

0% song guaranteed



Pôle National de Données de Biodiversité



Galaxy-E: Ecological data analysis, citizen science and biodiversity indicators production!

@ColineRoyaux #PNDB @Yvan2935

Coline Royaux

Yvan Le Bras

yvan.le-bras@mnhn.fr



Context – We need Atomization

Currently, in ecology ...

One R script for one input datafile

```
      ,direction="wide")
tab[is.na(tab)] <- 0
# filename <- "touverUnNom"
# chemin <- paste(rep,filename,sep="/")
# write.table(tab, chemin)
colnames(tab) <- sub("nombre.", "", colnames(tab))
  return(tab)
}

## sous jeux de donnees si choix d espece d annee ou d un pourcentage de carres
makeSousTab <- function(tab,vecSp=NULL,echantillon=1,
  methodeEchantillon="carre",vecannees=NULL) {
  cat(" -- Fabrication du sous jeu de donnees --\n")
  flush.console()
  ## reduction de la table à certaine espèces
  if(!is.null(vecSp)) {
    cat(" selection",length(vecSp),"espece(s):\n -> ")
    cat(vecSp)
    cat("\n")
    tab <- data.frame(carre = tab$carre,annee = tab$annee,tab[,vecSp])
    colnames(tab) <- c("carre","annee",vecSp)
  }
  ## reduction de la table pour certaines annees
  if(!is.null(vecannees)) {
    tab <- subset(tab,annee>=vecannees[1] & annee <= vecannees[2])
  }
  ## reduction de la table par une proportion de carre suivie
  if(echantillon != 1) {
    if(echantillon < 1 & echantillon >0) {
      nbinit <- nrow(tab)
      if(methodeEchantillon == "global") {
        nb <- round(nrow(tab)*echantillon)
        cat(" echantillonage",echantillon*100,
          "% des donnees par la methode",methodeEchantillon,"\n")
        cat(" -> conservation de",nb,"lignes sur",nbinit,"\n")
        flush.console()
        tab <- tab[sample(1:nrow(tab))[1:nb],]
      } else {
        if (methodeEchantillon == "carre") {
          cat(" echantillonage",echantillon*100,
            "% des carrees par la methode",methodeEchantillon,"\n")
          nbcarreinit <- length(unique(tab$carre))
          chat=sample(unique(tab$carre),
            length(unique(tab$carre))*echantillon,replace=F)
          cat(" -> conservation de",length(chat),"carrees sur",
            nbcarreinit)
          tab=subset(tab, subset = carre %in% chat)
          cat(" (",nrow(tab)," lignes sur ",nbinit,")\n",sep="")
        } else {
          stop("Methode d echantillonage non reconnue")
        }
      }
    }
  }
}
```

Context – We need Atomization

Currently, in ecology ...

One R script for one input datafile

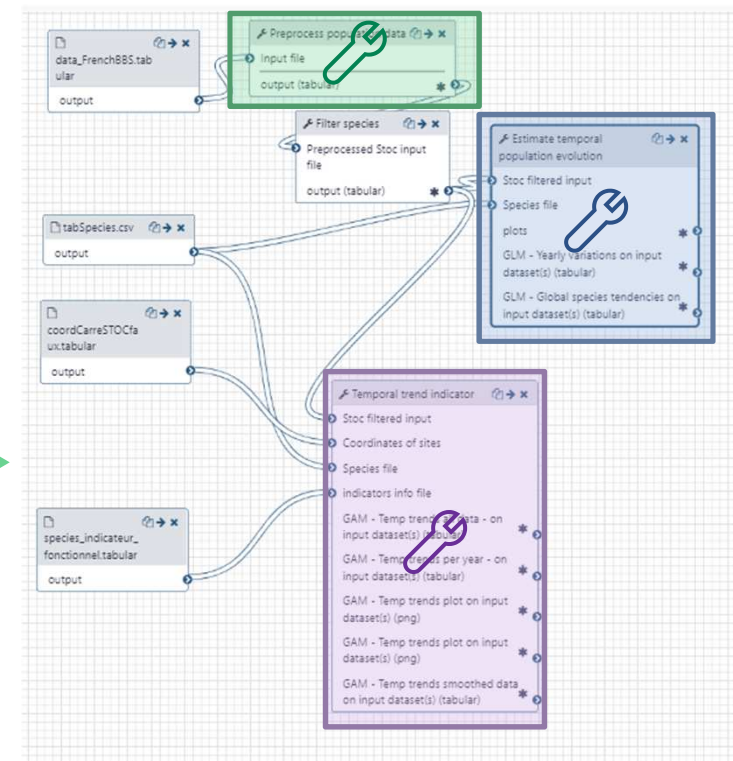
```
tab[is.na(tab),direction="wide")
tab[is.na(tab)] <- 0
# filename <- "touverUnNom"
# chemin <- paste(rep,filename,sep="/")
# write.table(tab, chemin)
colnames(tab) <- sub("nombre.", "", colnames(tab))
return(tab)

## sous jeux de donnees si choix d espece d annee ou d un pourcentage de carres
makesousstab <- function(tab,vecsp=NULL,echantillon=1,
                         methodeEchantillon="carre",vecannees=NULL) {
  cat(" -- Fabrication du sous jeu de donnees --\n")
  flush.console()
  ## reduction de la table à certaines espèces
  if(!is.null(vecSp)) {
    cat(" selection",length(vecSp), "espece(s):\n -> ")
    cat(vecSp)
    cat("\n")
    tab <- data.frame(carre = tab$carre,annee = tab$annee,tab[,vecSp])
    colnames(tab) <- c("carre","annee",vecSp)
  }
  ## reduction de la table pour certaines annees
  if(!is.null(vecannees)) {
    tab <- subset(tab,annee>=vecannees[1] & annee <= vecannees[2])
  }
  ## reduction de la table par une proportion de carres suivis
  if(echantillon != 1) {
    if(echantillon < 1 & echantillon > 0) {
      nbinit <- nrow(tab)
      if(methodeEchantillon == "global") {
        nb <- round(nrow(tab)*echantillon)
        cat(" echantillonnage",echantillon*100,
            "% des donnees par la methode",methodeEchantillon,"\n")
        cat(" -> conservation de",nb,"lignes sur",nbinit,"\n")
      } else {
        if (methodeEchantillon == "carre") {
          cat(" echantillonnage",echantillon*100,
              "% des carrees par la methode",methodeEchantillon,"\n")
          nbcarrereinit <- length(unique(tab$carre))
          chat=sample(unique(tab$carre),
                      length(unique(tab$carre))*echantillon,replace=F)
          cat(" -> conservation de",length(chat),"carrees sur",
              nbcarrereinit)
          tab=subset(tab, subset = carre %in% chat)
          cat(" (",nrow(tab)," lignes sur ",nbinit,")\n",sep="")
        } else {
          stop("Methode d echantillonnage non reconnue")
        }
      }
    }
  }
}
```



With Galaxy...

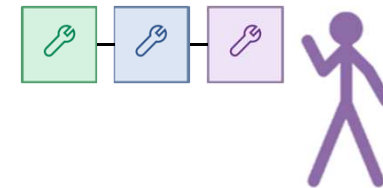
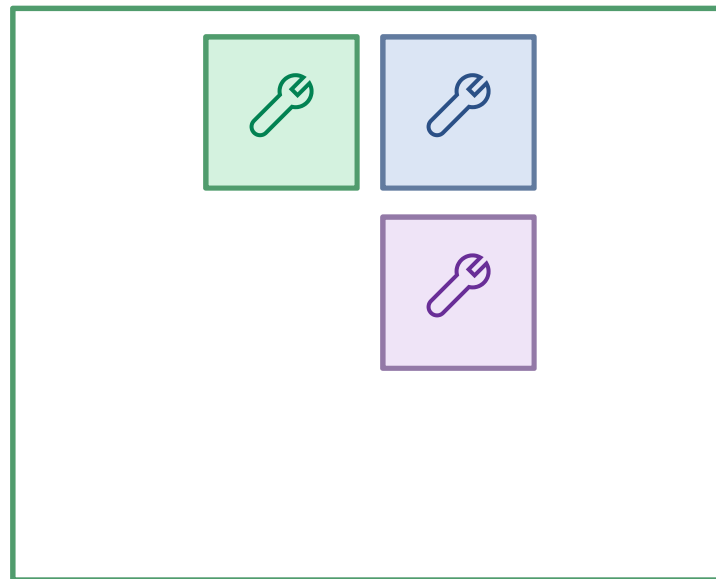
Several atomized R scripts for several input datafiles



Context – We need Sharing & Generalization

	participation	Nuit	num_micro	groupe	espece	nb_contacts
1	55de2cd52121b1000d27430e	2015-07-26	0	bat	Barbar	1
2	55de2cd52121b1000d27430e	2015-07-26	0	bush-cricket	Barfis	1
3	55de2cd52121b1000d27430e	2015-07-26	0	noise	noise	5022
4	55de2cd52121b1000d27430e	2015-07-26	0	bush-cricket	Decalb	5
5	55de2cd52121b1000d27430e	2015-07-26	0	bush-cricket	Tyllil	18
6	55de2cd52121b1000d27430e	2015-07-26	0	bat	Nyclei	1
7	55de2cd52121b1000d27430e	2015-07-26	0	bush-cricket	Phanan	269

Toolshed

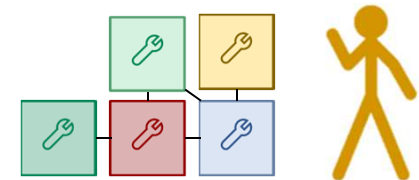
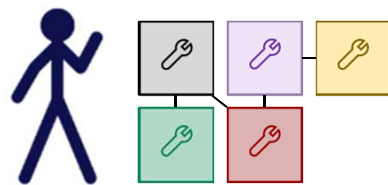
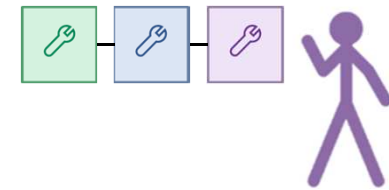
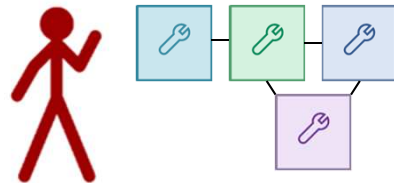
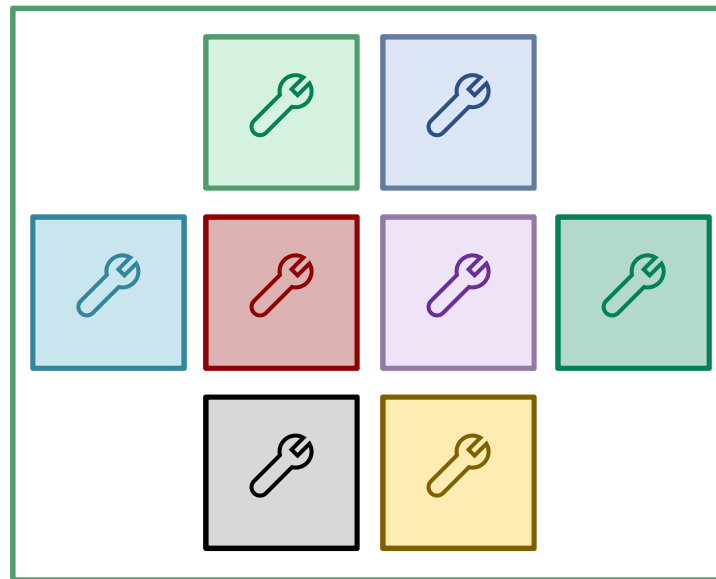


Context – We need Sharing & Generalization

	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
7	183	2018	ACCGEN	0
8	183	2019	ACCGEN	0

	participation	Nuit	num_micro	groupe	espece	nb_contacts
1	55de2cd52121b1000d27430e	2015-07-26	0	bat	Barbar	1
2	55de2cd52121b1000d27430e	2015-07-26	0	bush-cricket	Barfis	1
3	55de2cd52121b1000d27430e	2015-07-26	0	noise	noise	5022
4	55de2cd52121b1000d27430e	2015-07-26	0	bush-cricket	Decalb	5
5	55de2cd52121b1000d27430e	2015-07-26	0	bush-cricket	Tyllil	18
6	55de2cd52121b1000d27430e	2015-07-26	0	bat	Nyclei	1
7	55de2cd52121b1000d27430e	2015-07-26	0	bush-cricket	Phanan	269

Toolshed



	Unit	bs	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
5	AS140099	2		Parumult	-999	-999	P	-999	1
6	AS140088	1		Varilout	-999	-999	P	-999	1
7	AS140088	2		Gompvari	-999	-999	P	-999	2
8	AS140041	1		Nasosp.	-999	-999	P	-999	5

	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	0.088277
6	BITS	1991	1	22	126417	Clupea harengus	95	0.037892
7	BITS	1991	1	22	126417	Clupea harengus	100	0.063293
8	BITS	1991	1	22	126417	Clupea harengus	105	0.012492
9	BITS	1991	1	22	126417	Clupea harengus	110	0.618357

And some trainings! => <https://training.galaxyproject.org/>

Climate

Learn to analyze climate data through Galaxy.

You can view the tutorial materials in different languages by clicking the dropdown icon next to the slides and tutorial buttons below.

Requirements

Before diving into this topic, we recommend you to have a look at:

- [Introduction to Galaxy Analyses](#)

Material

Search

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Introduction to climate data					
Functionally Assembled Terrestrial Ecosystem Simulator (FATES) interactive-tools					
Functionally Assembled Terrestrial Ecosystem Simulator (FATES) with Galaxy Climate JupyterLab interactive-tools					
Getting your hands-on climate data					
Pangeo ecosystem 101 for everyone - Introduction to Xarray Galaxy Tools pangeo					
Pangeo Notebook in Galaxy - Introduction to Xarray pangeo interactive-tools jupyter-notebook jupyter-notebook					
Visualize Climate data with Panoply netCDF viewer interactive-tools					

And some trainings! => <https://training.galaxyproject.org/>

Climate

Learn to analyze climate data through Galaxy.

You can view the tutorial materials in different languages by clicking the dropdown icon next to the slides (📄) and tutorial (📖) buttons below.

Requirements

Before diving into this topic, we recommend you to have a look at:

- [Introduction to Galaxy Analyses](#)

Material

Lesson

Introduction to climate data

Functionally Assembled Terrestrial Ecosystem Simulator (FATES)

interactive-tools

Functionally Assembled Terrestrial Ecosystem Simulator (FATES) with Galaxy Climate JupyterLab

interactive-tools

Getting your hands-on climate data

Pangeo ecosystem 101 for everyone - Introduction to Xarray Galaxy Tools

pangeo

Pangeo Notebook in Galaxy - Introduction to Xarray

pangeo

interactive-tools

jupyter-notebook

jupyter-notebook

Visualize Climate data with Panoply netCDF viewer

interactive-tools

Slides

Ecology

Learn to analyse Ecological data through Galaxy.

You can view the tutorial materials in different languages by clicking the dropdown icon next to the slides (📄) and tutorial (📖) buttons below.

Requirements

Before diving into this topic, we recommend you to have a look at:

- [Introduction to Galaxy Analyses](#)

Material

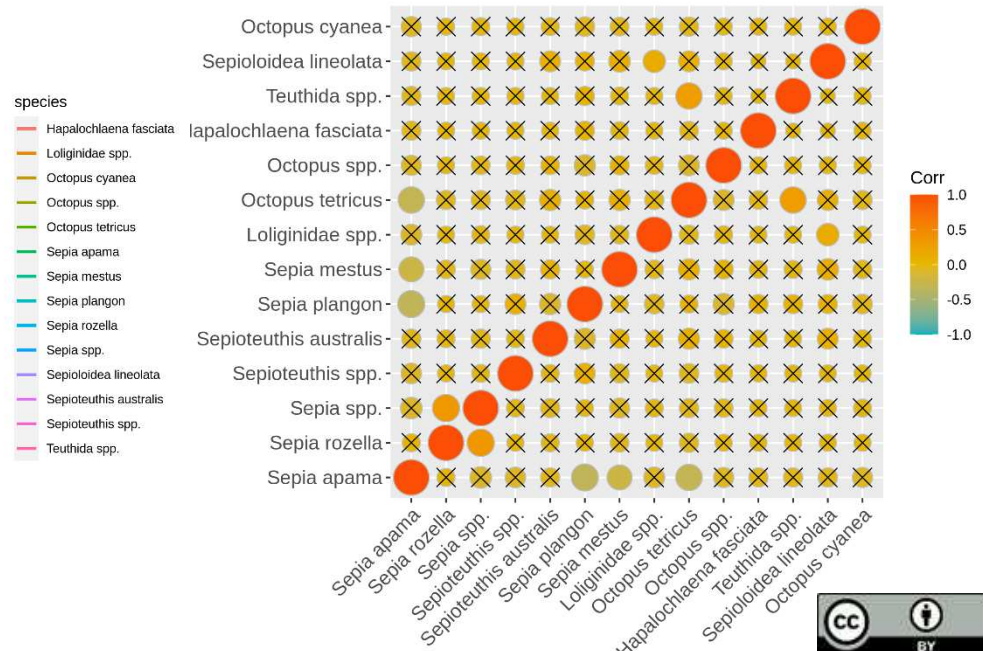
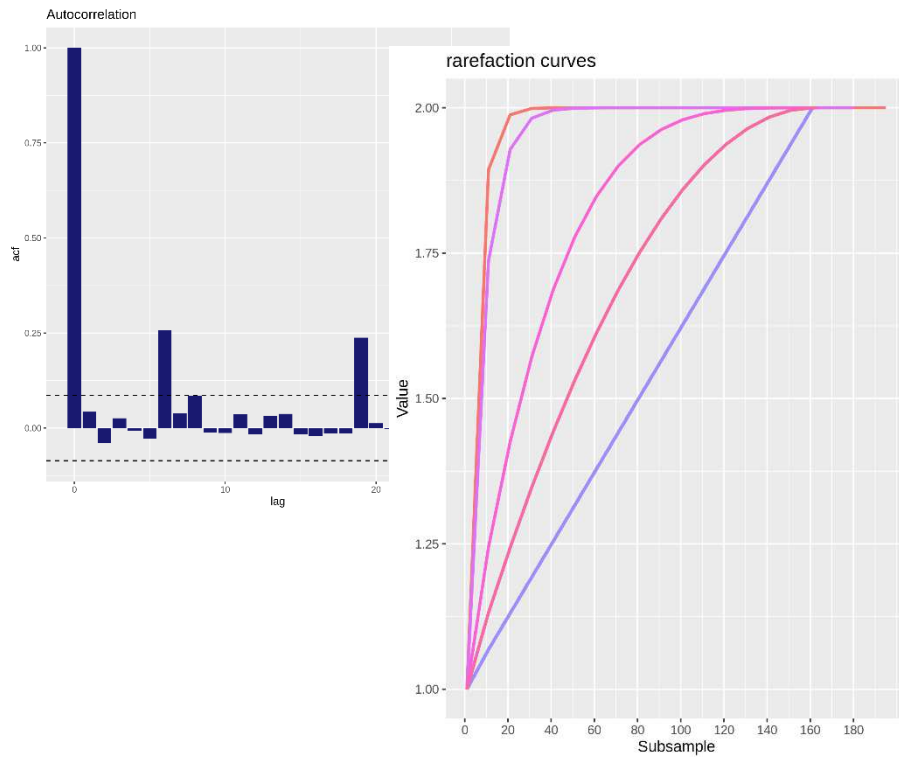
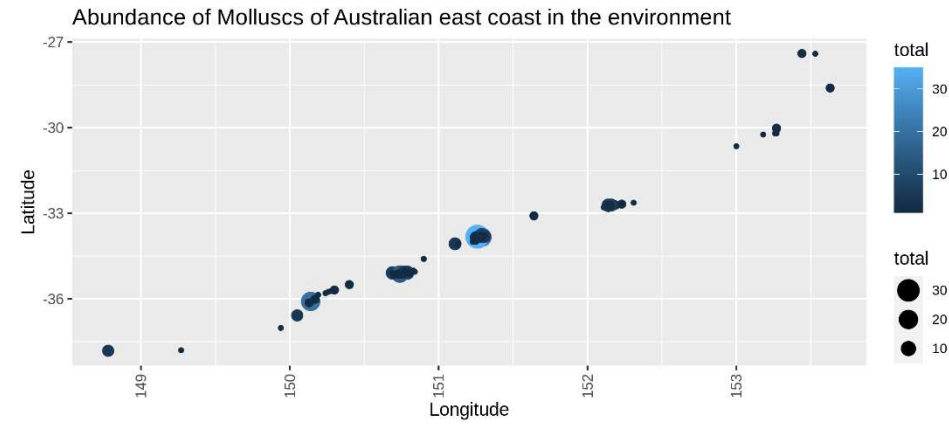
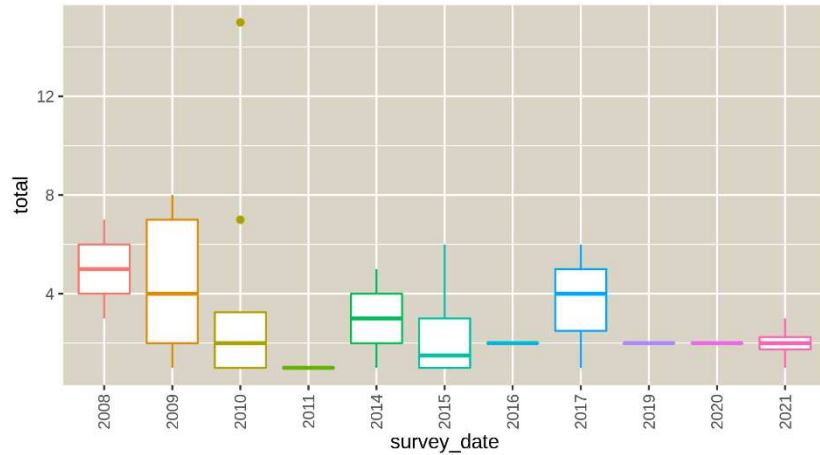
Search

Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Biodiversity data exploration					
Compute and analyze biodiversity metrics with PAMPA toolsuite					
Metabarcoding/eDNA through Obitoools					
RAD-Seq de-novo data analysis RAD-seq					
RAD-Seq Reference-based data analysis RAD-seq					
RAD-Seq to construct genetic maps RAD-seq					
Regional GAM					
Species distribution modeling interactive-tools					

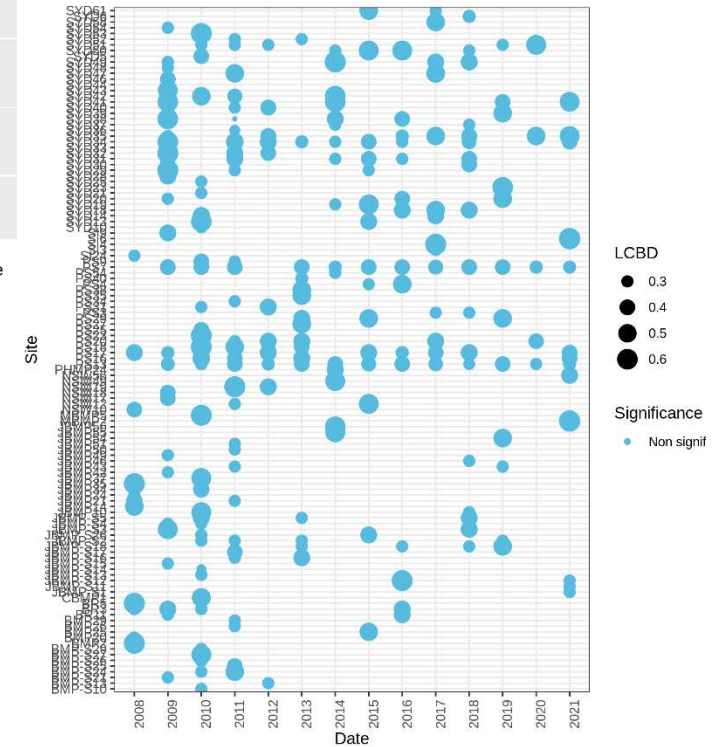
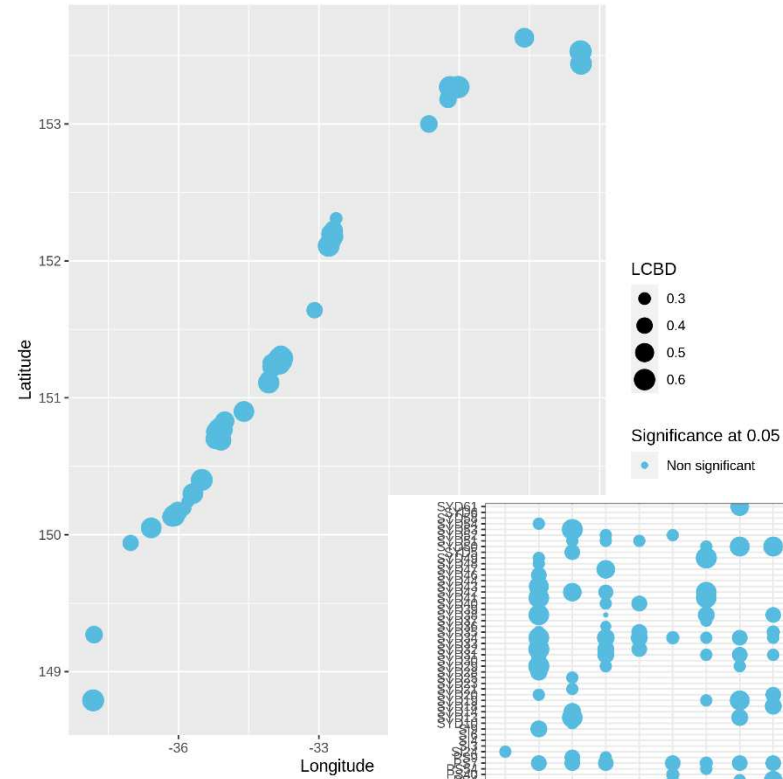
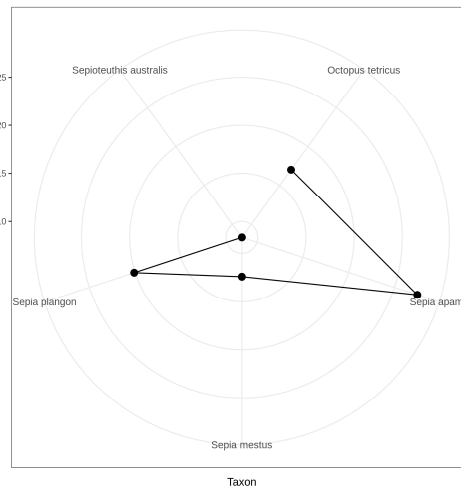
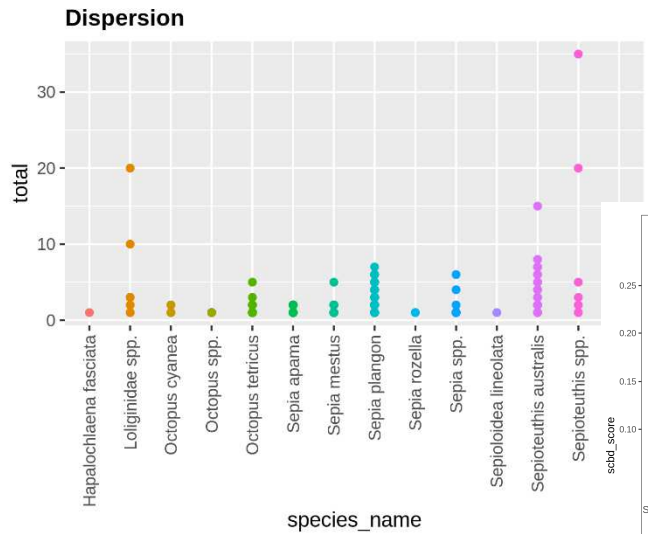
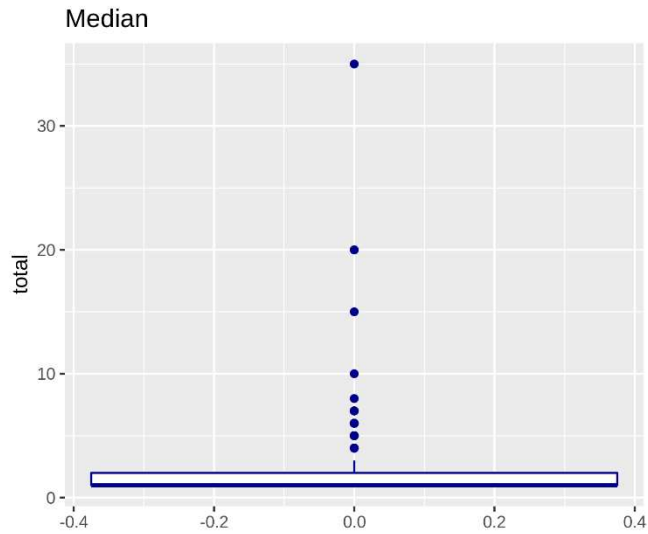
Galaxy-E Killer workflows

- Biodiversity exploration tools
- Biodiversity metrics & indicators production
- Dealing with GIS and netcdf files on Galaxy-E

Biodiversity exploration tools



Biodiversity exploration tools



Dealing with GIS and netcdf files

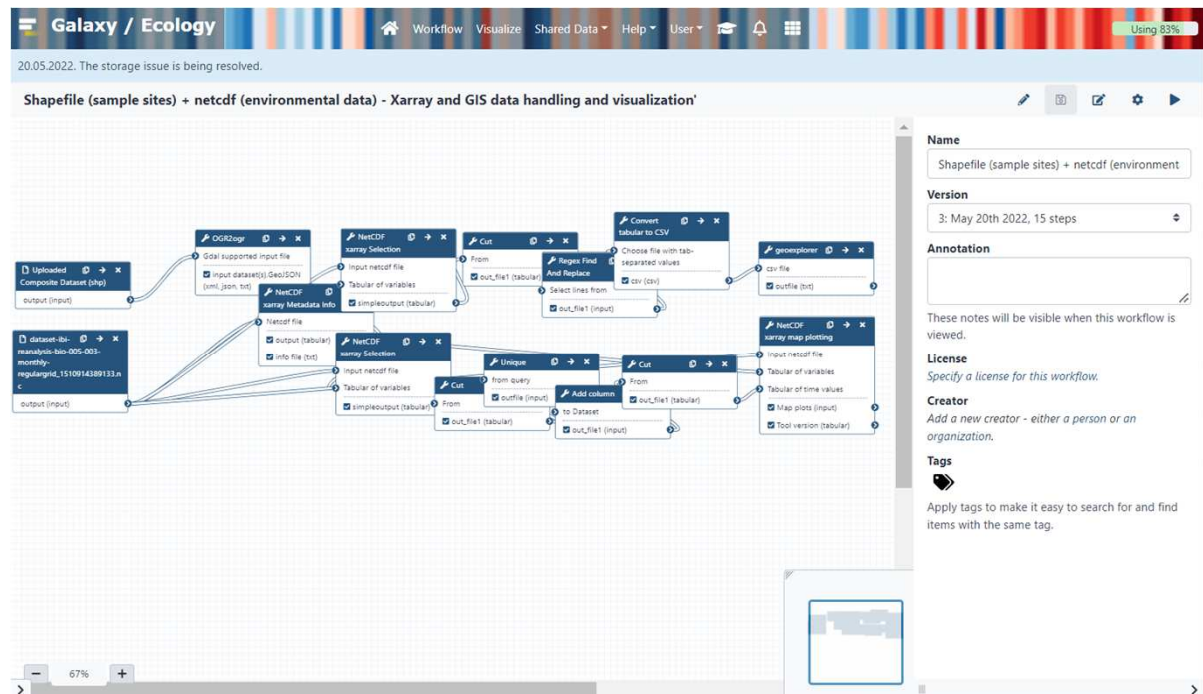
Fouilloux *et al.* EGU22 Pangeo for everyone with Galaxy

A “Classical” data processing:

Sampling sites information in GIS data file
(often shapefile)

Environmental information in netCDF file

Create a file with environmental
information on sampling sites!
Visualize maps of environmental
parameters on sampling sites



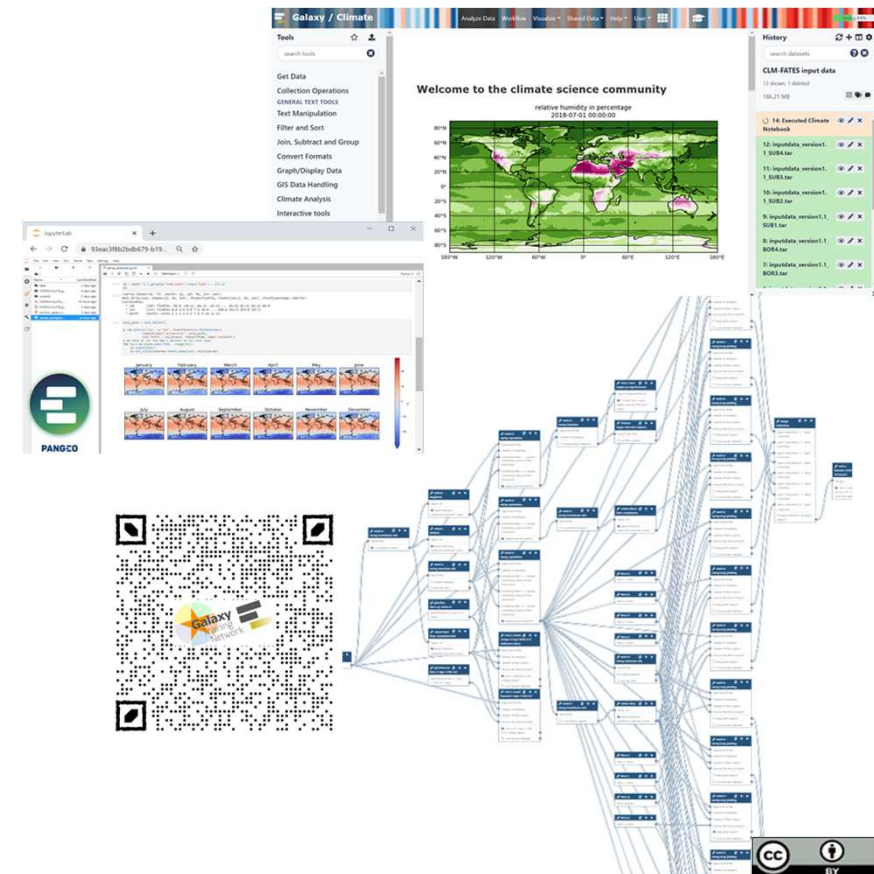
Until now: R + QGIS + a lot of manual manipulation

Now: a **Galaxy workflow** mixes scripts, GDAL & Xarray tools making it easily accessible and (re)-runnable.

Pangeo for everyone through Galaxy

Galaxy open-source platform for FAIR data analysis offers:

- Pangeo notebook deployment (local dask) **available to everyone (free registration)**;
- Pangeo Galaxy Tools for fully **automated workflows**;
- **GUI** for users with **no programming skills**;
- Self-Paced **Learning material** and organisation of online training events with the Galaxy Training Network;
- **Training Infrastructure as a Service** is a free and ready to use with private queues where only training's jobs run.



Pangeo for everyone through Galaxy

Galaxy open-source platform for FAIR data analysis offers:

- Pangeo notebook deployment (local dask) **available to everyone (free registration)**;
- Pangeo Galaxy Tools for fully **automated workflows**;
- **GUI** for users with **no programming skills**;
- Self-Paced **Learning material** and organisation of online training events with the Galaxy Training Network;
- **Training Infrastructure as a Service** is a free and ready to use with private queues where only training's jobs run.

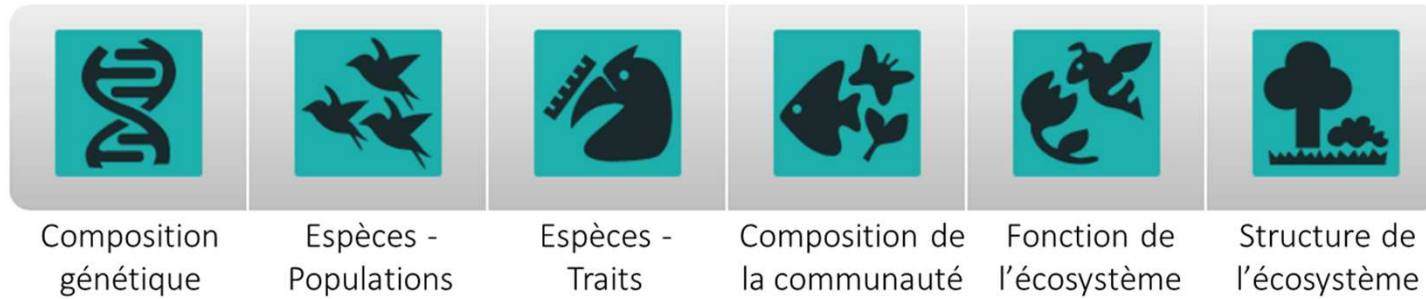


Amazing basis for  eosc | FAIR-EASE

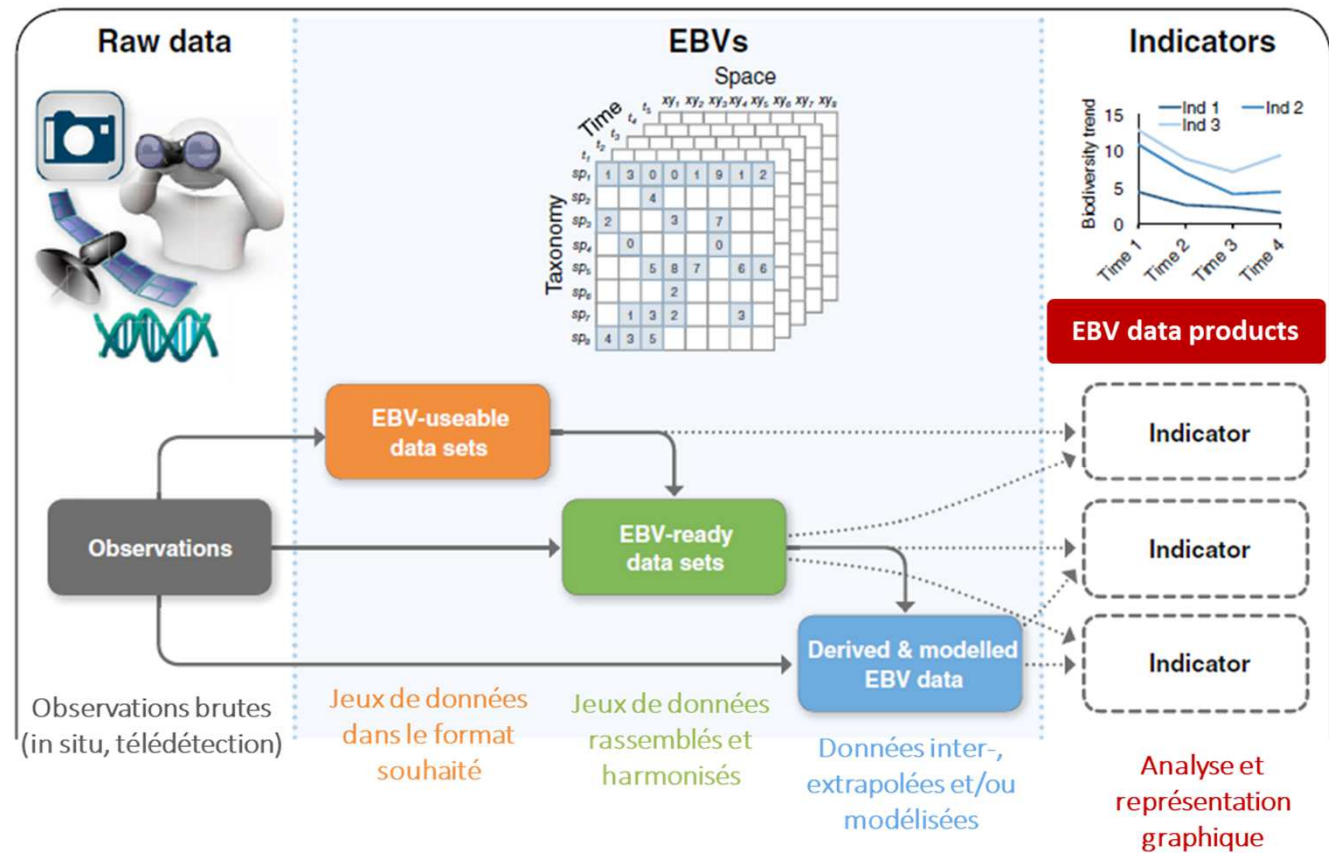
PNDB / Galaxy-E work for next months



Essential Biodiversity Variables workflows



Kissling *et al.* 2017



EBV workflows: STOC

<https://bit.ly/3AdFDFK>



1

Biodiversity data

Preprocess data

2

Filter data

Preprocess population data for evolution trend analyzes (Galaxy Version 0.0.1)

Input file
No tabular dataset available.
Population count file, with location, date, species and abundance.

STOC preprocess population data

Filter species with rare and low abundances (Galaxy Version 0.0.1)

Preprocessed Stoc input file
No tabular dataset available.
Output file from the "Preprocess population data tool"

STOC Filter species with rare and low abundances



Community

Analyze community indexes



Species - population

3

Analyze species abundance

Temporal trend indicator using Gini

Stoc filtered input
No tabular dataset available.

Input species tabular file, with 3 columns: Species ID, species name, species scientific name, specialization (status).

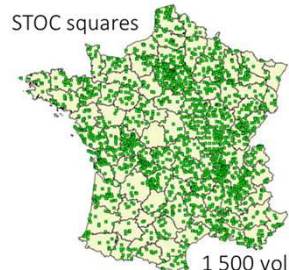
Indicators only file
No tabular dataset available.

Input indicator tabular file, with 3 columns: ID, indicator name, indicator status.

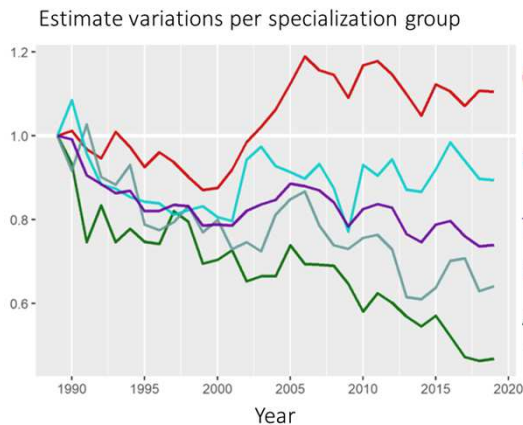
Choose the index you want to compute
Gini

Specify advanced parameters
No, use program defaults.

STOC Temporal population trend indicator



1 500 volunteers



Estimate temporal population evolution by specialization group (Galaxy Version 0.0.1)

Yearly variation dataset
No tabular dataset available.

Output from the "Estimate temporal population evolution by species" tool.

Global tendencies dataset
No tabular dataset available.

Output from the "Estimate temporal population evolution by species" tool.

Species file
No tabular dataset available.

Input species tabular file, with 5 columns: species ID, species name, species scientific name, specialization (status).

Specify advanced parameters
No, use program defaults.

Email notification
Yes No

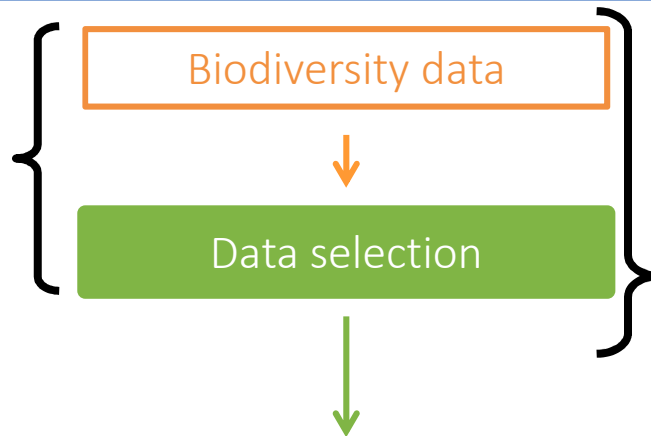
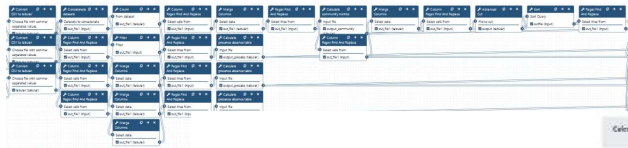
Send an email notification when the job completes.

STOC Estimate species population evolution

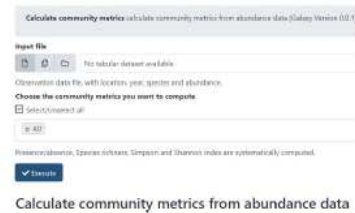


EBV workflows: PAMPA

Existing accessible & reuseable Galaxy tools
 convert / concatenate / Column
 Regex Find and Replace / Merge
 Columns / Filter / Count / Regex Find
 and Replace / Advanced Cut

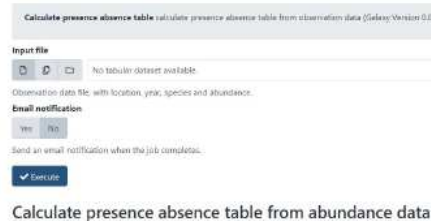


Pre-processed data



1
 Compute community metrics

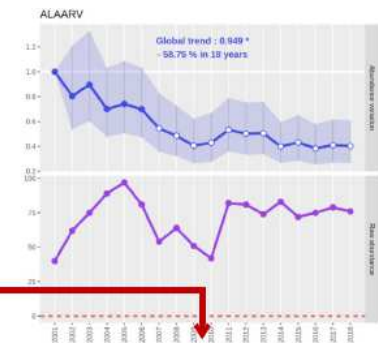
Compute population metrics



2
 GLM on community metrics

Metric ~ site + year + habitat

GLM on population metrics



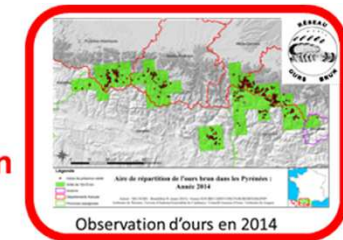
3
 Time-series plot from GLM results

Essential Biodiversity Variables workflows

Help BONs to identify gaps & reuse EBV workflows

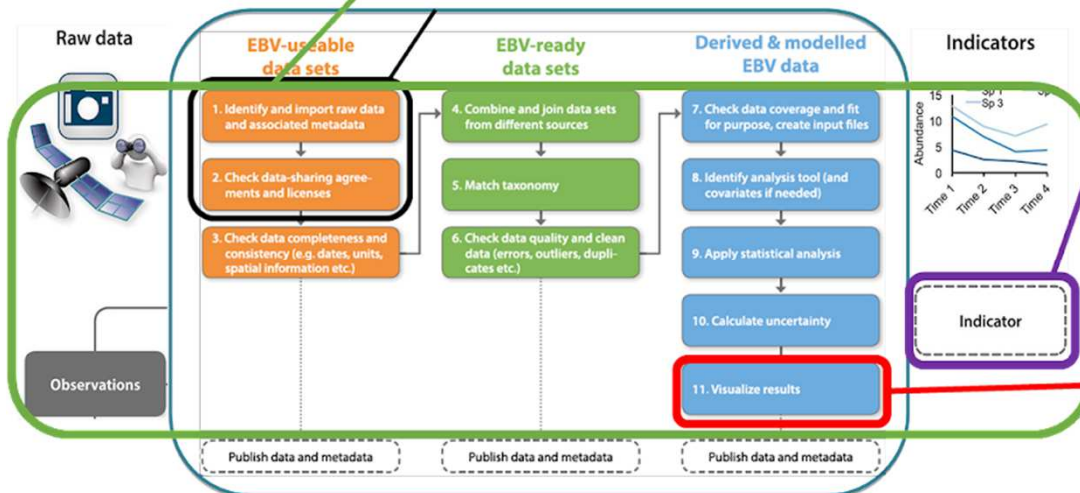
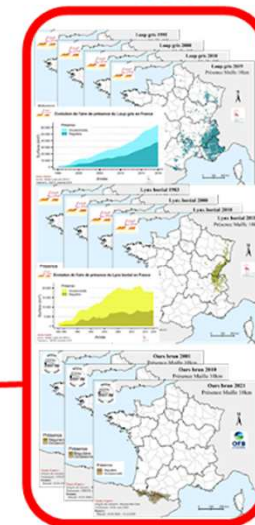
Help production and reuse of peer-reviewed Ebv workflows

Have an exhaustive vision of monitoring programs to identify gaps



?

Derived & modeled EBV file



Essential Biodiversity Variables workflows

Global Open Science Cloud (GOSC)

Help BONs to identify gaps & reuse EBV workflows

Case Studies

Amazing basis for **EBVOSC**



PNDB / Galaxy-E work for next months

Do you think this can help create a national biodiversity network on your country (Germany, Australian, ..) ? Contact us!

Not only for data analysis

=> Also for Research data management

Ecological research data management

Get species occurrences data from GBIF, ALA, INAT and others (Galaxy Version 0.9.0)

Scientific name of the species

Genus species format, eg : Canis lupus

Data source to get data from

Select/Unselect all

Any combination of gbif, bison, inat, ebird, antweb, ala, idigbio, obis, ecoengine and/or vertnet

Number of records to return

500

This is passed across all sources

Email notification

No

Send an email notification when the job completes.

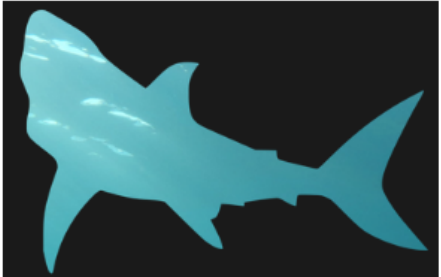
Get species occurrences data

What it does

Search species occurrences across a single or many data sources.

Import Biodiversity occurrences data

Generate metadata



MetaShARK

Generate Metadata from data using EML standard

Upload data and metadata

Convert metadata

xm1starlet convert a metadata XML file in one standard to another (Galaxy Version 1.6.1)

input xml file to convert

18: xm1starlet on data 2 and data 17: Converted xml

A xml file corresponding to a xsd schema you want to convert in another.

input xsl conversion file

2: iso2eml_all_in_one.xsl

A xsl file describing the mapping between a first xsd specification to another.

Email notification

No

Send an email notification when the job completes.

What it does

This tool converts a xml file to another using a xsl conversion file to specify the translation to be done, from a xsd schema to another.

Inputs

A xml metadata file using a standard (for example EML, ISO19115,...) and a xsl file describing the mapping between the standard terms from input searched output standard.

Outputs

A xml metadata file using a new standard (for example ISO19115, EML, ...).

NDB Pôle National de Données de Biodiversité Data Catalog

DATA PORTALS SUMMARY ABOUT Jump to: DOIx or IIT Go SIGN IN

Search Search phrase

Filter by:

- Data attribute
- Data files
- Creator
- Year
- Identifier
- Taxon
- Location
- Access

Datasets 1 to 25 of 68 Sort by Most recent

1 2 3 Next

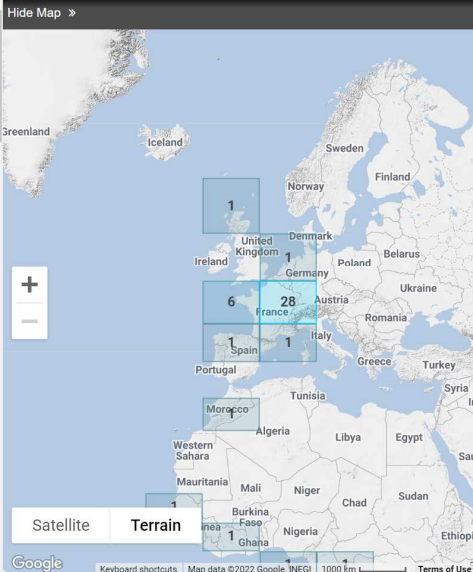
Camille Leroux, Christian Kerbiriou, Isabelle Le Viol, Nicolas Valet, and Kévin Barré. 2022. **Data from: Distance to hedgerows drives local repulsion and attraction of wind turbines on bats: implications for spatial siting.** PNDB Data Repository. urn:uuid:4267c75d-1707-41f0-8fe6-5e13489b2d4e.

Constance Blary, Kévin Barré, Christian Kerbiriou, and Isabelle Le Viol. 2021. **Assessing the importance of field margins for bat species and communities in intensive agricultural landscapes - Data.** PNDB Data Repository. urn:uuid:cb192b3b-dd23-4f6c-abd6-d0e3964c4b79.

Lorraine Coché, Elie Arnaud, Bouveret Laurent, Romain David, Eric Foulquier, et al. 2021. **Kakila database of marine mammal observation data around the French archipelago of Guadeloupe in the AGOA sanctuary - French Antilles.** PNDB Data Repository. doi:10.48502/8bb5-pk85.

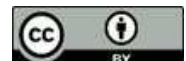
Institut de Recherche pour le Développement, UMR DIADE, France .. SouthGreen Development Platform, Agropolis Campus, Montpellier, France .. Africa Rice Center, Benin .. CEA, Institut de Biologie Française Jacob, Genoscope, Evry, France .. CNRS, UMR 8030, Evry, France .. et al. 2019. **African rice population genomics dataset or title of the article : "The Rise and Fall of African Rice Cultivation Revealed by Analysis of 246 New Genomes"**. PNDB Data Repository. doi:10.48502/xcah-3w69.

Hide Map



Satellite Terrain

Keyboard shortcuts Map data ©2022 Google, INEGI 1000 km Terms of Use



Ecological research data management

Get species occurrences data from GBIF, ALA, INAT and others (Galaxy Version 0.9.0)

Scientific name of the species

Genus species format, eg : Canis lupus

Data source to get data from

Number of records to return

500

Execute

Get species occurrences data
What it does
Search species occurrences across a single or many data sources.

Amazing basis for

PNDB / Galaxy-E work for next months



Convert metadata

xm1starlet convert a metadata XML file in one standard to another (Galaxy Version 1.6.1)

input xml file to convert

18: xm1starlet on data 2 and data 17: Converteu xmi

input xsl conversion file

2: iso2eml_all_in_one.xsl

Execute

What it does

This tool converts a xml file to another using a xsl conversion file to specify the translation to be done, from a xsd schema to another.

Inputs

A xml metadata file using a standard (for example EML, ISO19115,...) and a xsl file describing the mapping between the standard terms from input searched output standard.

Outputs

A xml metadata file using a new standard (for example ISO19115, EML, ...).

Pôle National de Données de Biodiversité Data Catalog

Search

Filter by: Data attribute, Data files, Creator, Year, Identifier, Taxon, Location, Access

Datasets 1 to 25 of 68

1 2 3 Next

Camille Leroux, Christian Kerbiriou, Isabelle Le Viol, Nicolas Valet, and Kévin Barré. 2022. Data from: Distance to hedgerows drives local repulsion and attraction of wind turbines on bats: implications for spatial siting. PNDB Data Repository. um:uuid:4267c75d-1707-41f0-8fe6-5e13489b2d4e.

Constance Blary, Kévin Barré, Christian Kerbiriou, and Isabelle Le Viol. 2021. Assessing the importance of field margins for bat species and communities in intensive agricultural landscapes - Data. PNDB Data Repository. um:uuid:cb192b3b-dd23-4f6c-abd6-d0e3964c4b79.

Lorraine Coché, Elie Arnaud, Bouveret Laurent, Romain David, Eric Foulquier, et al. 2021. Kakila database of marine mammal observation data around the French archipelago of Guadeloupe in the AGOA sanctuary - French Antilles. PNDB Data Repository. doi:10.48502/8bb5-pk85.

Institut de Recherche pour le Développement, UMR DIADE, France .. SouthGreen Development Platform, Agropolis Campus, Montpellier, France .. Africa Rice Center, Benin .. CEA, Institut de Biologie Français Jacob, Genoscope, Evry, France .. CNRS, UMR 8030, Evry, France .. et al. 2019. African rice population genomics dataset or title of the article : "The Rise and Fall of African Rice Cultivation Revealed by Analysis of 246 New Genomes". PNDB Data Repository. doi:10.48502/xcah-3w69.

Map showing distribution of datasets across Europe and Africa with numbered markers.

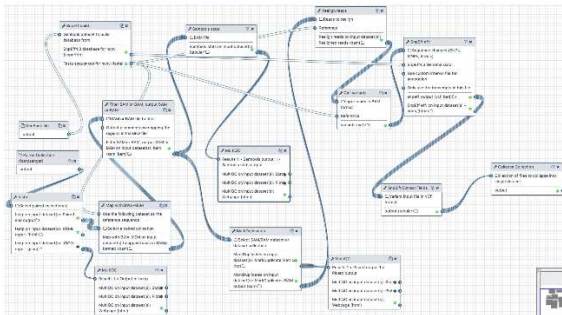


Not only asynchronously

=> Thanks to Galaxy interactive tools GxIT

One specific workflow goal

Galaxy as workflow engine workflows (the async ones)



```
class: GalaxyWorkflow
doc: |
  Simple workflow that no-op cats a file
inputs:
  the_input:
    type: File
    doc: input doc
outputs:
  the_output:
    outputSource: cat/out_file1
steps:
  cat:
    tool_id: cat1
    doc: cat doc
    in:
      input1: the_input
```

workflows (the async ones)

The screenshot shows the Galaxy Live interface. On the left, a Jupyter Notebook titled 'Lorenz.jynb' is open, displaying the Lorenz attractor differential equations and a 3D plot of the attractor. The code in the notebook includes:

```
def solve_lorenz(sigma=10.0, beta=8./3, rho=28.0):
    """Solve the Lorenz differential equations"""
    max_time = 4.0
    dt = 0.01
    fig = plt.figure()
    ax = fig.add_subplot(1, 1, 1, projection='3d')
    ax.axis('off')
    # prepare the axes labels
    ax.set_xlabel('x')
    ax.set_ylabel('y')
    ax.set_zlabel('z')
    def lorenz_deriv(x,y,z, t0, sigma=sigma, beta=beta, rho=rho):
        """Compute the time-derivative of a Lorenz system"""
        return [sigma*(y-x), x*(rho-z)-y, x+y-beta*z]
    # Choose random starting points, uniformly distributed from -18 to 18
    np.random.seed(1)
    x0 = -18 + 36 * np.random.random((N, 3))
    # Solve for the trajectories
```

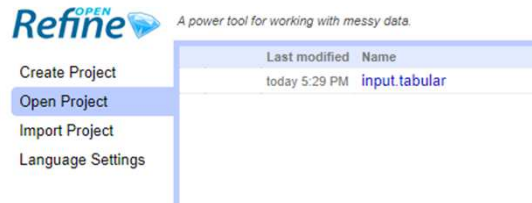
Below the notebook, the Galaxy dashboard is visible, featuring several tool icons: Ethercalc (A web spreadsheet), PHINCH (Visualise large biological datasets (microbiomes, metagenomes, etc)), MetaSHARK (Generate Metadata from data using EML standard), Wallace, iSEE, and WILSON.

Interactive tools

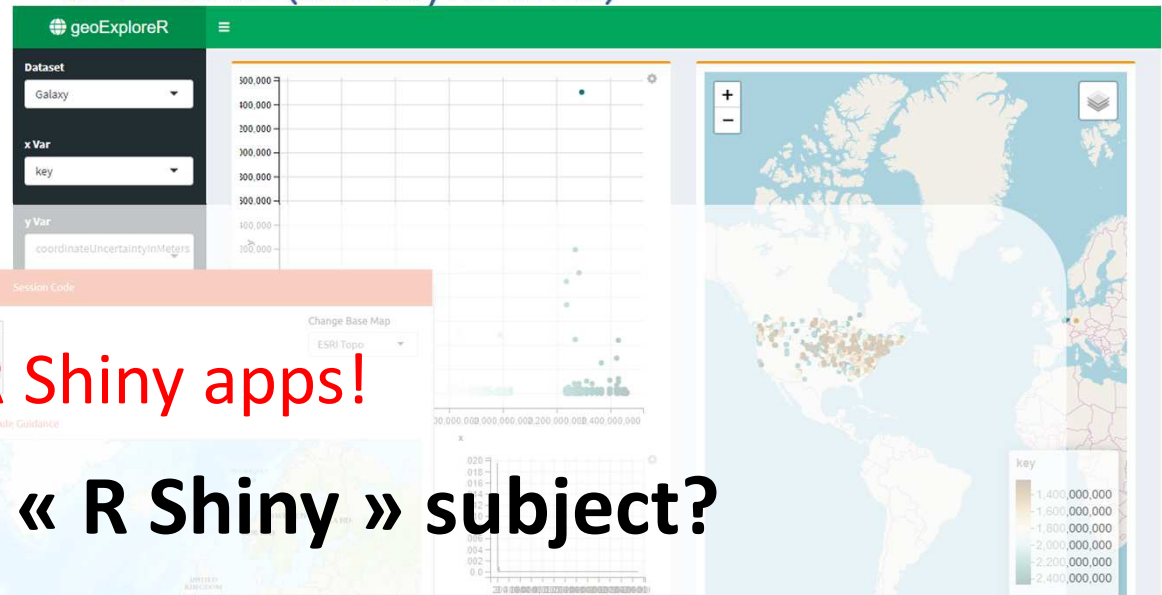
- Interactive Jupyter Notebook**
- GPU enabled Interactive Jupyter Notebook for Machine Learning**
- Interactive Climate Notebook**
- Interactive Pangeo Notebook**
- RStudio**
- Pyiron Interactive Jupyter Notebook**
- HiGlass** an interactive Hi-C data visualizer.
- OpenRefine** Working with messy data
- Ubuntu XFCE Desktop**
- Panoply** interactive plotting tool for geo-referenced data
- AskOmics** a visual SPARQL query builder
- Interactive CellXgene Environment**
- bam.iobio visualisation**
- VCF (ioBio) Visualisation**
- Neo4j (Graph Database)**
- Pinch Visualisation**
- Paraview**
- Wilson** Webbased Interactive Omics visualization
- Wallace** Webbased Interactive modeling of species niches and distributions
- geoexplorer** An interactive spatial analysis platform using ggvis and Leaflet
- radiant** Data analytics using Radiant R Shiny app
- EtherCalc**
- VRM Editor** interactive tool for creating Variable Resolution Mesh for NorESM/CESM
- SimText** Interactive shiny app to explore SimText output data
- iSEE**
- metashark** Metadata Shiny Automated Resource and Knowledge

One specific workflow goal

Galaxy as workflow engine

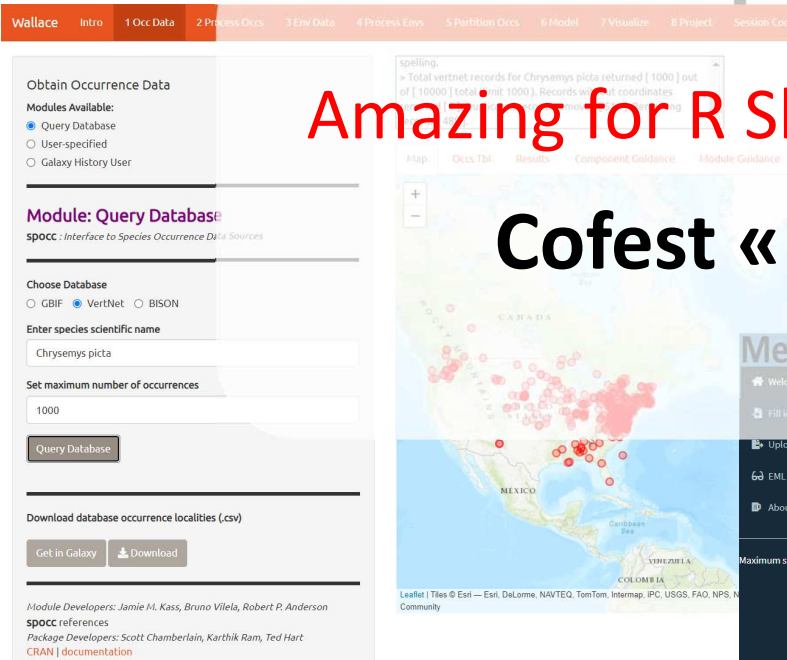


workflows (the async ones)



Amazing for R Shiny apps!

Cofest « R Shiny » subject?



One specific workflow goal

Galaxy as workflow eng **What is making this guy ?**



One specific workflow goal

Galaxy as workflow eng **What is making this guy ?**



The screenshot displays the 'Audio Labeler App' interface. At the top, the filename is '1_1_SMU05115_20211104_064902_start_39_16.wav'. Below this is a spectrogram with frequency markers at 6kHz, 8kHz, and 10kHz. A large red text overlay reads 'Analysing audio files on Galaxy!'. The interface includes a 'Save Selection' button, a 'Meta Information' section, a 'Class Label Selection: Species' section with buttons for 'Eurasian Magpie', 'European Goldfinch', 'House Sparrow', 'Noise', and 'Other', a 'Type in additional category:' text input, an 'Additional Notes:' text input, a 'Play audio: (selected)' section with a play button and a progress bar showing 0:11 / 0:11, a 'Playback Speed:' dropdown set to '1x', a 'Reset Plot' button, a 'Call Type:' text input, and a control panel with 'Add category', 'Remove category', 'Reset categories', 'Save to List', 'Delete Selection', and 'Undo Deletion' buttons. At the bottom, there are sliders for 'Audio Frequency Range:' (set to 0 to 32) and 'Label Confidence:' (set to 0 to 1).

Not only for scientists

- Crowdsourcing through Galaxy webhooks
- Data / Biodiversity literacy through Galaxy-Bricks

Galaxy-Bricks

VIGIENATURE École

Galaxy for pupils !
bricks.vigienature-ecole.fr

The screenshot displays the Galaxy Bricks web interface. At the top, there is a navigation bar with the following items: "Le projet", "Analyser les données", "Visualiser les données", "Exemples de scenarios", and "Forum".

The left sidebar contains the following menu items:

- Question de recherche
- Importer des données
- Manipuler des données
 - Convertir des dates
 - Opérations sur des lignes (2)
 - Résumer des données
 - Sélectionner des lignes
 - Rechercher/Remplacer
 - Operations sur des lignes (WIP-repeat)
 - Fusionner le contenu de deux colonnes
- Visualiser
 - Plot w ggplot2
 - Représenter
- Tests statistiques
 - Régression linéaire multiple
 - Regression linéaire simple

The central area features the Galaxy Bricks logo and the text "Bienvenue dans Galaxy Bricks !". Below this is a green button that says "Commencer une question de recherche !".

The right-hand panel shows a research question: "Effet de l'environnement sur les oiseaux". Below the question is a list of tasks, with the first one being "7 Représenter (sur Résumer des données on data 3)". A warning message is displayed below the task list: "Warning message: NAs introduits lors de la conversion automatique. Erreur : Must use either variable name or expression when facetting. Exécution arrêtée".

Galaxy-Bricks

VIGIENATURE École

Galaxy for pupils !
bricks.vigienature-ecole.fr

The screenshot displays the Galaxy-Bricks web interface. On the left, a sidebar lists various data processing tasks under three main categories: 'Question de recherche' (Research question), 'Manipuler des données' (Manipulate data), and 'Visualiser' (Visualize). The 'Manipuler des données' section includes options like 'Convertir des dates', 'Opérations sur des lignes (2)', 'Résumer des données', 'Sélectionner des lignes', 'Rechercher/Remplacer', 'Opérations sur des lignes (WIP-repeat)', and 'Fusionner le contenu de deux colonnes'. The 'Visualiser' section includes 'Plot w ggplot2' and 'Représenter'. The 'Tests statistiques' section includes 'Régression linéaire multiple' and 'Regression linéaire simple'. The central workspace features a large 'Commencer une question de recherche !' button. The right-hand panel shows a scenario titled 'Effet de l'environnement sur les oiseaux' with a question: 'Quel est l'effet de l'environnement sur la diversité des oiseaux ?'. Below the question, there are numbered steps (7 and 6) for representing data, and a warning message: 'Warning message: NAs introduits lors de la conversion automatique. Erreur : Must use either variable name or expression when facetting. Exécution arrêtée'.

Amazing basis for
Cofest « vue.js » GUI?

Crowd sourcing through Galaxy



GAPARS project

MOODA concept (Massively Open Online Data Analysis)

Crowdsourcing with hoverflies (syphres) images from SPIPOLL project



A screenshot of the Galaxy web interface. The main window displays a "Citizen Science Project" titled "Citizen Science Experiment!". It shows a photograph of a hoverfly on a purple flower. Below the image are four radio button options: "Male", "Likely male" (which is selected and highlighted with a green box), "Cannot See", "Likely female", and "Female". A "Submit" button is at the bottom. The left sidebar shows a "Tools" menu with various categories like "Get Data", "Collection Operations", and "GENERAL TEXT TOOLS". The right sidebar shows a "History" panel with a list of recent jobs, including "GlimmTMB - Temp trends plot on data 1 0, data 11, and others".

63 320 classifications in 2,5 years !



Crowd sourcing through Galaxy



MOODA concept (Massively Open Online Data Analysis)

Crowdsourcing with hoverflies (syphres) images from SPIPOLL project

The screenshot displays the Galaxy web interface for a project named 'Production / SpiPoll Fly'. The top navigation bar includes 'repo', 'Dashboard', 'Batches', 'Tasks', 'Project settings', 'Data upload', 'Import', 'Data exports', and 'Logout'. The main content area is divided into several sections:

- Overview (All time):** A dashboard with six key metrics: Batches (6), Active tasks (4.76k), Training tasks (400), Average run (2.16), Players (2.00), and Classifications (73.59k).
- Classifications (Last 30 days):** A bar chart showing the number of classifications per day from 2022-09-04 to 2022-10-03. The y-axis ranges from 0 to 220. The chart shows a peak of approximately 200 classifications on 2022-09-30.
- Players (Last 30 days):** A bar chart showing the number of active players per day. The y-axis ranges from 0 to 1. The chart shows a consistent number of players, mostly 1, with some days having 2 players.
- Classification Task Interface:** A task titled 'GimmTMB - Temp trends plot on data 1 0, data 11, and others'. It shows a 'History' panel with a list of tasks (e.g., '39: GimmTMB - Temp trends plot on data 1 0, data 11, and others'). Below the history, there is a 'Rechercher des données' search bar and a 'ff' filter. The main task area shows a 'Format: png, génome de référence: ?' and 'Method: gimmTMB'. The description includes 'Estimation de la variation annuelle' and 'Estimation de la tendance'. A 'Submit' button is visible at the bottom.

63 320 classifications in 2,5 years !



Crowd sourcing through Galaxy



MOODA concept (Massively Open Online Data Analysis)

Crowdsourcing with hoverflies (sypbres) images from SPIPOLL project

The screenshot shows the Galaxy web interface for a project named 'Production / SpiPoll Fly'. The top navigation bar includes 'repo', 'Dashboard', 'Batches', 'Tasks', 'Project settings', 'Data upload', 'Import', 'Data exports', and 'Logout'. The main content area is divided into several sections:

- Overview (All time):** A dashboard with six key metrics: Batches (6), Active tasks (4.76k), Training tasks (400), Average run (2.16), Players (2.00), and Classifications (73.59k).
- Classifications (Last 30 days):** A bar chart showing the number of classifications per day from 2022-09-04 to 2022-10-03. The y-axis ranges from 0 to 220. The chart shows a significant peak in late September and early October.
- Players (Last 30 days):** A bar chart showing the number of active players per day. The y-axis ranges from 0 to 1. The chart shows a high density of players during the same period as the classification peak.
- Task History:** A list of tasks on the right side, including '29: GlimmTMB - Temp trends per year - on data 10, data 11, and others', '28: GlimmTMB - Temp trends per year - on data 10, data 11, and others', '27: GlimmTMB - Temp trends plot on data 10, data 11, and others', '26: GlimmTMB - Temp trends per year - on data 10, data 11, and others', and '16: GAM - Temp tren'.

Overlaid on the screenshot is the text: **Amazing basis for Cofest « citizen science » webhook?**

63 320 classifications in 2,5 years !



Crowd sourcing through Galaxy

MOODA concept (Massively Open Online Data Analysis)

<https://tinyurl.com/galaxymooda>

The screenshot displays the 'Audio Labeler App' interface. At the top, the filename is '1_Michael Jackson - Heal The World _Official Video_.wav (1/3)'. The main area features a spectrogram with a frequency range from 0kHz to 20kHz and a time axis from 0 to 150 seconds. A cyan box highlights a section of the spectrogram, labeled 'Amazing Galaxy community!'. Below the spectrogram, there is a 'Save Selection' button and a 'Meta Information' section. The 'Class Label Selection: Species' section includes buttons for 'Eurasian Magpie', 'European Goldfinch', 'House Sparrow', 'Noise', 'Other', and 'Amazing Galaxy community!'. A text input field for 'Type in additional category:' contains 'Amazing Galaxy community!'. The 'Additional Notes:' section has a text area. The 'Audio Frequency Range:' section has a slider from 0 to 32. The 'Label Confidence:' section has a slider from 0 to 1. The bottom right corner features a Creative Commons BY license icon.

Thank you !

PNDB team

Coline Royaux – engineer R / Galaxy dev
(workflows to compute biodiversity indicators)

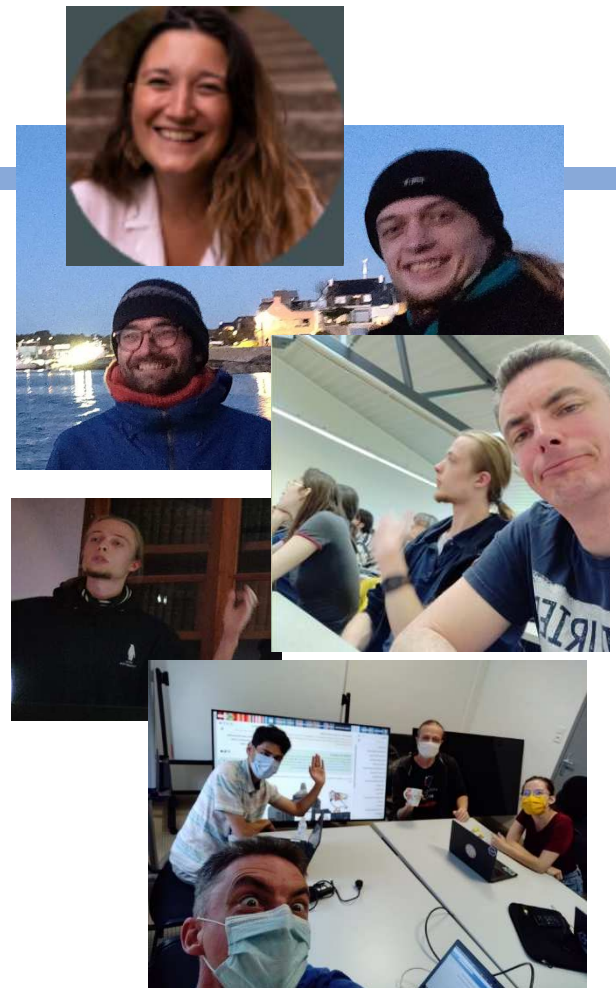
Elie Arnaud – engineer R Shiny /
knowledge – metadata dev

Marie Jossé – engineer R / Galaxy dev

Julien Sananikone – engineer DevOps /
sys admin / web dev

Olivier Norvez – animation coordinator

Yvan Le Bras – Beta tester yvan.le-bras@mnhn.fr



<https://www.pndb.fr/>

PNDB « bricks »:

MetaShARK Metadata work:

<https://youtu.be/OVViSMzRGtw>

Data metadata portal:

<https://youtu.be/STwsYDHET2A>

Galaxy Europe demo:

- <https://youtu.be/HeIAHggX6D4>

- [Essential biodiversity variables on Galaxy: implementing PAMPA](#)

- [Producing biodiversity indicators from citizen science projects](#)

