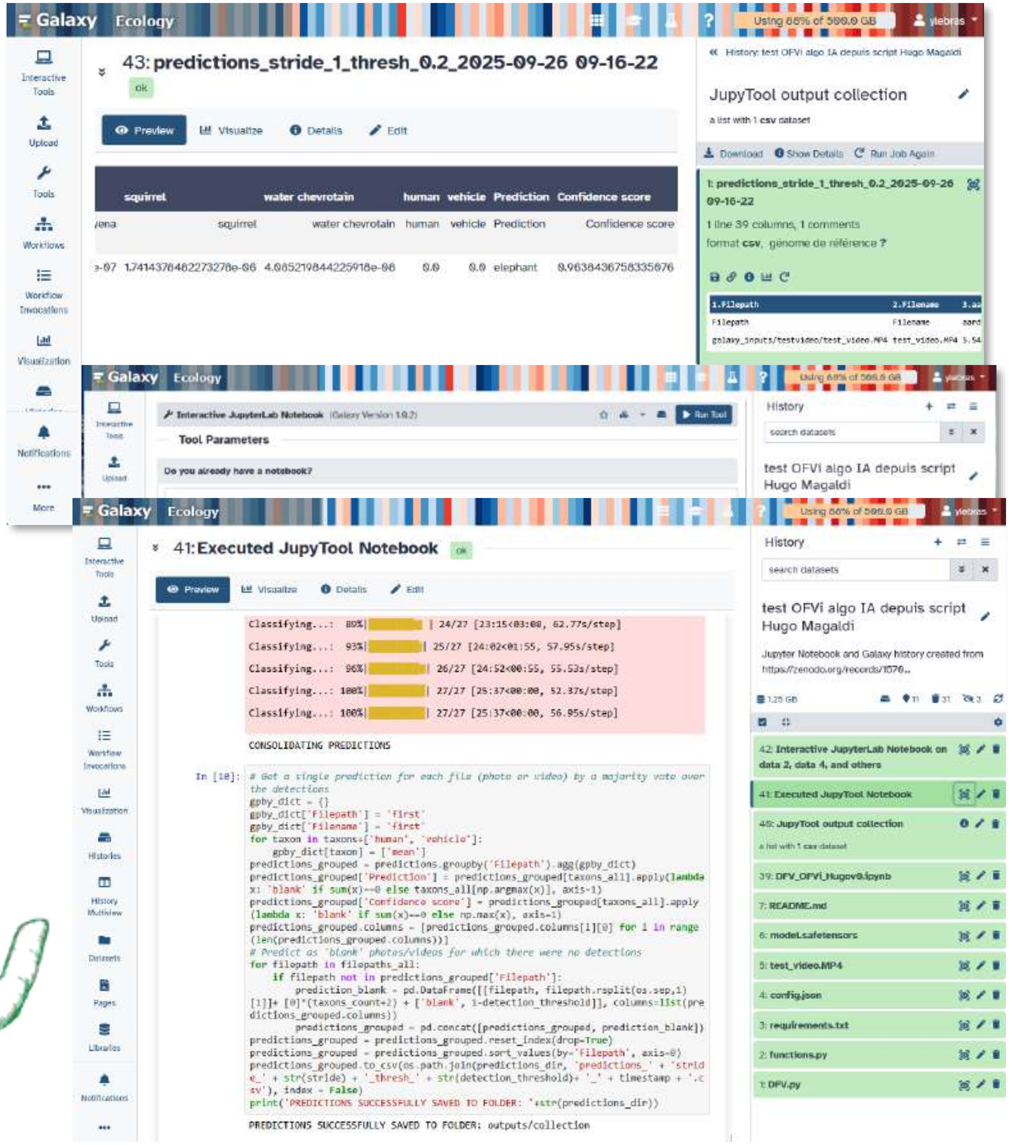
MOOREV

SCIENCES SORBONNE UNIVERSITÉ

Annotations & AI

Using Jupyter notebook

From camera trap video to species detection with megadetector & pytorch-wildlife

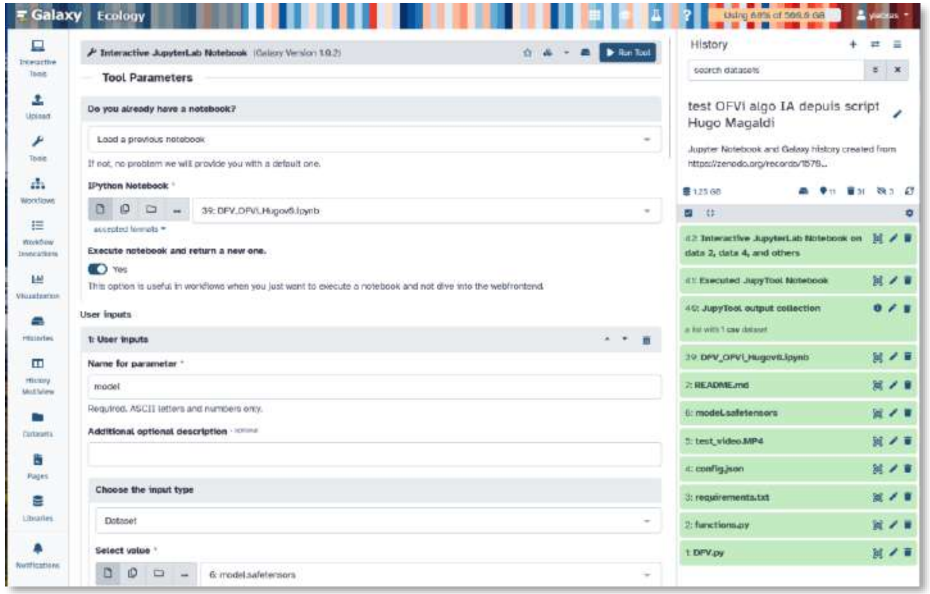


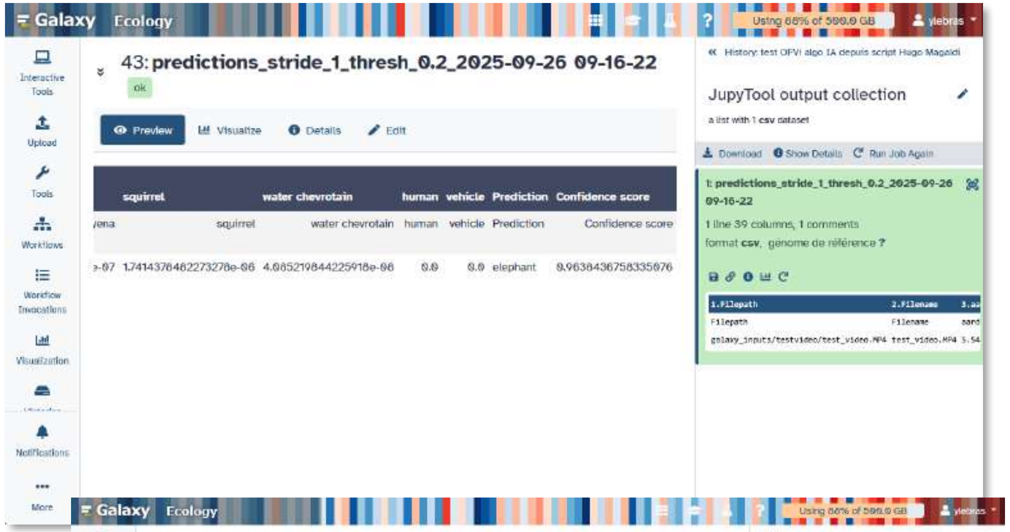
ONE FOREST VISION

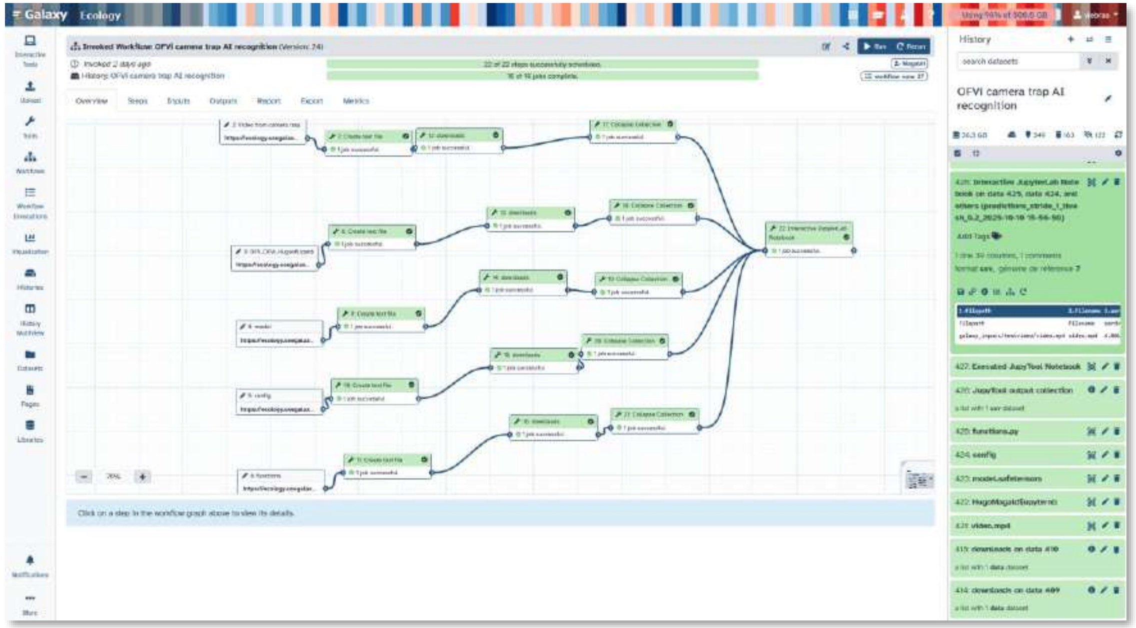
Annotations & AI

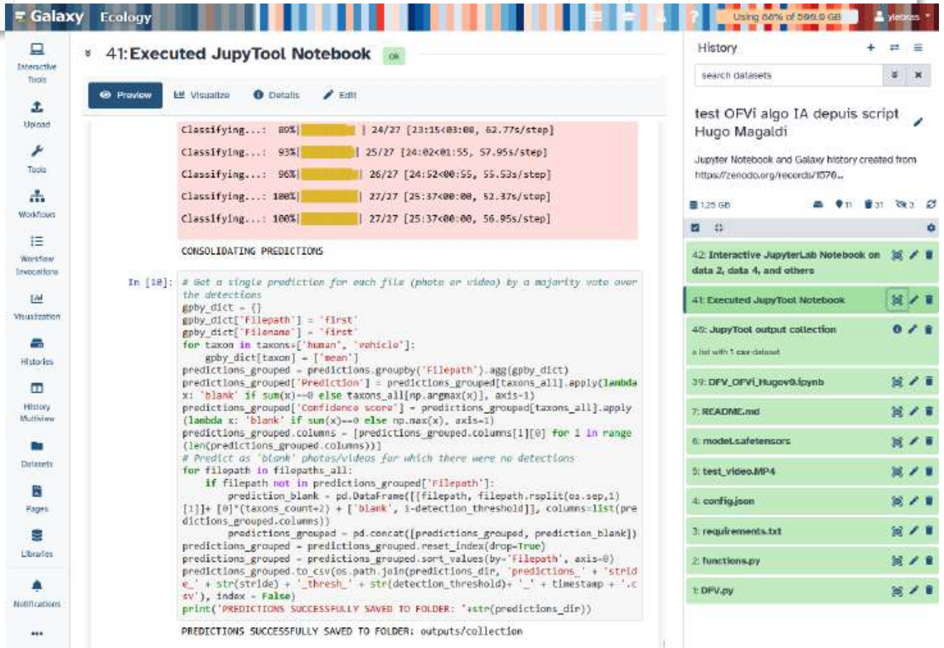
Using Jupyter notebook

From camera trap video to species detection with megadetector & pytorch-wildlife



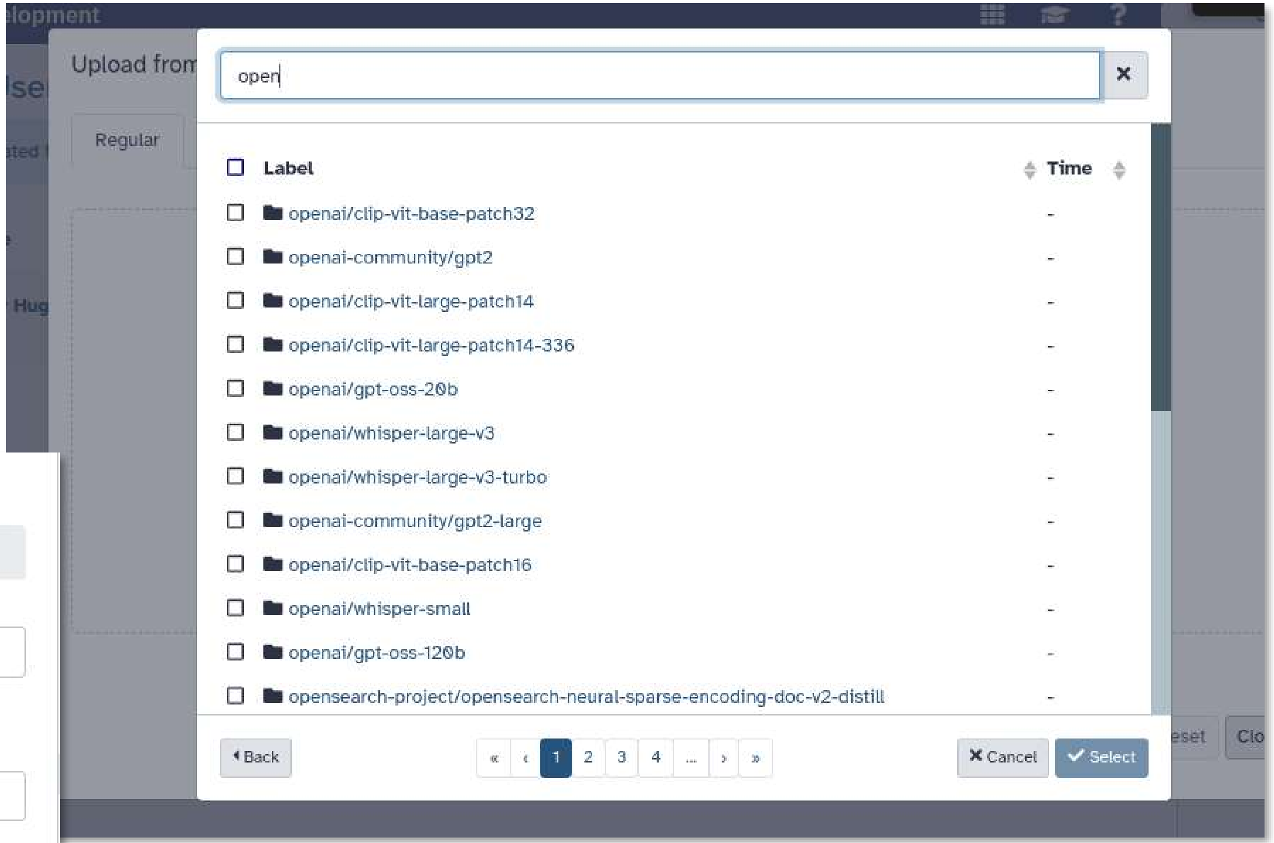


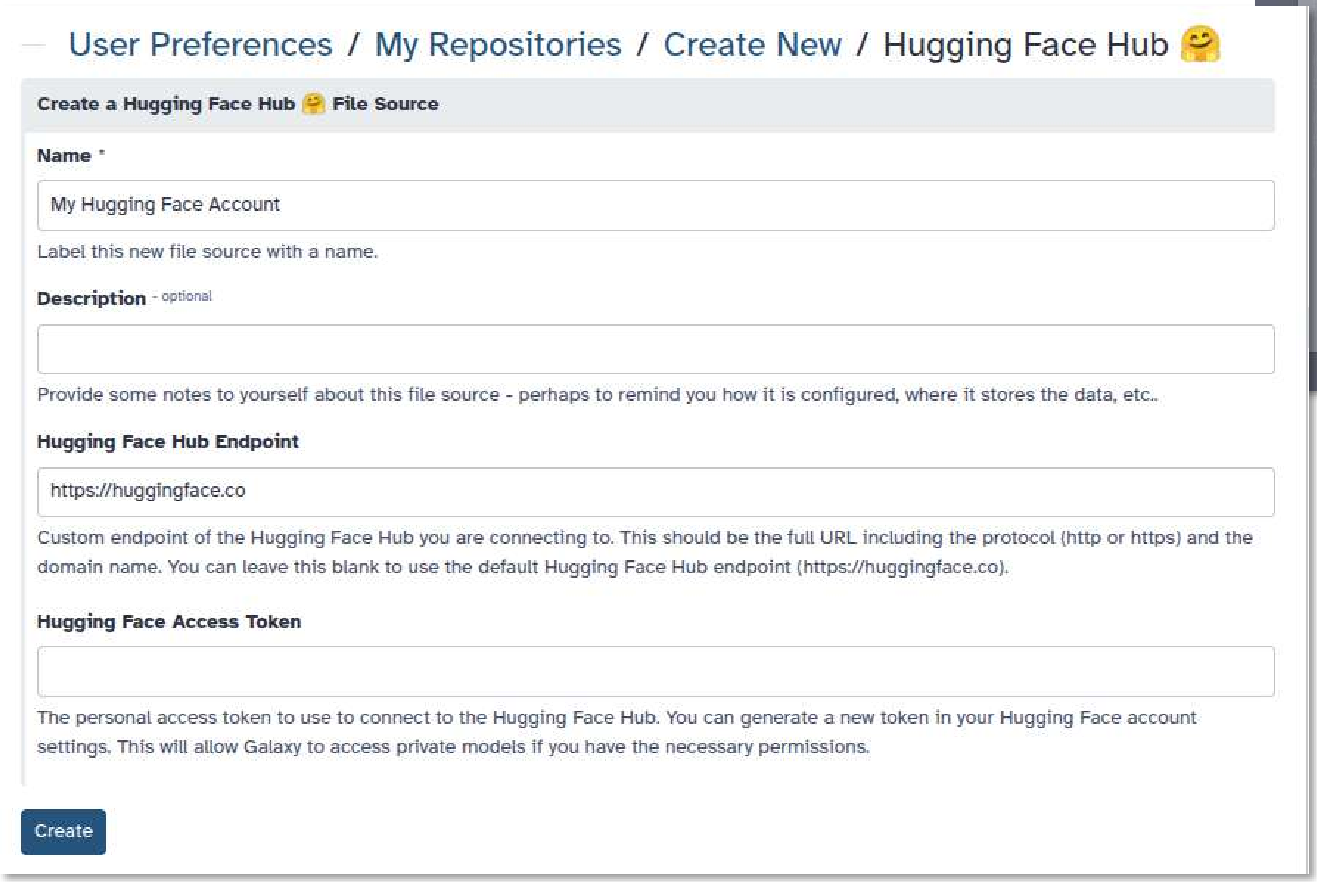




Annotations & AI

New: **Using Hugging Face models in Galaxy Ecology**





Galaxy Ecology initiative: A lot of young contributors!

Alan Amossé

Clara Urfer

Elisa Michon

Timothée Virgoulay

Valentin Chambon

Eloïse Trigodet

Claire Dussin

Elouan Le Mestric

Jean Le Cras

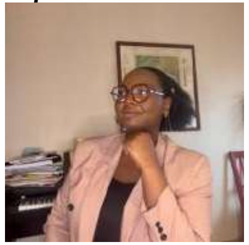
Olivier Norvez Elie Arnaud

Yassine Ankerl

Molène Mahé

Kévin Payet

Najat Amoukou



Simon Bénateau

Benjamin Yguel

Coline Royaux

Marie Jossé

Pauline Seguineau

Triskell Cumunel

Arthur Barreau

2018 2019 2020 2021 2022 2023 2024 2025