Galaxy pour le système terre – Perspectives

DATA TERRA

GAIA Data

eosc

cea cirad cnrs IGN INRAE Ifremer SORBONNE UNIVERSITÉ Université Clermont Auvergne UNIVERSITÉ DE MONTPELLIER Université de Strasbourg

brgm Cerema cnes FRB INERIS METEO FRANCE OFB SHOM UGA Université Grenoble Alpes Université de Lille

Du monde de la biodiversité au système terre

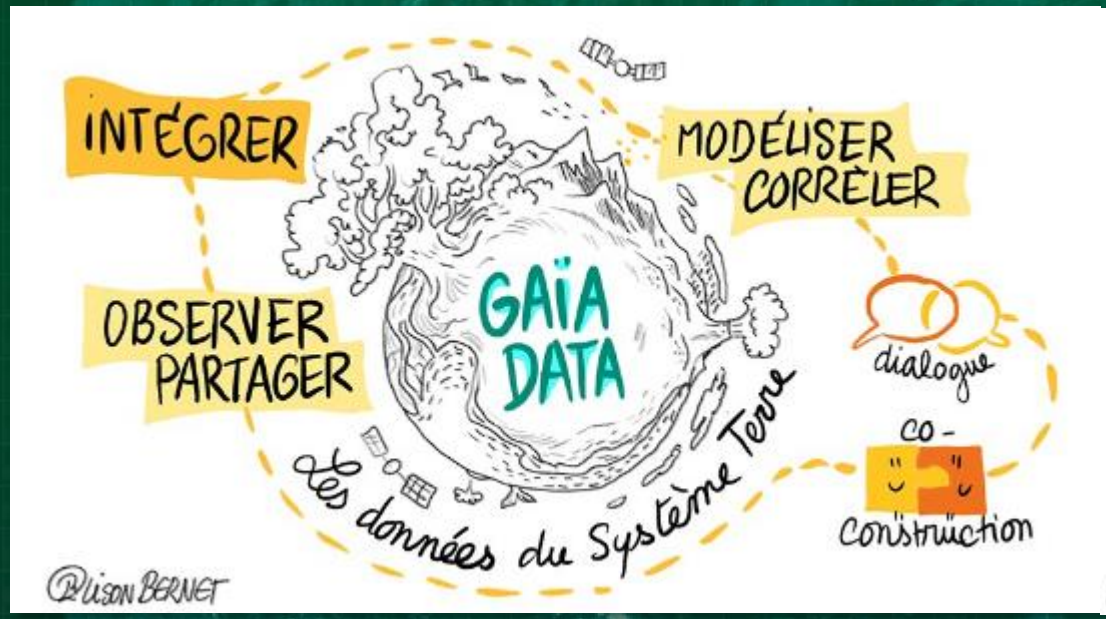
Du PNDB au projet GAIA DATA

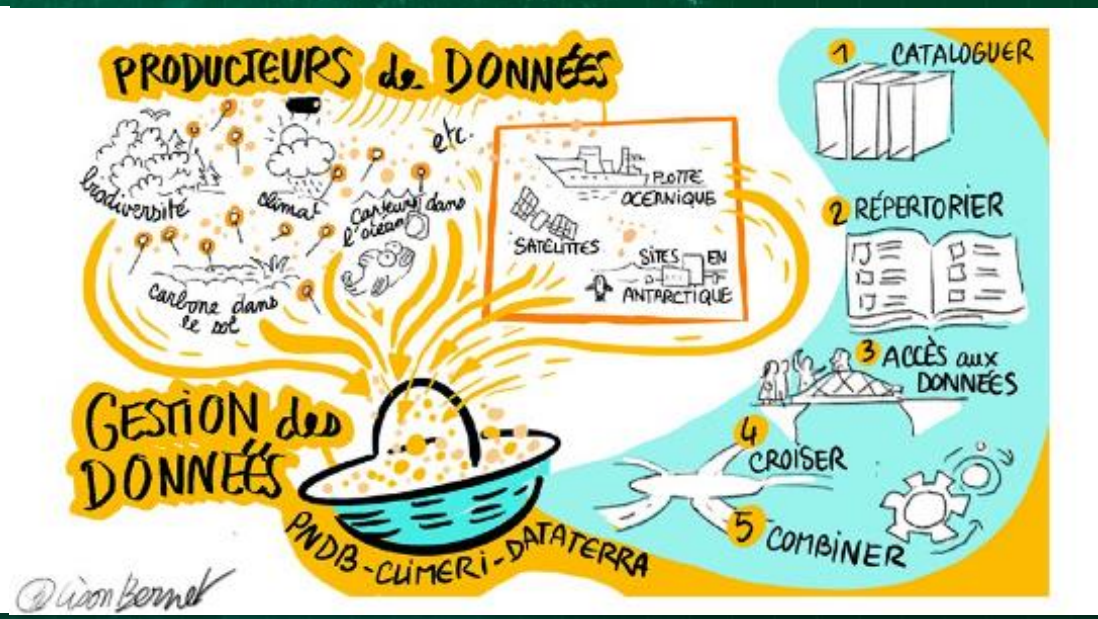
2022->

Du monde de la biodiversité au système terre

Du PNDB au projet GAIA DATA

2022->

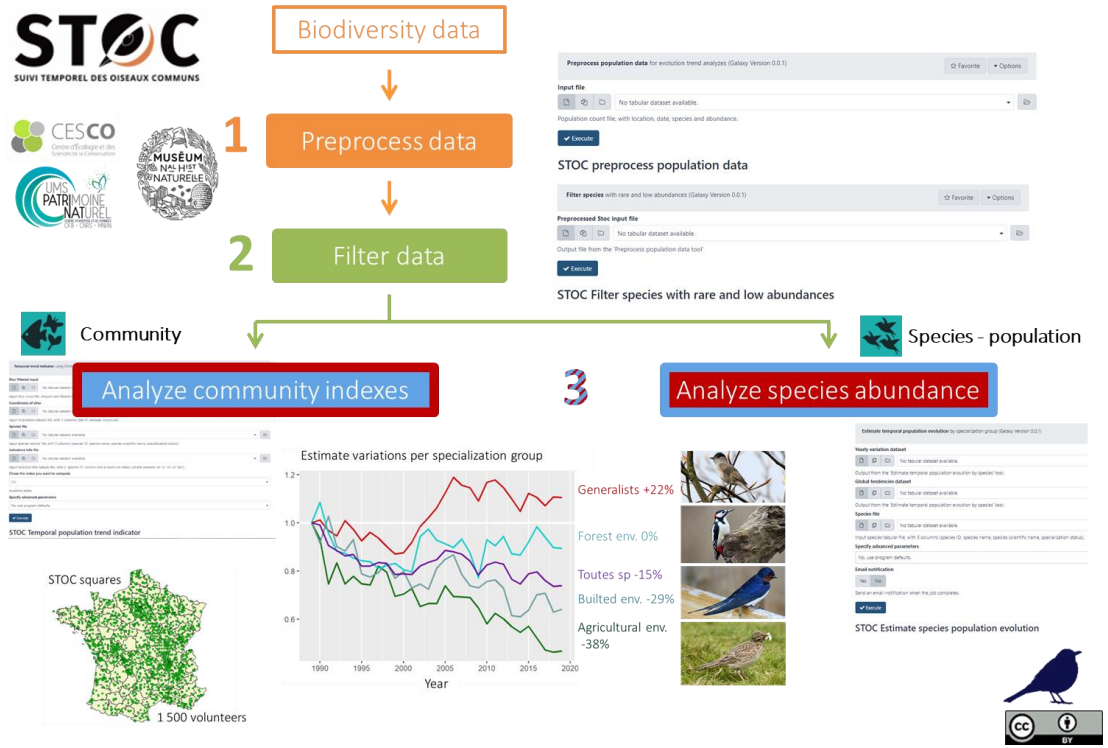


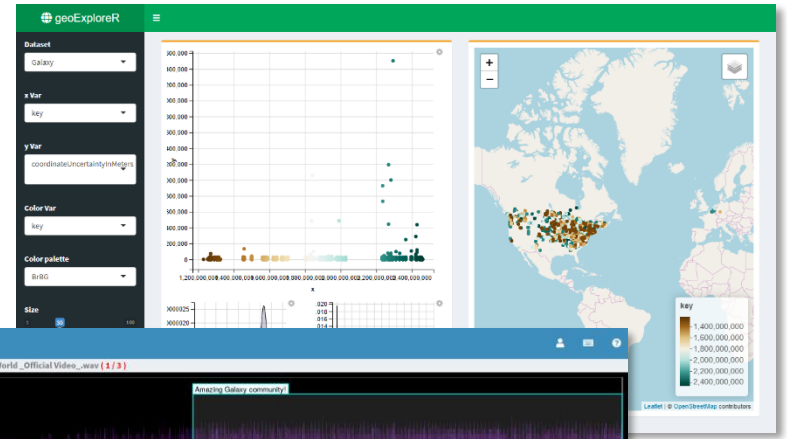


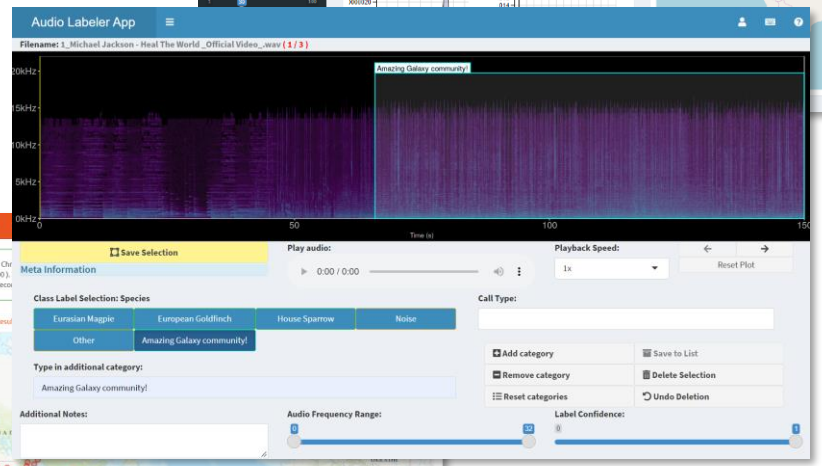
Partager des outils, mobiliser les utilisateurs

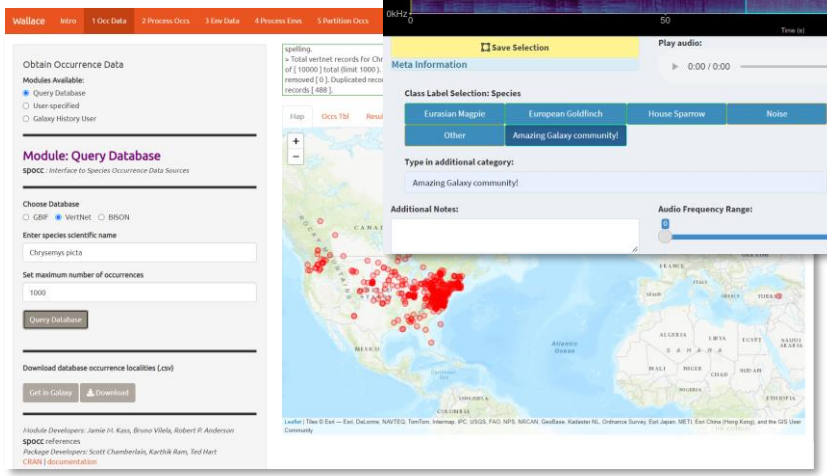
Galaxy-E comme un « démonstrateur »

2023









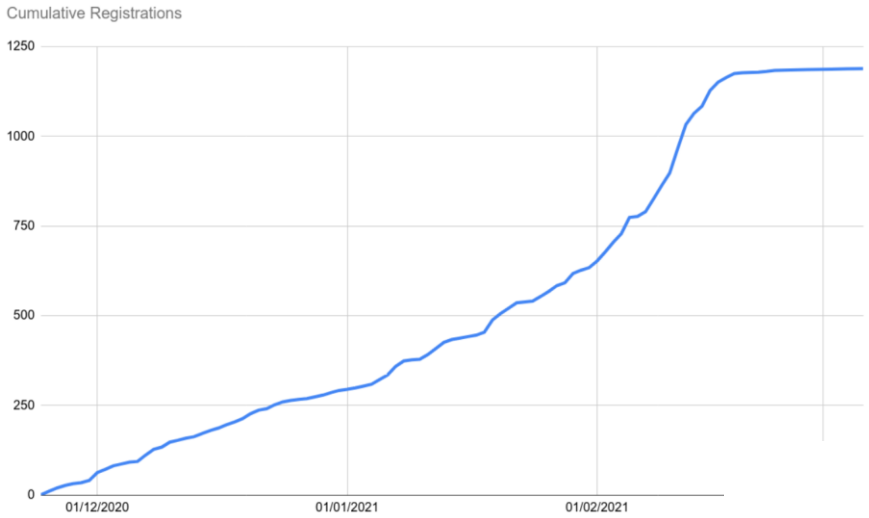
Pôle National de Données de Biodiversité

~100 tools ~10 000 execution / year ~140 users / month ~1650 uniq users

Brique essentielle pour contribuer à des initiatives de grande ampleur

Projet européen sur la culture de la donnée et de son analyse

2019-2023

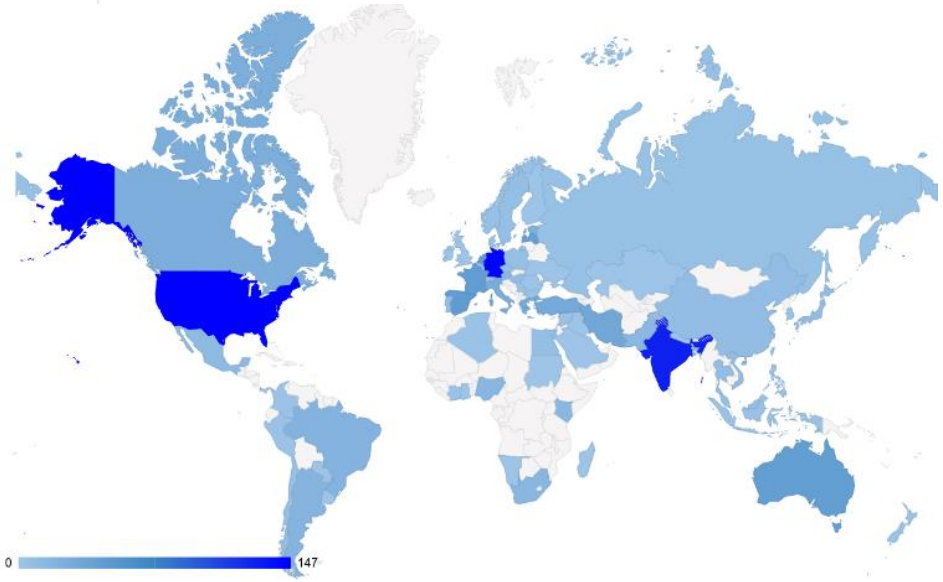


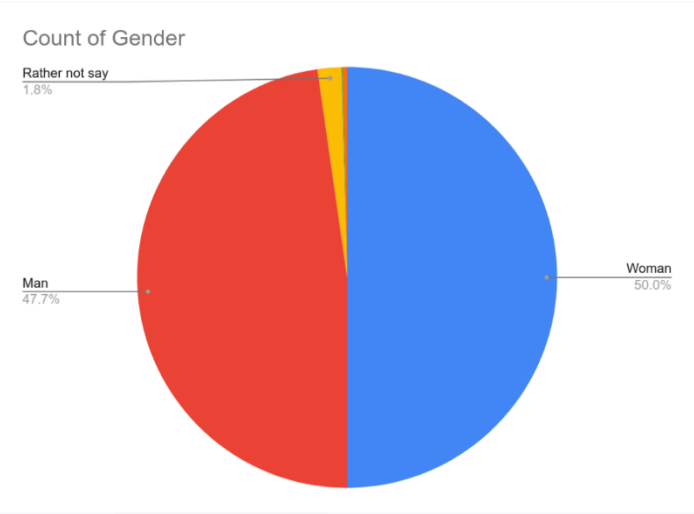
Gallantries

Good Practice

With the support of the Erasmus+ Programme of the European Union

This project (2020-1-NL01-KA203-064717) is funded with the support of the Erasmus+ programme of the European Union.





Massive training week

- 1200 registrations 2021
- 3300 registrations 2022!

Brique essentielle pour contribuer à des initiatives de grande ampleur

Des projets pour la mise en place d'un cloud Européen science ouverte

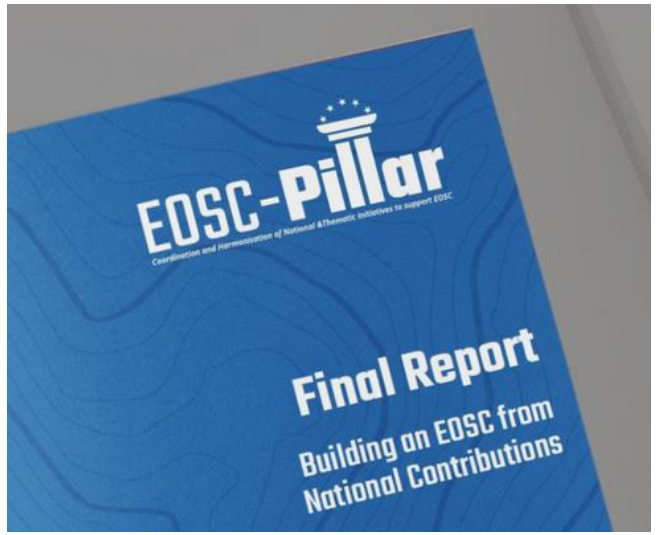
2022->2025

eosc | FAIR-EASE
Building Interoperable Earth Science & Environmental Services

Brique essentielle pour contribuer à des initiatives de grande ampleur

Des projets pour la mise en place d'un cloud Européen science ouverte

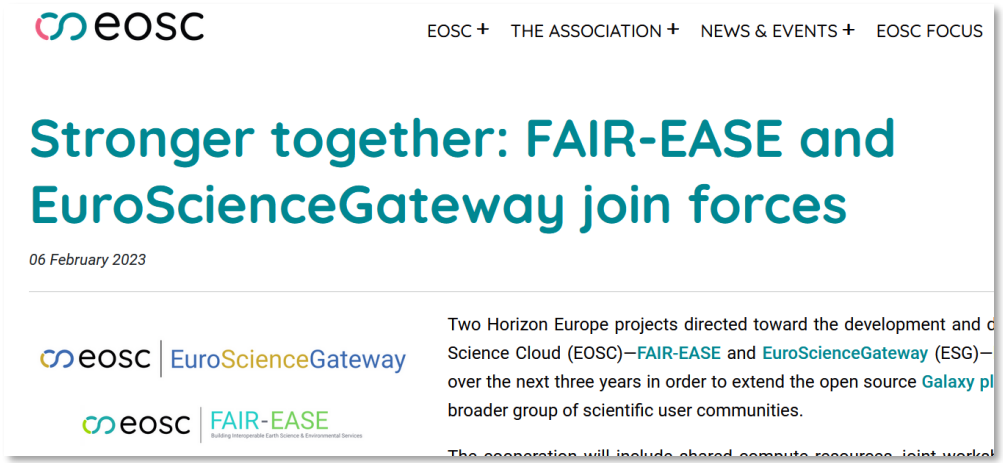
2019->2025



eosc | FAIR-EASE
Building Interoperable Earth Science & Environmental Services

Scientific Collaboration Agreement between EOSC-Pillar and EOSC-Life

eosc | EuroScienceGateway



eosc EOSC + THE ASSOCIATION + NEWS & EVENTS + EOSC FOCUS

Stronger together: FAIR-EASE and EuroScienceGateway join forces

06 February 2023

eosc | EuroScienceGateway

eosc | FAIR-EASE

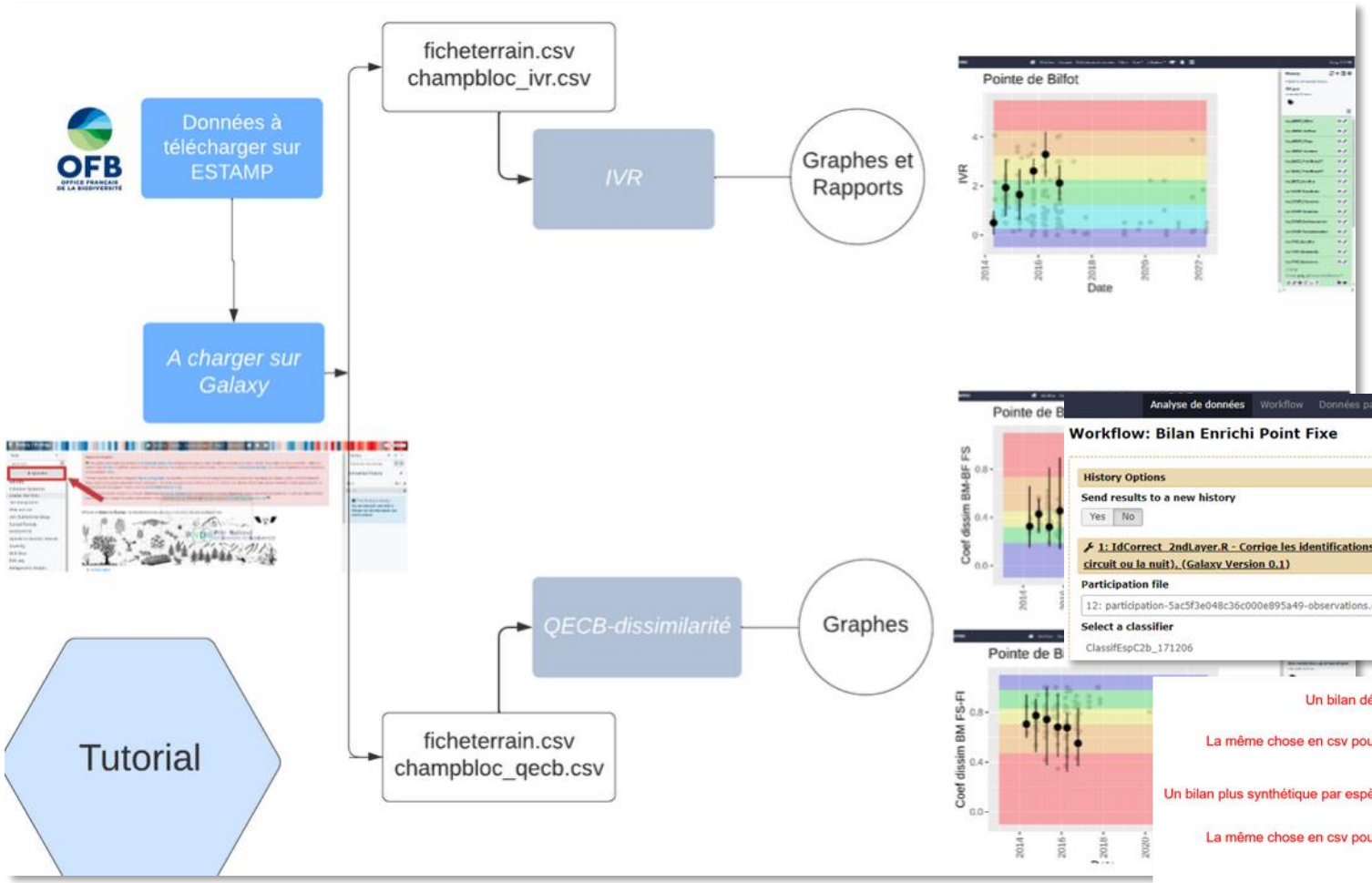
Two Horizon Europe projects directed toward the development and c
Science Cloud (EOSC)—FAIR-EASE and EuroScienceGateway (ESG)—
over the next three years in order to extend the open source Galaxy pl
broader group of scientific user communities.

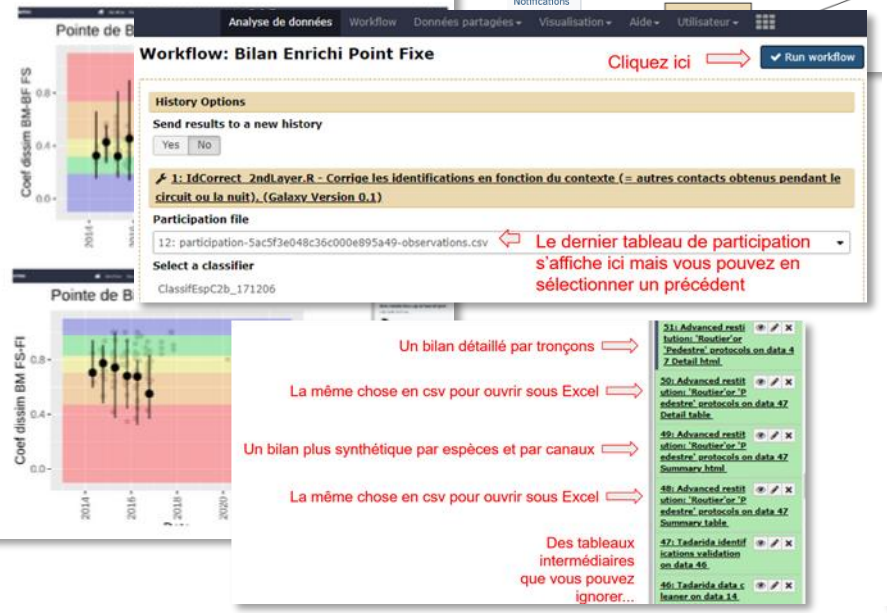
Galaxy as a cross-EOSC platform

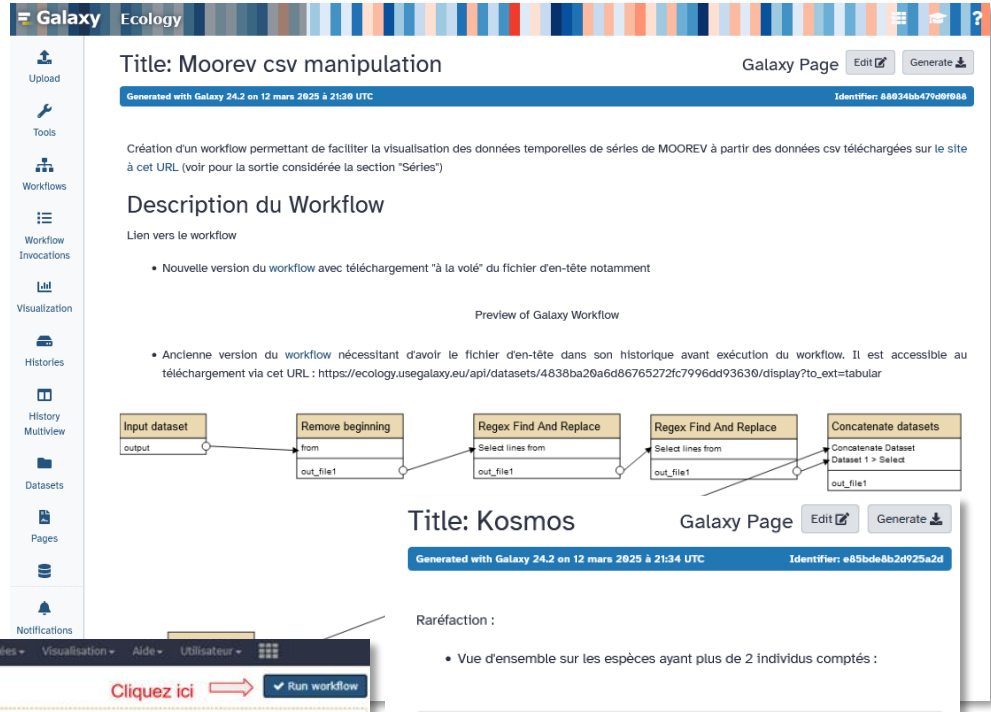
New scientific collaboration agreement between EOSC-Life and EOSC-Pillar

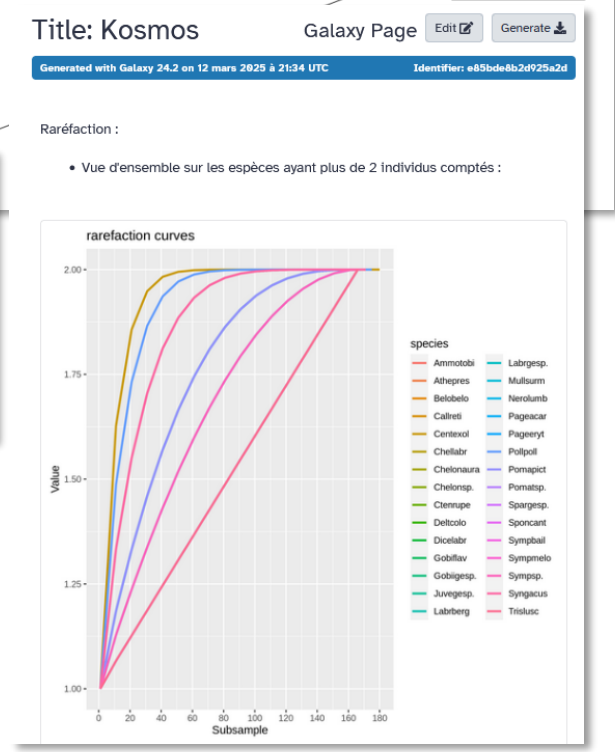
Science participatives

Support aux projets de sciences participatives









Science participatives

Crowdsourcing / annotation de données par la foule

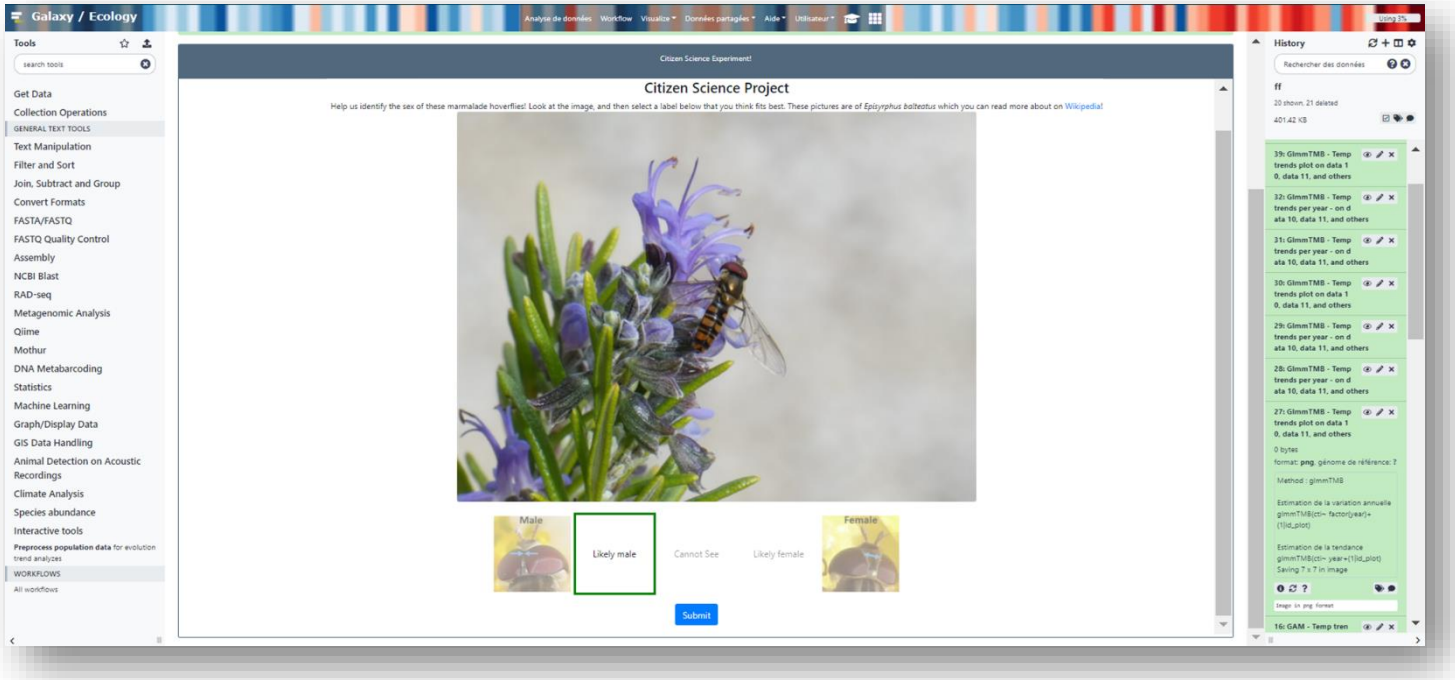
2020-2022

Citizen Science for Serious Gamers

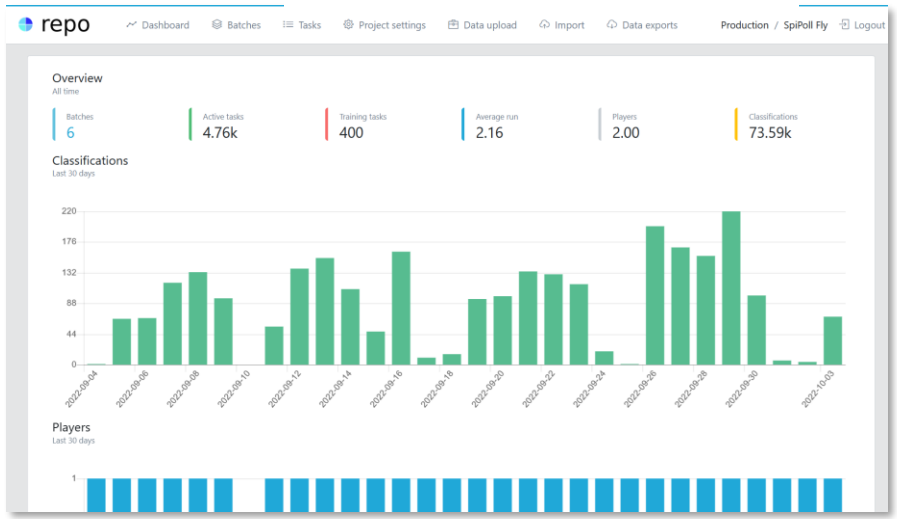
GAPARS project

Gamification of participatory science for training and education purposes

MOODA concept (Massively Open Online Data Analysis)



Crowdsourcing with hoverflies (syphres) images from **SPIPOLL** project



203 690 classifications

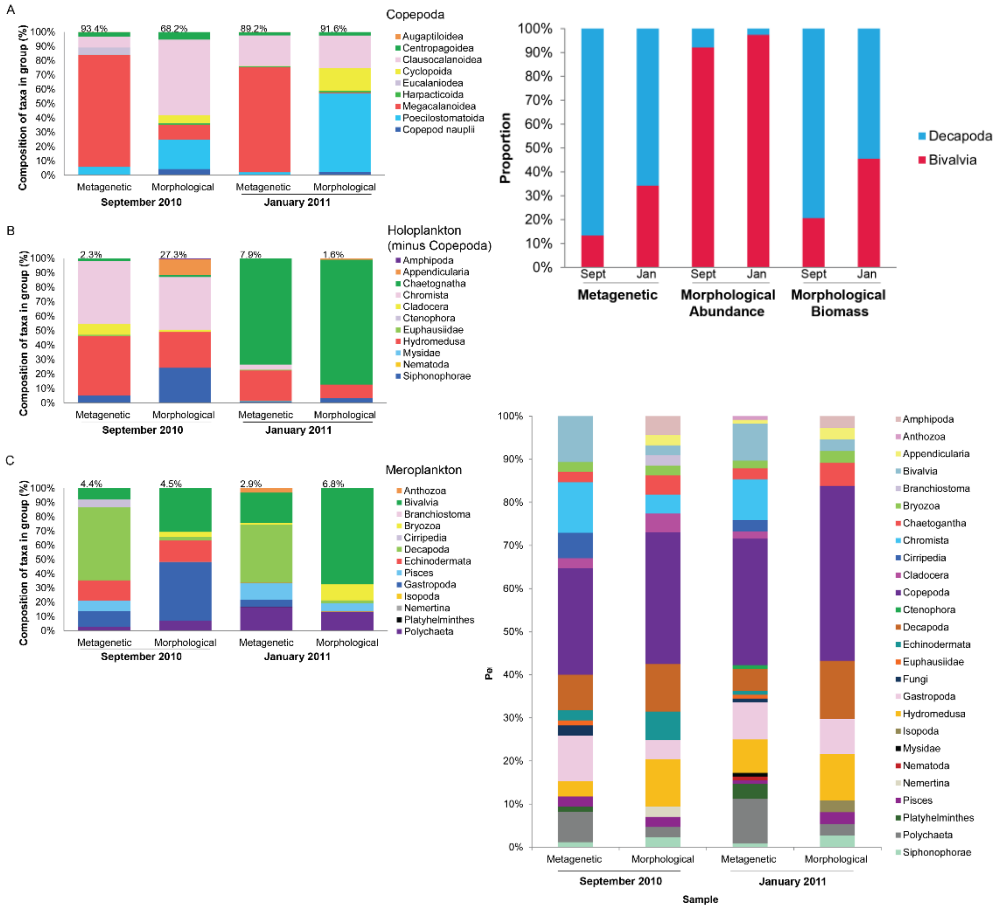
Science participatives

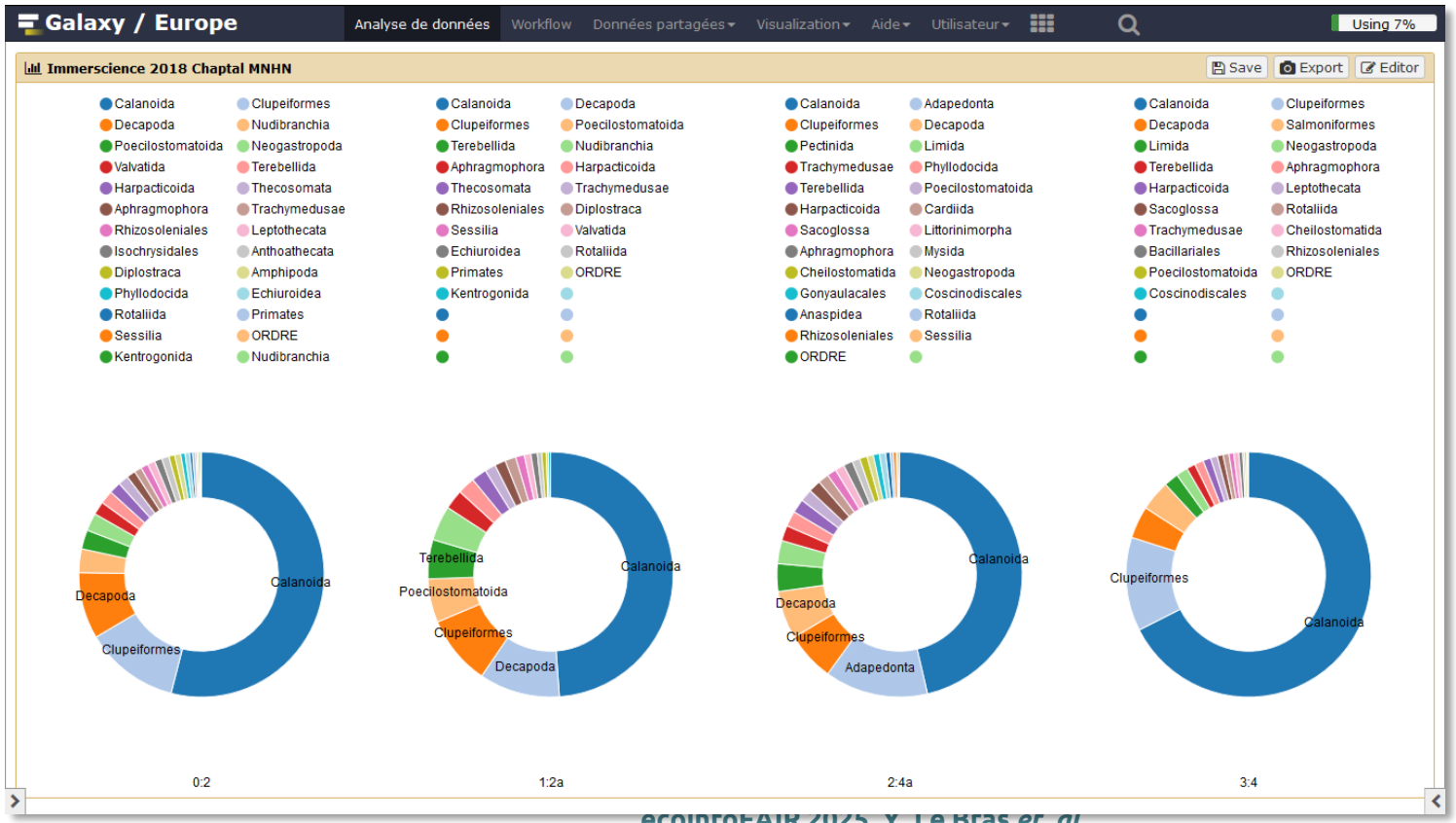
Analyse complexe de données par les citoyens

Next Generation Sequencing Reveals the Hidden Diversity of Zooplankton Assemblages

Penelope K. Lindeque, Helen E. Parry, Rachel A. Harmer, Paul J. Somerfield, Angus Atkinson

Published: November 7, 2013 • https://doi.org/10.1371/journal.pone.0081327

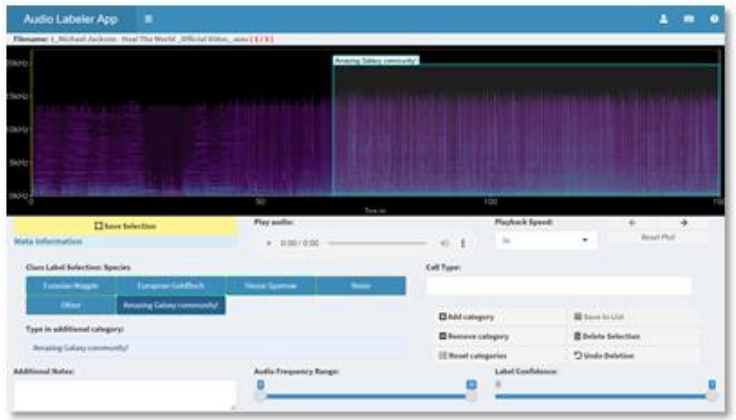


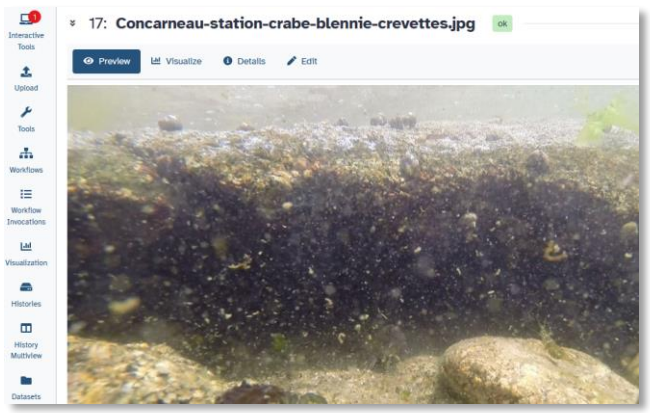


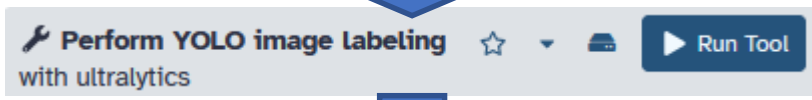
ecoinfoFAIR 2025, Y. Le Bras *et. al*

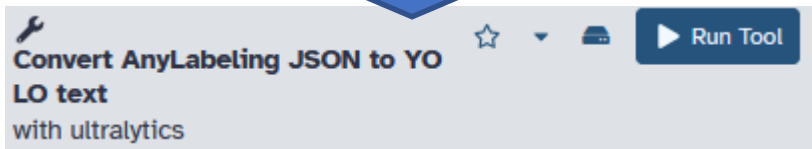
Annotations et IA

Adapter des outils existant pour nos cas d'usage







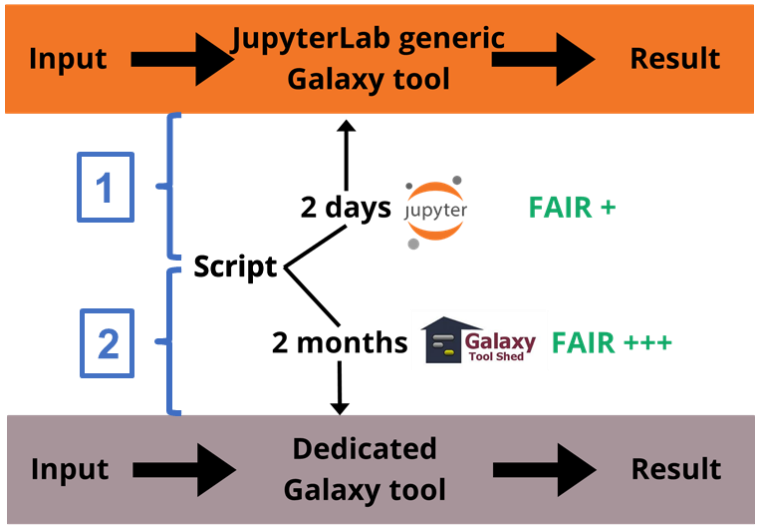


MOOREV

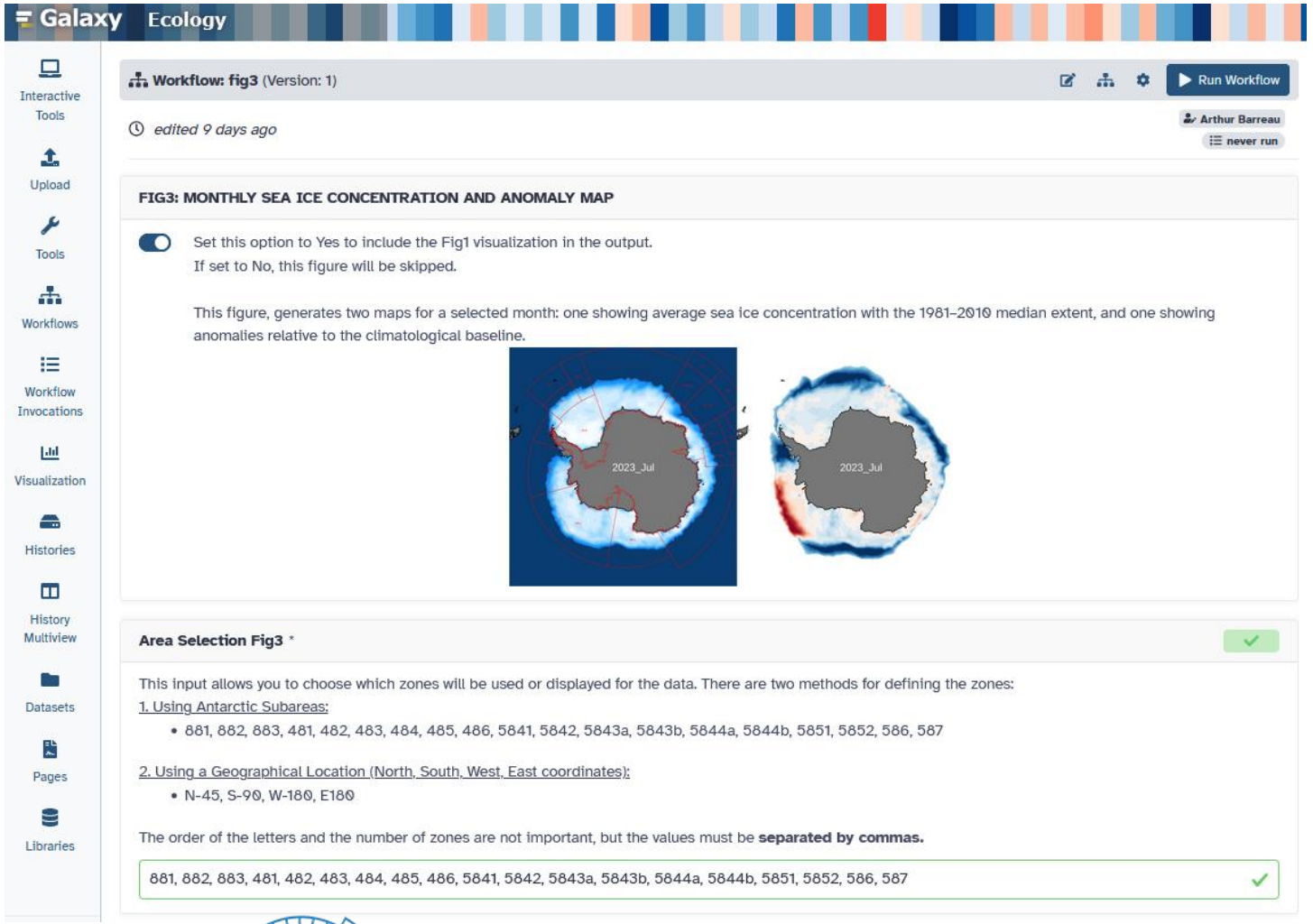
Reproduire les analyses à partir d'article de recherche

Nouvelle méthode pour créer et partager des workflows

Une nouvelle approche à 2 étapes pour proposer des workflows Galaxy à partir de scripts bruts / codes sources



Un workflow pour automatiser la génération d'un rapport sur l'état des écosystèmes Antarctique

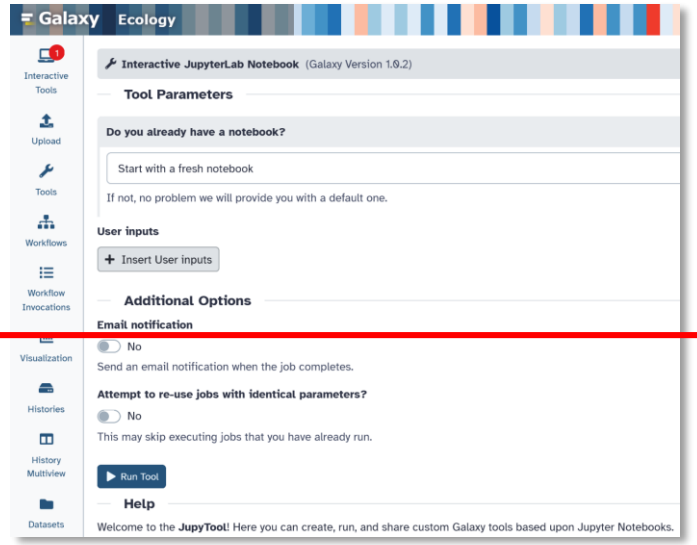


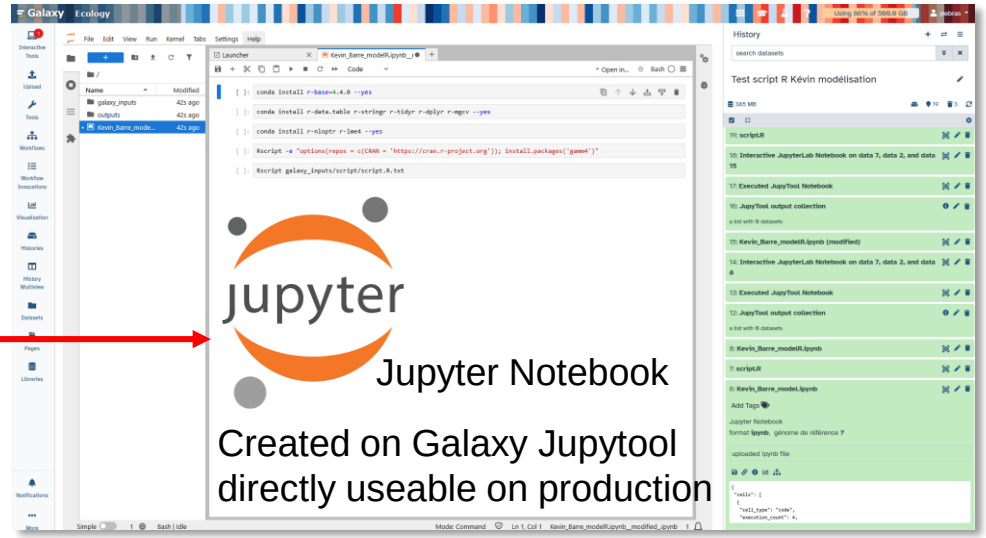
CCAMLR

ecoinfoFAIR 2025, Y. Le Bras *et. al*

Nouvelle méthode pour créer et partager des workflows

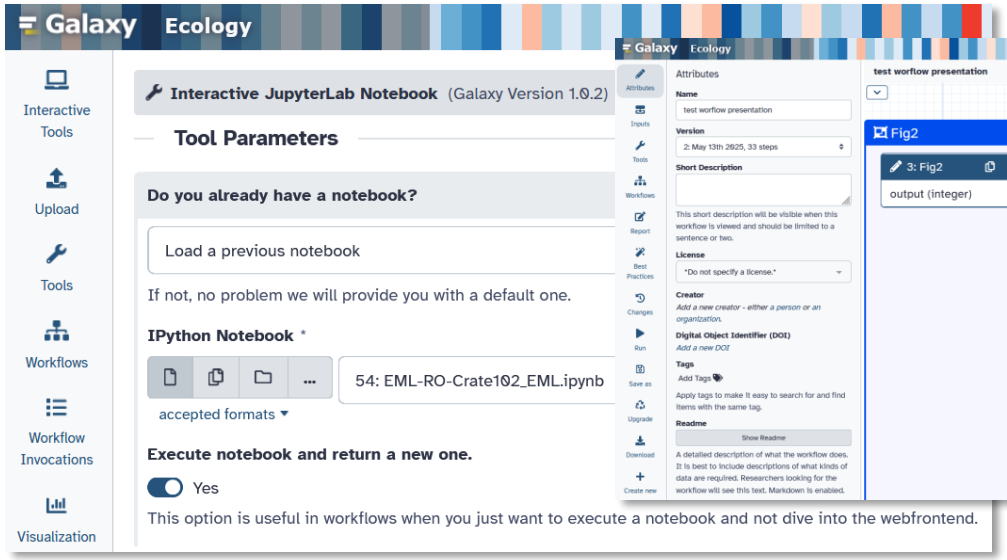
```
library(data.table)
library(stringr)
library(tidyr)
library(dplyr)
library(mgcv)
library(gamm4)

# ** dataset  ----
data <- fread("galaxy_inputs/P10/P10_ER90_Final_COMPLETED.csv.txt")
data <- data[,-1]
data <- as.data.frame(data)
data <- data[-which(data$id_site_annee %in% c("81124-E04_2021", "81124-E03_2021")),]
colnames(data)[25] <- "diametre_rotor"
colnames(data)[26] <- "hauteur_nacelle"

data$MRE <- as.numeric(data$Pippip_90)+
  as.numeric(data$Pippyg_90)+
  as.numeric(data$Pipnat_90)+
  as.numeric(data$Hypsav_90)+
  as.numeric(data$Pipkuh_90)+
  as.numeric(data$Minsch_90)

data$MRE_binom <- rep(NA,dim(data)[1])
data$MRE_binom[which(data$MRE > 0)] = 1
data$MRE_binom[which(data$MRE == 0)] = 0

data$materiel_tot <- paste0(data$materiel_corrige,"_",data$`trigger dB`)

data_10 <- data %>%
  drop_na(jday,
          pourcentage_nuit,
          rpm_60min,
          vent_60min,
          temperature_60min,
          rr_60min,
          annee,
          Code_eolienne_meteo,
          materiel_tot)

# ** Model  ----

m <- gamm(MRE_binom ~ s(jday, k=4), family = binomial, data = data_10, random=list(annee=~1))
m <- gamm4(MRE_binom ~ s(jday, k=4), family = binomial, data = data_10, random=~(1|annee))
```

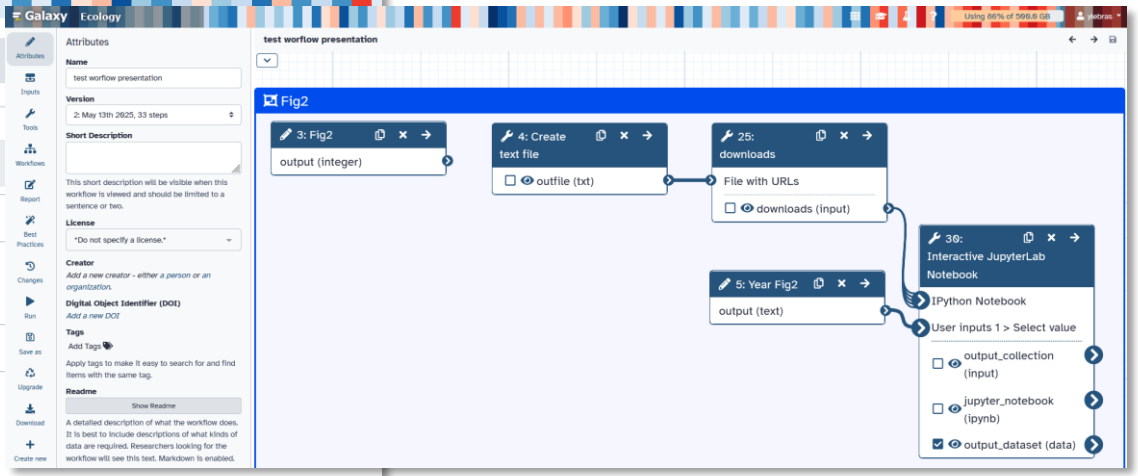


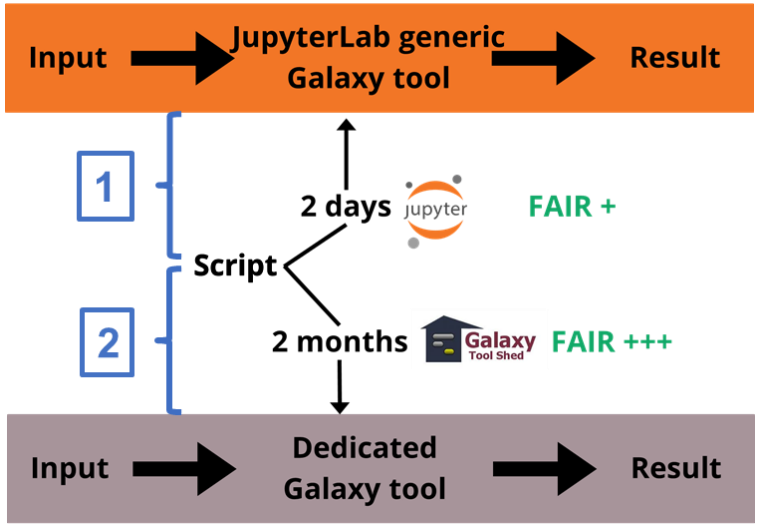


Exécution de Jupyter Notebook

Exécution de Jupyter Notebook dans workflows







Tutoriels en ligne dédiés système terre

GTN:S00127	Slides	An earth-system RO-Crate earth-system ro-crate ocean marine omics	FAIR Data, Workflows, and Research	1H
GTN:T00455	Hands-on	Ecoregionalization workflow tutorial ecology taxonomic data EBV workflow modeling gbif ocean earth-system interactive-tools biodiversity	Ecology	2H
GTN:T00450	Hands-on	Marine Omics identifying biosynthetic gene clusters earth-system ocean marine omics biodiversity	Ecology	3H
GTN:T00367	Hands-on	QGIS Web Feature Services earth-system GIS Geographical Information System WFS Spatial data Maps OGC biodiversity	Ecology	0H30M
GTN:T00400	Hands-on	From NDVI data with OpenEO to time series visualisation with Holoviews earth-system land degradation NDVI copernicus holoviews	Ecology	1H
GTN:T00332	Hands-on	Obis marine indicators earth-system ocean marine omics biodiversity	Ecology	1H
GTN:T00529	Hands-on	Adding file-sources to Galaxy earth-system ocean data work-in-progress	Galaxy Server	
GTN:T00473	Hands-on	Create a subdomain for your community on UseGalaxy.eu subdomain community earth-system interactive-tools		

GTN:T00476	Hands-on	Make your tools available on your subdomain subdomain earth-system interactive-tools	Galaxy Community Building	30m
GTN:T00375	Hands-on	Ocean Data View (ODV) earth-system ocean geographical information system ODV netcdf data maps marine data	Climate	1H
GTN:T00457	Hands-on	Getting your hands-on earth data earth-system ocean land atmosphere biodiversity	Climate	1H
GTN:T00376	Hands-on	Ocean's variables study earth-system ocean netcdf data	Climate	1H
GTN:T00396	Hands-on	Sentinel 5P data visualisation earth-system volcano satellite data	Climate	1H
GTN:T00548	Hands-on	Ocean's variables study earth-system ocean biogeochemical work-in-progress	Climate	1H
GTN:T00426	Hands-on	Analyse Argo data earth-system ocean pangeo argo	Climate	1H

https://tinyurl.com/galaxy-es

Equipe Galaxy système terre

Alessandro Rizzo (IRD)

Marie Jossé (CNRS)

Erwan Bodéré (IFREMER)

Jérôme Détoc (IFREMER)

Sylvain Grellet (BRGM / One Water)

Eric Lecaude (BRGM / One Water)

Gwenaël Le Caër (IFREMER)

CDD BRGM One Water

Yvan Le Bras (MNHN)

Pauline Seguineau (MNHN)

Karim Ramage (IPSL)

Annexes

Présentation martyre / provisoire !

et pourquoi pas vous ?

Navigateurs, scientifiques, citoyens et décideurs tous unis pour l'Océan

EN MER

Collecte, numérisation, transmission

FairScope

We,EXPLORE.

Plankton Planet

A TERRE

Réception, traitement et analyse

EcoTaxa 2.7

DATA TERRA

OneWeb

FDJ UNITED

PlayforPlankton

Présentation martyre / provisoire !

EXPLORE.

PROGRAMME SCIENCES CITOYENNES

Etapes du process & parties prenantes

COLLECTE PLANCTON → NUMERISATION RESULTATS → TRANSMISSION DONNEES → RECEPTION DONNEES → TRAITEMENT DONNEES → ANALYSES DONNEES → DIFFUSION DONNEES

We,EXPLORE.

FairScope

Plankton Planet

esa European Space Agency

CONNECTED BY eutelsat ONEWEB

EcoTaxa 2.7

COMMUNAUTE SCIENTIFIQUE

GAIA Data

OBIS OCEAN BIODIVERSITY INFORMATION SYSTEM

CITOYENS Application

PLAY PLANKTON

FDJ UNITED

SORBONNE UNIVERSITÉ

cnrs

MUSÉUM NAL HIST NATURELLE

SCIENTIFIQUES & GRAND PUBLIC

2025

ROUTE DU RHUM DESTINATION GUADELOUPE

2026

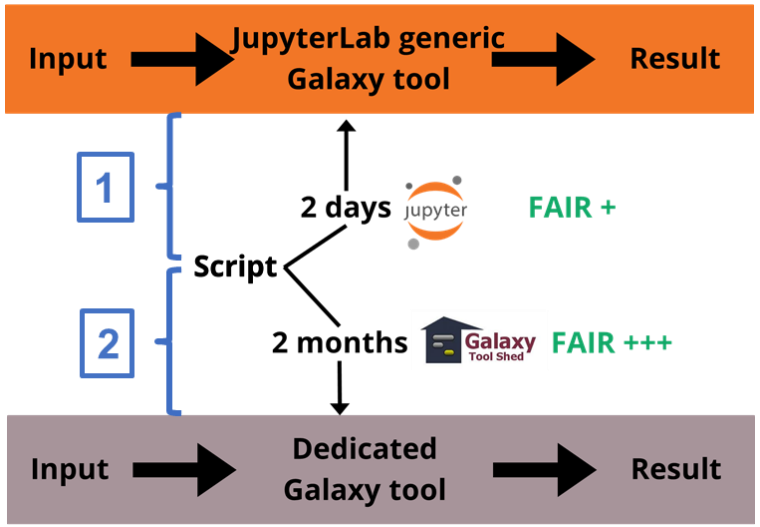
Support numérique

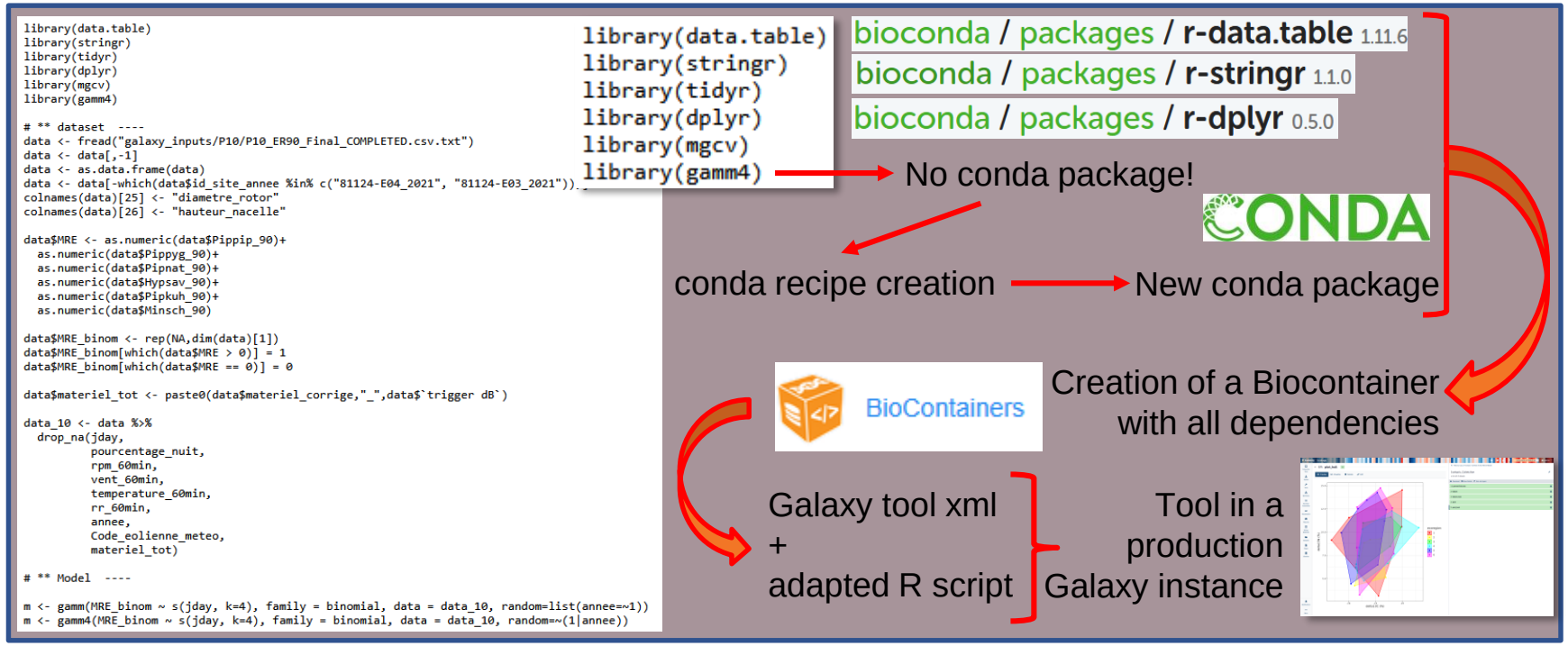
DATA TERRA

Outils et services

Nouvelle méthode pour créer et partager des workflows

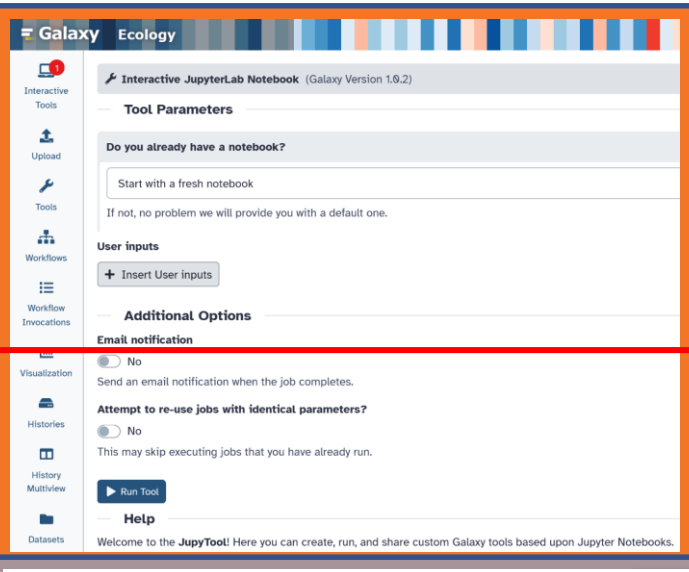
Une nouvelle approche à 2 étapes pour proposer des workflows Galaxy à partir de scripts bruts / codes sources

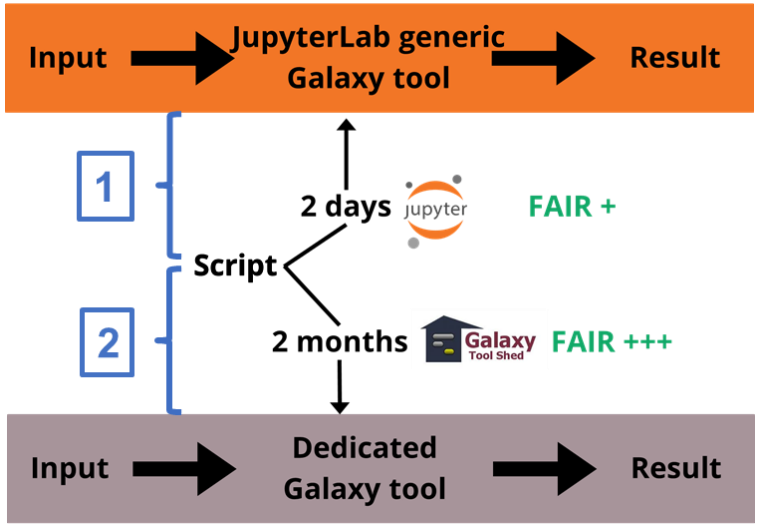




Nouvelle méthode pour créer et partager des workflows

```
library(data.table)
library(stringr)
library(tidyr)
library(dplyr)
library(mgcv)
library(gamm4)

# ** dataset  ----
data <- fread("galaxy_inputs/P10/P10_ER90_Final_COMPLETED.csv.txt")
data <- data[,-1]
data <- as.data.frame(data)
data <- data[-which(data$id_site_annee %in% c("81124-E04_2021", "81124-E03_2021")),]
colnames(data)[25] <- "diametre_rotor"
colnames(data)[26] <- "hauteur_nacelle"

data$MRE <- as.numeric(data$Pippip_90)+
  as.numeric(data$Pippyg_90)+
  as.numeric(data$Pipnat_90)+
  as.numeric(data$Hypsav_90)+
  as.numeric(data$Pipkuh_90)+
  as.numeric(data$Minsch_90)

data$MRE_binom <- rep(NA,dim(data)[1])
data$MRE_binom[which(data$MRE > 0)] = 1
data$MRE_binom[which(data$MRE == 0)] = 0

data$materiel_tot <- paste0(data$materiel_corrige,"_",data$`trigger dB`)

data_10 <- data %>%
  drop_na(jday,
          pourcentage_nuit,
          rpm_60min,
          vent_60min,
          temperature_60min,
          rr_60min,
          annee,
          Code_eolienne_meteo,
          materiel_tot)

# ** Model  ----

m <- gamm(MRE_binom ~ s(jday, k=4), family = binomial, data = data_10, random=list(annee=~1))
m <- gamm4(MRE_binom ~ s(jday, k=4), family = binomial, data = data_10, random=~(1|annee))
```






```
library(data.table)
library(stringr)
library(tidyr)
library(dplyr)
library(mgcv)
library(gamm4)

# ** dataset  ----
data <- fread("galaxy_inputs/P10/P10_ER90_Final_COMPLETED.csv.txt")
data <- data[,-1]
data <- as.data.frame(data)
data <- data[-which(data$id_site_annee %in% c("81124-E04_2021", "81124-E03_2021")),]
colnames(data)[25] <- "diametre_rotor"
colnames(data)[26] <- "hauteur_nacelle"

data$MRE <- as.numeric(data$Pippip_90)+
  as.numeric(data$Pippyg_90)+
  as.numeric(data$Pipnat_90)+
  as.numeric(data$Hypsav_90)+
  as.numeric(data$Pipkuh_90)+
  as.numeric(data$Minsch_90)

data$MRE_binom <- rep(NA,dim(data)[1])
data$MRE_binom[which(data$MRE > 0)] = 1
data$MRE_binom[which(data$MRE == 0)] = 0

data$materiel_tot <- paste0(data$materiel_corrige,"_",data$`trigger dB`)

data_10 <- data %>%
  drop_na(jday,
          pourcentage_nuit,
          rpm_60min,
          vent_60min,
          temperature_60min,
          rr_60min,
          annee,
          Code_eolienne_meteo,
          materiel_tot)

# ** Model  ----

m <- gamm(MRE_binom ~ s(jday, k=4), family = binomial, data = data_10, random=list(annee=~1))
m <- gamm4(MRE_binom ~ s(jday, k=4), family = binomial, data = data_10, random=~(1|annee))
```

```
library(data.table)
library(stringr)
library(tidyr)
library(dplyr)
library(mgcv)
library(gamm4)
```

bioconda / packages / **r-data.table** 1.11.6

bioconda / packages / **r-stringr** 1.1.0

bioconda / packages / **r-dplyr** 0.5.0

No conda package!

CONDA

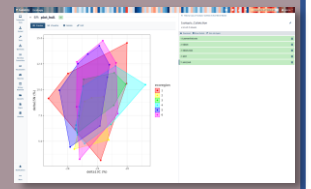
conda recipe creation → New conda package

BioContainers

Creation of a Biocontainer with all dependencies

Galaxy tool xml
+
adapted R script

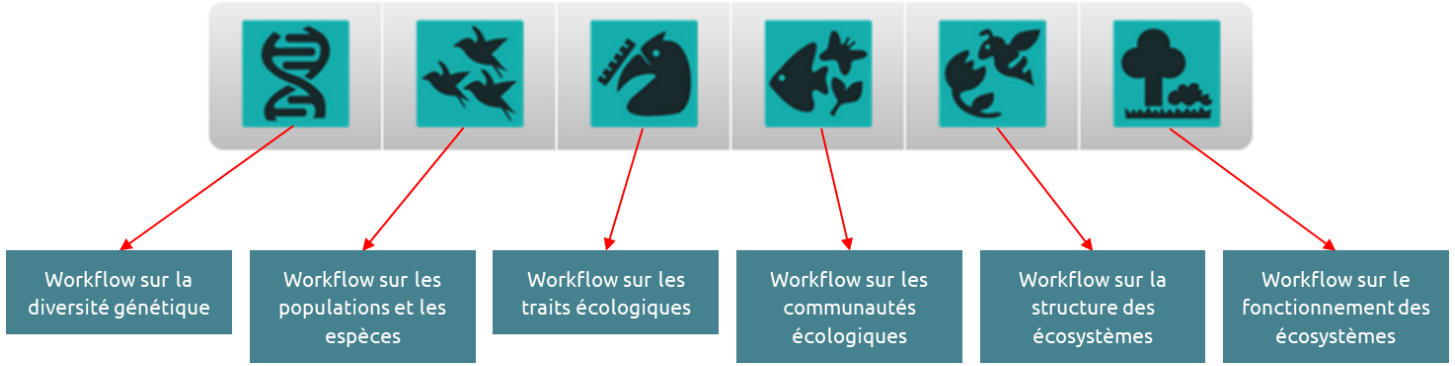
Tool in a production Galaxy instance



Besoins utilisateurs vs outils « analyse / traitement de données »

Opérationnalisation des EBVs

Mise à disposition workflows de production de métriques de biodiversité



https://www.pndb.fr/fr/activites/projets-techniques-et-scientifiques/ebvosc#Workflow%20examples

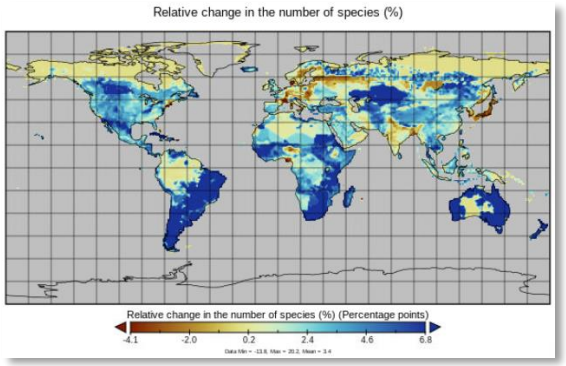
Accès banques de données internationales (occurrence, génétiques, satellites, climat..)

Lien recherche - expertise

Formation

Visualisation de données

Analyse scientifique de données



Ecosystème « Galaxy Ecology »

Galaxy / Ecology

Galaxy Training!

RStudio

jupyter

Besoins utilisateurs vs outils « analyse / traitement de données »

Opérationnalisation des EBVs

Mise à disposition workflows de production de métriques de biodiversité

Accès banques de données internationales (occurrence, génétiques, satellites, climat..)

Lien recherche - expertise

Formation

Visualisation de données

Analyse scientifique de données

Get species occurrences data from GBIF, OBIS, ALA, iNAT and others

OBIS occurences retrieve data

Argo data access for global ocean in situ observing system

SdmPredictors List Layers from environmental predictors for species distribution modeling

Adds environment variables From Copernicus and etopo given geolocalized and timestamped observations

Retrieve bold Functions to search in Bold and download the available sequences of each subtaxa (get_fasta)

Download and Extract Reads in FASTQ format from NCBI SRA

Sentinel-5P Level 2	Observations from the Sentinel-5 Precursor satellite of the Copernicus Earth Observation Programme. It contains a polar orbiting satellite that completes 14 orbits of the Earth a day.
Coupled Model Intercomparison Project 6	The sixth phase of global coupled ocean-atmosphere general circulation model ensemble
CMIP6 GCMs downscaled using WRF	High-resolution historical and future climate simulations from 1980-2100
GBIF European region public datasets	The Global Biodiversity Information Facility is an international network and data infrastructure aimed at providing anyone, anywhere, open access to data about all types of life on Earth.
NOAA Global Forecast System (GFS)	The Global Forecast System (GFS) is a weather forecast model produced by the National Centers for Environmental Prediction (NCEP).
NOAA Unified Forecast System Subseasonal to Seasonal Prototype 5	The Unified Forecast System Subseasonal to Seasonal prototype 5 (UFS S2Sp5) dataset is reforecast data from the UFS atmosphere-ocean.
Copernicus Digital Elevation Model (DEM)	The Copernicus DEM is a Digital Surface Model (DSM) which represents the surface of the Earth including buildings, infrastructure and vegetation.
ECMWF ERA5 Reanalysis	ERA5 is the fifth generation of ECMWF atmospheric reanalyses of the global climate created with the best available observation data from satellites, in-situ stations and ECMWF's Integrated Forecast System. The dataset provides atmospheric meteorological parameters.
Atmospheric Models from Météo-France	Global and high-resolution regional atmospheric models from Météo-France.

Galaxy / Ecology

Galaxy Training!

RStudio

jupyter

Besoins utilisateurs vs outils « analyse / traitement de données »

Opérationnalisation des EBVs

Mise à disposition workflows de production de métriques de biodiversité

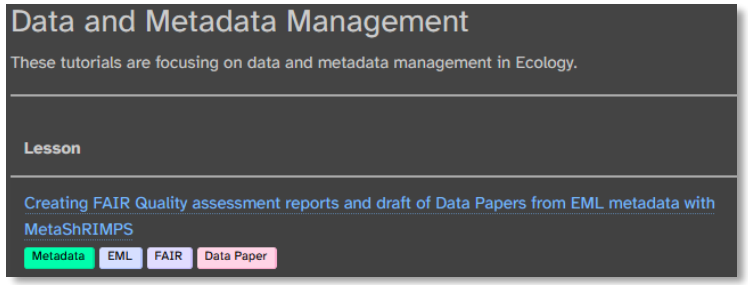
Accès banques de données internationales (occurrence, génétiques, satellites, climat..)

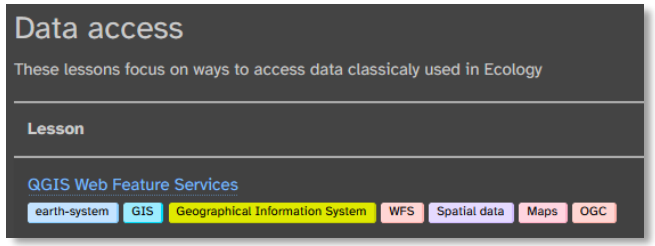
Lien recherche - expertise

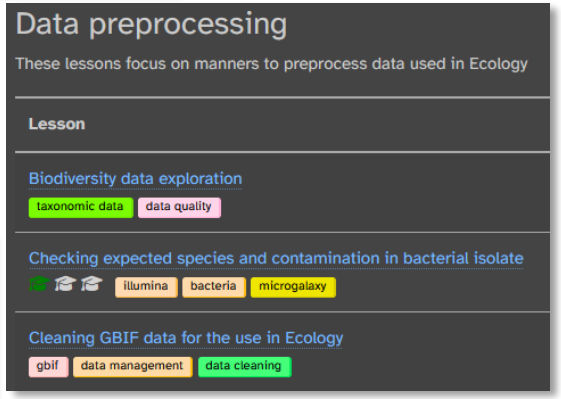
Formation

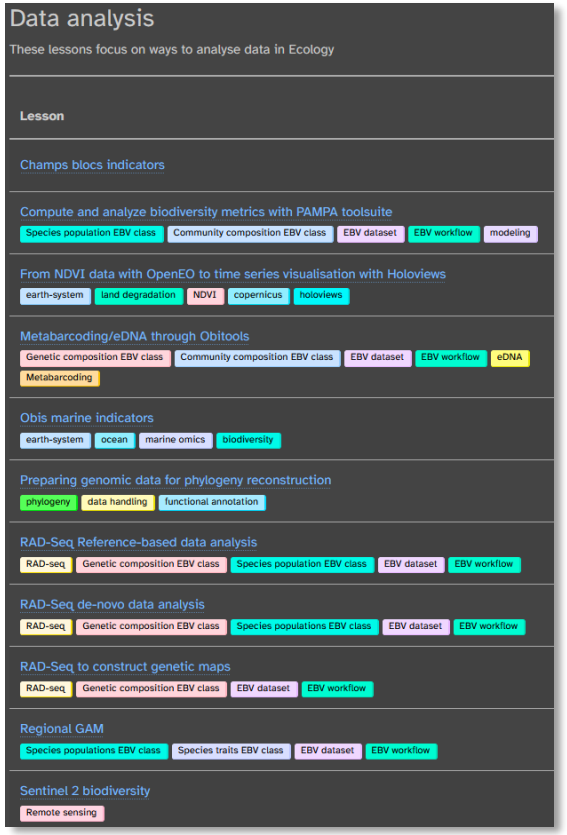
Visualisation de données

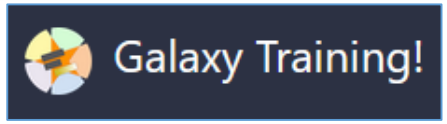
Analyse scientifique de données

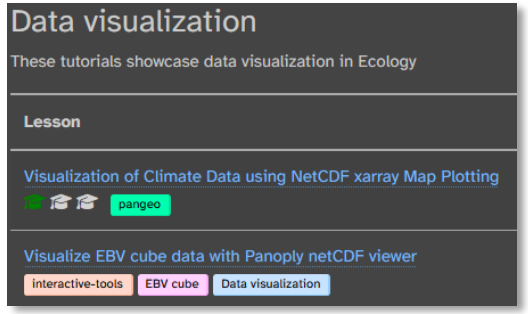


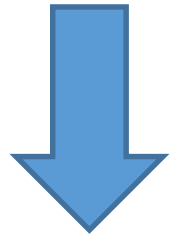












https://training.galaxyproject.org/training-material/topics/ecology/

Besoins utilisateurs vs outils « analyse / traitement de données »

Opérationnalisation des EBVs

Mise à disposition workflows de production de métriques de biodiversité

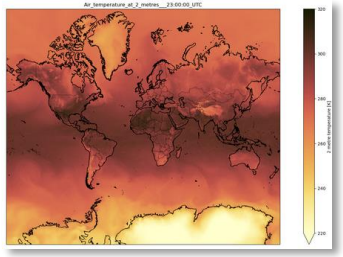
Accès banques de données internationales (occurrence, génétiques, satellites, climat..)

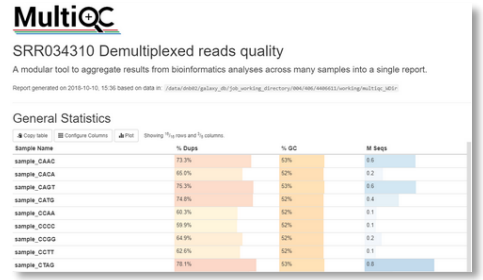
Lien recherche - expertise

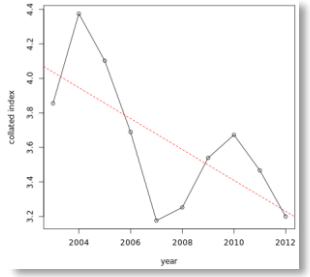
Formation

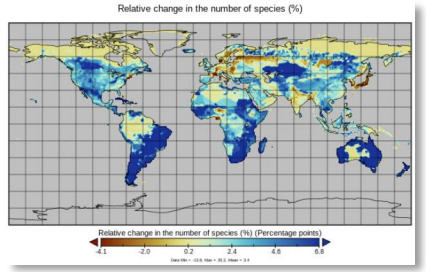
Visualisation de données

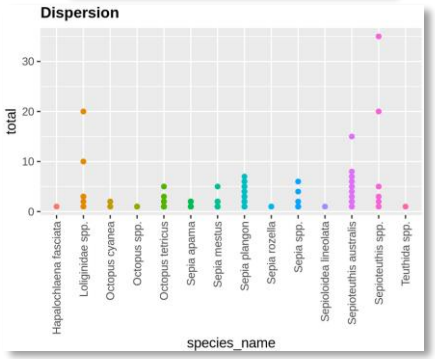
Analyse scientifique de données

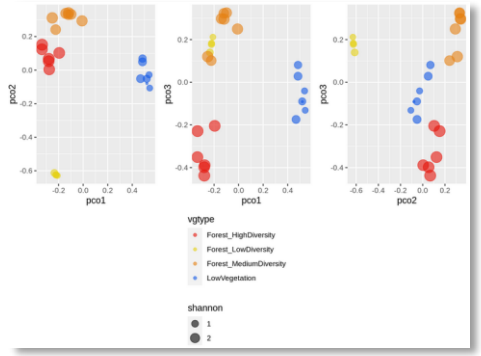


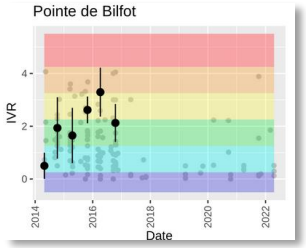


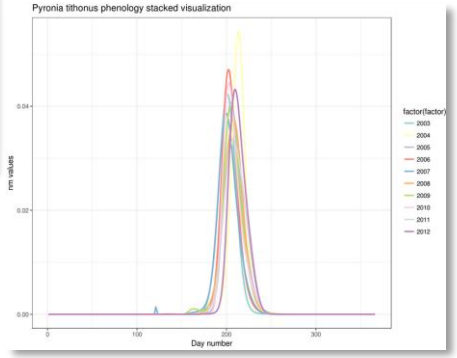


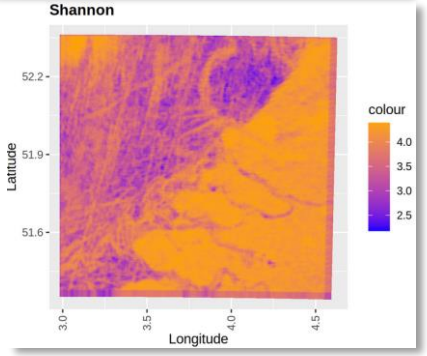






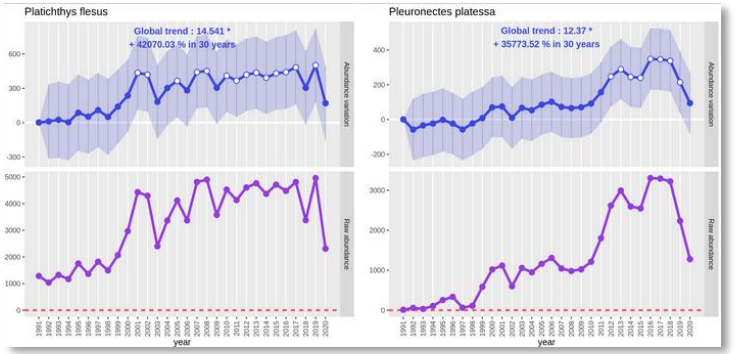


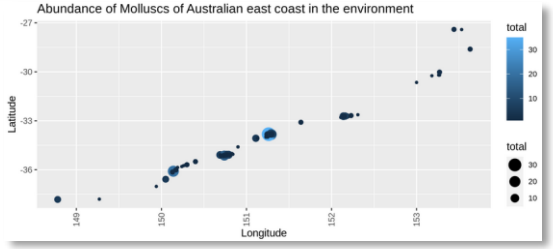




Galaxy / Ecology

Galaxy Training!

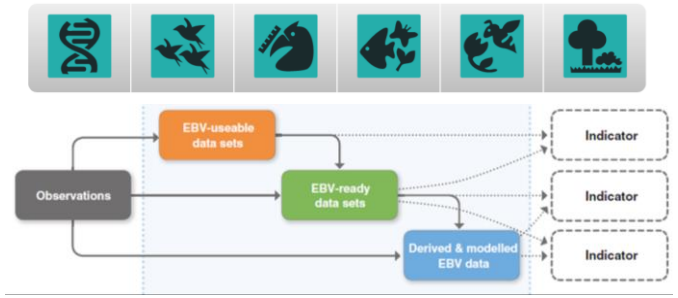




MOTTO

Rely on what already exists and has proven efficiency

Concepts and principles



GO FAIR

Adapt and apply
to help identify gaps and bias

Data

Use and reuse
to exploit full potential

Analytical processes

Atomize and generalize
to get plasticity and reproducibility

Technical solutions & related networks

Galaxy PROJECT

DataONE

Share and contribute
to enhance communication and networks

Global Open Science Cloud | **EBVOSC case study**

Build National BONs → **DigitalBON -> GBiOS**