BEeS
Heraklion
30 June - 3 July 2025

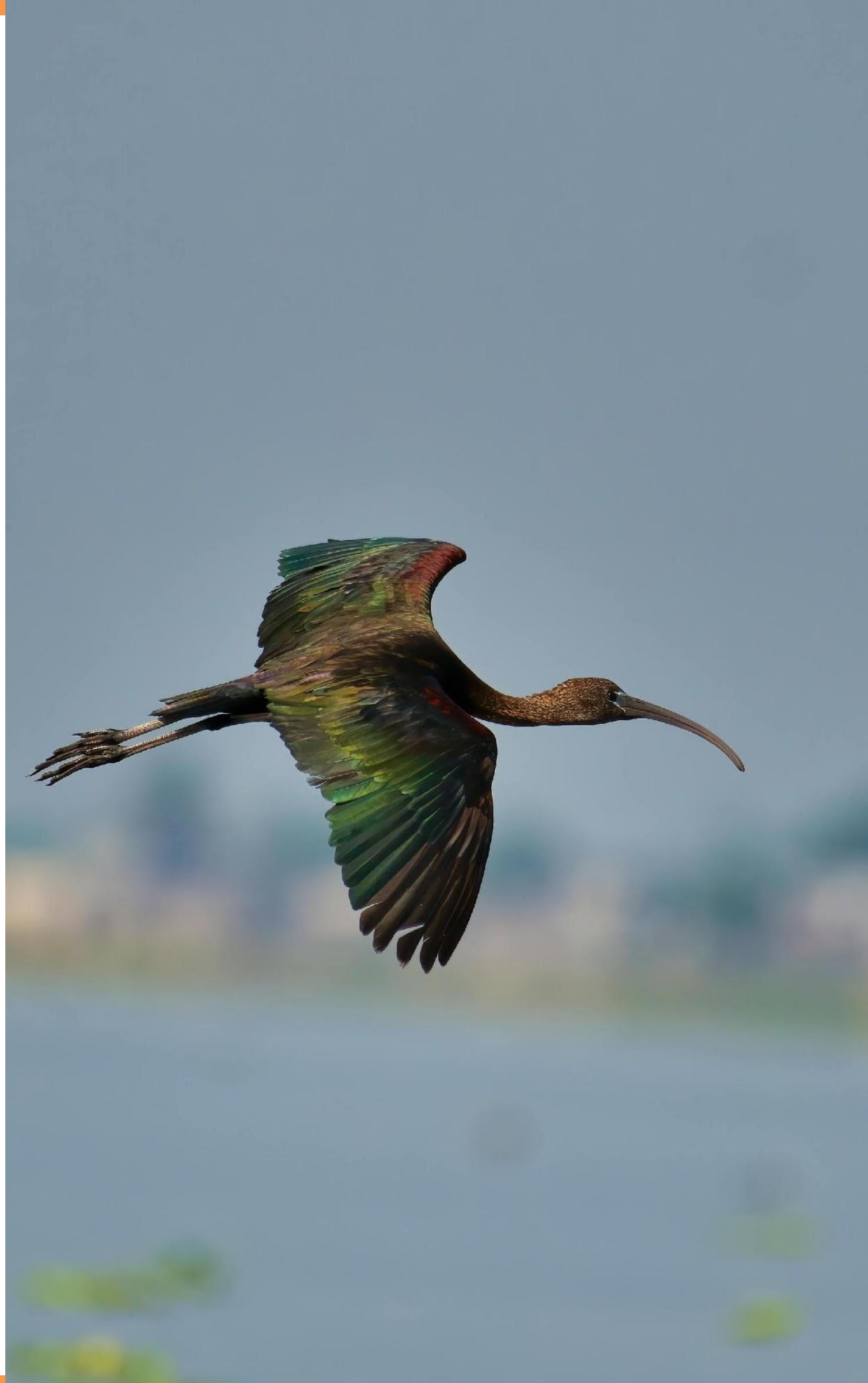
LifeWatch ERIC

hcmr
ΕΛΚΕΘΕ

iMBBC
ΙΘΑΒΒΥΚ

Session: Mapping life on planet: Biogeography in a changing world

1 July 2025 | 11:30-13:00



BEeS
Heraklion
30 June - 3 July 2025
LifeWatch ERIC

Cataloguing for findability:

software tools, workflows, and learning materials for biodiversity and ecology

Presenter: ***Matúš Kalaš*** *(University of Bergen & ELIXIR Norway)*

elixir

The ELIXIR Biodiversity Community

4 Use Cases:
Good examples of how molecular data can inform biodiversity understanding

7 Recommendations:
For the wider field of biodiversity research to benefit from the vast quantity of molecular data that will be generated in the coming years

OPINION ARTICLE

REVISED Recommendations for connecting molecular sequence and biodiversity research infrastructures through ELIXIR [version 2; peer review: 2 approved]

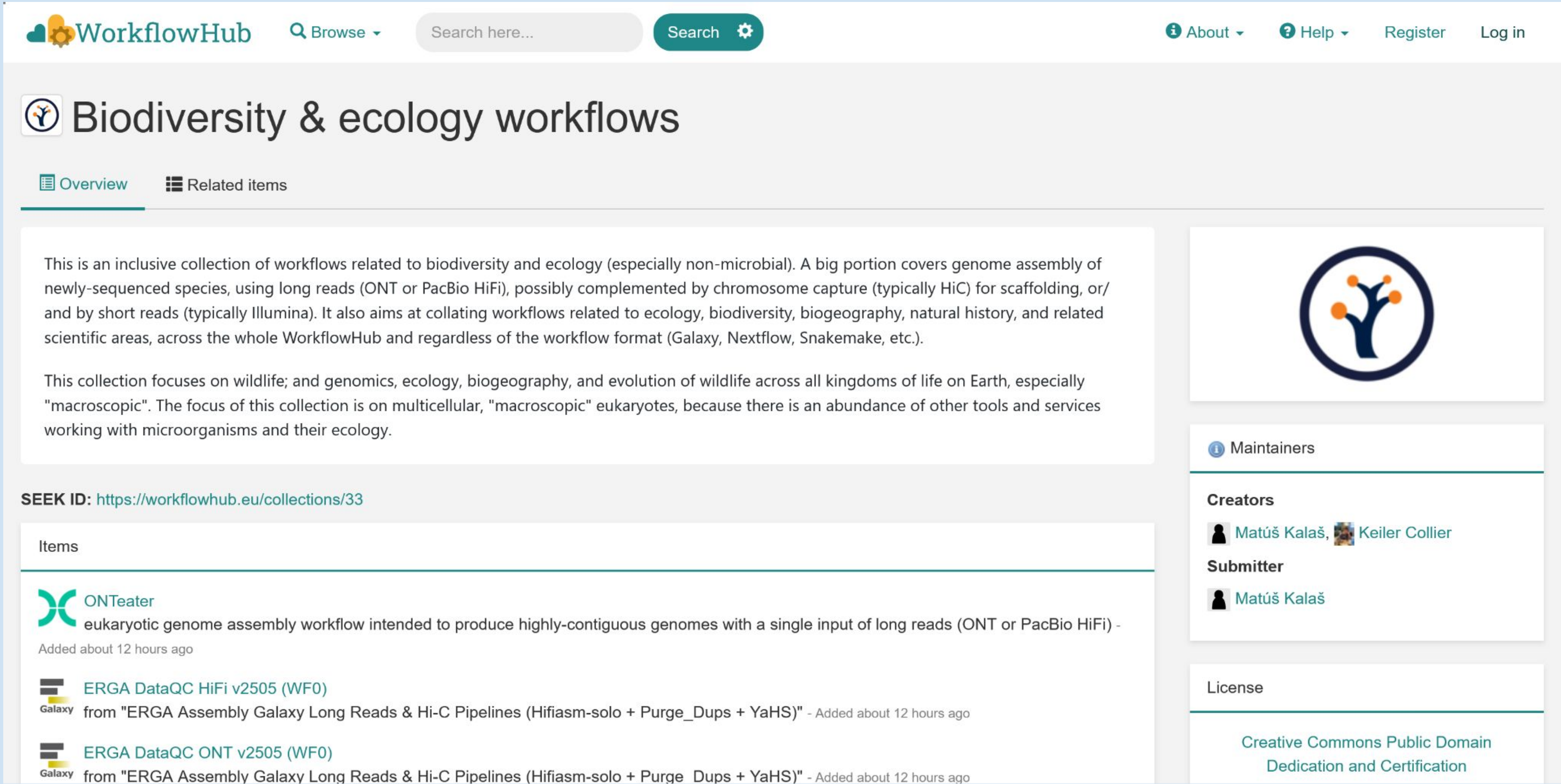
Robert M. Waterhouse[1], Anne-Françoise Adam-Blondon[2], Donat Agosti[3], Petr Baldrian[4], Bachir Balech[5], Erwan Corre[6], Robert P. Davey[7], Henrik Lantz[8], Graziano Pesole[5,9], Christian Quast[10], Frank Oliver Glöckner[11,12], Niels Raes[13], Anna Sandionigi[14], Monica Santamaria[5], Wouter Addink[15], Jiri Vohradsky[16], Amandine Nunes-Jorge[10], Nils Peder Willassen[17], Jerry Lanfear[18]

Author details

2021-2022

This article is included in the ELIXIR gateway.

https://workflowhub.eu/collections/33



Biodiversity & ecology workflows

This is an inclusive collection of workflows related to biodiversity and ecology (especially non-microbial). A big portion covers genome assembly of newly-sequenced species, using long reads (ONT or PacBio HiFi), possibly complemented by chromosome capture (typically HiC) for scaffolding, or/and by short reads (typically Illumina). It also aims at collating workflows related to ecology, biodiversity, biogeography, natural history, and related scientific areas, across the whole WorkflowHub and regardless of the workflow format (Galaxy, Nextflow, Snakemake, etc.).

This collection focuses on wildlife; and genomics, ecology, biogeography, and evolution of wildlife across all kingdoms of life on Earth, especially "macroscopic". The focus of this collection is on multicellular, "macroscopic" eukaryotes, because there is an abundance of other tools and services working with microorganisms and their ecology.

SEEK ID: https://workflowhub.eu/collections/33

Items

ONTeater
eukaryotic genome assembly workflow intended to produce highly-contiguous genomes with a single input of long reads (ONT or PacBio HiFi) -
Added about 12 hours ago

Galaxy ERGA DataQC HiFi v2505 (WF0)
from "ERGA Assembly Galaxy Long Reads & Hi-C Pipelines (Hifiasm-solo + Purge_Dups + YaHS)" - Added about 12 hours ago

Galaxy ERGA DataQC ONT v2505 (WF0)
from "ERGA Assembly Galaxy Long Reads & Hi-C Pipelines (Hifiasm-solo + Purge_Dups + YaHS)" - Added about 12 hours ago

Maintainers

Creators

Matúš Kalaš, Keiler Collier

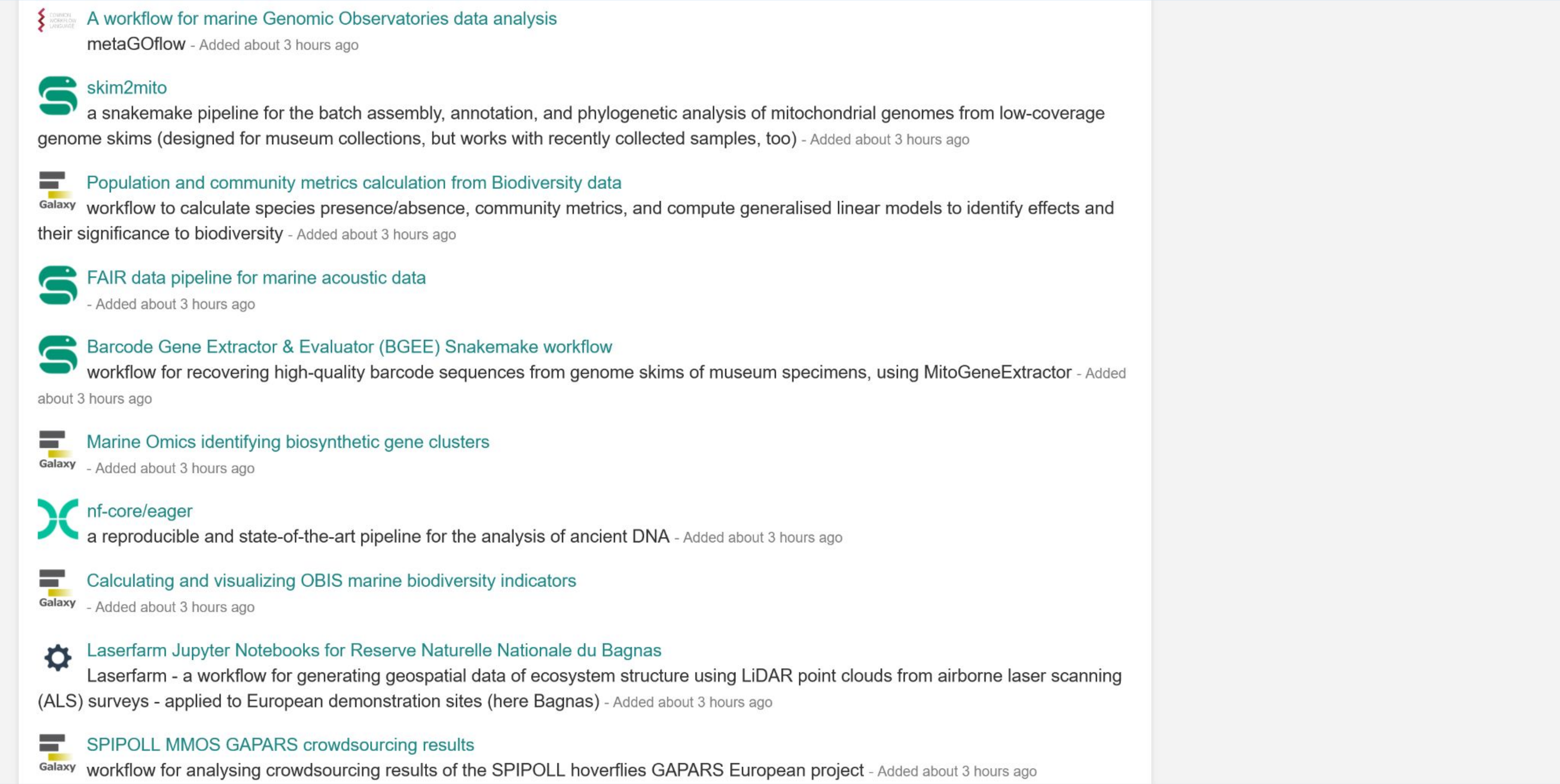
Submitter

Matúš Kalaš

License

Creative Commons Public Domain Dedication and Certification

https://workflowhub.eu/collections/33



A workflow for marine Genomic Observatories data analysis
metaGOflow - Added about 3 hours ago

skim2mito
a snakemake pipeline for the batch assembly, annotation, and phylogenetic analysis of mitochondrial genomes from low-coverage genome skims (designed for museum collections, but works with recently collected samples, too) - Added about 3 hours ago

Population and community metrics calculation from Biodiversity data
workflow to calculate species presence/absence, community metrics, and compute generalised linear models to identify effects and their significance to biodiversity - Added about 3 hours ago

FAIR data pipeline for marine acoustic data
- Added about 3 hours ago

Barcode Gene Extractor & Evaluator (BGEE) Snakemake workflow
workflow for recovering high-quality barcode sequences from genome skims of museum specimens, using MitoGeneExtractor - Added about 3 hours ago

Marine Omics identifying biosynthetic gene clusters
- Added about 3 hours ago

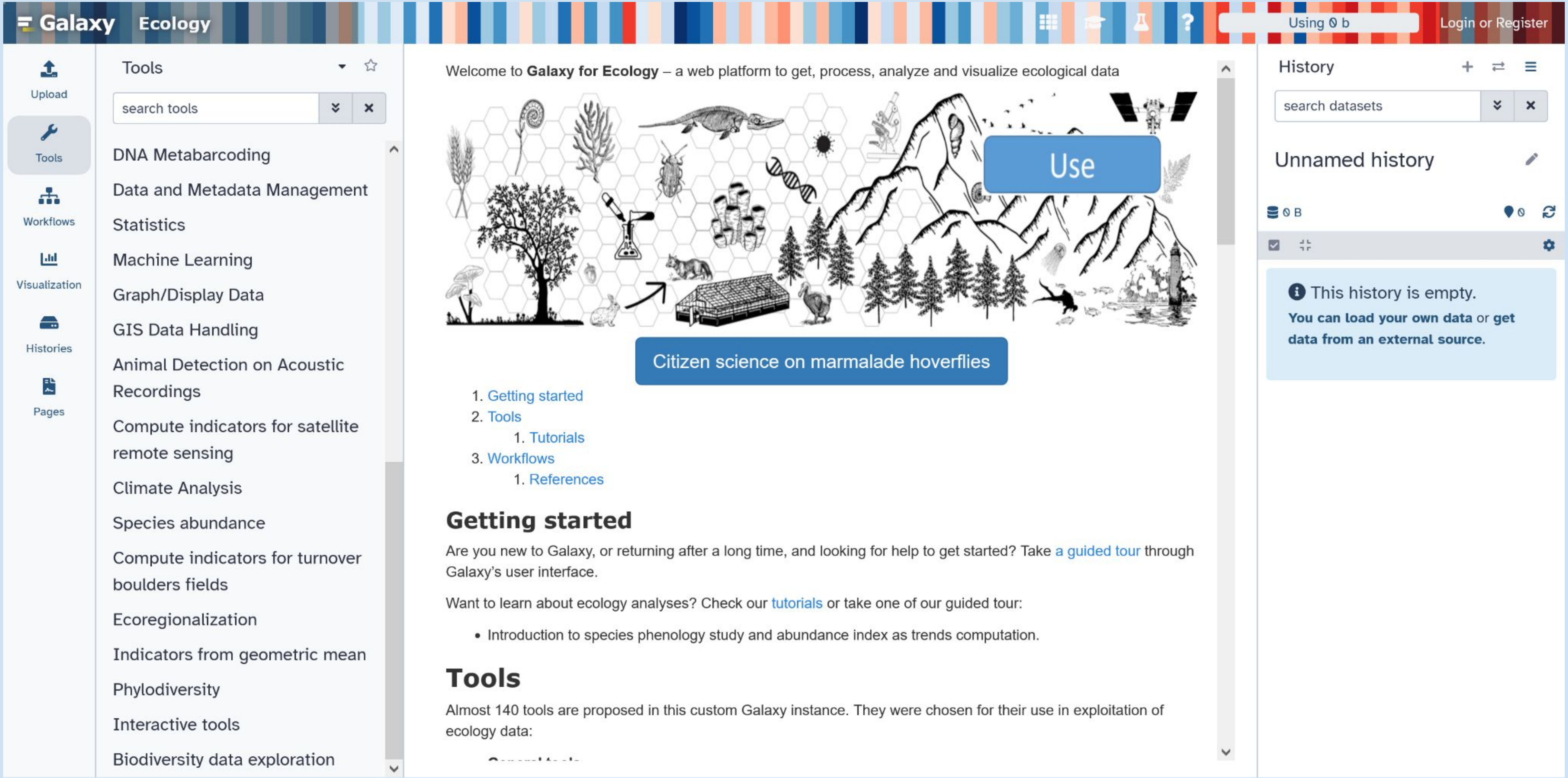
nf-core/eager
a reproducible and state-of-the-art pipeline for the analysis of ancient DNA - Added about 3 hours ago

Calculating and visualizing OBIS marine biodiversity indicators
- Added about 3 hours ago

Laserfarm Jupyter Notebooks for Reserve Naturelle Nationale du Bagnas
Laserfarm - a workflow for generating geospatial data of ecosystem structure using LiDAR point clouds from airborne laser scanning (ALS) surveys - applied to European demonstration sites (here Bagnas) - Added about 3 hours ago

SPIPOLL MMOS GAPARS crowdsourcing results
workflow for analysing crowdsourcing results of the SPIPOLL hoverflies GAPARS European project - Added about 3 hours ago

https://ecology.usegalaxy.eu



Welcome to **Galaxy for Ecology** – a web platform to get, process, analyze and visualize ecological data

1. Getting started
2. Tools
    1. Tutorials
3. Workflows
    1. References

Getting started

Are you new to Galaxy, or returning after a long time, and looking for help to get started? Take a guided tour through Galaxy's user interface.

Want to learn about ecology analyses? Check our tutorials or take one of our guided tour:

- Introduction to species phenology study and abundance index as trends computation.

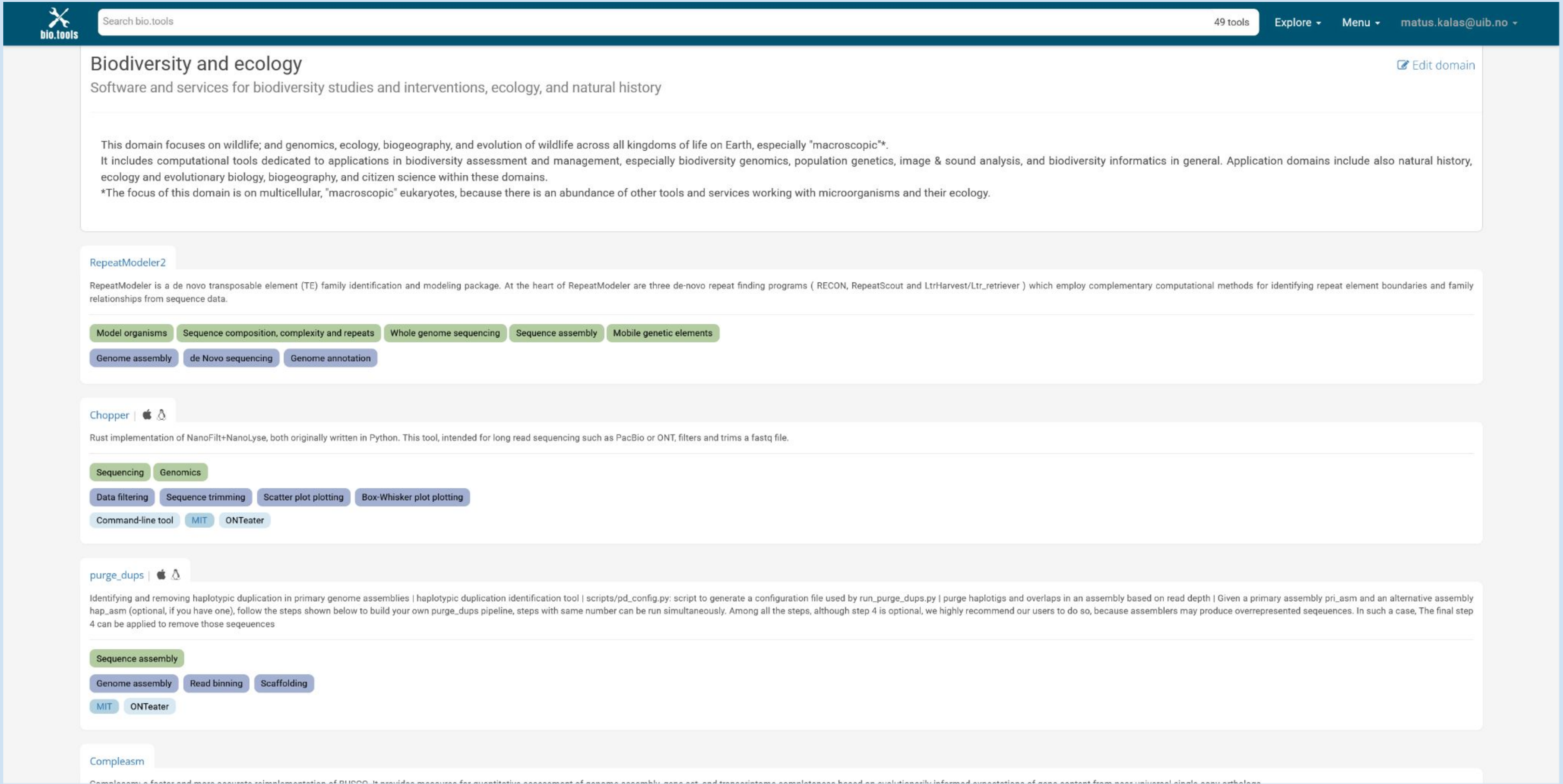
Tools

Almost 140 tools are proposed in this custom Galaxy instance. They were chosen for their use in exploitation of ecology data:

This history is empty.

You can load your own data or get data from an external source.

https://bio.tools/t?domain=biodiversity



Biodiversity and ecology

Software and services for biodiversity studies and interventions, ecology, and natural history

This domain focuses on wildlife; and genomics, ecology, biogeography, and evolution of wildlife across all kingdoms of life on Earth, especially "macroscopic"*.
It includes computational tools dedicated to applications in biodiversity assessment and management, especially biodiversity genomics, population genetics, image & sound analysis, and biodiversity informatics in general. Application domains include also natural history, ecology and evolutionary biology, biogeography, and citizen science within these domains.
*The focus of this domain is on multicellular, "macroscopic" eukaryotes, because there is an abundance of other tools and services working with microorganisms and their ecology.

RepeatModeler2

RepeatModeler is a de novo transposable element (TE) family identification and modeling package. At the heart of RepeatModeler are three de-novo repeat finding programs (RECON, RepeatScout and LtrHarvest/Ltr_retriever) which employ complementary computational methods for identifying repeat element boundaries and family relationships from sequence data.

Model organisms | Sequence composition, complexity and repeats | Whole genome sequencing | Sequence assembly | Mobile genetic elements

Genome assembly | de Novo sequencing | Genome annotation

Chopper

Rust implementation of NanoFilt+NanoLyse, both originally written in Python. This tool, intended for long read sequencing such as PacBio or ONT, filters and trims a fastq file.

Sequencing | Genomics

Data filtering | Sequence trimming | Scatter plot plotting | Box-Whisker plot plotting

Command-line tool | MIT | ONTeater

purge_dups

Identifying and removing haplotypic duplication in primary genome assemblies | haplotypic duplication identification tool | scripts/pd_config.py: script to generate a configuration file used by run_purge_dups.py | purge haplotigs and overlaps in an assembly based on read depth | Given a primary assembly pri_asm and an alternative assembly hap_asm (optional, if you have one), follow the steps shown below to build your own purge_dups pipeline, steps with same number can be run simultaneously. Among all the steps, although step 4 is optional, we highly recommend our users to do so, because assemblers may produce overrepresented seqeuences. In such a case, The final step 4 can be applied to remove those seqeuences

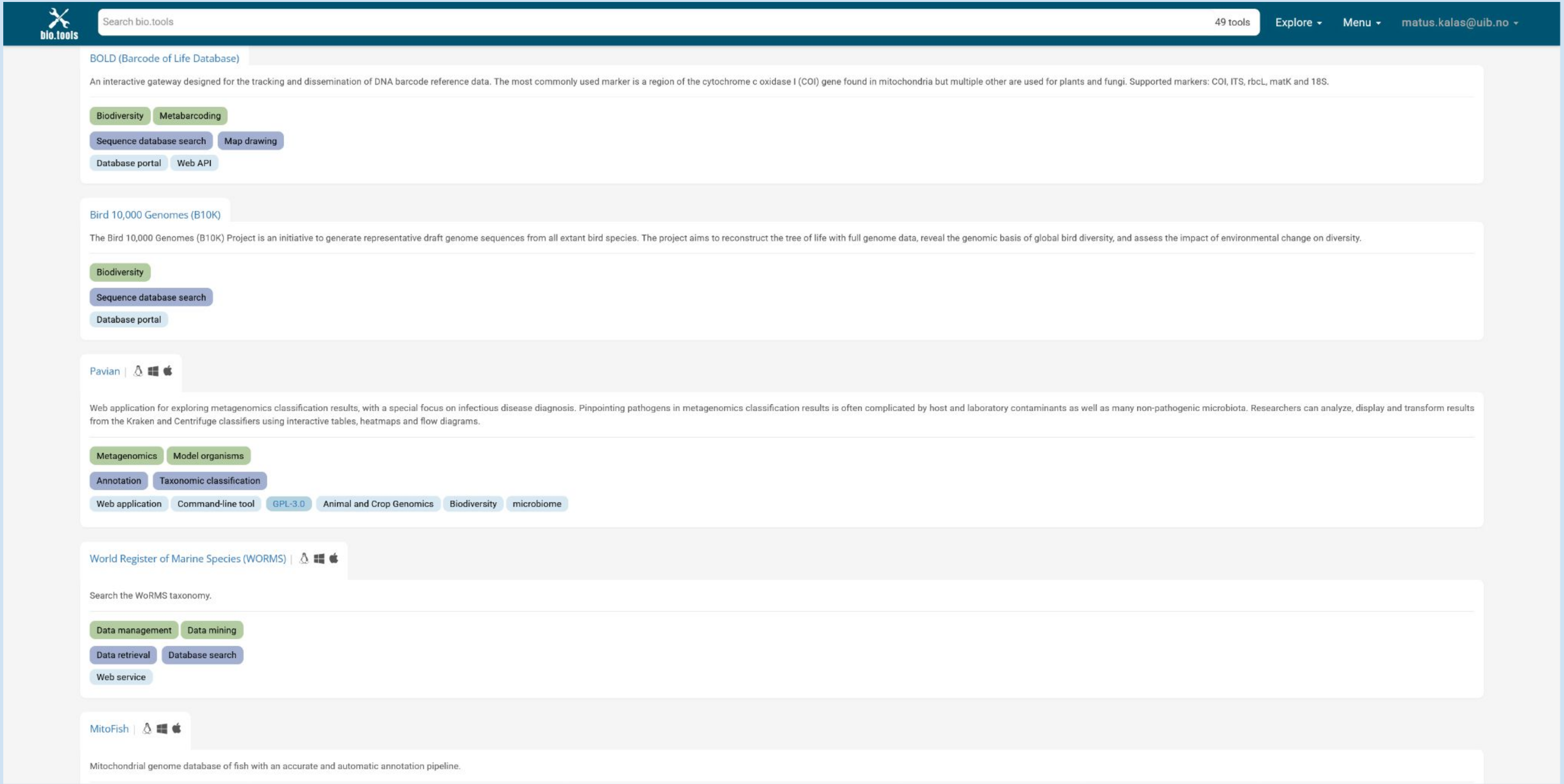
Sequence assembly

Genome assembly | Read binning | Scaffolding

MIT | ONTeater

Compleasm

https://bio.tools/t?domain=biodiversity



BOLD (Barcode of Life Database)

An interactive gateway designed for the tracking and dissemination of DNA barcode reference data. The most commonly used marker is a region of the cytochrome c oxidase I (COI) gene found in mitochondria but multiple other are used for plants and fungi. Supported markers: COI, ITS, rbcL, matK and 18S.

Biodiversity | Metabarcoding
Sequence database search | Map drawing
Database portal | Web API

Bird 10,000 Genomes (B10K)

The Bird 10,000 Genomes (B10K) Project is an initiative to generate representative draft genome sequences from all extant bird species. The project aims to reconstruct the tree of life with full genome data, reveal the genomic basis of global bird diversity, and assess the impact of environmental change on diversity.

Biodiversity
Sequence database search
Database portal

Pavian

Web application for exploring metagenomics classification results, with a special focus on infectious disease diagnosis. Pinpointing pathogens in metagenomics classification results is often complicated by host and laboratory contaminants as well as many non-pathogenic microbiota. Researchers can analyze, display and transform results from the Kraken and Centrifuge classifiers using interactive tables, heatmaps and flow diagrams.

Metagenomics | Model organisms
Annotation | Taxonomic classification
Web application | Command-line tool | GPL-3.0 | Animal and Crop Genomics | Biodiversity | microbiome

World Register of Marine Species (WORMS)

Search the WoRMS taxonomy.

Data management | Data mining
Data retrieval | Database search
Web service

MitoFish

Mitochondrial genome database of fish with an accurate and automatic annotation pipeline.

EDAM ontology

Topic

Ecology

Metatranscriptomics

DNA barcoding

Operation

Genome assembly

Taxonomic classification

Data management planning

Data

Phylogenetic tree

Geospatial coordinate

DOI

Format

FASTQ

GFF3

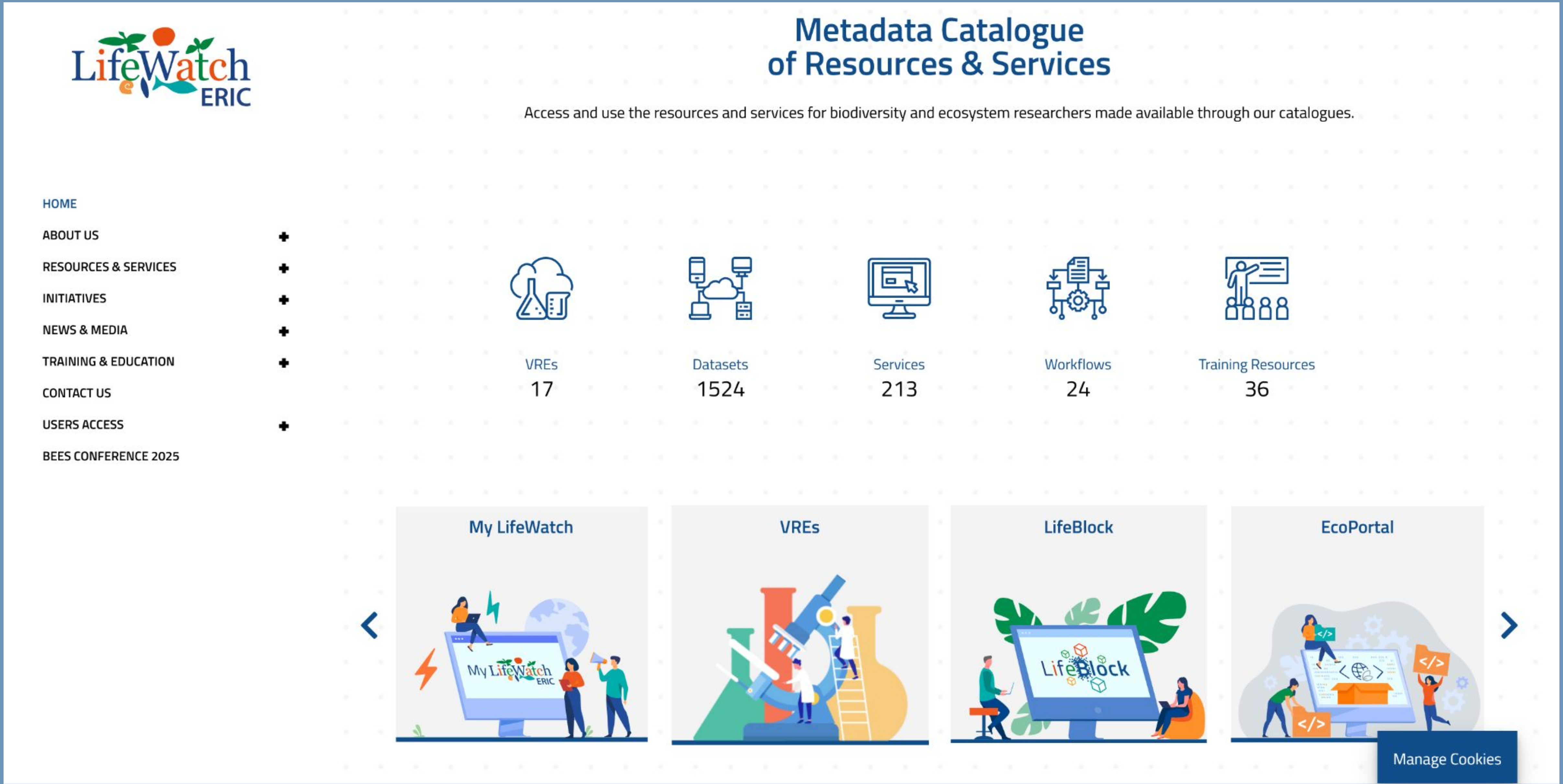
JSON-LD

Usage areas of EDAM

- Finding tools, data, learning materials, workflows, standards, …
- Data provenance (metadata)
- Tools and data integration
- Text mining
- Choosing terminology

BEeS
Heraklion
30 June - 3 July 2025

Let’s add the awesome LifeWatch resources, too!



Metadata Catalogue of Resources & Services

Access and use the resources and services for biodiversity and ecosystem researchers made available through our catalogues.

- HOME
- ABOUT US ✚
- RESOURCES & SERVICES ✚
- INITIATIVES ✚
- NEWS & MEDIA ✚
- TRAINING & EDUCATION ✚
- CONTACT US
- USERS ACCESS ✚
- BEES CONFERENCE 2025

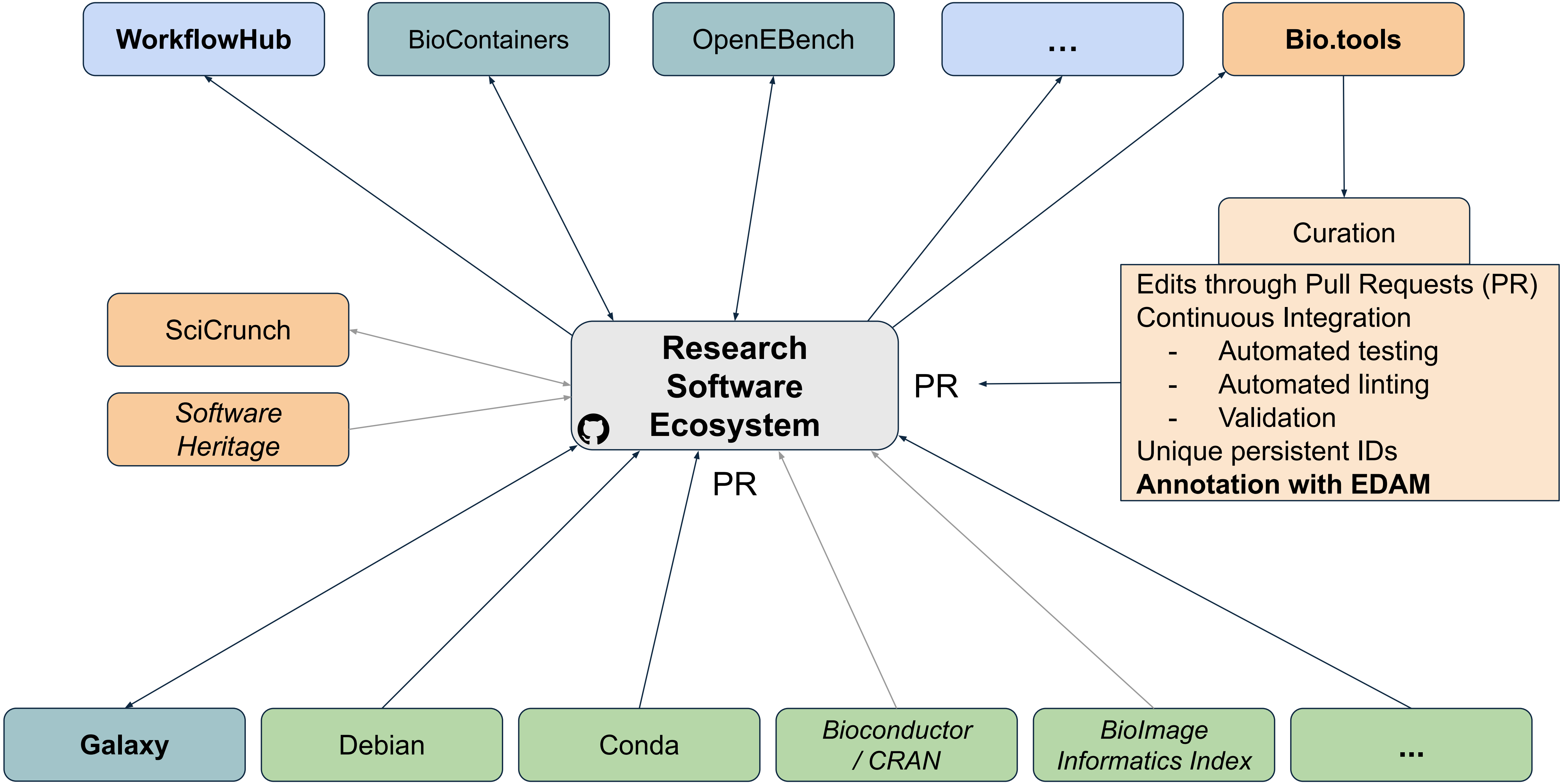
VREs	Datasets	Services	Workflows	Training Resources
17	1524	213	24	36

My LifeWatch

VREs

LifeBlock

EcoPortal



WorkflowHub

BioContainers

OpenEBench

…

Bio.tools

Curation

- Edits through Pull Requests (PR)
- Continuous Integration
  - Automated testing
  - Automated linting
  - Validation
- Unique persistent IDs
- **Annotation with EDAM**

SciCrunch

Software Heritage

Research Software Ecosystem

PR

PR

Galaxy

Debian

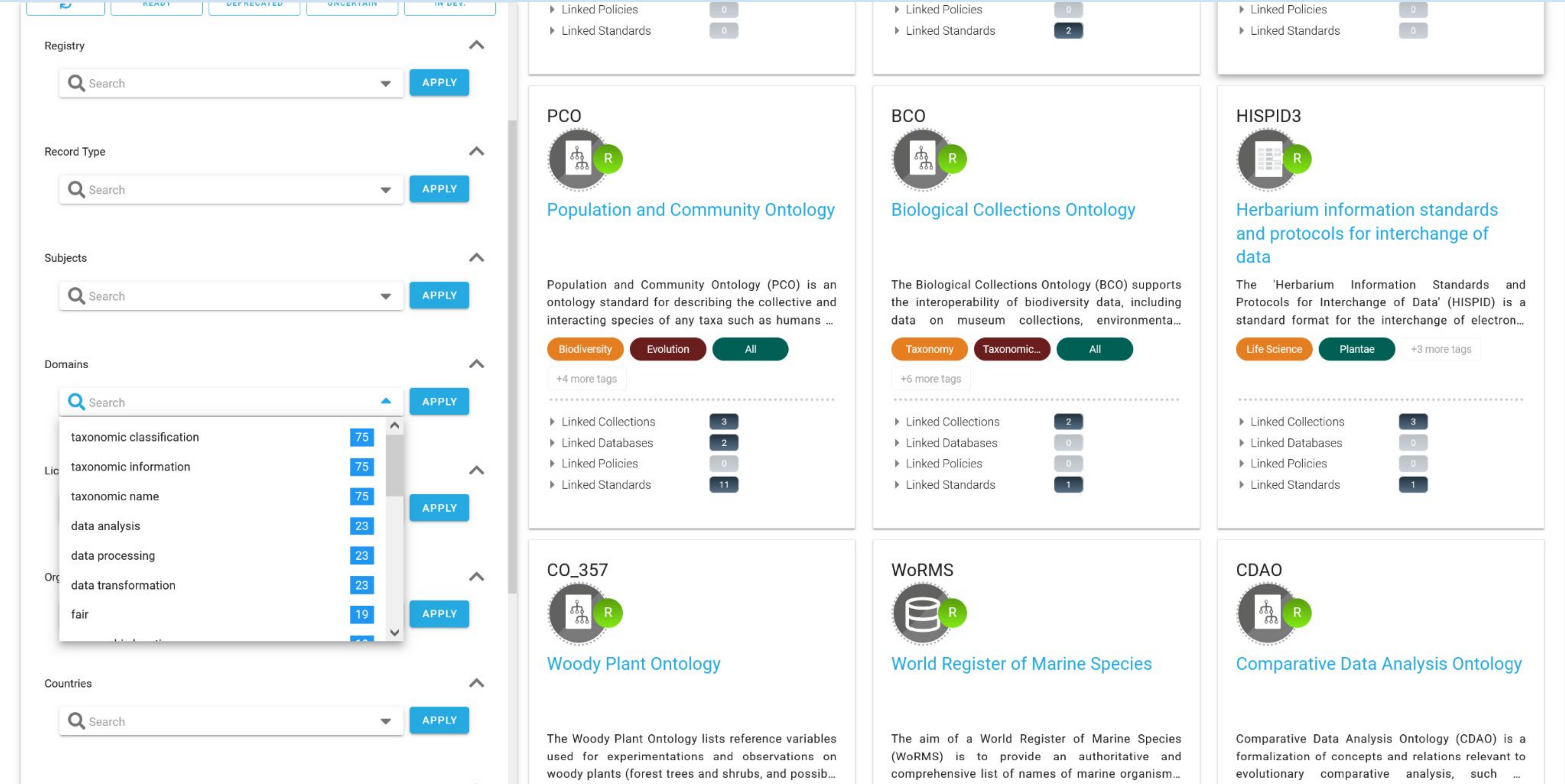
Conda

Bioconductor / CRAN

BioImage Informatics Index

…

fairsharing.org/ELIXIR-Biodiversity



PCO

Population and Community Ontology

Population and Community Ontology (PCO) is an ontology standard for describing the collective and interacting species of any taxa such as humans …

BCO

Biological Collections Ontology

The Biological Collections Ontology (BCO) supports the interoperability of biodiversity data, including data on museum collections, environmenta…

HISPID3

Herbarium information standards and protocols for interchange of data

The 'Herbarium Information Standards and Protocols for Interchange of Data' (HISPID) is a standard format for the interchange of electron…

CO_357

Woody Plant Ontology

The Woody Plant Ontology lists reference variables used for experimentations and observations on woody plants (forest trees and shrubs, and possib…

WoRMS

World Register of Marine Species

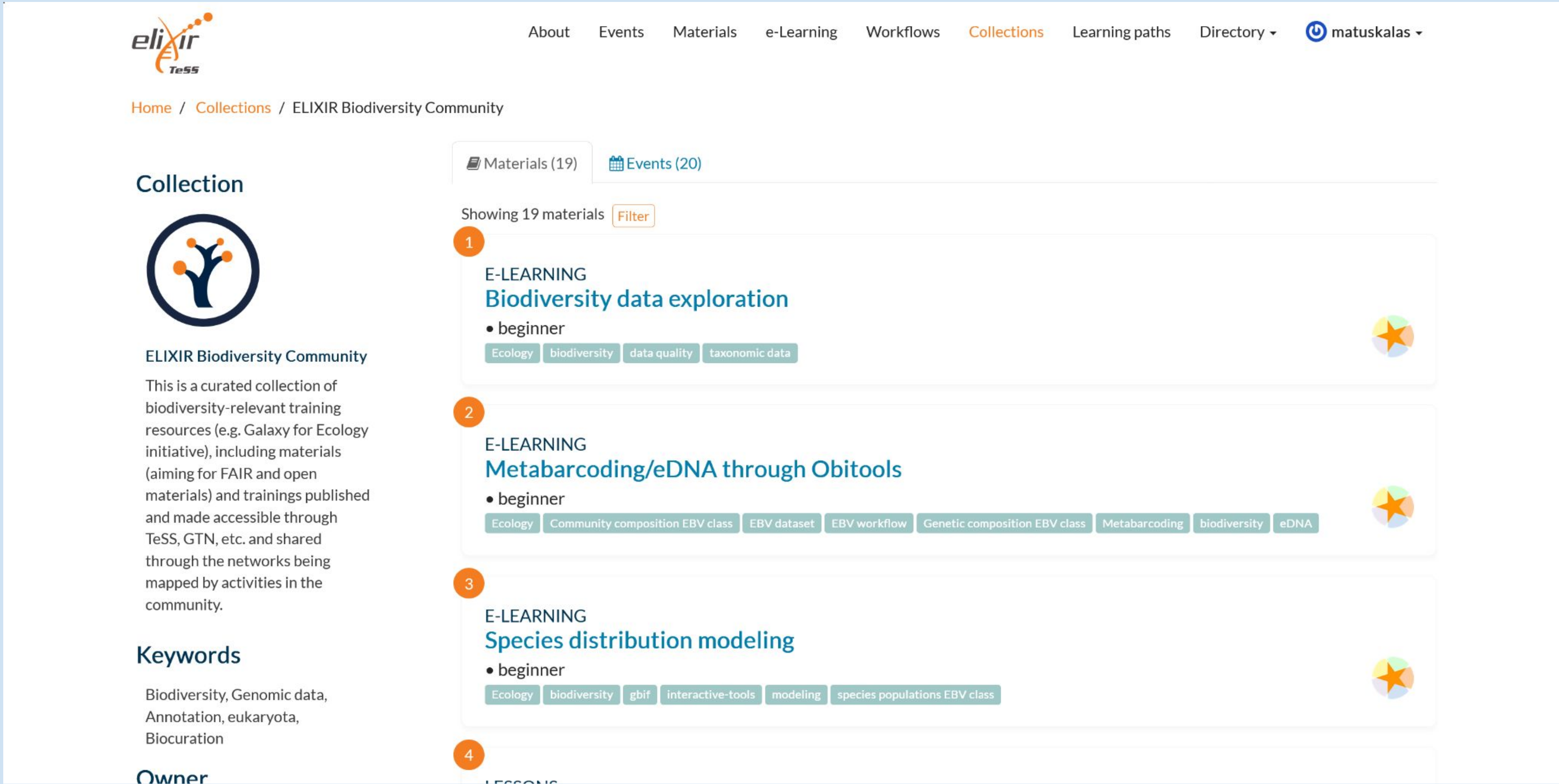
The aim of a World Register of Marine Species (WoRMS) is to provide an authoritative and comprehensive list of names of marine organism…

CDAO

Comparative Data Analysis Ontology

Comparative Data Analysis Ontology (CDAO) is a formalization of concepts and relations relevant to evolutionary comparative analysis, such …

tess.elixir-europe.org/collections/elixir-biodiversity-community



Home / Collections / ELIXIR Biodiversity Community

Collection

ELIXIR Biodiversity Community

This is a curated collection of biodiversity-relevant training resources (e.g. Galaxy for Ecology initiative), including materials (aiming for FAIR and open materials) and trainings published and made accessible through TeSS, GTN, etc. and shared through the networks being mapped by activities in the community.

Keywords

Biodiversity, Genomic data, Annotation, eukaryota, Biocuration

Owner

Materials (19) Events (20)

Showing 19 materials Filter

1

E-LEARNING
Biodiversity data exploration
- beginner

Ecology biodiversity data quality taxonomic data

2

E-LEARNING
Metabarcoding/eDNA through Obitools
- beginner

Ecology Community composition EBV class EBV dataset EBV workflow Genetic composition EBV class Metabarcoding biodiversity eDNA

3

E-LEARNING
Species distribution modeling
- beginner

Ecology biodiversity gbif interactive-tools modeling species populations EBV class

4

Thank you!

Questions?
drmatuskalas@duck.com