Journées Calcul et Données JCAD 2020 (03 décembre 2020)

Pôle national de données de Biodiversité

de la donnée de biodiversité au calcul scientifique via la métadonnée

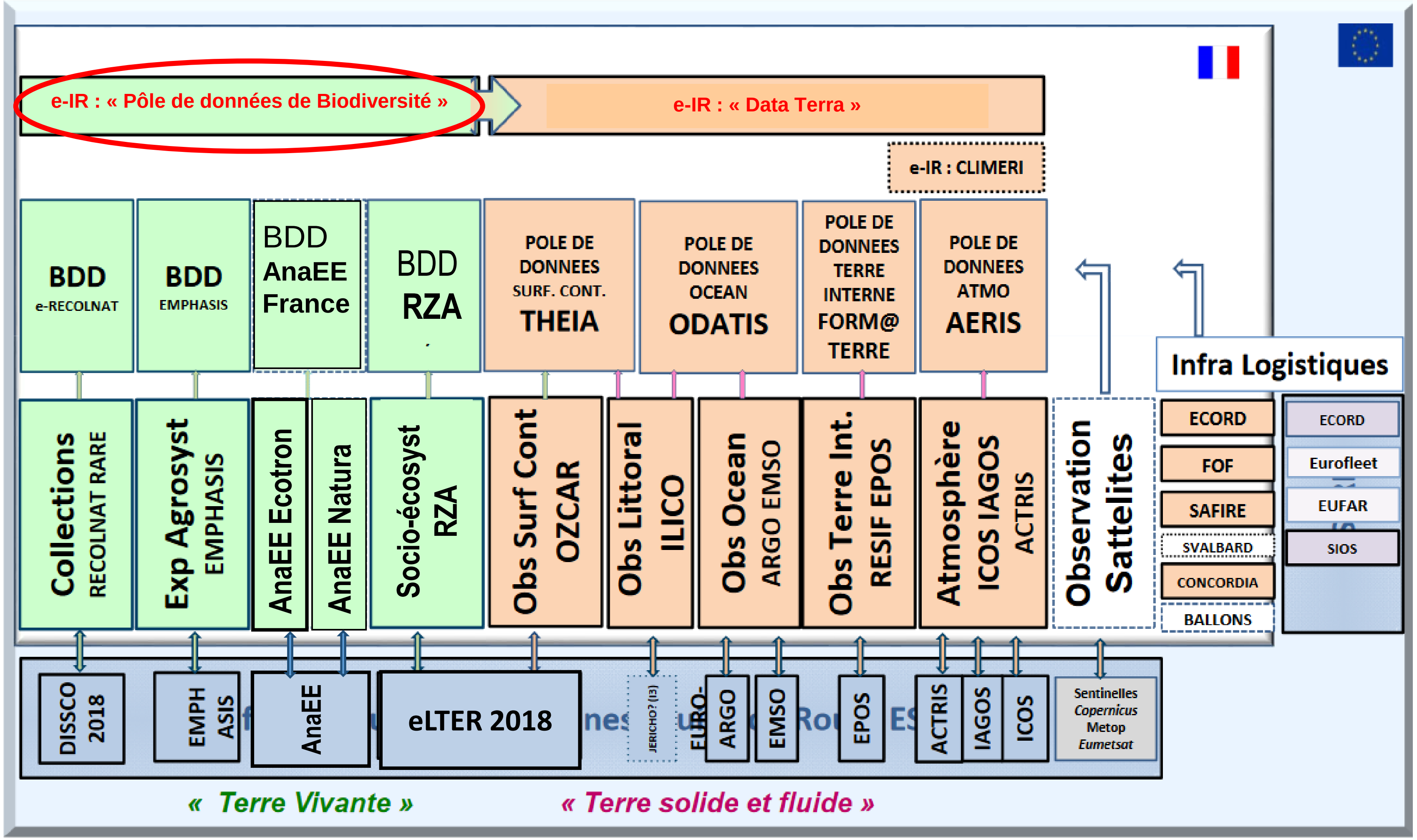
French BON

Biodiversity & Ecosystems Action Group

Yvan LE BRAS, Chef de projet (UMS PatriNat)
Sandrine Pavoine, MNHN (UMR CESCO)
Anne-Sophie ARCHAMBEAU, GBIF-France (UMS PatriNat)
Cécile CALLOU, Dir UMS BBEES (CNRS-MNHN)
Aurélie DELAVAUD (FRB)
Dominique JOLY, DAS CNRS (INEE)
Thomas Milon, Chef de projet "Système d'information sur la biodiversité« (UMS PatriNat)
Laurent PONCET, Dir. UMS PatriNat, en charge du Centre de données (MNHN)
Jean-Denis VIGNE, DGD-Recherche, expertise, valorisation, enseignement MNHN

#PNDB @Yvan2935 @earnaud @ColineRoyaux @jusana_photos
Yvan.le-bras@mnhn.fr, elie.arnaud@mnhn.fr,
coline.royaux@mnhn.fr, julien.sananikone@mnhn.fr

BiodiFAIRse



e-IR : « Pôle de données de Biodiversité »

e-IR : « Data Terra »

e-IR : CLIMERI

BDD e-RECOLNAT

BDD EMPHASIS

BDD AnaEE France

BDD RZA

POLE DE DONNEES SURF. CONT. THEIA

POLE DE DONNEES OCEAN ODATIS

POLE DE DONNEES TERRE INTERNE FORM@TERRE

POLE DE DONNEES ATMO AERIS

Infra Logistiques

Collections RECOLNAT RARE

Exp Agrosyst EMPHASIS

AnaEE Ecotron

AnaEE Natura

Socio-écosyst RZA

Obs Surf Cont OZCAR

Obs Littoral ILICO

Obs Ocean ARGO EMSO

Obs Terre Int. RESIF EPOS

Atmosphère ICOS IAGOS ACTRIS

Observation Satellites

ECORD

FOF

SAFIRE

SVALBARD

CONCORDIA

BALLONS

ECORD

Eurofleet

EUFAR

SIOS

DISSCO 2018

EMPH ASIS

AnaEE

eLTER 2018

JERICHO? (I3)

EURO-

ARGO

EMSO

EPOS

ACTRIS

IAGOS

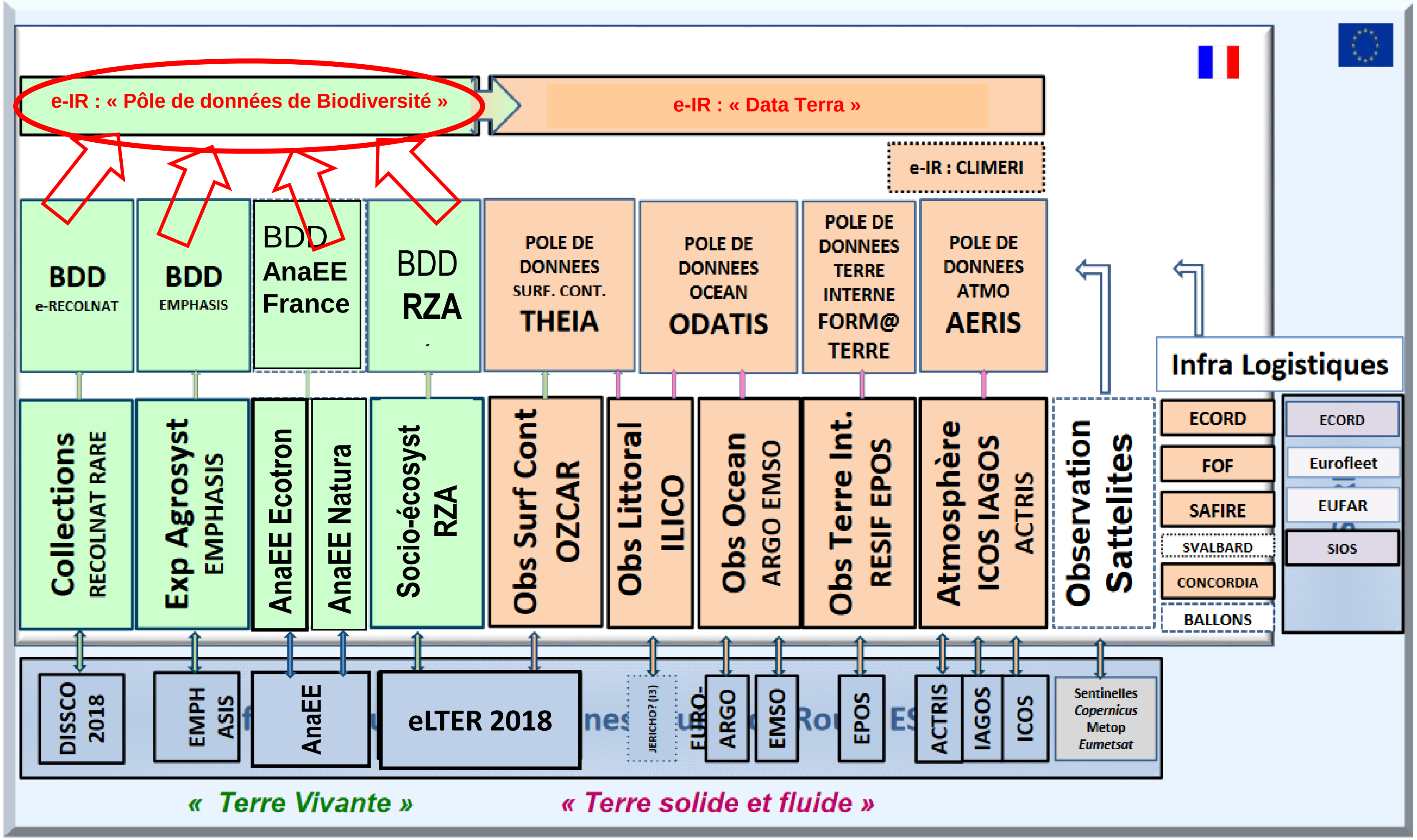
ICOS

Sentinelles
Copernicus
Metop
Eumetsat

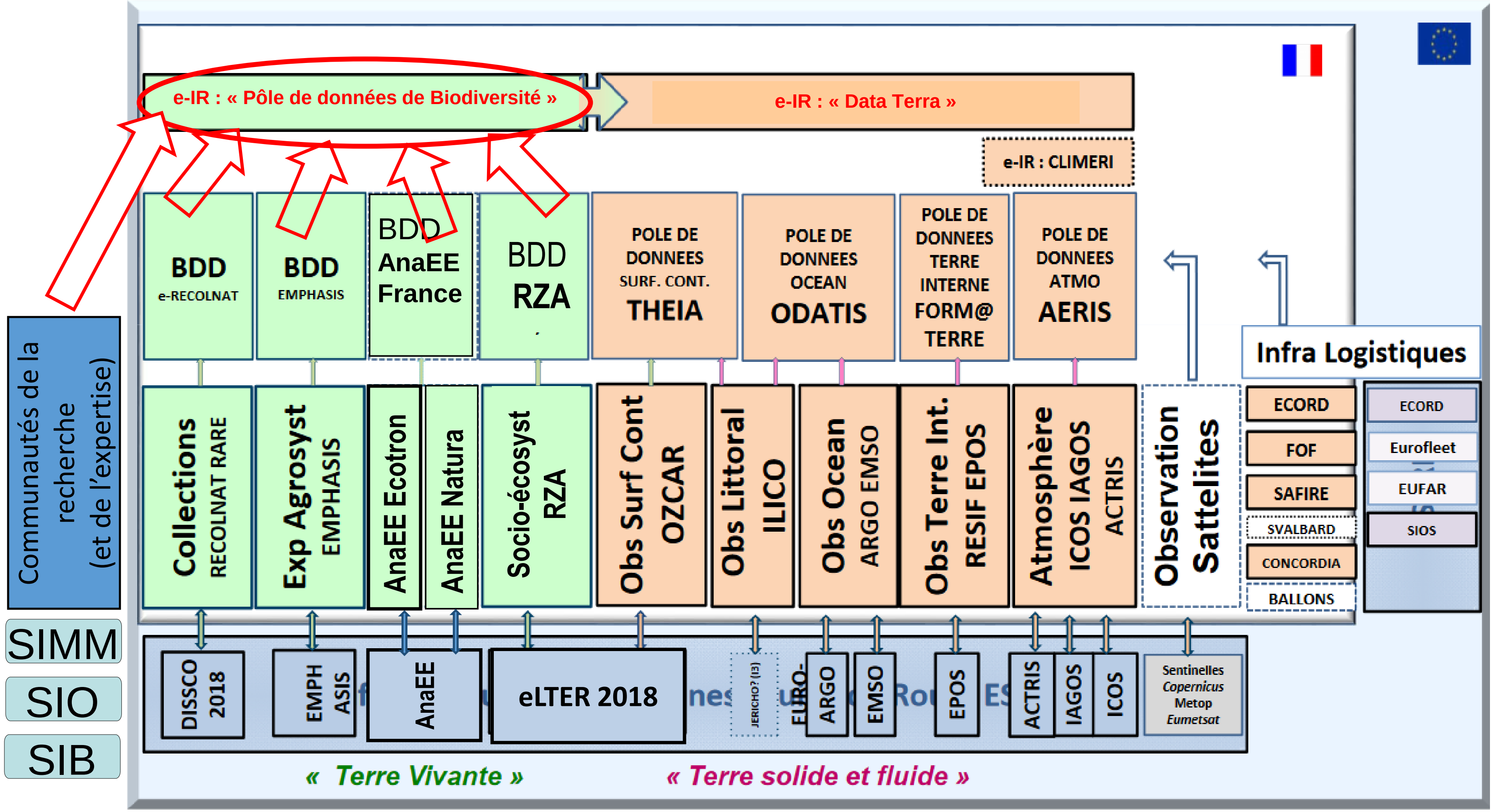
« Terre Vivante »

« Terre solide et fluide »

Interface « terre vivante » IR



Interface « terre vivante » IR + chercheurs surtout via organismes + SI fédérateurs OFB



e-IR : « Pôle de données de Biodiversité »

e-IR : « Data Terra »

e-IR : CLIMERI

BDD e-RECOLNAT

BDD EMPHASIS

BDD AnaEE France

BDD RZA

POLE DE DONNEES SURF. CONT. THEIA

POLE DE DONNEES OCEAN ODATIS

POLE DE DONNEES TERRE INTERNE FORM@TERRE

POLE DE DONNEES ATMO AERIS

Infra Logistiques

Collections RECOLNAT RARE

Exp Agrosyst EMPHASIS

AnaEE Ecotron

AnaEE Natura

Socio-écosyst RZA

Obs Surf Cont OZCAR

Obs Littoral ILICO

Obs Ocean ARGO EMSO

Obs Terre Int. RESIF EPOS

Atmosphère ICOS IAGOS ACTRIS

Observation Sattelites

ECORD

FOF

SAFIRE

SVALBARD

CONCORDIA

BALLONS

ECORD

Eurofleet

EUFAR

SIOS

Communautés de la recherche (et de l'expertise)

SIMM

SIO

SIB

DISSCO 2018

EMPHASIS

AnaEE

eLTER 2018

JERICHO? (I3)

EURO-ARGO

EMSO

EPOS

ACTRIS

IAGOS

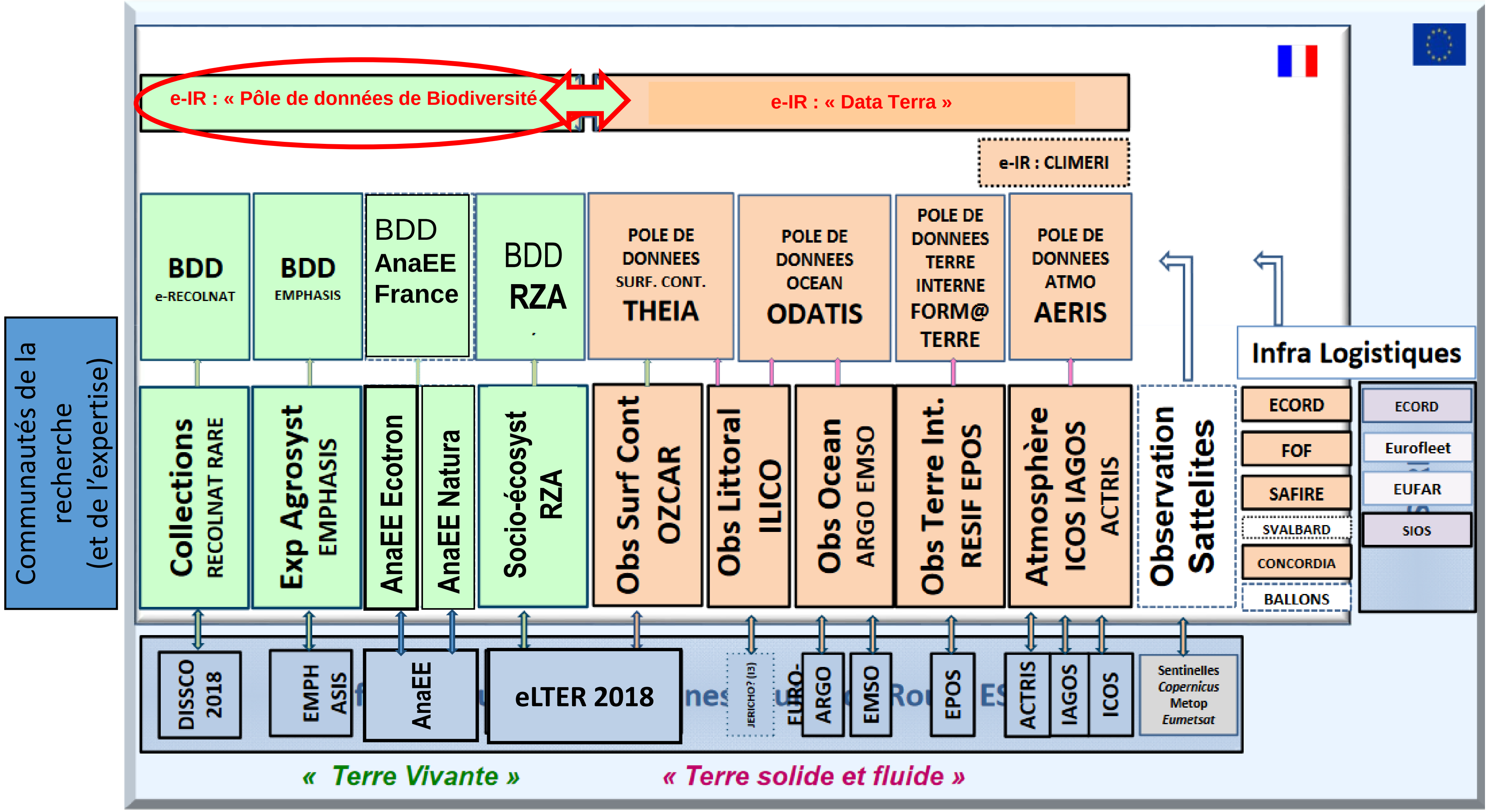
ICOS

Sentinelles *Copernicus* Metop *Eumetsat*

« Terre Vivante »

« Terre solide et fluide »

Interface « terre solide » IR



Communautés de la recherche (et de l'expertise)

e-IR : « Pôle de données de Biodiversité

e-IR : « Data Terra »

e-IR : CLIMERI

BDD e-RECOLNAT

BDD EMPHASIS

BDD AnaEE France

BDD RZA

POLE DE DONNEES SURF. CONT. THEIA

POLE DE DONNEES OCEAN ODATIS

POLE DE DONNEES TERRE INTERNE FORM@TERRE

POLE DE DONNEES ATMO AERIS

Infra Logistiques

Collections RECOLNAT RARE

Exp Agrosyst EMPHASIS

AnaEE Ecotron

AnaEE Natura

Socio-écosyst RZA

Obs Surf Cont OZCAR

Obs Littoral ILICO

Obs Ocean ARGO EMSO

Obs Terre Int. RESIF EPOS

Atmosphère ICOS IAGOS ACTRIS

Observation Sattelites

ECORD

FOF

SAFIRE

SVALBARD

CONCORDIA

BALLONS

ECORD

Eurofleet

EUFAR

SIOS

DISSCO 2018

EMPH ASIS

AnaEE

eLTER 2018

JERICHO? (I3)

EURO-

ARGO

EMSO

EPOS

ACTRIS

IAGOS

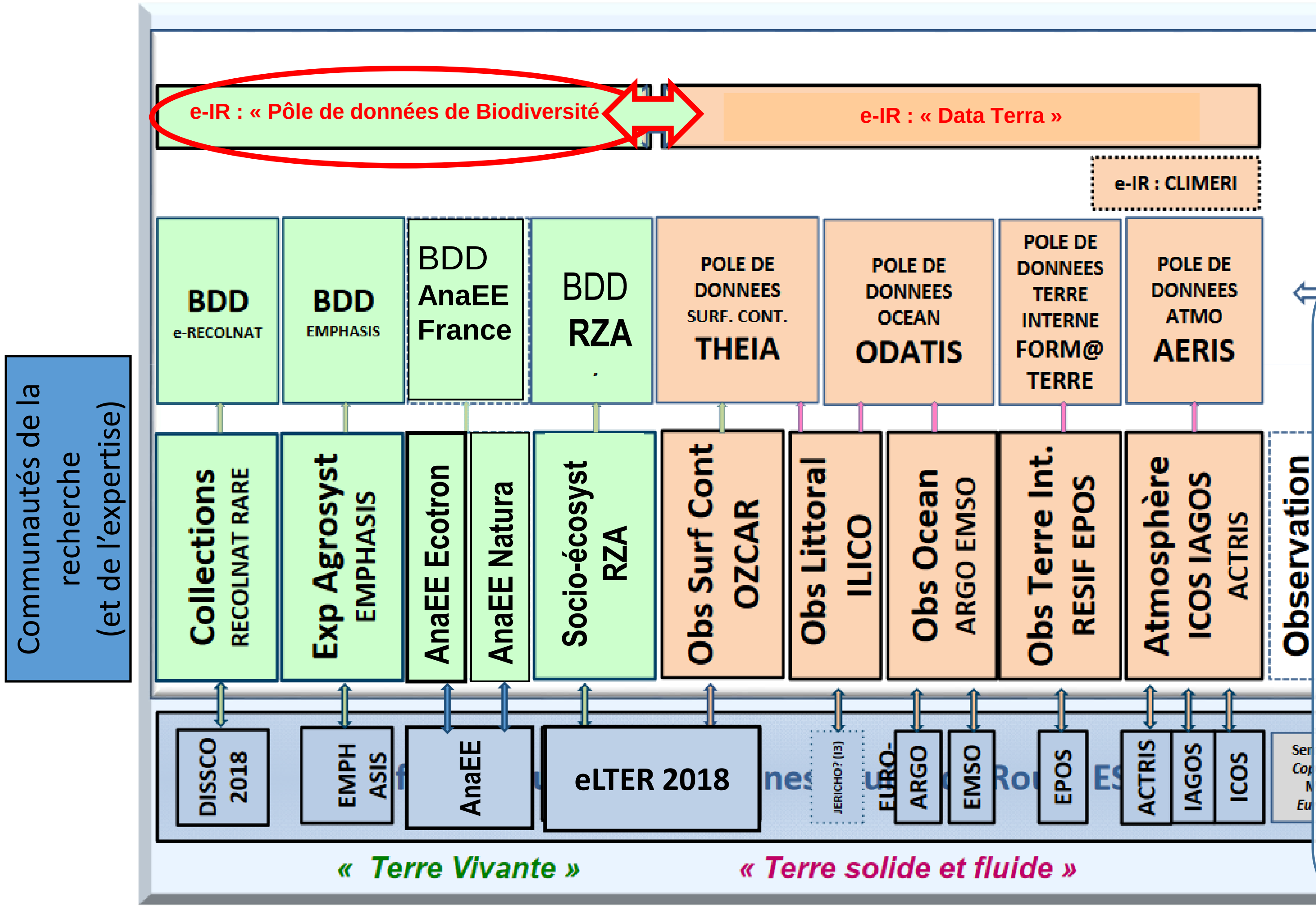
ICOS

Sentinelles *Copernicus* Metop *Eumetsat*

« Terre Vivante »

« Terre solide et fluide »

Cohérence avec les dispositifs internationaux


e-IR : « Pôle de données de Biodiversité »
e-IR : « Data Terra »
e-IR : CLIMERI
BDD e-RECOLNAT
BDD EMPHASIS
BDD AnaEE France
BDD RZA
POLE DE DONNEES SURF. CONT. THEIA
POLE DE DONNEES OCEAN ODATIS
POLE DE DONNEES TERRE INTERNE FORM@TERRE
POLE DE DONNEES ATMO AERIS
Communautés de la recherche (et de l'expertise)
Collections RECOLNAT RARE
Exp Agrosyst EMPHASIS
AnaEE Ecotron
AnaEE Natura
Socio-écosyst RZA
Obs Surf Cont OZCAR
Obs Littoral ILICO
Obs Ocean ARGO EMSO
Obs Terre Int. RESIF EPOS
Atmosphère ICOS IAGOS ACTRIS
Observation
DISSCO 2018
EMPHASIS
AnaEE
eLTER 2018
JERICHO? (I3)
EURO-ARGO
EMSO
EPOS
ACTRIS
IAGOS
ICOS
Sentinelles Copernicus Metop Eumetsat
« Terre Vivante »
« Terre solide et fluide »

GBIF	GBIF France - PNDB
iBOL2	BOLD France - PNDB
GEO BON	PNDB - French BON EBVs operationalization pilot
GO FAIR	GO-FAIR IN BiodiFAIRse
RDA	DMP common - standard
H2020	EOSC Pillar infraeosc GAPARS SwafS ERASMUS+

AGENCE FRANÇAISE POUR LA BIODIVERSITÉ ÉTABLISSEMENT PUBLIC DE L'ÉTAT · brgm · cirad · cnrs · Ifremer · INERIS · INRA · IRD · irstea · MUSÉUM · FRB

Le PNDB, e-Infrastructure nationale de recherche

- **Infrastructure numérique** inscrite sur la feuille de route du MESRI depuis mars 2018

- **Consortium maître d'ouvrage :**
  (22 partenaires institutionnels, avec le soutien financier du MESRI

- **Maître d'œuvre :**
  MNHN depuis mars 2018 (UMS PatriNat)

- **Gouvernance :**
  Comité de pilotage
  Conseil scientifique
  Comité exécutif

Comités d'experts scientifique :

1. Variabilité génétique des populations domestiques ou sauvages
2. Espèces, traits fonctionnels, communautés
3. Ecosystèmes et socio-écosystèmes, variables SHS : structures et fonctions
4. Données, scénarios, modélisation ; télédétection

Cahier des charges du PNDB

10 lignes de force

1. Orientation FAIR (aussi FAIR et *Open* que possible)
2. Relier/mutualiser avec les composantes existantes du Syst. Terre-Env.
3. Articuler/mutualiser avec le SIB-SIMM
4. Offrir des services à valeur ajoutée
5. Privilégier la qualité (*quality data*), au-delà de la quantité (*big data*),
6. Promouvoir la flexibilité des services (« à la carte »)
7. Développer, faciliter et favoriser la description fine des données
8. Viser une portée internationale (couverture & utilisation)
9. Articuler le PNDB avec les initiatives européennes et internationales
10. Pour commencer, s'appuyer sur un petit nombre de cas d'étude

    (preuve de concept)

3 volets

1. **Accès aux métadonnées et données**
2. **Animation et accompagnement**
3. **Accès aux outils de traitement, de couplage, d'analyse (calcul)**

Journées Calcul et Données JCAD 2020 (03 décembre 2020)

Le paysage **(méta)données** via l' *Ecological Metadata Language*

Données + Métadonnées → Datapackage

GeoNetwork

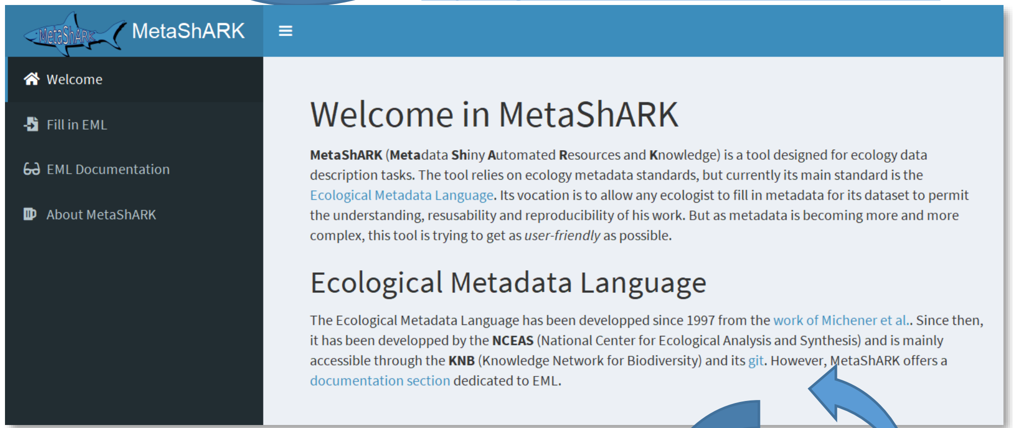
SI externes
- Infrastructures
- organismes

Dataverse

MetaSHARK

https://github.com/earnaud/MetaShARK-v2

Metacat



Welcome in MetaShARK

MetaShARK (**Meta**data **Sh**iny **A**utomated **R**esources and **K**nowledge) is a tool designed for ecology data description tasks. The tool relies on ecology metadata standards, but currently its main standard is the Ecological Metadata Language. Its vocation is to allow any ecologist to fill in metadata for its dataset to permit the understanding, resusability and reproducibility of his work. But as metadata is becoming more and more complex, this tool is trying to get as *user-friendly* as possible.

Ecological Metadata Language

The Ecological Metadata Language has been developped since 1997 from the work of Michener et al.. Since then, it has been developped by the **NCEAS** (National Center for Ecological Analysis and Synthesis) and is mainly accessible through the **KNB** (Knowledge Network for Biodiversity) and its git. However, MetaShARK offers a documentation section dedicated to EML.

dataset1

dataset2

script1

article1

protocol1

EML Assembly Line

https://ediorg.github.io/EMLassemblyline/index.html

EML

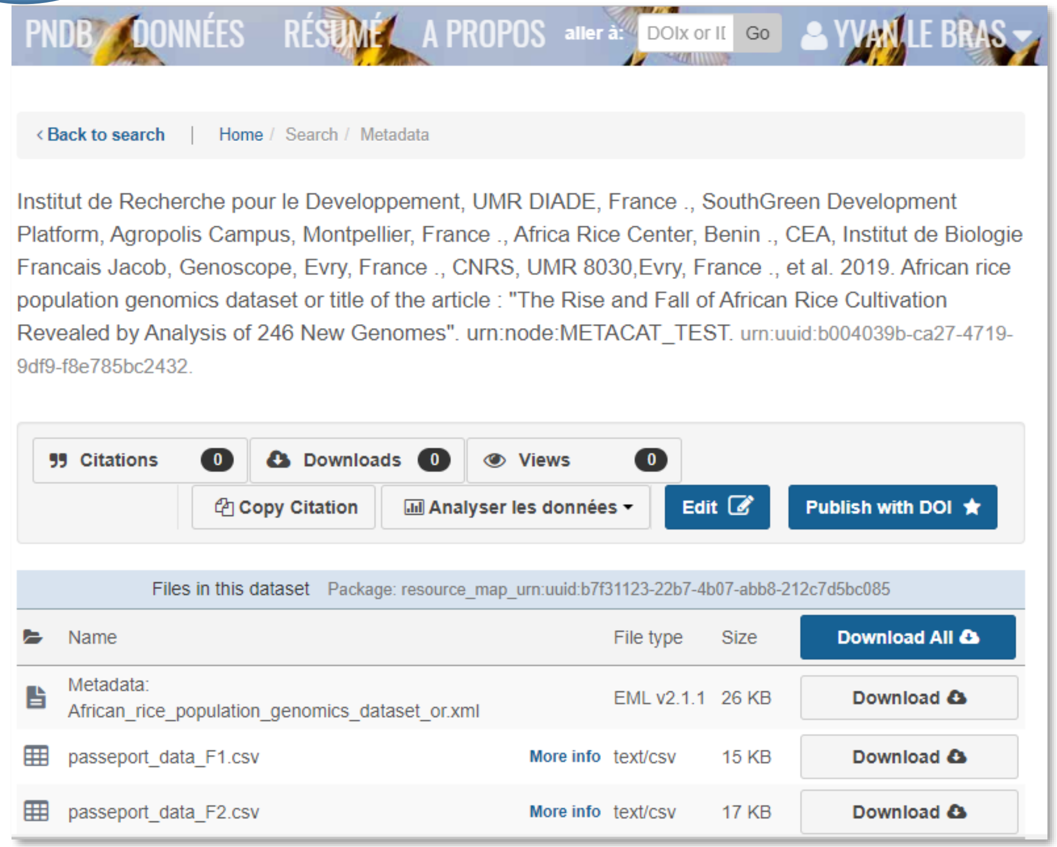
Morpho->Outdated!

Semantics enrichment

cedarr

CEDAR R package for API linking in an R interface.

https://github.com/earnaud/cedarr


PNDB DONNÉES RÉSUMÉ A PROPOS aller à: DOIx or I[Go
YVAN LE BRAS
‹ Back to search | Home / Search / Metadata

Institut de Recherche pour le Developpement, UMR DIADE, France ., SouthGreen Development Platform, Agropolis Campus, Montpellier, France ., Africa Rice Center, Benin ., CEA, Institut de Biologie Francais 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". urn:node:METACAT_TEST. urn:uuid:b004039b-ca27-4719-9df9-f8e785bc2432.

Files in this dataset Package: resource_map_urn:uuid:b7f31123-22b7-4b07-abb8-212c7d5bc085

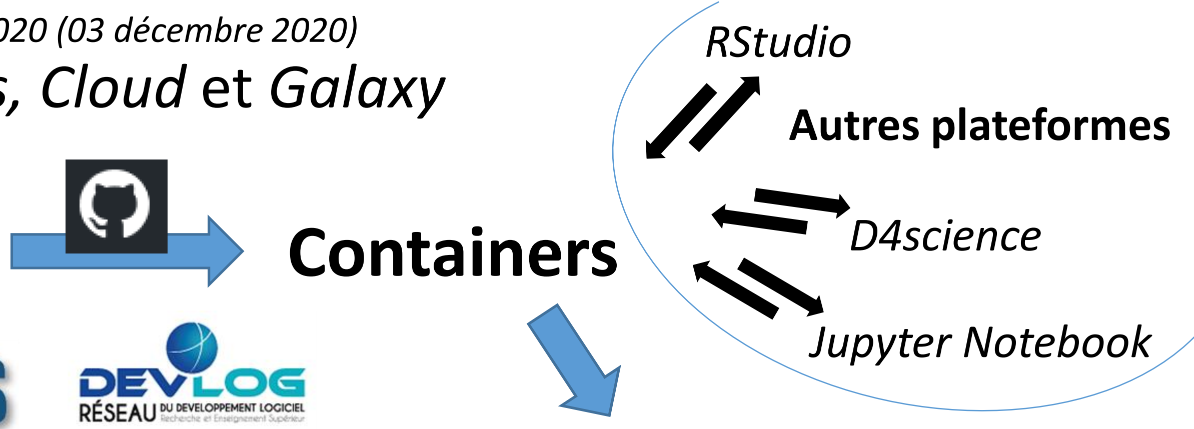
Name		File type	Size	Download All
Metadata: African_rice_population_genomics_dataset_or.xml		EML v2.1.1	26 KB	Download
passeport_data_F1.csv	More info	text/csv	15 KB	Download
passeport_data_F2.csv	More info	text/csv	17 KB	Download

Portail de données métadonnées :
https://openstack-192-168-100-101.genouest.org/metacatui
http://data.test.pndb.fr/data https://data.pndb.fr/data

Journées Calcul et Données JCAD 2020 (03 décembre 2020)

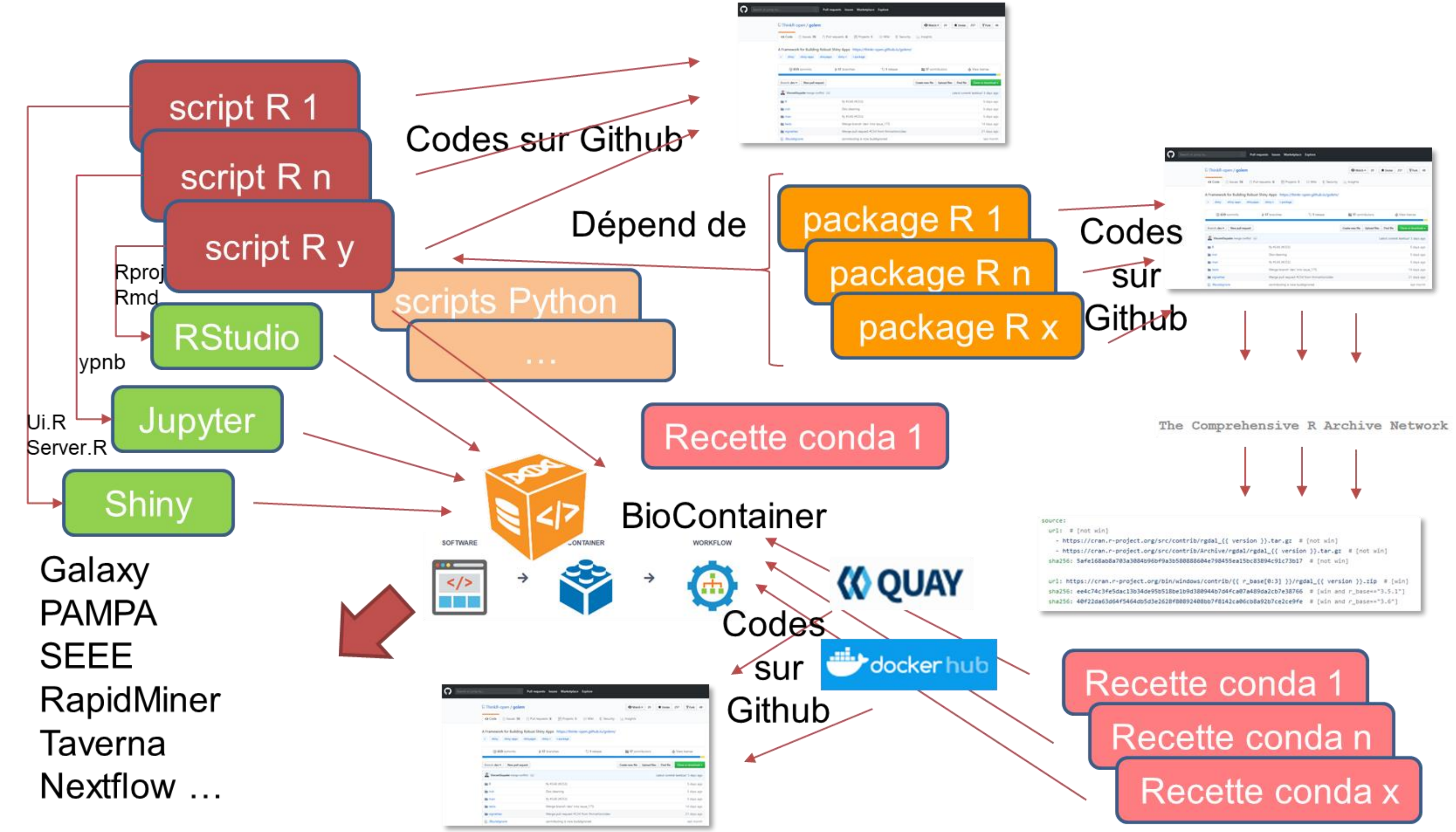
Le paysage **analyse** via *Github, Conda, Containers, Cloud* et *Galaxy*

Codes + dépendances



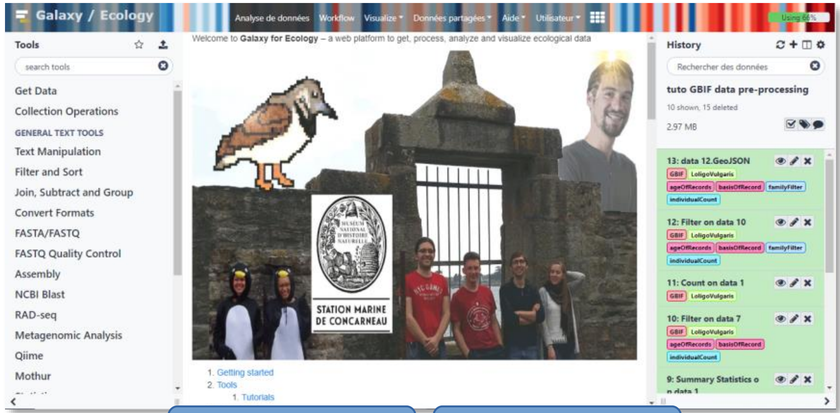
Machines virtuelles (local ou cloud)

Galaxy



```
source:
  url:  # [not win]
    - https://cran.r-project.org/src/contrib/rgdal_{{ version }}.tar.gz  # [not win]
    - https://cran.r-project.org/src/contrib/Archive/rgdal/rgdal_{{ version }}.tar.gz  # [not win]
  sha256: 5afe168ab8a703a3084b96bf9a3b580888604e798455ea15bc83894c91c73b17  # [not win]

  url: https://cran.r-project.org/bin/windows/contrib/{{ r_base[0:3] }}/rgdal_{{ version }}.zip  # [win]
  sha256: ee4c74c3fe5dac13b34de95b518be1b9d380944b7d4fca07a489da2cb7e38766  # [win and r_base=="3.5.1"]
  sha256: 40f22da63d64f5464db5d3e2628f80892408bb7f8142ca06cb8a92b7ce2ce9fe  # [win and r_base=="3.6"]
```

Galaxy
PAMPA
SEEE
RapidMiner
Taverna
Nextflow …



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

1. Getting started
2. Tools
   1. Tutorials

tuto GBIF data pre-processing

10 shown, 15 deleted

2.97 MB

13: data 12.GeoJSON

12: Filter on data 10

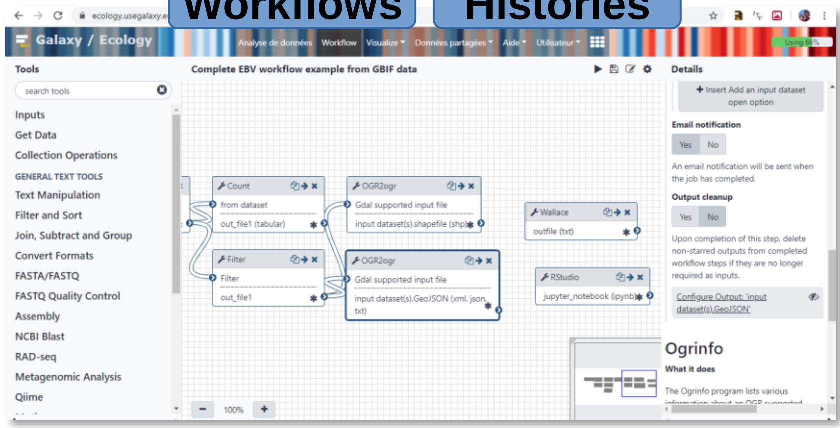
11: Count on data 1

10: Filter on data 7

9: Summary Statistics on data 1

Workflows

Histories



Complete EBV workflow example from GBIF data

+ Insert Add an input dataset open option

Email notification: Yes No

An email notification will be sent when the job has completed.

Output cleanup: Yes No

Upon completion of this step, delete non-starred outputs from completed workflow steps if they are no longer required as inputs.

Configure Output: 'input dataset(s).GeoJSON'

Ogrinfo

What it does

The Ogrinfo program lists various

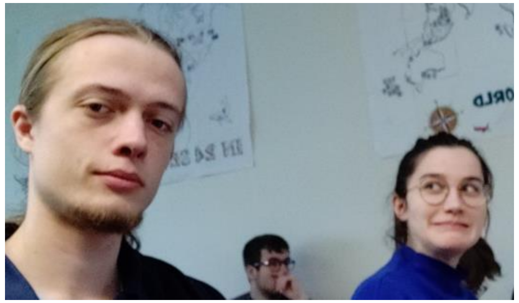
Les projets

FNSO - OpenMetaPaper

ANR Challenge IA & Biodiversité

GBIF	**GBIF France - PNDB**
iBOL2	**BOLD France - PNDB**
GEO BON	**PNDB - French BON**
	EBVs operationalization pilot
GO FAIR	**GO-FAIR IN BiodiFAIRse**
RDA	**DMP common - standard**
H2020	**EOSC Pillar infraeosc**
	GAPARS SwafS
	ERASMUS+

Merci de votre attention



PNDB team

Coline Royaux – Ingénieure développement scientifique R / Galaxy (workflows Galaxy pour calcul indicateurs espèces / communautés)

Elie Arnaud – Ingénieur développement scientifique via R Shiny / connaissance métadonnées

Julien Sananikone – Ingénieur DevOps / administration système & réseau / développeur web

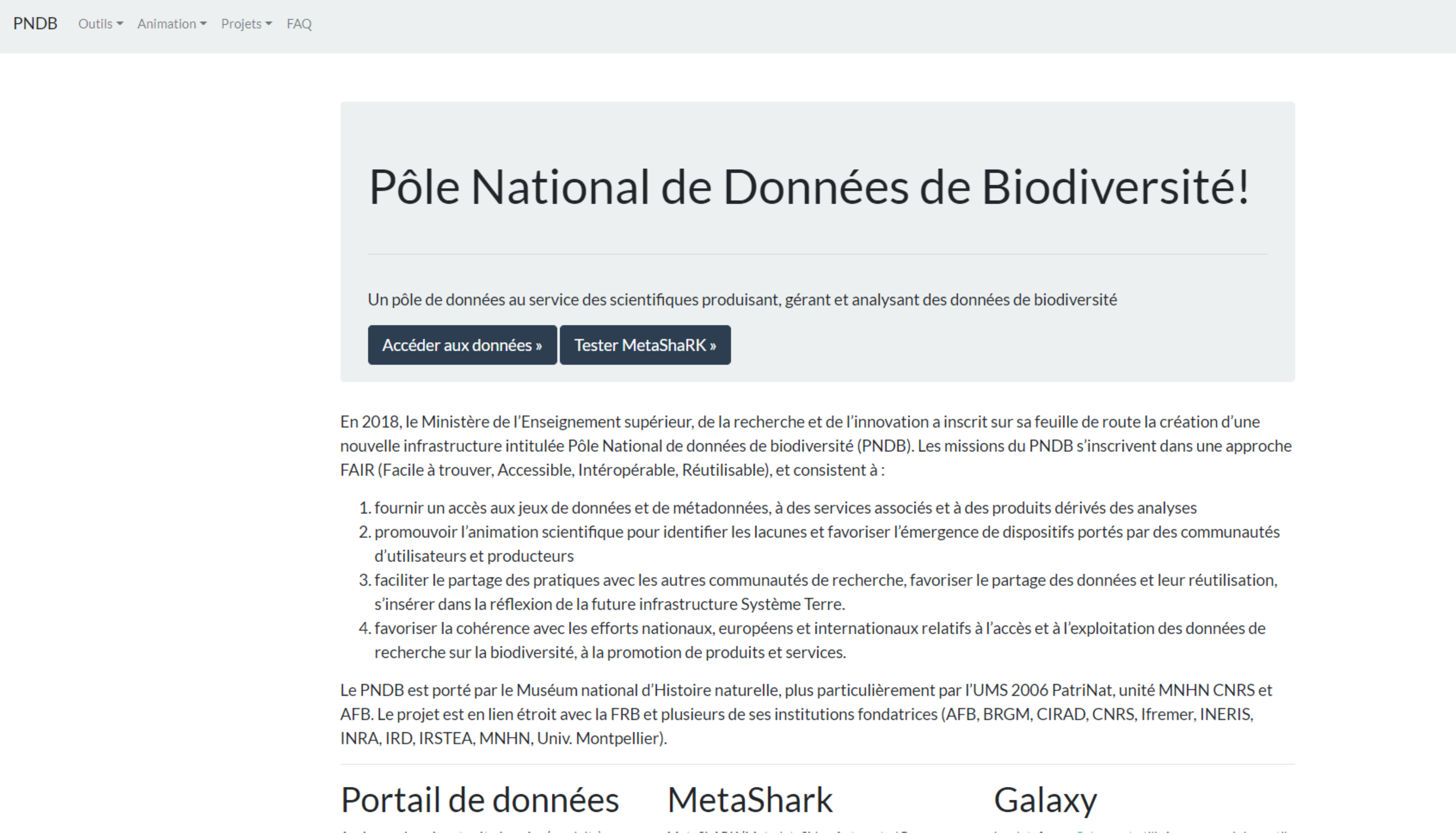
Yvan Le Bras – Béta testeur

Yvan.le-bras@mnhn.fr

https://www.pndb.fr/

Démo saisie et publication données / métadonnées MetaShARK
https://youtu.be/OVViSMzRGtw
Démo portail de données et métadonnées
https://youtu.be/STwsYDHEt2A
Démo Galaxy Europe
- https://youtu.be/HelAHggX6D4
- Essential biodiversity variables on Galaxy: implementing the PAMPA application
- Producing biodiversity indicators from citizen science projects: update of birds and bats monitoring schemes on Galaxy-E

Visitez la version préliminaire du site web du PNDB pour suivre le projet
https://www.pndb.fr/



Pôle National de Données de Biodiversité!

Un pôle de données au service des scientifiques produisant, gérant et analysant des données de biodiversité

En 2018, le Ministère de l'Enseignement supérieur, de la recherche et de l'innovation a inscrit sur sa feuille de route la création d'une nouvelle infrastructure intitulée Pôle National de données de biodiversité (PNDB). Les missions du PNDB s'inscrivent dans une approche FAIR (Facile à trouver, Accessible, Intéropérable, Réutilisable), et consistent à :

1. fournir un accès aux jeux de données et de métadonnées, à des services associés et à des produits dérivés des analyses
2. promouvoir l'animation scientifique pour identifier les lacunes et favoriser l'émergence de dispositifs portés par des communautés d'utilisateurs et producteurs
3. faciliter le partage des pratiques avec les autres communautés de recherche, favoriser le partage des données et leur réutilisation, s'insérer dans la réflexion de la future infrastructure Système Terre.
4. favoriser la cohérence avec les efforts nationaux, européens et internationaux relatifs à l'accès et à l'exploitation des données de recherche sur la biodiversité, à la promotion de produits et services.

Le PNDB est porté par le Muséum national d'Histoire naturelle, plus particulièrement par l'UMS 2006 PatriNat, unité MNHN CNRS et AFB. Le projet est en lien étroit avec la FRB et plusieurs de ses institutions fondatrices (AFB, BRGM, CIRAD, CNRS, Ifremer, INERIS, INRA, IRD, IRSTEA, MNHN, Univ. Montpellier).

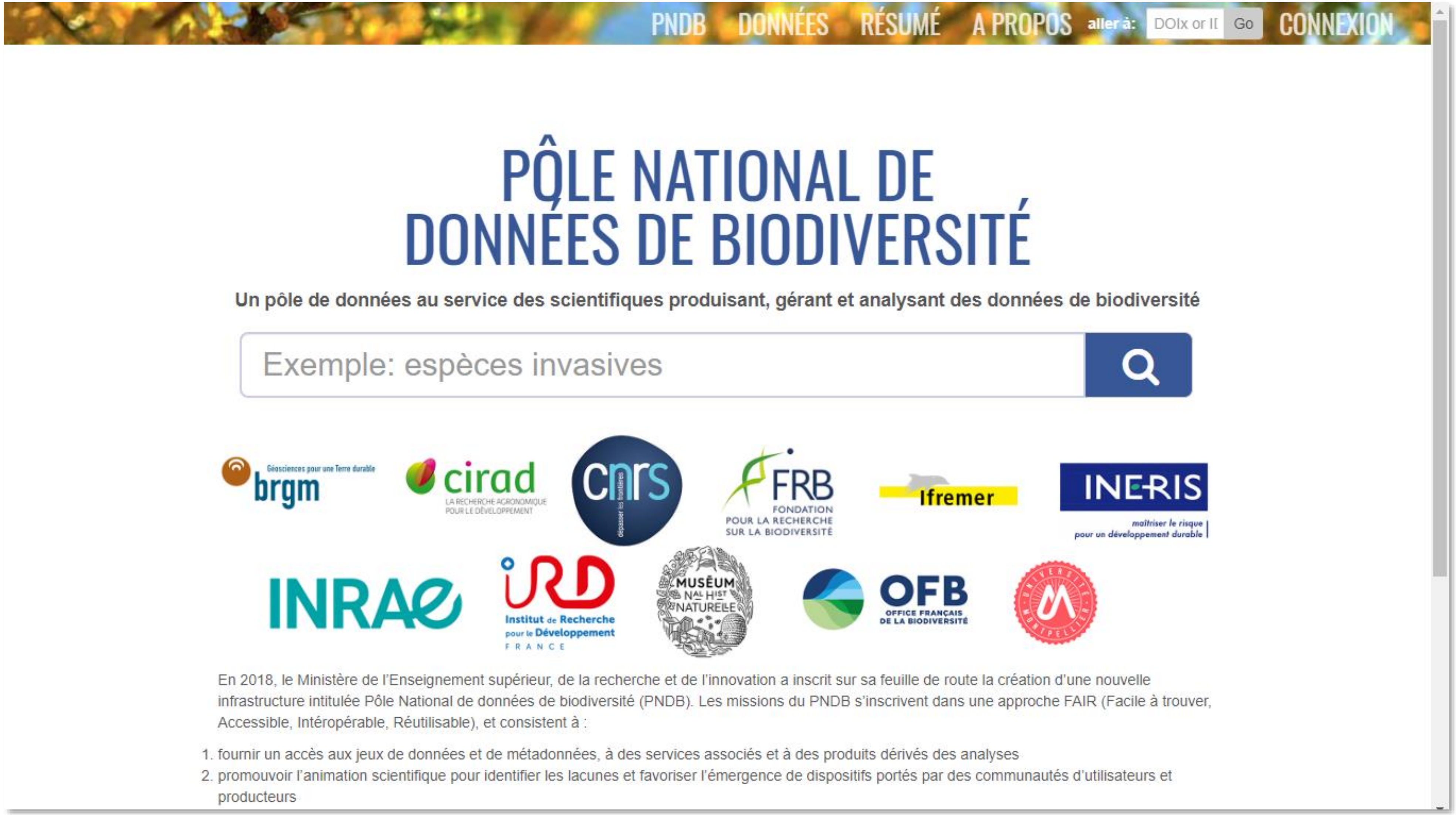
Portail de données

MetaShark

Galaxy

Visitez le portail de données du PNDB et jouez avec les premiers jeux de données-métadonnées

https://data.pndb.fr/



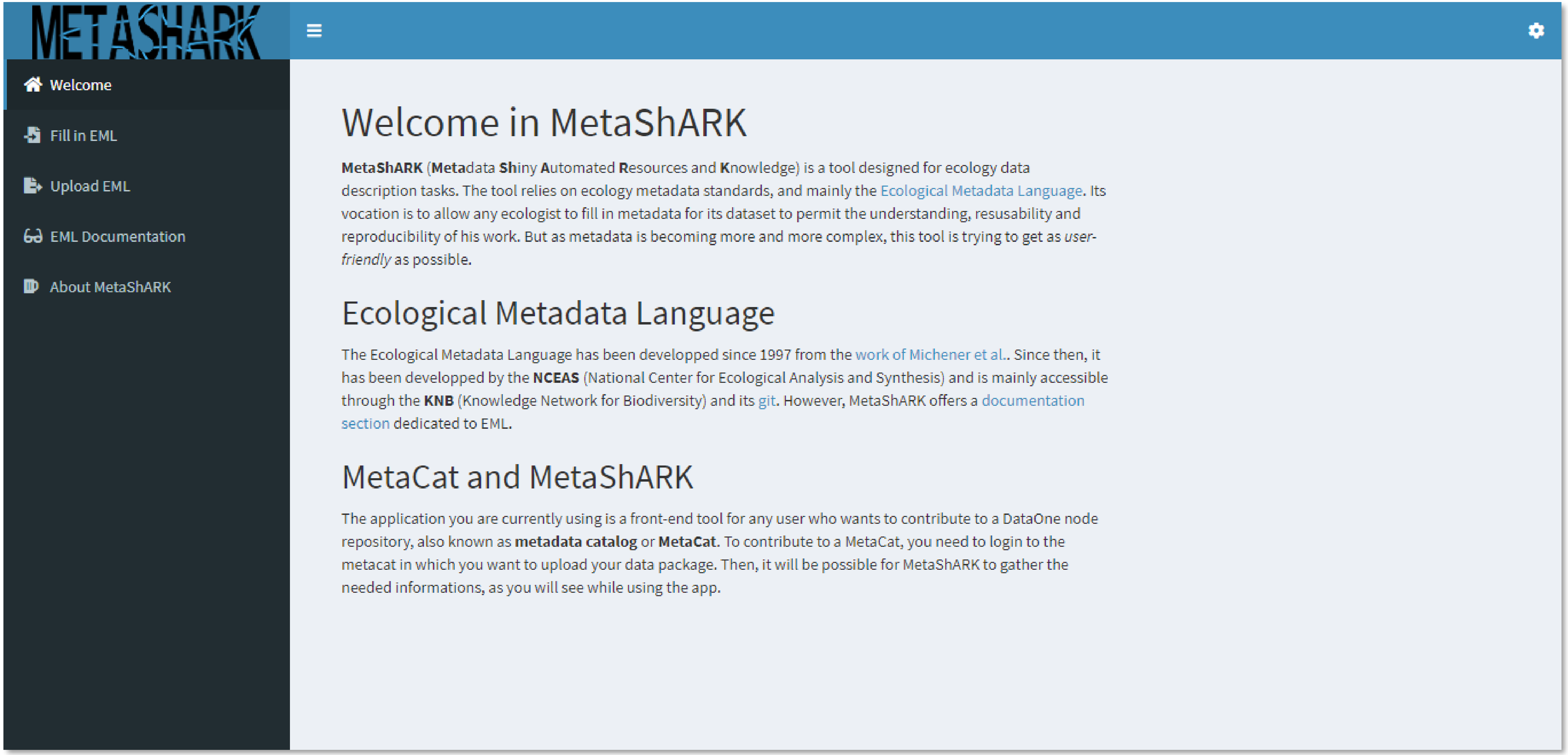
PÔLE NATIONAL DE DONNÉES DE BIODIVERSITÉ

Un pôle de données au service des scientifiques produisant, gérant et analysant des données de biodiversité

En 2018, le Ministère de l'Enseignement supérieur, de la recherche et de l'innovation a inscrit sur sa feuille de route la création d'une nouvelle infrastructure intitulée Pôle National de données de biodiversité (PNDB). Les missions du PNDB s'inscrivent dans une approche FAIR (Facile à trouver, Accessible, Intéropérable, Réutilisable), et consistent à :

1. fournir un accès aux jeux de données et de métadonnées, à des services associés et à des produits dérivés des analyses
2. promouvoir l'animation scientifique pour identifier les lacunes et favoriser l'émergence de dispositifs portés par des communautés d'utilisateurs et producteurs

Testez l’interface de saisie de données et métadonnées du PNDB
https://metashark.test.pndb.fr/



Welcome in MetaShARK

MetaShARK (**Meta**data **Sh**iny **A**utomated **R**esources and **K**nowledge) is a tool designed for ecology data description tasks. The tool relies on ecology metadata standards, and mainly the Ecological Metadata Language. Its vocation is to allow any ecologist to fill in metadata for its dataset to permit the understanding, resusability and reproducibility of his work. But as metadata is becoming more and more complex, this tool is trying to get as *user-friendly* as possible.

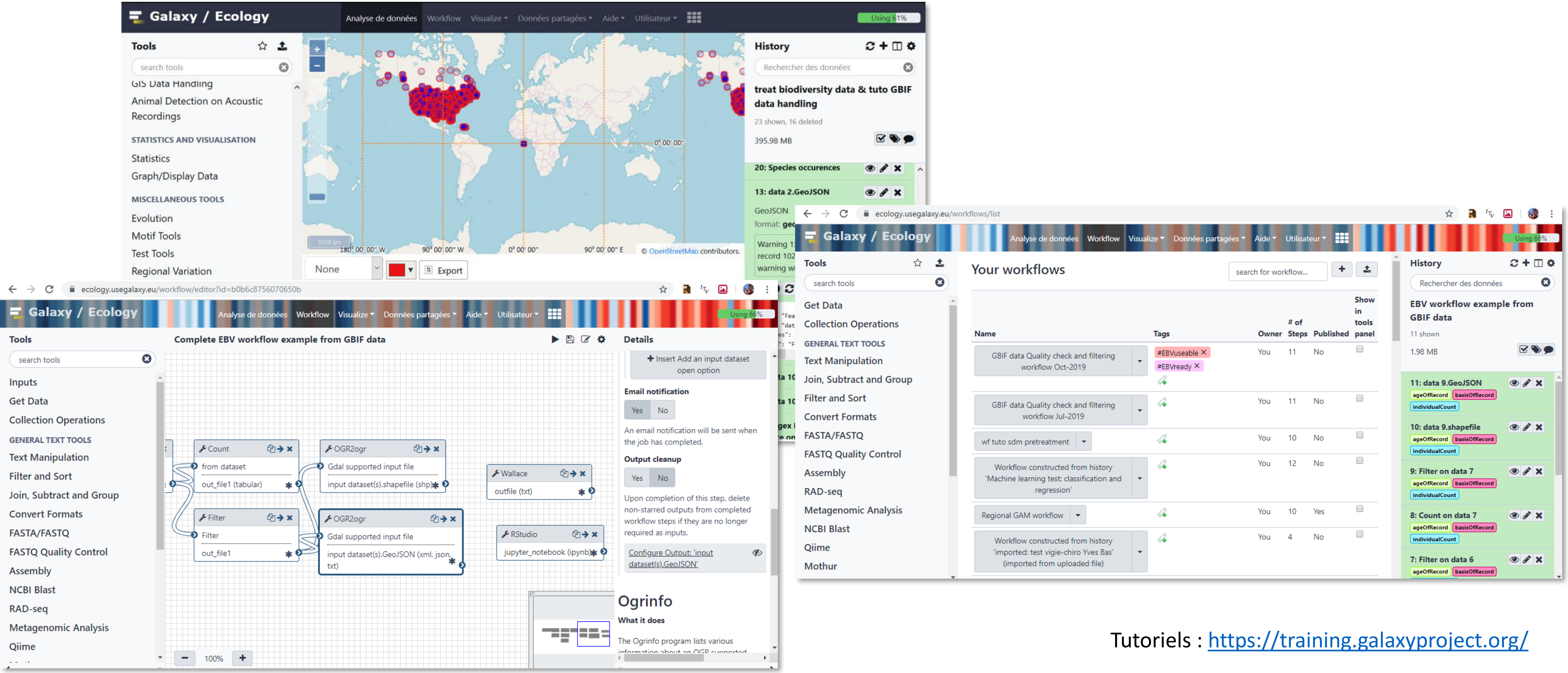
Ecological Metadata Language

The Ecological Metadata Language has been developped since 1997 from the work of Michener et al.. Since then, it has been developped by the **NCEAS** (National Center for Ecological Analysis and Synthesis) and is mainly accessible through the **KNB** (Knowledge Network for Biodiversity) and its git. However, MetaShARK offers a documentation section dedicated to EML.

MetaCat and MetaShARK

The application you are currently using is a front-end tool for any user who wants to contribute to a DataOne node repository, also known as **metadata catalog** or **MetaCat**. To contribute to a MetaCat, you need to login to the metacat in which you want to upload your data package. Then, it will be possible for MetaShARK to gather the needed informations, as you will see while using the app.

Testez la plateforme d'analyse / couplage de données du PNDB https://ecology.usegalaxy.eu/



An email notification will be sent when the job has completed.

Upon completion of this step, delete non-starred outputs from completed workflow steps if they are no longer required as inputs.

Ogrinfo

What it does

The Ogrinfo program lists various information about an OGR supported

Your workflows

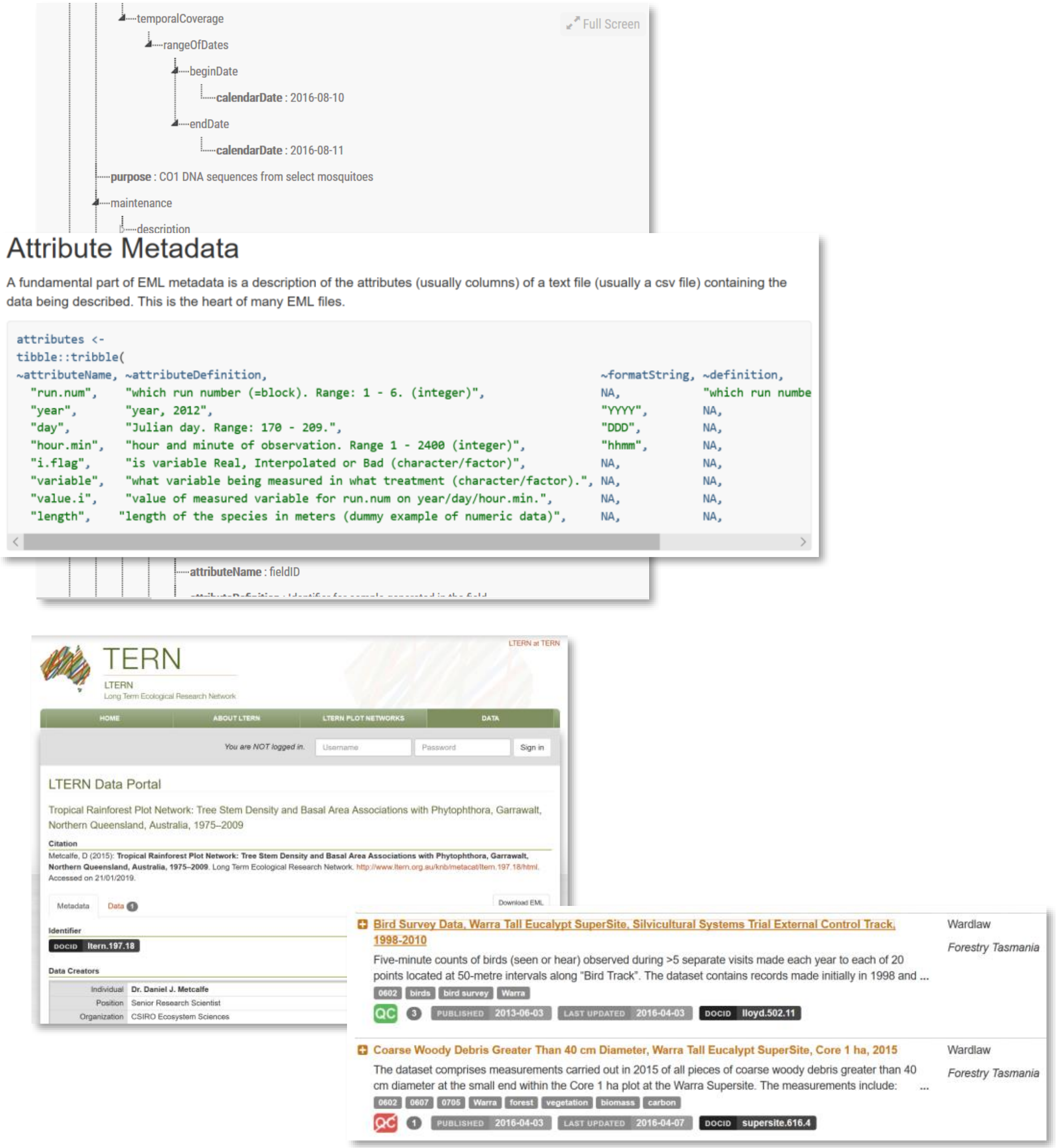
Name	Tags	Owner	# of Steps	Published	Show in tools panel
GBIF data Quality check and filtering workflow Oct-2019	#EBVuseable #EBVready	You	11	No	□
GBIF data Quality check and filtering workflow Jul-2019		You	11	No	□
wf tuto sdm pretreatment		You	10	No	□
Workflow constructed from history 'Machine learning test: classification and regression'		You	12	No	□
Regional GAM workflow		You	10	Yes	□
Workflow constructed from history 'imported: test vigie-chiro Yves Bas' (imported from uploaded file)		You	4	No	□

Tutoriels : https://training.galaxyproject.org/

Codes sources : https://github.com/65MO/Galaxy-E
https://github.com/galaxyecology/tools-ecology

Pourquoi ce fameux *Ecological Metadata Language* ?

Les LTER -> THE exemple US & Australie



Full Screen

```
temporalCoverage
  rangeOfDates
    beginDate
      calendarDate : 2016-08-10
    endDate
      calendarDate : 2016-08-11
purpose : CO1 DNA sequences from select mosquitoes
maintenance
  description
```

Attribute Metadata

A fundamental part of EML metadata is a description of the attributes (usually columns) of a text file (usually a csv file) containing the data being described. This is the heart of many EML files.

```
attributes <-
tibble::tribble(
~attributeName, ~attributeDefinition,                                               ~formatString, ~definition,
  "run.num",    "which run number (=block). Range: 1 - 6. (integer)",              NA,            "which run numbe
  "year",       "year, 2012",                                                      "YYYY",        NA,
  "day",        "Julian day. Range: 170 - 209.",                                   "DDD",         NA,
  "hour.min",   "hour and minute of observation. Range 1 - 2400 (integer)",        "hhmm",        NA,
  "i.flag",     "is variable Real, Interpolated or Bad (character/factor)",        NA,            NA,
  "variable",   "what variable being measured in what treatment (character/factor).", NA,         NA,
  "value.i",    "value of measured variable for run.num on year/day/hour.min.",    NA,            NA,
  "length",    "length of the species in meters (dummy example of numeric data)",  NA,            NA,
```

attributeName : fieldID

TERN

LTERN

Long Term Ecological Research Network

LTERN at TERN

HOME | ABOUT LTERN | LTERN PLOT NETWORKS | DATA

You are NOT logged in. Username Password Sign in

LTERN Data Portal

Tropical Rainforest Plot Network: Tree Stem Density and Basal Area Associations with Phytophthora, Garrawalt, Northern Queensland, Australia, 1975–2009

Citation

Metcalfe, D (2015): **Tropical Rainforest Plot Network: Tree Stem Density and Basal Area Associations with Phytophthora, Garrawalt, Northern Queensland, Australia, 1975–2009**. Long Term Ecological Research Network. http://www.ltern.org.au/knb/metacat/ltern.197.18/html. Accessed on 21/01/2019.

Metadata | Data 1 | Download EML

Identifier

DOCID ltern.197.18

Data Creators

Individual	Dr. Daniel J. Metcalfe
Position	Senior Research Scientist
Organization	CSIRO Ecosystem Sciences

Bird Survey Data, Warra Tall Eucalypt SuperSite, Silvicultural Systems Trial External Control Track, 1998-2010 — Wardlaw, *Forestry Tasmania*

Five-minute counts of birds (seen or hear) observed during >5 separate visits made each year to each of 20 points located at 50-metre intervals along "Bird Track". The dataset contains records made initially in 1998 and ...

0602 | birds | bird survey | Warra

QC 3 PUBLISHED 2013-06-03 LAST UPDATED 2016-04-03 DOCID lloyd.502.11

Coarse Woody Debris Greater Than 40 cm Diameter, Warra Tall Eucalypt SuperSite, Core 1 ha, 2015 — Wardlaw, *Forestry Tasmania*

The dataset comprises measurements carried out in 2015 of all pieces of coarse woody debris greater than 40 cm diameter at the small end within the Core 1 ha plot at the Warra Supersite. The measurements include: ...

0602 | 0607 | 0705 | Warra | forest | vegetation | biomass | carbon

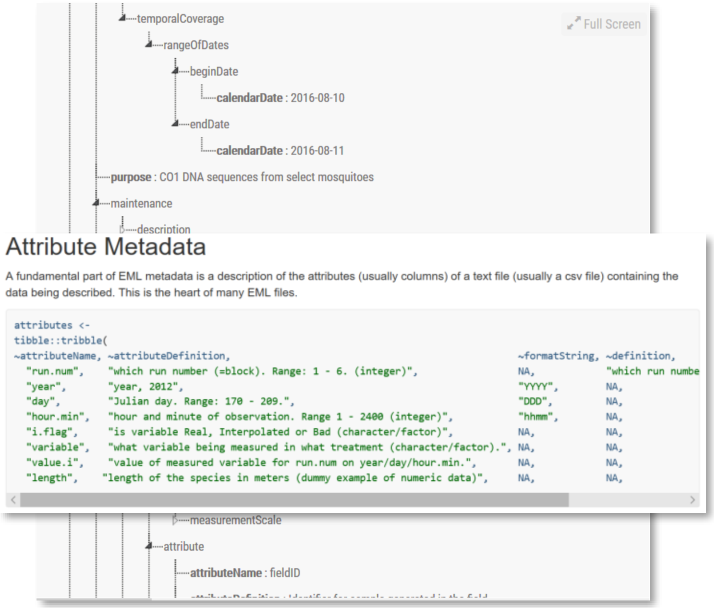
QC 1 PUBLISHED 2016-04-03 LAST UPDATED 2016-04-07 DOCID supersite.616.4

Pourquoi ce fameux *Ecological Metadata Language* ?

Les LTER -> THE exemple US & Australie

EML (MTD std) https://ropensci.github.io/EML/articles/creating-EML.html
EML Assembly Line https://ediorg.github.io/EMLassemblyline/index.html
Metacat (data portal) https://github.com/NCEAS/metacatui
OAI-PMH (harvesting)
NCEAS / EDI / KNB
DataOne

NCEAS



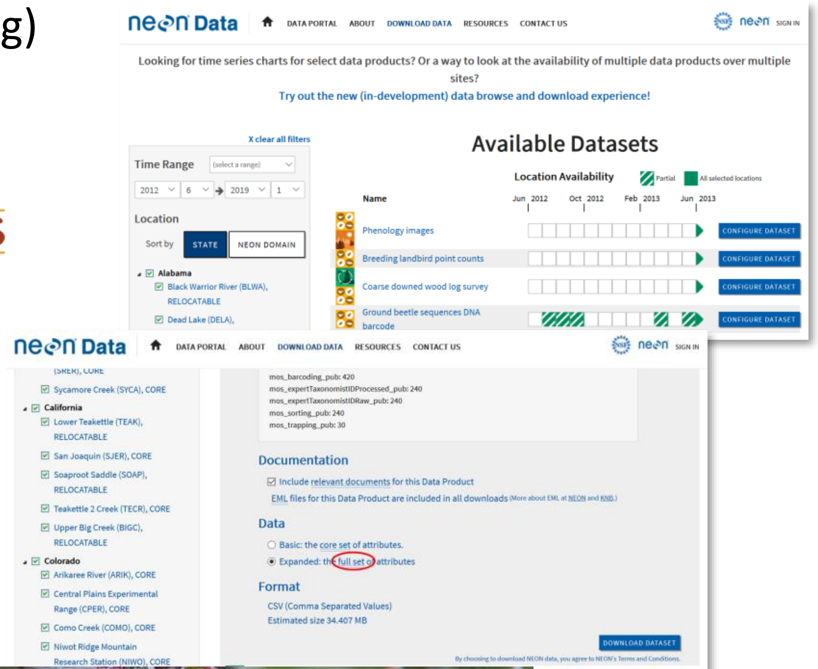
```
temporalCoverage
  rangeOfDates
    beginDate
      calendarDate : 2016-08-10
    endDate
      calendarDate : 2016-08-11
purpose : CO1 DNA sequences from select mosquitoes
maintenance
  description
```

Attribute Metadata

A fundamental part of EML metadata is a description of the attributes (usually columns) of a text file (usually a csv file) containing the data being described. This is the heart of many EML files.

```
attributes <-
tibble::tribble(
~attributeName, ~attributeDefinition,                                                  ~formatString, ~definition,
  "run.num",    "which run number (=block). Range: 1 - 6. (integer)",                  NA,            "which run numbe
  "year",       "year, 2012",                                                          "YYYY",        NA,
  "day",        "Julian day. Range: 170 - 209.",                                       "DDD",         NA,
  "hour.min",   "hour and minute of observation. Range 1 - 2400 (integer)",            "hhmm",        NA,
  "i.flag",     "is variable Real, Interpolated or Bad (character/factor)",            NA,            NA,
  "variable",   "what variable being measured in what treatment (character/factor).", NA,            NA,
  "value.i",    "value of measured variable for run.num on year/day/hour.min.",        NA,            NA,
  "length",    "length of the species in meters (dummy example of numeric data)",      NA,            NA,
```

```
measurementScale
attribute
  attributeName : fieldID
  attributeDefinition : Identifier for sample generated in the field
```



Looking for time series charts for select data products? Or a way to look at the availability of multiple data products over multiple sites?

Try out the new (in-development) data browse and download experience!

X clear all filters

Time Range (select a range)

2012 | 6 → 2019 | 1

Location

Sort by STATE | NEON DOMAIN

- ☑ Alabama
  - ☑ Black Warrior River (BLWA), RELOCATABLE
  - ☑ Dead Lake (DELA),

Available Datasets

Location Availability ▨ Partial 🟩 All selected locations

Name — Jun 2012, Oct 2012, Feb 2013, Jun 2013

- Phenology images — CONFIGURE DATASET
- Breeding landbird point counts — CONFIGURE DATASET
- Coarse downed wood log survey — CONFIGURE DATASET
- Ground beetle sequences DNA barcode — CONFIGURE DATASET

(SRER), CORE

- ☑ Sycamore Creek (SYCA), CORE
- ☑ California
  - ☑ Lower Teakettle (TEAK), RELOCATABLE
  - ☑ San Joaquin (SJER), CORE
  - ☑ Soaproot Saddle (SOAP), RELOCATABLE
  - ☑ Teakettle 2 Creek (TECR), CORE
  - ☑ Upper Big Creek (BIGC), RELOCATABLE
- ☑ Colorado
  - ☑ Arikaree River (ARIK), CORE
  - ☑ Central Plains Experimental Range (CPER), CORE
  - ☑ Como Creek (COMO), CORE
  - ☑ Niwot Ridge Mountain Research Station (NIWO), CORE

mos_barcoding_pub: 420
mos_expertTaxonomistIDProcessed_pub: 240
mos_expertTaxonomistIDRaw_pub: 240
mos_sorting_pub: 240
mos_trapping_pub: 30

Documentation

☑ Include relevant documents for this Data Product

EML files for this Data Product are included in all downloads (More about EML at NEON and KNB.)

Data

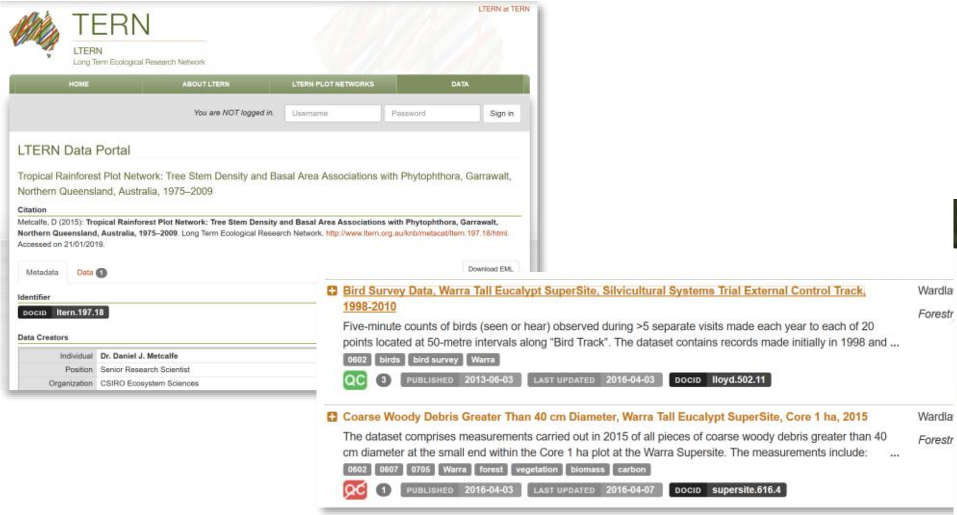
○ Basic: the core set of attributes.
● Expanded: the full set of attributes

Format

CSV (Comma Separated Values)
Estimated size 34.407 MB

DOWNLOAD DATASET

By choosing to download NEON data, you agree to NEON's Terms and Conditions.



You are NOT logged in.

LTERN Data Portal

Tropical Rainforest Plot Network: Tree Stem Density and Basal Area Associations with Phytophthora, Garrawalt, Northern Queensland, Australia, 1975–2009

Citation

Metcalfe, D (2015): **Tropical Rainforest Plot Network: Tree Stem Density and Basal Area Associations with Phytophthora, Garrawalt, Northern Queensland, Australia, 1975–2009**. Long Term Ecological Research Network. http://www.ltern.org.au/knb/metacat/ltern.197.18/html. Accessed on 21/01/2019.

Metadata | Data 1 | Download EML

Identifier

DOCID ltern.197.18

Data Creators

Individual	Dr. Daniel J. Metcalfe
Position	Senior Research Scientist
Organization	CSIRO Ecosystem Sciences

Bird Survey Data, Warra Tall Eucalypt SuperSite, Silvicultural Systems Trial External Control Track, 1998-2010

Five-minute counts of birds (seen or hear) observed during >5 separate visits made each year to each of 20 points located at 50-metre intervals along "Bird Track". The dataset contains records made initially in 1998 and ...

0602 | birds | bird survey | Warra

QC 3 | PUBLISHED 2013-06-03 | LAST UPDATED 2016-04-03 | DOCID lloyd.502.11

Wardla
Forestr

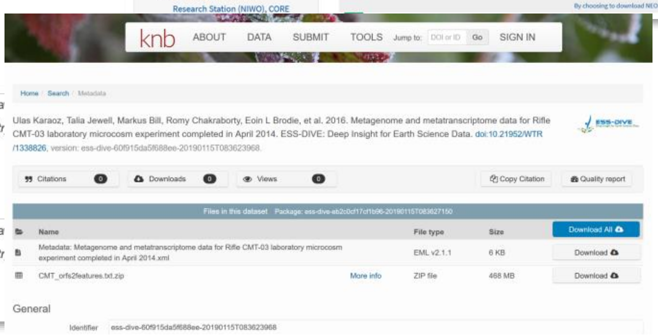
Coarse Woody Debris Greater Than 40 cm Diameter, Warra Tall Eucalypt SuperSite, Core 1 ha, 2015

The dataset comprises measurements carried out in 2015 of all pieces of coarse woody debris greater than 40 cm diameter at the small end within the Core 1 ha plot at the Warra Supersite. The measurements include: ...

0602 | 0607 | 0705 | Warra | forest | vegetation | biomass | carbon

QC 1 | PUBLISHED 2016-04-03 | LAST UPDATED 2016-04-07 | DOCID supersite.616.4

Wardla
Forestr



Home / Search / Metadata

Ulas Karaoz, Talia Jewell, Markus Bill, Romy Chakraborty, Eoin L Brodie, et al. 2016. Metagenome and metatranscriptome data for Rifle CMT-03 laboratory microcosm experiment completed in April 2014. ESS-DIVE: Deep Insight for Earth Science Data. doi:10.21952/WTR/1338826, version: ess-dive-60f915da5f688ee-20190115T083623968.

Citations 0 | Downloads 0 | Views 0 | Copy Citation | Quality report

Files in this dataset Package: ess-dive-ab2c0cf17cf1b96-20190115T083627150

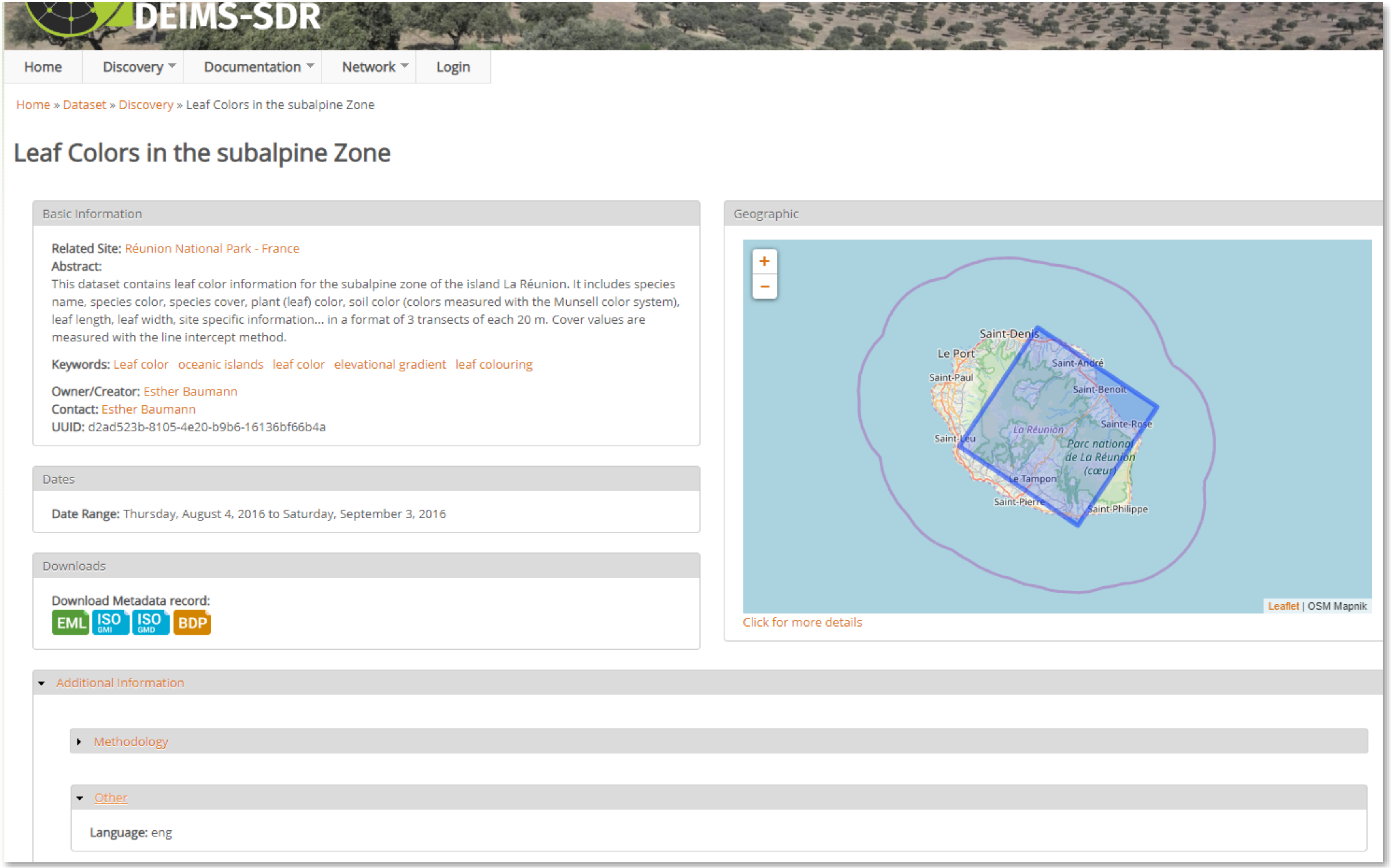
Name		File type	Size	Download All
Metadata: Metagenome and metatranscriptome data for Rifle CMT-03 laboratory microcosm experiment completed in April 2014.xml		EML v2.1.1	6 KB	Download
CMT_orfs2features.txt.zip	More info	ZIP file	468 MB	Download

General

Identifier ess-dive-60f915da5f688ee-20190115T083623968

Pourquoi ce fameux *Ecological Metadata Language* ?

Les LTER -> THE exemple Europe



Home » Dataset » Discovery » Leaf Colors in the subalpine Zone

Leaf Colors in the subalpine Zone

Basic Information

Related Site: Réunion National Park - France
Abstract:
This dataset contains leaf color information for the subalpine zone of the island La Réunion. It includes species name, species color, species cover, plant (leaf) color, soil color (colors measured with the Munsell color system), leaf length, leaf width, site specific information... in a format of 3 transects of each 20 m. Cover values are measured with the line intercept method.

Keywords: Leaf color oceanic islands leaf color elevational gradient leaf colouring

Owner/Creator: Esther Baumann
Contact: Esther Baumann
UUID: d2ad523b-8105-4e20-b9b6-16136bf66b4a

Dates

Date Range: Thursday, August 4, 2016 to Saturday, September 3, 2016

Downloads

Download Metadata record:
EML ISO GMI ISO GMD BDP

Geographic

Click for more details

Additional Information

Methodology

Other

Language: eng

Pourquoi ce fameux *Ecological Metadata Language* ?

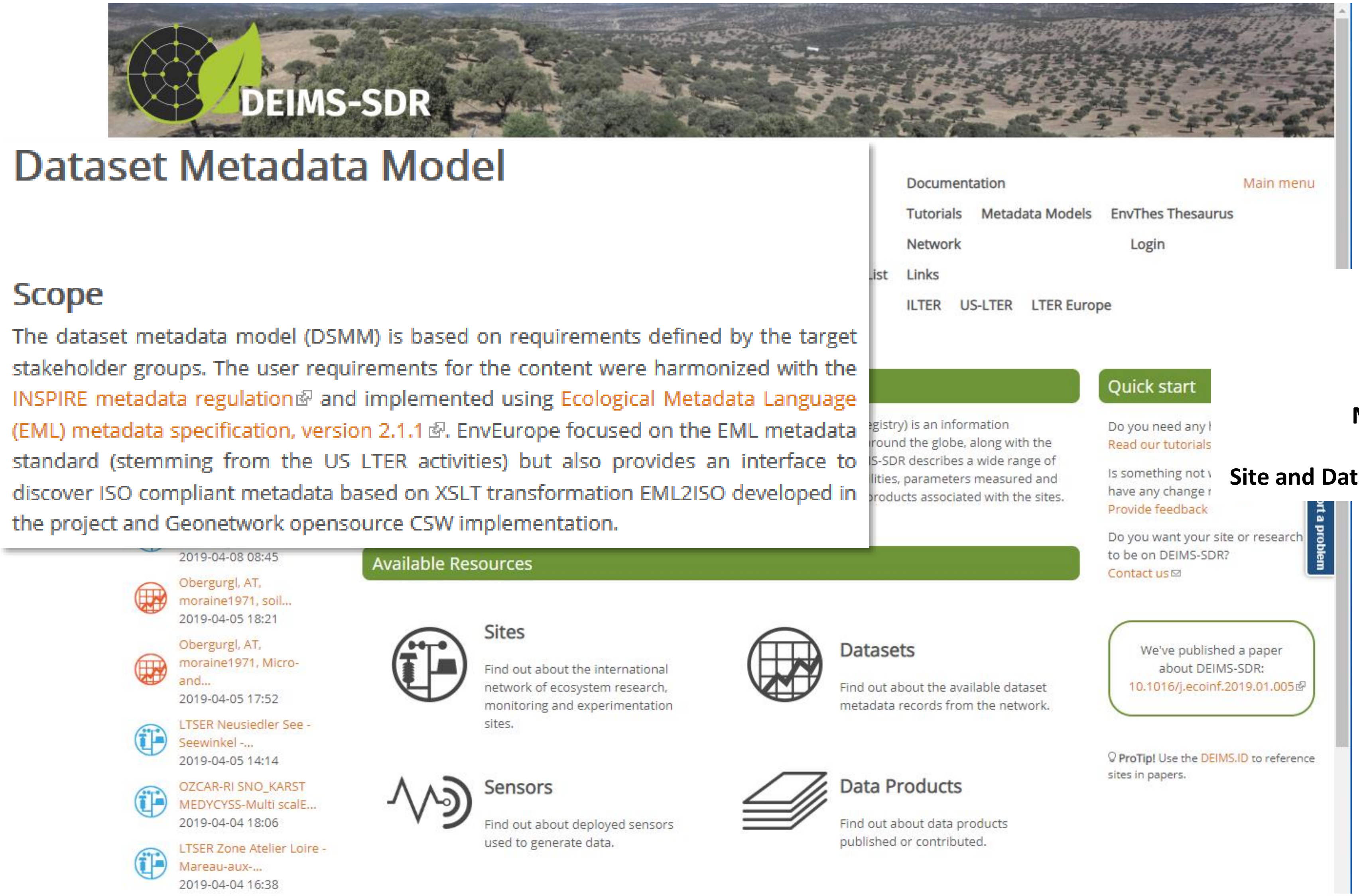
Les LTER -> THE exemple Europe

~2010

Drupal Ecological Information Management System

2018

DEIMS-SDR
Dynamic
Ecological
Information
Management
System
Site and Dataset Registry



Dataset Metadata Model

Scope

The dataset metadata model (DSMM) is based on requirements defined by the target stakeholder groups. The user requirements for the content were harmonized with the INSPIRE metadata regulation and implemented using Ecological Metadata Language (EML) metadata specification, version 2.1.1. EnvEurope focused on the EML metadata standard (stemming from the US LTER activities) but also provides an interface to discover ISO compliant metadata based on XSLT transformation EML2ISO developed in the project and Geonetwork opensource CSW implementation.

egistry) is an information
round the globe, along with the
S-SDR describes a wide range of
lities, parameters measured and
products associated with the sites.

Do you need any
Read our tutorials

Is something not
have any change
Provide feedback

Do you want your site or research to be on DEIMS-SDR?
Contact us

2019-04-08 08:45

Obergurgl, AT, moraine1971, soil...
2019-04-05 18:21

Obergurgl, AT, moraine1971, Micro- and...
2019-04-05 17:52

LTSER Neusiedler See - Seewinkel -...
2019-04-05 14:14

OZCAR-RI SNO_KARST MEDYCYSS-Multi scalE...
2019-04-04 18:06

LTSER Zone Atelier Loire - Mareau-aux-...
2019-04-04 16:38

Find out about the international network of ecosystem research, monitoring and experimentation sites.

Find out about the available dataset metadata records from the network.

Find out about deployed sensors used to generate data.

Find out about data products published or contributed.

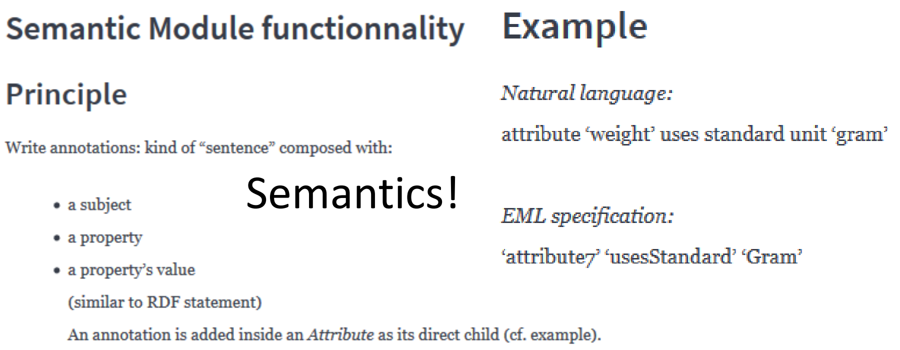
We've published a paper about DEIMS-SDR: 10.1016/j.ecoinf.2019.01.005

ProTip! Use the DEIMS.ID to reference sites in papers.

Pourquoi ce fameux *Ecological Metadata Language* ?

Amazing EML

Language + Modules -> flexibility!



Semantic Module functionnality

Principle

Write annotations: kind of "sentence" composed with:

- a subject
- a property
- a property's value

(similar to RDF statement)

An annotation is added inside an *Attribute* as its direct child (cf. example).

Example

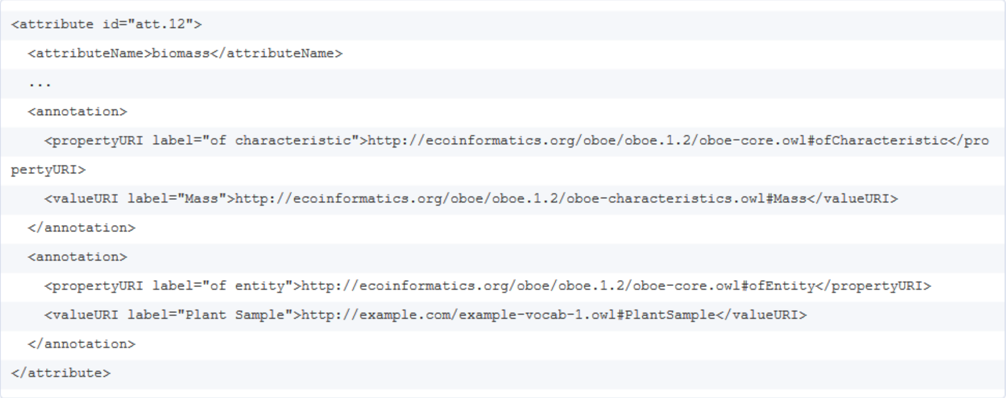
Natural language:

attribute 'weight' uses standard unit 'gram'

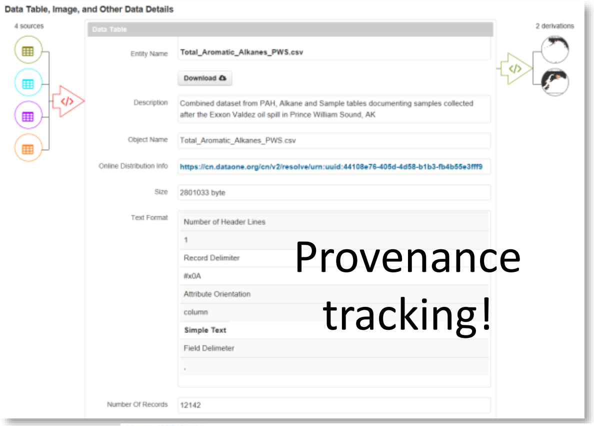
EML specification:

'attribute7' 'usesStandard' 'Gram'

Semantics!



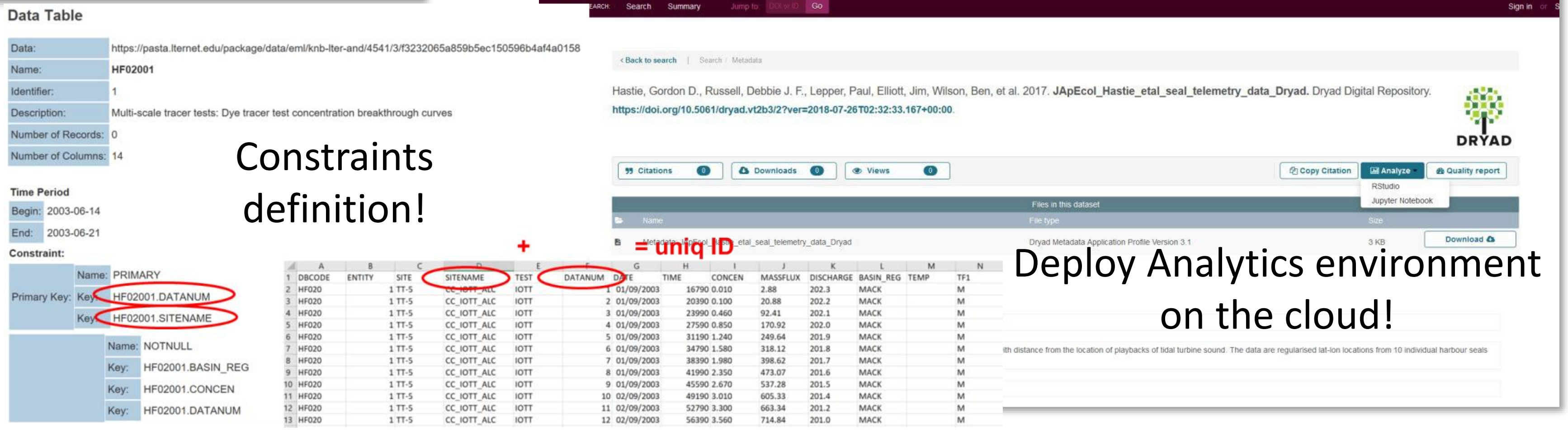
```
<attribute id="att.12">
  <attributeName>biomass</attributeName>
  ...
  <annotation>
    <propertyURI label="of characteristic">http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#ofCharacteristic</pro
pertyURI>
    <valueURI label="Mass">http://ecoinformatics.org/oboe/oboe.1.2/oboe-characteristics.owl#Mass</valueURI>
  </annotation>
  <annotation>
    <propertyURI label="of entity">http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#ofEntity</propertyURI>
    <valueURI label="Plant Sample">http://example.com/example-vocab-1.owl#PlantSample</valueURI>
  </annotation>
</attribute>
```



Data Table, Image, and Other Data Details

Entity Name	Total_Aromatic_Alkanes_PWS.csv
Description	Combined dataset from PAH, Alkane and Sample tables documenting samples collected after the Exxon Valdez oil spill in Prince William Sound, AK
Object Name	Total_Aromatic_Alkanes_PWS.csv
Online Distribution Info	https://cn.dataone.org/cn/v2/resolve/urn:uuid:44108e76-405d-4d58-b1b3-fb4b55e3fff9
Size	2801033 byte
Text Format	Number of Header Lines 1 Record Delimiter #x0A Attribute Orientation column **Simple Text** Field Delimeter ,
Number Of Records	12142

Provenance tracking!



Data Table

Data:	https://pasta.lternet.edu/package/data/eml/knb-lter-and/4541/3/f3232065a859b5ec150596b4af4a0158
Name:	HF02001
Identifier:	1
Description:	Multi-scale tracer tests: Dye tracer test concentration breakthrough curves
Number of Records:	0
Number of Columns:	14

Time Period

Begin:	2003-06-14
End:	2003-06-21

Constraint:

Primary Key:	Name:	PRIMARY
	Key:	HF02001.DATANUM
	Key:	HF02001.SITENAME
	Name:	NOTNULL
	Key:	HF02001.BASIN_REG
	Key:	HF02001.CONCEN
	Key:	HF02001.DATANUM

Constraints definition!

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	DBCODE	ENTITY	SITE	SITENAME	TEST	DATANUM	DATE	TIME	CONCEN	MASSFLUX	DISCHARGE	BASIN_REG	TEMP	TF1
2	HF020	1	TT-5	CC_IOTT_ALC	IOTT	1	01/09/2003	16790	0.010	2.88	202.3	MACK		M
3	HF020	1	TT-5	CC_IOTT_ALC	IOTT	2	01/09/2003	20390	0.100	20.88	202.2	MACK		M
4	HF020	1	TT-5	CC_IOTT_ALC	IOTT	3	01/09/2003	23990	0.460	92.41	202.1	MACK		M
5	HF020	1	TT-5	CC_IOTT_ALC	IOTT	4	01/09/2003	27590	0.850	170.92	202.0	MACK		M
6	HF020	1	TT-5	CC_IOTT_ALC	IOTT	5	01/09/2003	31190	1.240	249.64	201.9	MACK		M
7	HF020	1	TT-5	CC_IOTT_ALC	IOTT	6	01/09/2003	34790	1.580	318.12	201.8	MACK		M
8	HF020	1	TT-5	CC_IOTT_ALC	IOTT	7	01/09/2003	38390	1.980	398.62	201.7	MACK		M
9	HF020	1	TT-5	CC_IOTT_ALC	IOTT	8	01/09/2003	41990	2.350	473.07	201.6	MACK		M
10	HF020	1	TT-5	CC_IOTT_ALC	IOTT	9	01/09/2003	45590	2.670	537.28	201.5	MACK		M
11	HF020	1	TT-5	CC_IOTT_ALC	IOTT	10	02/09/2003	49190	3.010	605.33	201.4	MACK		M
12	HF020	1	TT-5	CC_IOTT_ALC	IOTT	11	02/09/2003	52790	3.300	663.34	201.2	MACK		M
13	HF020	1	TT-5	CC_IOTT_ALC	IOTT	12	02/09/2003	56390	3.560	714.84	201.0	MACK		M

+ = uniq ID

Hastie, Gordon D., Russell, Debbie J. F., Lepper, Paul, Elliott, Jim, Wilson, Ben, et al. 2017. **JApEcol_Hastie_etal_seal_telemetry_data_Dryad.** Dryad Digital Repository. https://doi.org/10.5061/dryad.vt2b3/2?ver=2018-07-26T02:32:33.167+00:00.

Name	File type	Size
Metadata: JApEcol_Hastie_etal_seal_telemetry_data_Dryad	Dryad Metadata Application Profile Version 3.1	3 KB

ith distance from the location of playbacks of tidal turbine sound. The data are regularised lat-lon locations from 10 individual harbour seals

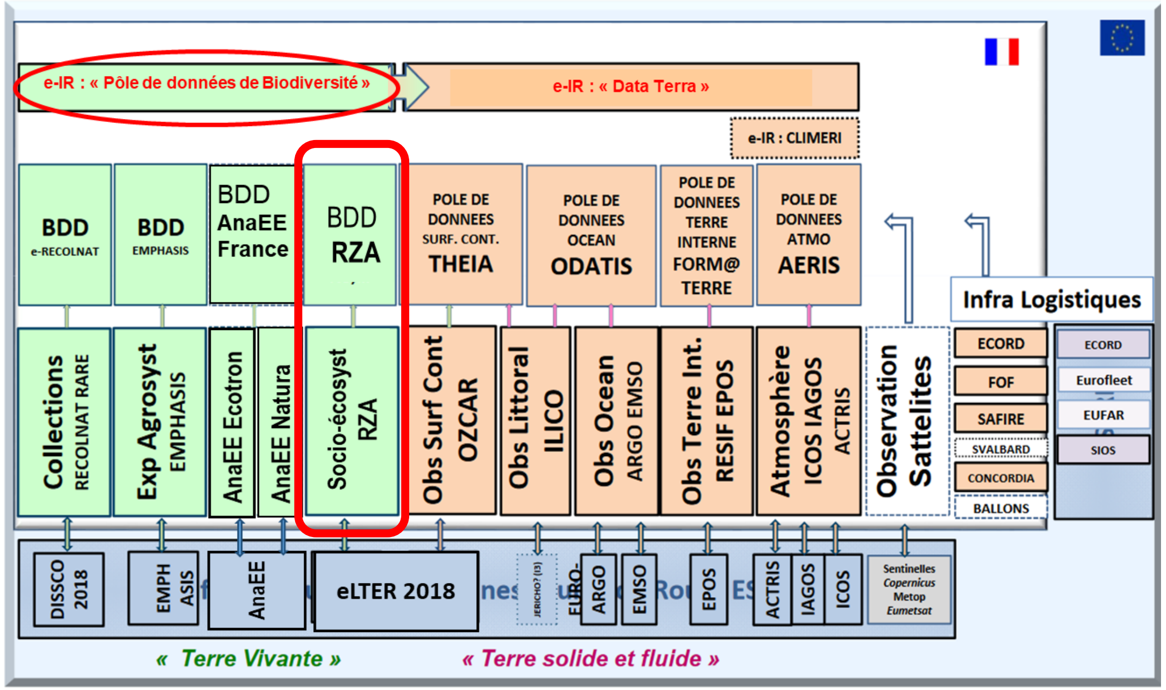
Deploy Analytics environment on the cloud!

Proposition de rôle via BED

- Liens communautés ZA -> PNDB

- Support technique PNDB

- Retours / conseils vis-à-vis initiatives internationales (GEO BON dont EBV / DataOne ..)



e-IR : « Pôle de données de Biodiversité »

e-IR : « Data Terra »

e-IR : CLIMERI

BDD e-RECOLNAT

BDD EMPHASIS

BDD AnaEE France

BDD RZA

POLE DE DONNEES SURF. CONT. THEIA

POLE DE DONNEES OCEAN ODATIS

POLE DE DONNEES TERRE INTERNE FORM@TERRE

POLE DE DONNEES ATMO AERIS

Infra Logistiques

Collections RECOLNAT RARE

Exp Agrosyst EMPHASIS

AnaEE Ecotron

AnaEE Natura

Socio-écosyst RZA

Obs Surf Cont OZCAR

Obs Littoral ILICO

Obs Ocean ARGO EMSO

Obs Terre Int. RESIF EPOS

Atmosphère ICOS IAGOS ACTRIS

Observation Satellites

ECORD

FOF

SAFIRE

SVALBARD

CONCORDIA

BALLONS

ECORD

Eurofleet

EUFAR

SIOS

DISSCO 2018

EMPH ASIS

AnaEE

eLTER 2018

JERICHO? (I3)

EURO-

ARGO

EMSO

EPOS

ACTRIS

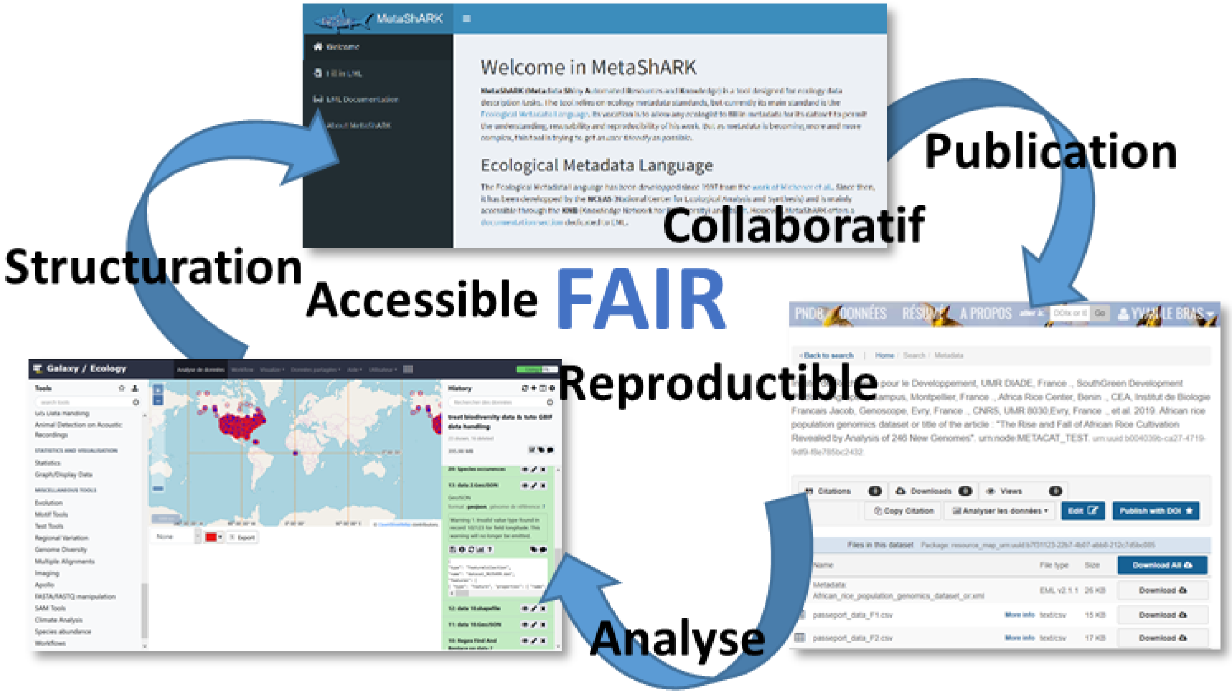
IAGOS

ICOS

Sentinelles *Copernicus* Metop *Eumetsat*

« Terre Vivante »

« Terre solide et fluide »



Structuration

Publication

Collaboratif

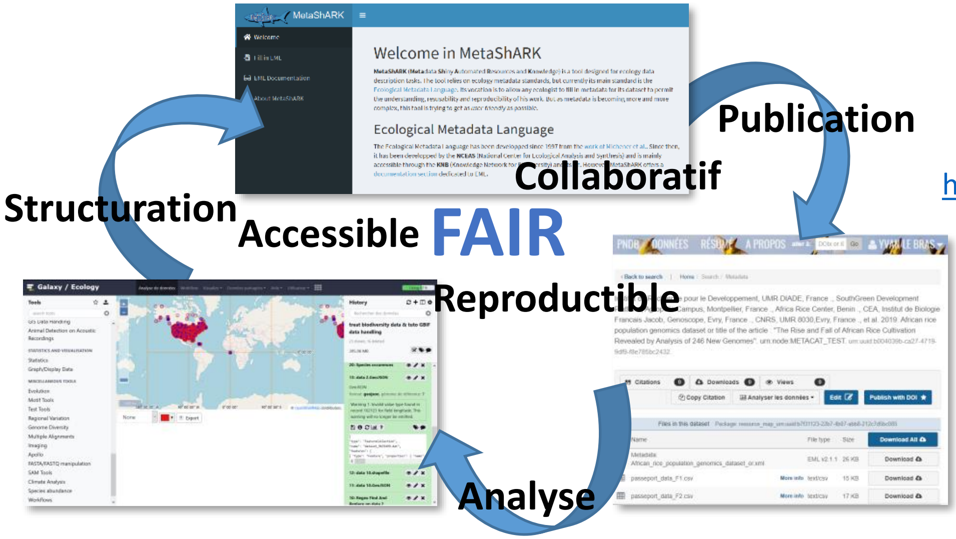
Accessible

FAIR

Reproductible

Analyse

Démonstrations !!!!!!!



Publication

Collaboratif

Structuration

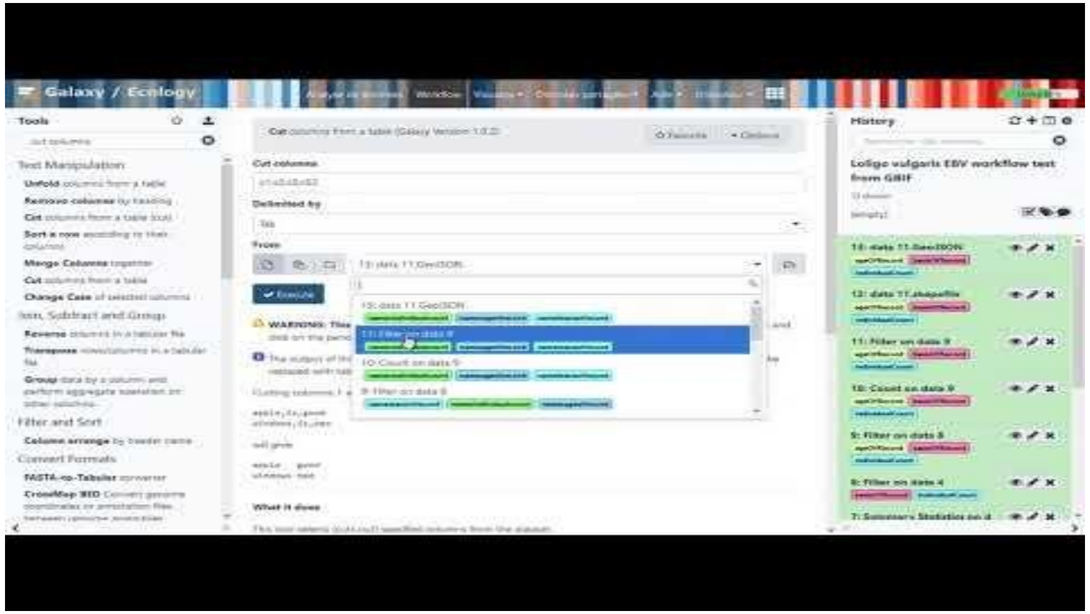
Accessible **FAIR**

Reproductible

Analyse

Démo saisie et publication données / métadonnées MetaShARK
https://youtu.be/OVViSMzRGtw

Démo portail de données et métadonnées basé sur Metacat/metacatui/Geoserver
https://youtu.be/STwsYDHEt2A



Démo plateforme d’analyse de données de biodiversité via Galaxy Europe
https://youtu.be/HelAHggX6D4

MetaShARK

Au début était l’EML

Ecological Metadata Language (by knb)

Au début était l’EML

Ecological Metadata Language (by knb)

Solutions techniques existantes

- MetaCAT (→ DataONE)
- Morpho (écriture métadonnées)

Au début était l’EML

Ecological Metadata Language (by knb)

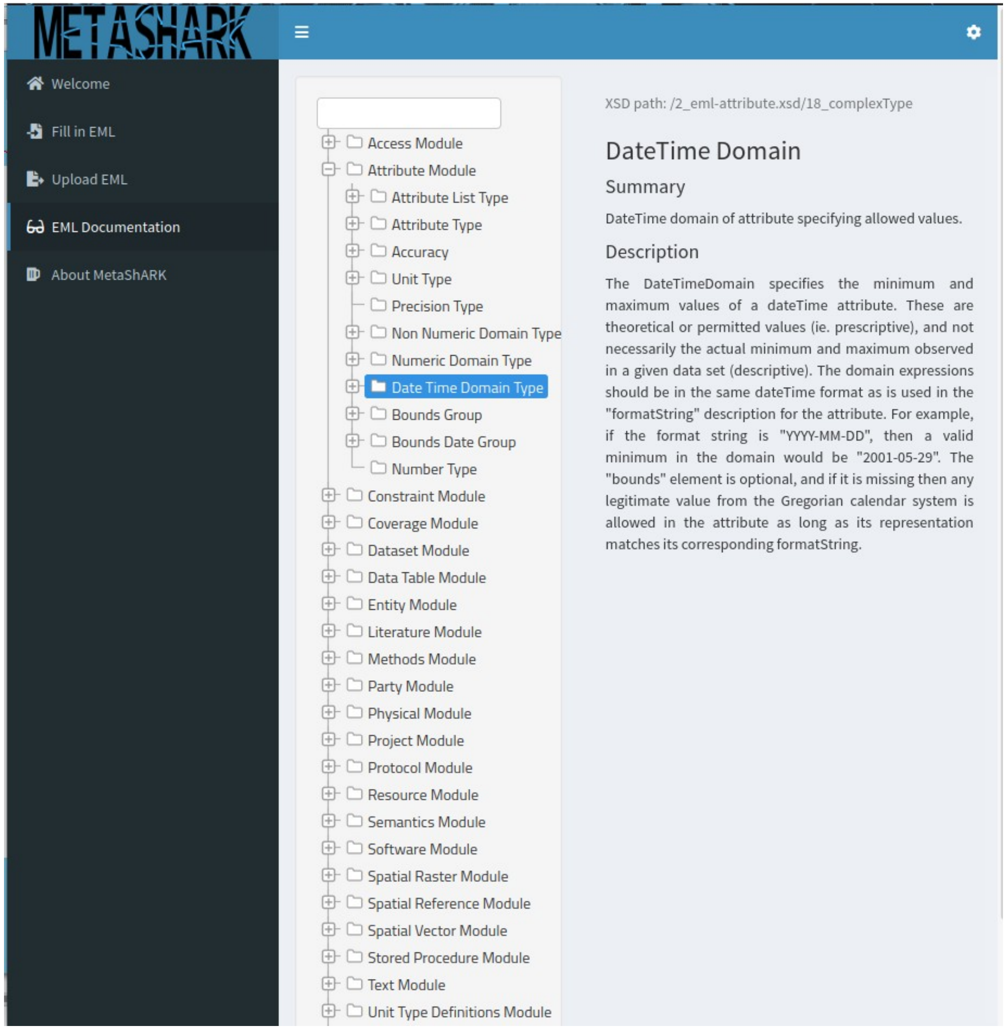
Solutions techniques existantes

- MetaCAT (→ DataONE)
- Morpho (écriture métadonnées)
  - **MetaShARK** =

    EML Assembly Line (by EDI) + Travail de stage 2019

Documentation

- Localisation
- Interactions
- Recherche
- Utilisation
- Infos techniques



XSD path: /2_eml-attribute.xsd/18_complexType

DateTime Domain

Summary

DateTime domain of attribute specifying allowed values.

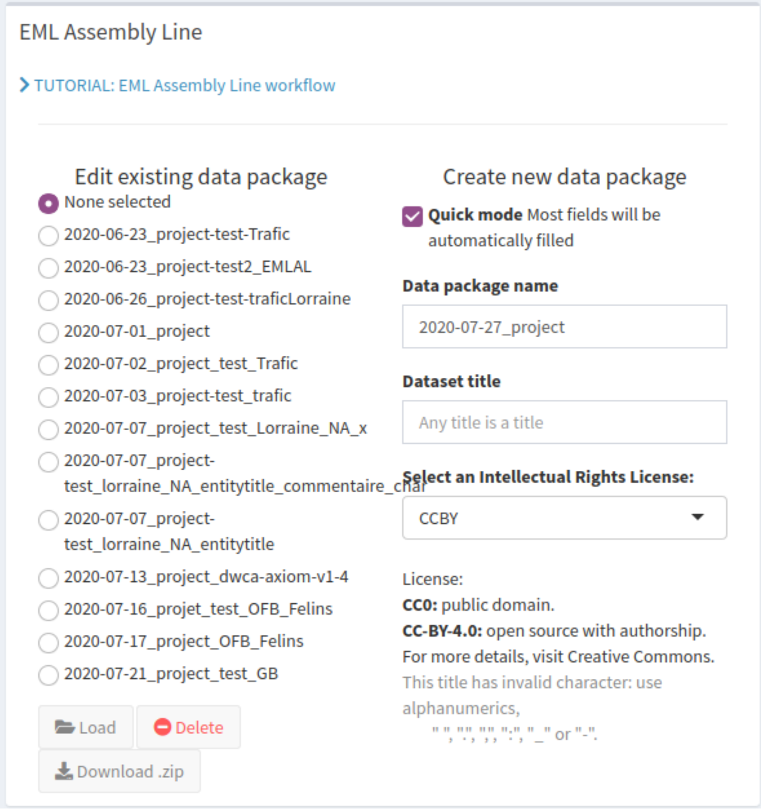
Description

The DateTimeDomain specifies the minimum and maximum values of a dateTime attribute. These are theoretical or permitted values (ie. prescriptive), and not necessarily the actual minimum and maximum observed in a given data set (descriptive). The domain expressions should be in the same dateTime format as is used in the "formatString" description for the attribute. For example, if the format string is "YYYY-MM-DD", then a valid minimum in the domain would be "2001-05-29". The "bounds" element is optional, and if it is missing then any legitimate value from the Gregorian calendar system is allowed in the attribute as long as its representation matches its corresponding formatString.

Saisie

Etape 1: sélection d’un projet

- Structuration en *data packages:* données + métadonnées
- Gestion de la propriété intellectuelle



EML Assembly Line

Edit existing data package

- ● None selected
- ○ 2020-06-23_project-test-Trafic
- ○ 2020-06-23_project-test2_EMLAL
- ○ 2020-06-26_project-test-traficLorraine
- ○ 2020-07-01_project
- ○ 2020-07-02_project_test_Trafic
- ○ 2020-07-03_project-test_trafic
- ○ 2020-07-07_project_test_Lorraine_NA_x
- ○ 2020-07-07_project-test_lorraine_NA_entitytitle_commentaire_char
- ○ 2020-07-07_project-test_lorraine_NA_entitytitle
- ○ 2020-07-13_project_dwca-axiom-v1-4
- ○ 2020-07-16_projet_test_OFB_Felins
- ○ 2020-07-17_project_OFB_Felins
- ○ 2020-07-21_project_test_GB

Create new data package

☑ **Quick mode** Most fields will be automatically filled

Data package name

2020-07-27_project

Dataset title

Any title is a title

Select an Intellectual Rights License:

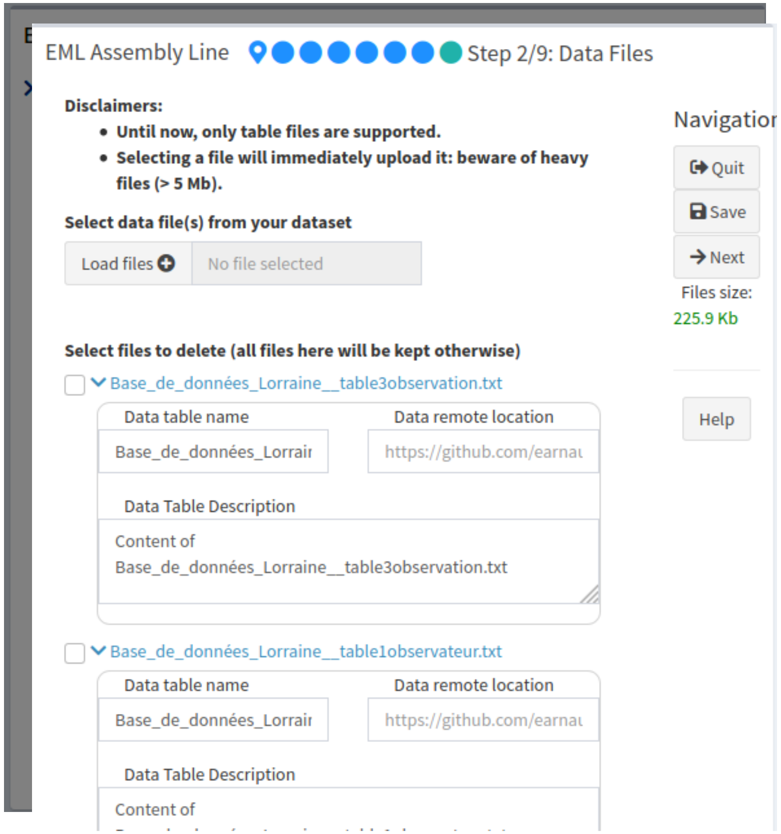
CCBY

License:
CC0: public domain.
CC-BY-4.0: open source with authorship.
For more details, visit Creative Commons.
This title has invalid character: use alphanumerics, " ", ".", ",", ":", "_" or "-".

Saisie

Etape 2: téléchargement des fichiers à décrire

- Fichiers tableurs (pour l'instant)
- Description de chaque fichier



Disclaimers:

- Until now, only table files are supported.
- Selecting a file will immediately upload it: beware of heavy files (> 5 Mb).

Select data file(s) from your dataset

No file selected

Files size: 225.9 Kb

Select files to delete (all files here will be kept otherwise)

□ Base_de_données_Lorraine__table3observation.txt

Data table name: Base_de_données_Lorrair

Data remote location: https://github.com/earnaı

Data Table Description: Content of Base_de_données_Lorraine__table3observation.txt

□ Base_de_données_Lorraine__table1observateur.txt

Data table name: Base_de_données_Lorrair

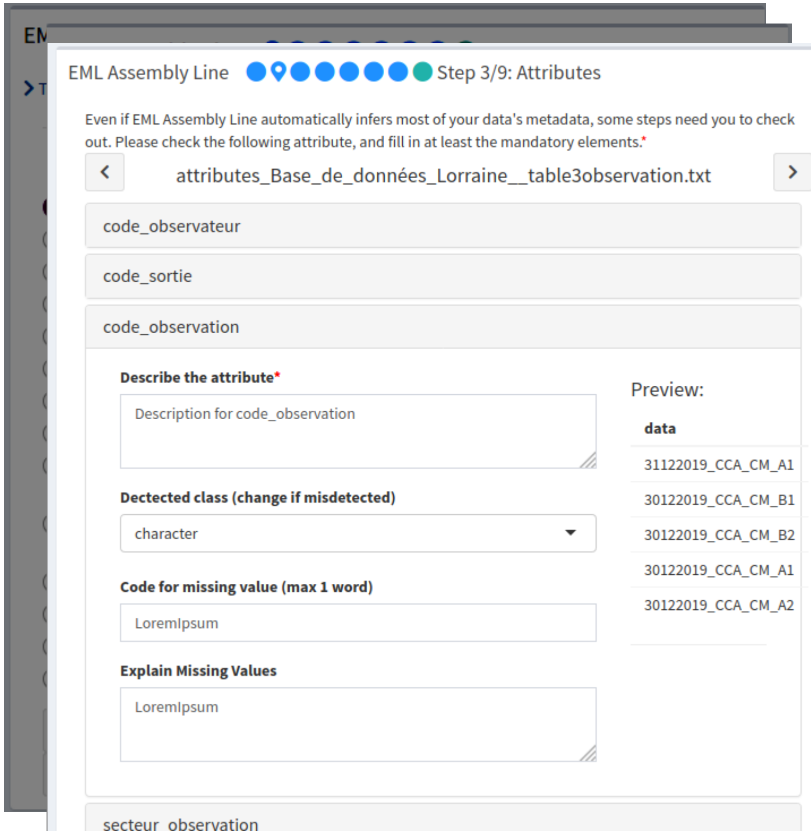
Data remote location: https://github.com/earnaı

Data Table Description: Content of

Saisie

Etape 3&4: attributs

- Typage + explication des attributs (=variables)
- Définition des modes possibles des variables catégorielles
- “Custom units”



EML Assembly Line ●●●●●●●● Step 3/9: Attributes

Even if EML Assembly Line automatically infers most of your data's metadata, some steps need you to check out. Please check the following attribute, and fill in at least the mandatory elements.*

‹ attributes_Base_de_données_Lorraine__table3observation.txt ›

code_observateur

code_sortie

code_observation

Describe the attribute*

Description for code_observation

Dectected class (change if misdetected)

character

Code for missing value (max 1 word)

LoremIpsum

Explain Missing Values

LoremIpsum

Preview:

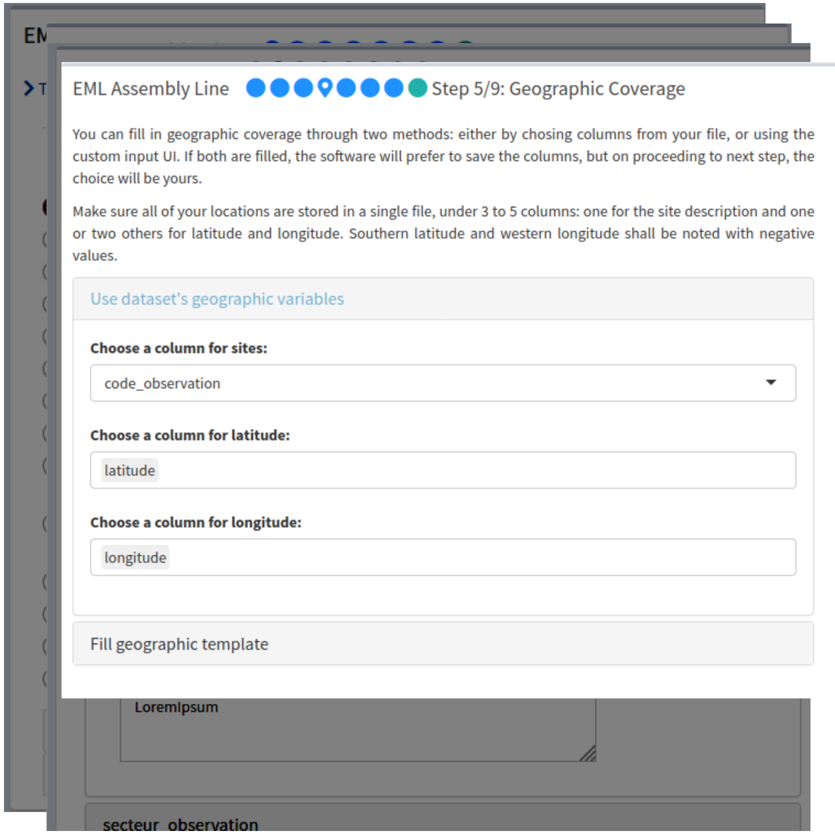
data
31122019_CCA_CM_A1
30122019_CCA_CM_B1
30122019_CCA_CM_B2
30122019_CCA_CM_A1
30122019_CCA_CM_A2

secteur_observation

Saisie

Etape 5: couverture géographique

- Définition en employant des colonnes du jeu de données
- Définition manuelle, site par site



EML Assembly Line ●●●●●●●● Step 5/9: Geographic Coverage

You can fill in geographic coverage through two methods: either by chosing columns from your file, or using the custom input UI. If both are filled, the software will prefer to save the columns, but on proceeding to next step, the choice will be yours.

Make sure all of your locations are stored in a single file, under 3 to 5 columns: one for the site description and one or two others for latitude and longitude. Southern latitude and western longitude shall be noted with negative values.

Use dataset's geographic variables

Choose a column for sites: code_observation

Choose a column for latitude: latitude

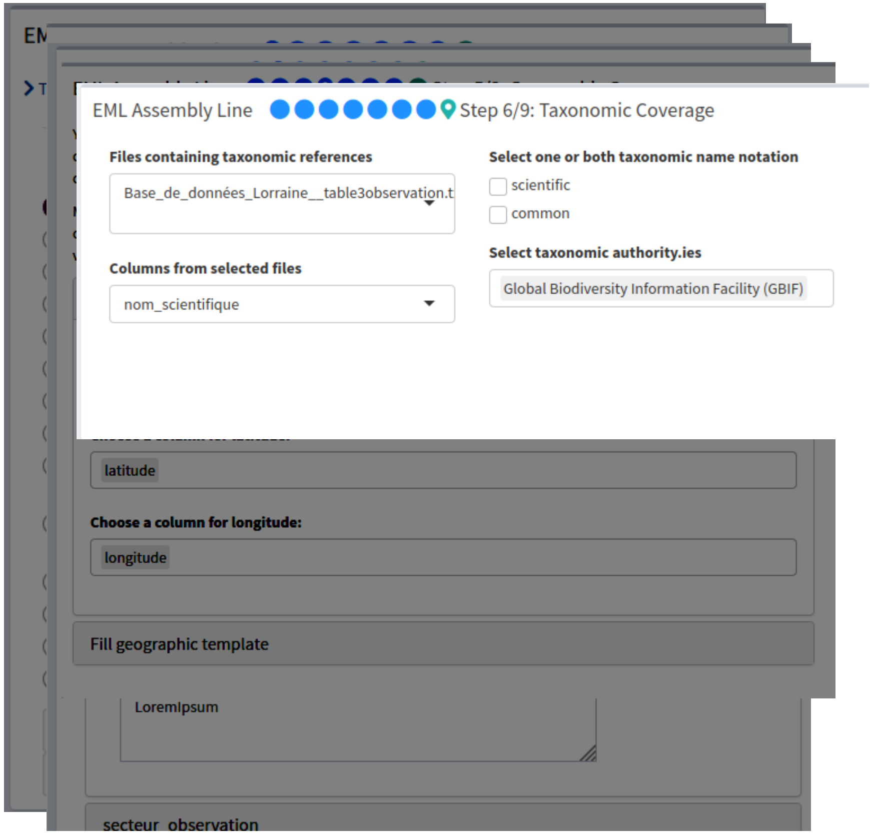
Choose a column for longitude: longitude

Fill geographic template

Saisie

Etape 6: couverture taxonomique

- Définition en employant des colonnes du jeu de données
- Connection à des ressources terminologiques (GBIF, ITIS)



EML Assembly Line ●●●●●●● Step 6/9: Taxonomic Coverage

Files containing taxonomic references

Base_de_données_Lorraine__table3observation.t

Columns from selected files

nom_scientifique

Select one or both taxonomic name notation

□ scientific

□ common

Select taxonomic authority.ies

Global Biodiversity Information Facility (GBIF)

latitude

Choose a column for longitude:

longitude

Fill geographic template

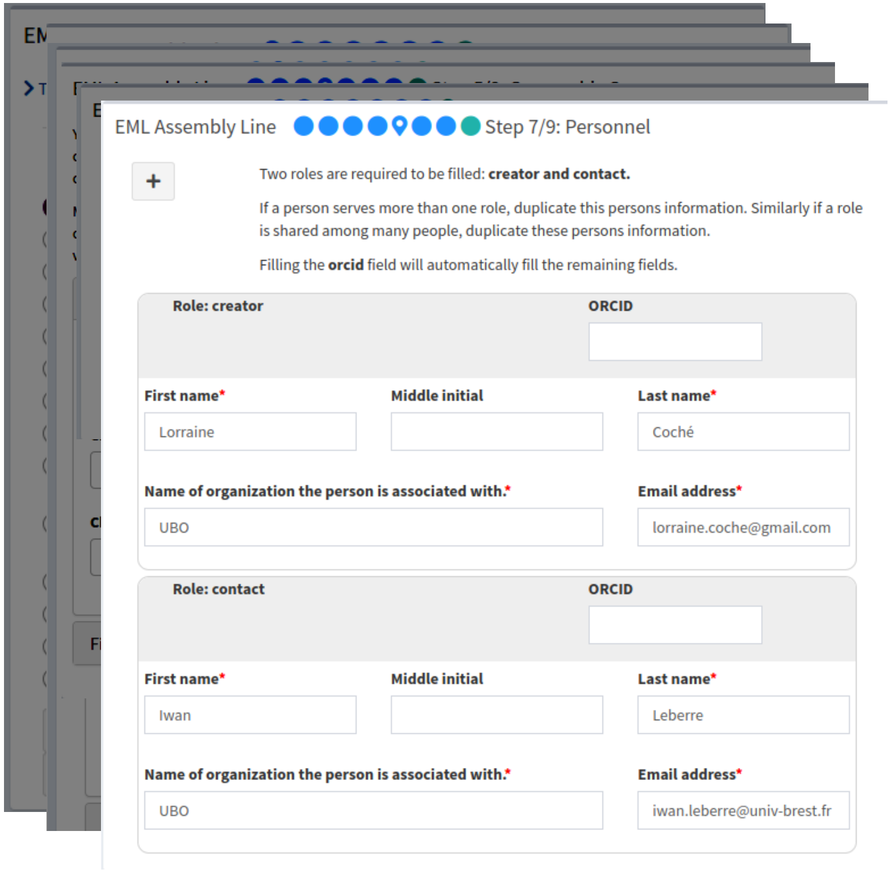
LoremIpsum

secteur_observation

Saisie

Etape 7: personnel & rôles

- Remplissage automatique par moissonnage ORCID
- Précision du rôle par personne



EML Assembly Line ●●●●⚲●●● Step 7/9: Personnel

+

Two roles are required to be filled: **creator and contact.**

If a person serves more than one role, duplicate this persons information. Similarly if a role is shared among many people, duplicate these persons information.

Filling the **orcid** field will automatically fill the remaining fields.

Role: creator **ORCID**

First name* Lorraine
Middle initial
Last name* Coché
Name of organization the person is associated with.* UBO
Email address* lorraine.coche@gmail.com

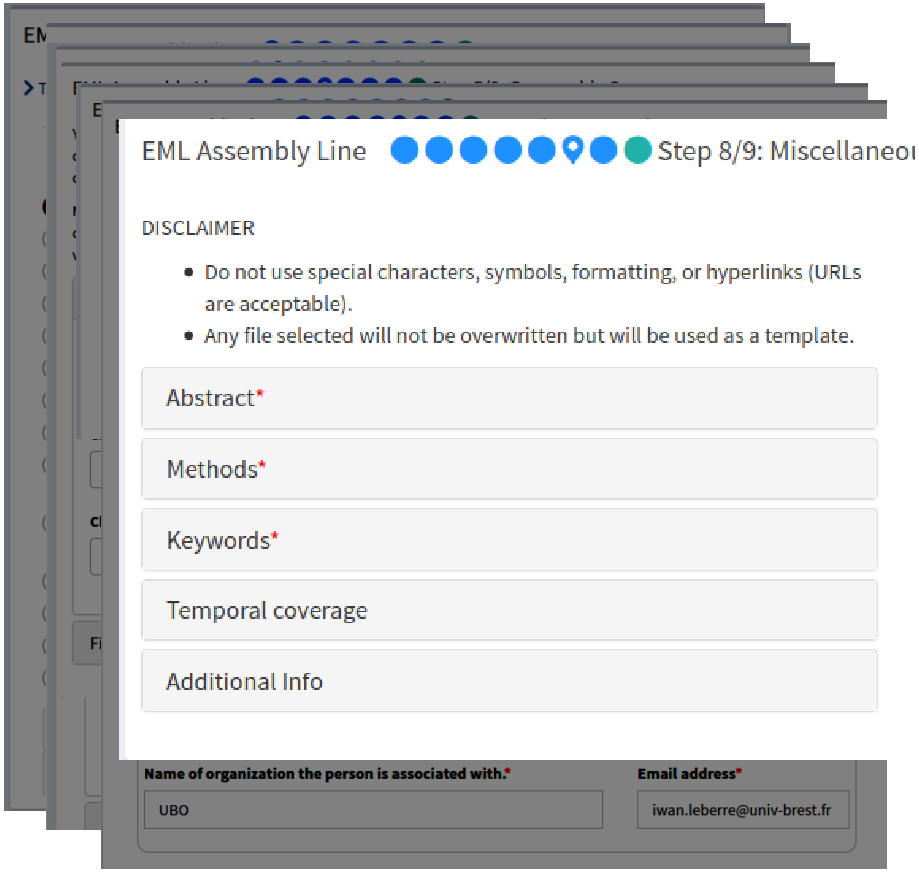
Role: contact **ORCID**

First name* Iwan
Middle initial
Last name* Leberre
Name of organization the person is associated with.* UBO
Email address* iwan.leberre@univ-brest.fr

Saisie

Etape 8: “détails”

- Abstract
- Méthodes
- Durée
- Mots-clés
- Informations additionnelles



EML Assembly Line ● ● ● ● ● ● ● ● Step 8/9: Miscellaneou

DISCLAIMER

- Do not use special characters, symbols, formatting, or hyperlinks (URLs are acceptable).
- Any file selected will not be overwritten but will be used as a template.

Abstract*

Methods*

Keywords*

Temporal coverage

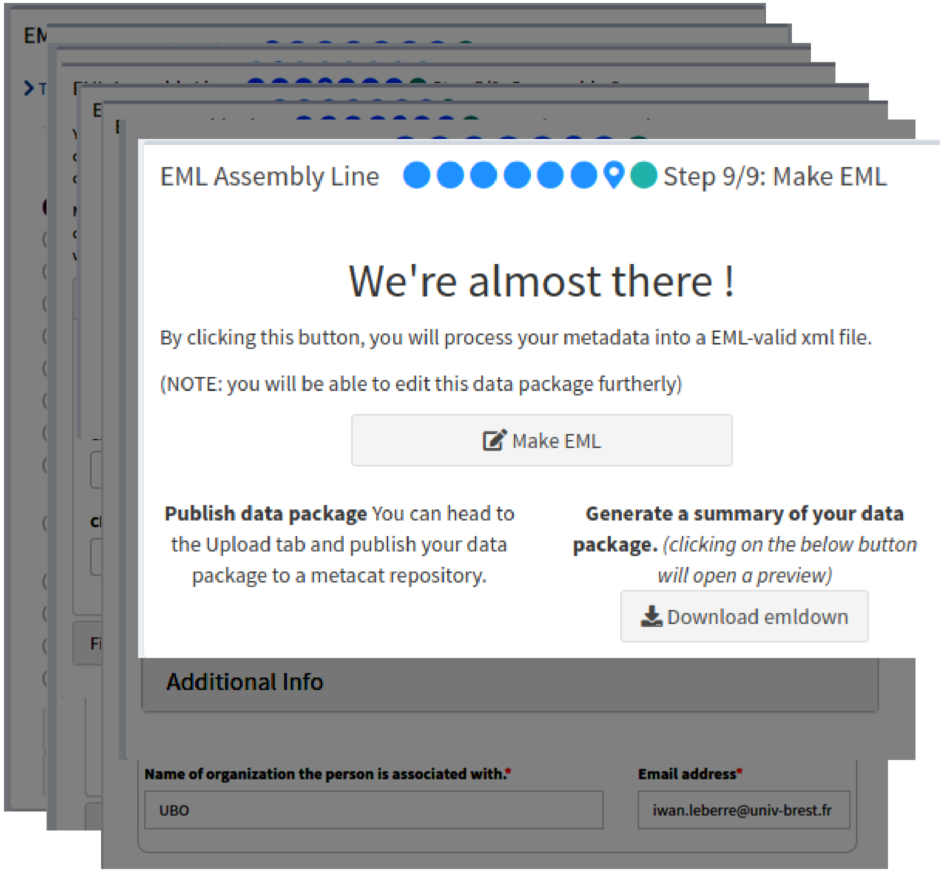
Additional Info

Name of organization the person is associated with.*	Email address*
UBO	iwan.leberre@univ-brest.fr

Saisie

Etape 9: écriture

- Génération de .xml “EML-valide”
- Possibilité de proto-data paper (*emldown*)



EML Assembly Line ●●●●●●○● Step 9/9: Make EML

We're almost there !

By clicking this button, you will process your metadata into a EML-valid xml file.

(NOTE: you will be able to edit this data package furtherly)

Publish data package You can head to the Upload tab and publish your data package to a metacat repository.

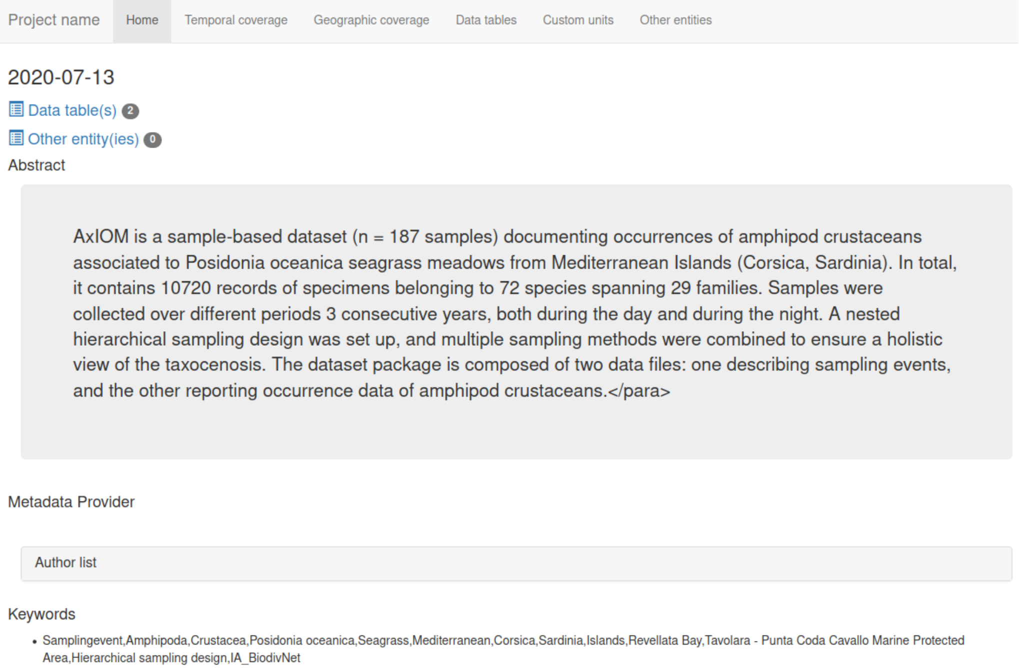
Generate a summary of your data package. *(clicking on the below button will open a preview)*

Additional Info

Name of organization the person is associated with.*	Email address*
UBO	iwan.leberre@univ-brest.fr

Résultats

- Data package téléchargeable
- *emldown* succinct



2020-07-13

Data table(s) 2

Other entity(ies) 0

Abstract

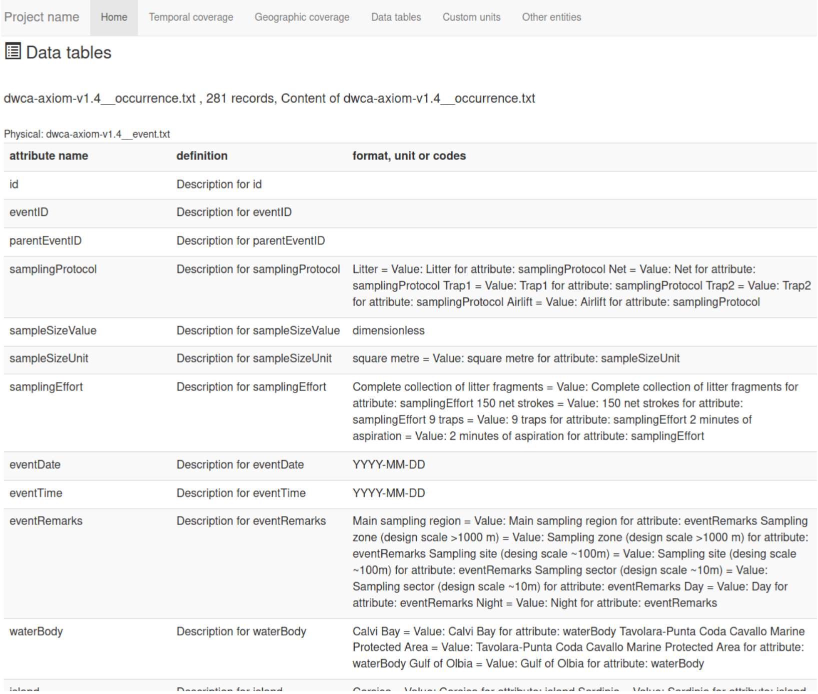
AxIOM is a sample-based dataset (n = 187 samples) documenting occurrences of amphipod crustaceans associated to Posidonia oceanica seagrass meadows from Mediterranean Islands (Corsica, Sardinia). In total, it contains 10720 records of specimens belonging to 72 species spanning 29 families. Samples were collected over different periods 3 consecutive years, both during the day and during the night. A nested hierarchical sampling design was set up, and multiple sampling methods were combined to ensure a holistic view of the taxocenosis. The dataset package is composed of two data files: one describing sampling events, and the other reporting occurrence data of amphipod crustaceans.</para>

Metadata Provider

Author list

Keywords

- Samplingevent,Amphipoda,Crustacea,Posidonia oceanica,Seagrass,Mediterranean,Corsica,Sardinia,Islands,Revellata Bay,Tavolara - Punta Coda Cavallo Marine Protected Area,Hierarchical sampling design,IA_BiodivNet



Data tables

dwca-axiom-v1.4__occurrence.txt , 281 records, Content of dwca-axiom-v1.4__occurrence.txt

Physical: dwca-axiom-v1.4__event.txt

attribute name	definition	format, unit or codes
id	Description for id	
eventID	Description for eventID	
parentEventID	Description for parentEventID	
samplingProtocol	Description for samplingProtocol	Litter = Value: Litter for attribute: samplingProtocol Net = Value: Net for attribute: samplingProtocol Trap1 = Value: Trap1 for attribute: samplingProtocol Trap2 = Value: Trap2 for attribute: samplingProtocol Airlift = Value: Airlift for attribute: samplingProtocol
sampleSizeValue	Description for sampleSizeValue	dimensionless
sampleSizeUnit	Description for sampleSizeUnit	square metre = Value: square metre for attribute: sampleSizeUnit
samplingEffort	Description for samplingEffort	Complete collection of litter fragments = Value: Complete collection of litter fragments for attribute: samplingEffort 150 net strokes = Value: 150 net strokes for attribute: samplingEffort 9 traps = Value: 9 traps for attribute: samplingEffort 2 minutes of aspiration = Value: 2 minutes of aspiration for attribute: samplingEffort
eventDate	Description for eventDate	YYYY-MM-DD
eventTime	Description for eventTime	YYYY-MM-DD
eventRemarks	Description for eventRemarks	Main sampling region = Value: Main sampling region for attribute: eventRemarks Sampling zone (design scale >1000 m) = Value: Sampling zone (design scale >1000 m) for attribute: eventRemarks Sampling site (desing scale ~100m) = Value: Sampling site (desing scale ~100m) for attribute: eventRemarks Sampling sector (design scale ~10m) = Value: Sampling sector (design scale ~10m) for attribute: eventRemarks Day = Value: Day for attribute: eventRemarks Night = Value: Night for attribute: eventRemarks
waterBody	Description for waterBody	Calvi Bay = Value: Calvi Bay for attribute: waterBody Tavolara-Punta Coda Cavallo Marine Protected Area = Value: Tavolara-Punta Coda Cavallo Marine Protected Area for attribute: waterBody Gulf of Olbia = Value: Gulf of Olbia for attribute: waterBody
island	Description for island	Corsica = Value: Corsica for attribute: island Sardinia = Value: Sardinia for attribute: island

Résultats

- Data package téléchargeable
- *emldown* succinct
- Mise en ligne possible vers MetaCAT

Exemple *Study of specific diversity of phytoplankton populations*, F. Rigault-Jalabert (lien)

WIPs

- EML Annotation: connection aux ontologies
- Mise en ligne: réparation en cours
- Galaxyfication
- Conversion vers d’autres standards (DwC, SINP, ISO 19k+, …)
- Sessions personnalisées + profilage des utilisateurs par communauté

Galaxy-E : une instance de Galaxy dédiée à l’analyse de données en Ecologie

Coline Royaux

Galaxy

Galaxy PROJECT

Plateforme web pour le partage et le traitement des données de recherche

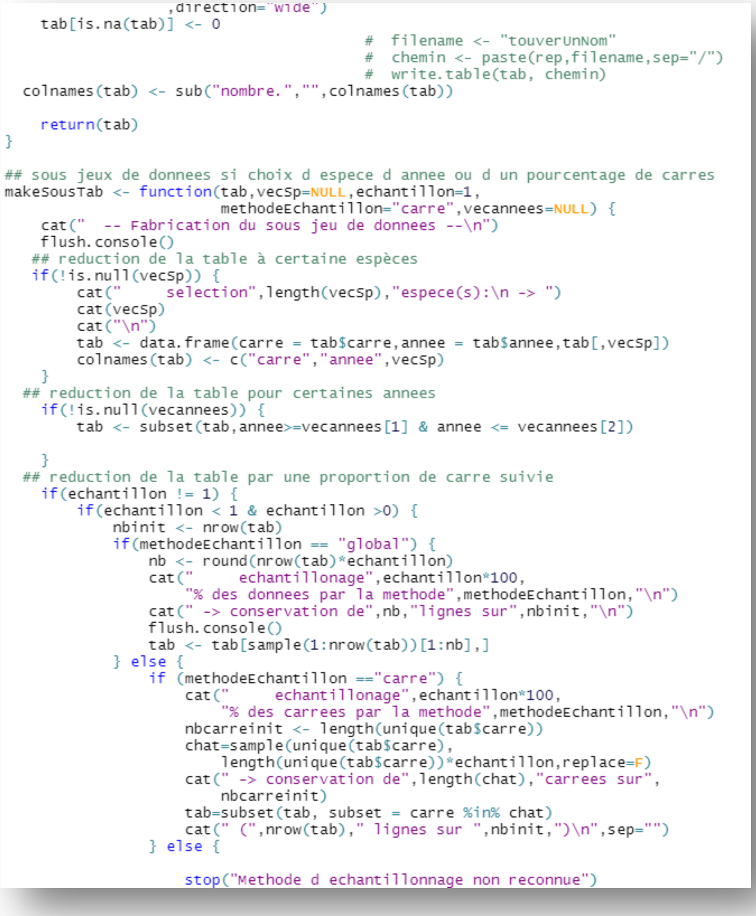
Permet un accès facilité à l'analyse cloud et au Calcul Haute Performance par l'interfaçage de n'importe quel langage informatique

Quatre principes clés

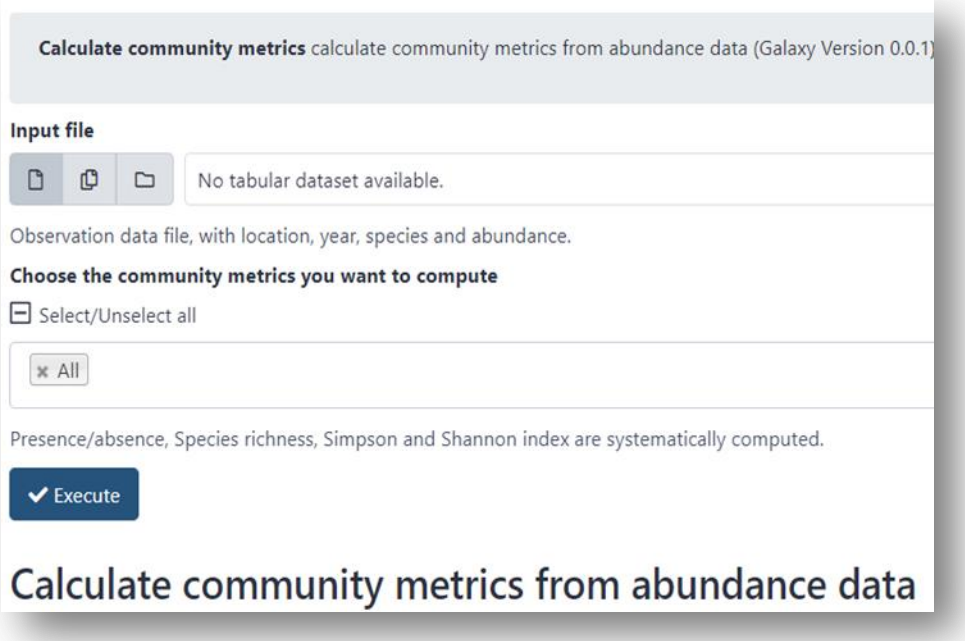
- ✓ Accessibilité
- ✓ Reproductibilité
- ✓ Transparence
- ✓ Peer review

Galaxy

Galaxy PROJECT



```
                  ,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 echantillonnage non reconnue")
```



Calculate community metrics calculate community metrics from abundance data (Galaxy Version 0.0.1)

Input file

No tabular dataset available.

Observation data file, with location, year, species and abundance.

Choose the community metrics you want to compute

⊟ Select/Unselect all

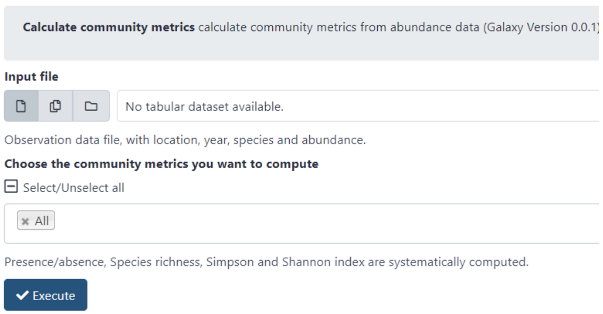
× All

Presence/absence, Species richness, Simpson and Shannon index are systematically computed.

✓ Execute

Calculate community metrics from abundance data

Le workflow Galaxy PAMPA



Calculate community metrics calculate community metrics from abundance data (Galaxy Version 0.0.1)

Input file

No tabular dataset available.

Observation data file, with location, year, species and abundance.

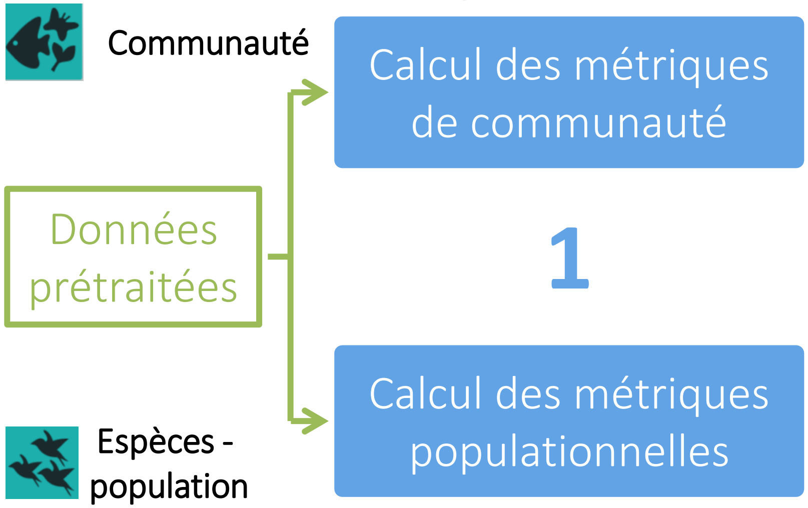
Choose the community metrics you want to compute

□ Select/Unselect all

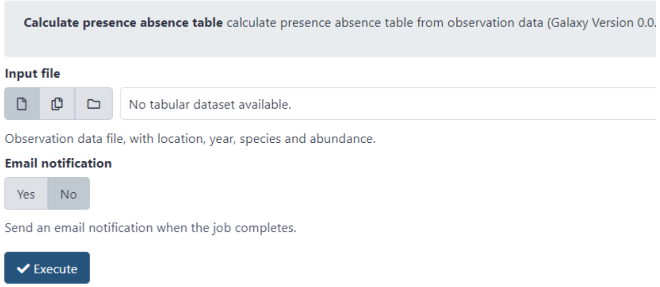
× All

Presence/absence, Species richness, Simpson and Shannon index are systematically computed.

Calculate community metrics from abundance data



Communauté
Calcul des métriques de communauté
Données prétraitées
1
Calcul des métriques populationnelles
Espèces - population



Calculate presence absence table calculate presence absence table from observation data (Galaxy Version 0.0.

Input file

No tabular dataset available.

Observation data file, with location, year, species and abundance.

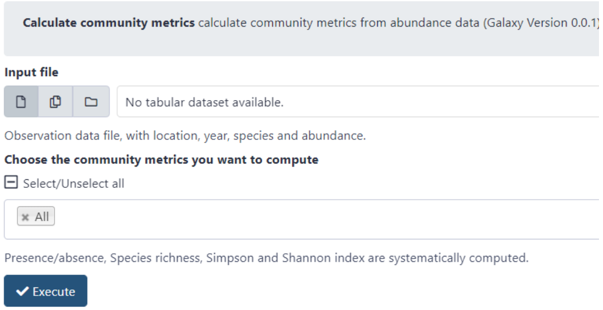
Email notification

Yes No

Send an email notification when the job completes.

Calculate presence absence table from abundance data

Le workflow Galaxy PAMPA



Calculate community metrics calculate community metrics from abundance data (Galaxy Version 0.0.1)

Input file

No tabular dataset available.

Observation data file, with location, year, species and abundance.

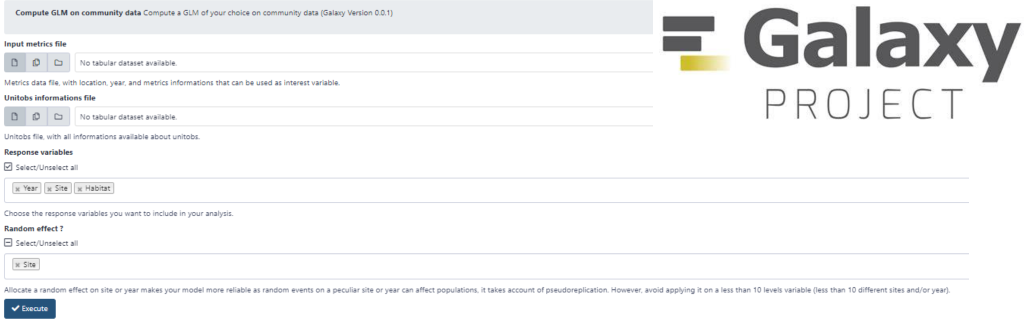
Choose the community metrics you want to compute

⊟ Select/Unselect all

× All

Presence/absence, Species richness, Simpson and Shannon index are systematically computed.

Calculate community metrics from abundance data



Compute GLM on community data Compute a GLM of your choice on community data (Galaxy Version 0.0.1)

Input metrics file

No tabular dataset available.

Metrics data file, with location, year, and metrics informations that can be used as interest variable.

Unitobs informations file

No tabular dataset available.

Unitobs file, with all informations available about unitobs.

Response variables

☑ Select/Unselect all

× Year × Site × Habitat

Choose the response variables you want to include in your analysis.

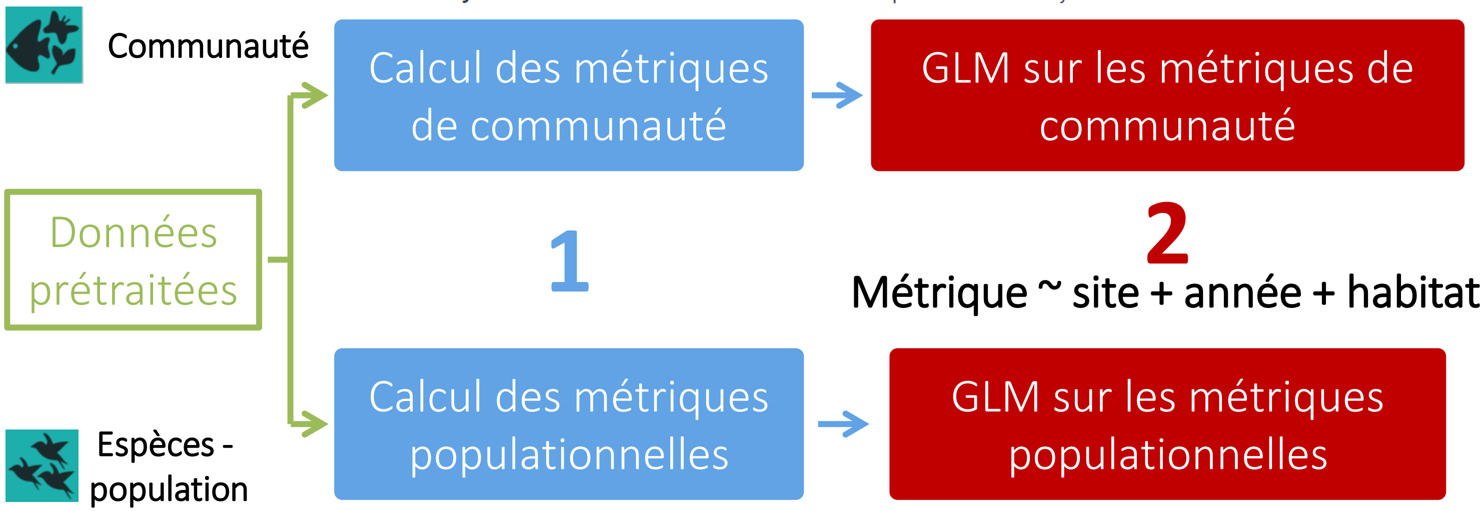
Random effect ?

⊟ Select/Unselect all

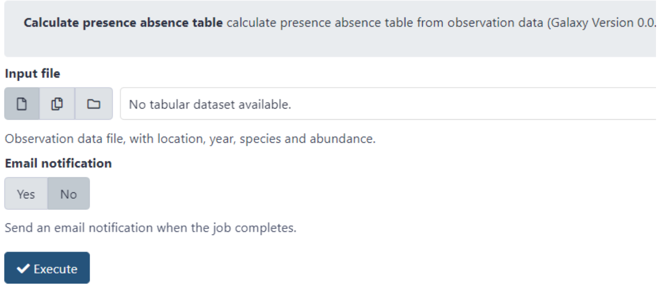
× Site

Allocate a random effect on site or year makes your model more reliable as random events on a peculiar site or year can affect populations, it takes account of pseudoreplication. However, avoid applying it on a less than 10 levels variable (less than 10 different sites and/or year).

Compute GLM on community data with selected interest variables



Communauté
Données prétraitées
Calcul des métriques de communauté
GLM sur les métriques de communauté
1
2
Métrique ~ site + année + habitat
Espèces - population
Calcul des métriques populationnelles
GLM sur les métriques populationnelles



Calculate presence absence table calculate presence absence table from observation data (Galaxy Version 0.0.

Input file

No tabular dataset available.

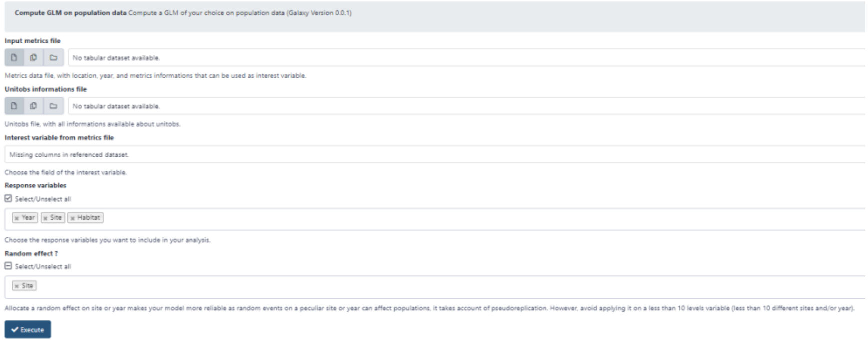
Observation data file, with location, year, species and abundance.

Email notification

Yes No

Send an email notification when the job completes.

Calculate presence absence table from abundance data



Compute GLM on population data Compute a GLM of your choice on population data (Galaxy Version 0.0.1)

Input metrics file

No tabular dataset available.

Metrics data file, with location, year, and metrics informations that can be used as interest variable.

Unitobs informations file

No tabular dataset available.

Unitobs file, with all informations available about unitobs.

Interest variable from metrics file

Missing columns in referenced dataset.

Choose the field of the interest variable.

Response variables

☑ Select/Unselect all

× Year × Site × Habitat

Choose the response variables you want to include in your analysis.

Random effect ?

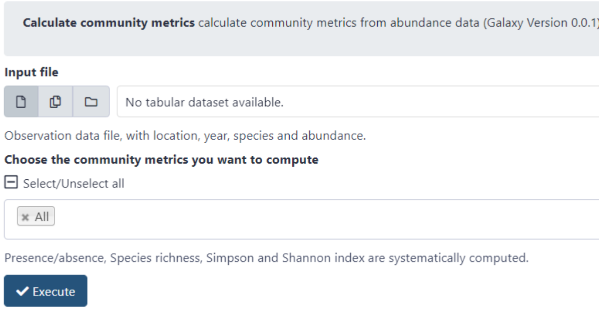
⊟ Select/Unselect all

× Site

Allocate a random effect on site or year makes your model more reliable as random events on a peculiar site or year can affect populations, it takes account of pseudoreplication. However, avoid applying it on a less than 10 levels variable (less than 10 different sites and/or year).

Compute GLM on population data with selected interest variables

Le workflow Galaxy PAMPA



Calculate community metrics calculate community metrics from abundance data (Galaxy Version 0.0.1)

Input file

No tabular dataset available.

Observation data file, with location, year, species and abundance.

Choose the community metrics you want to compute

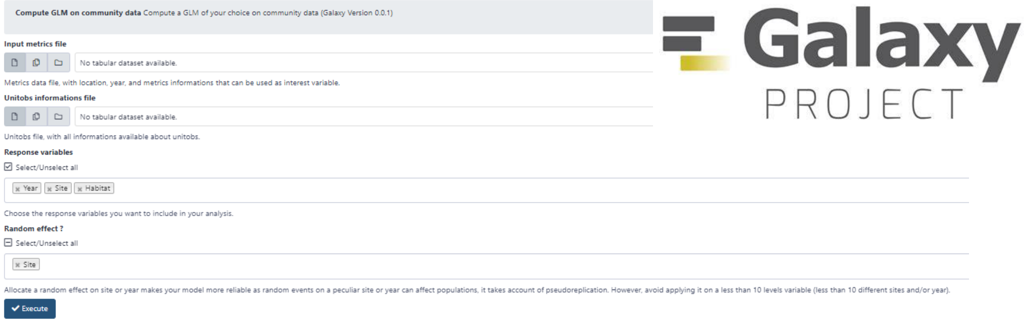
□ Select/Unselect all

× All

Presence/absence, Species richness, Simpson and Shannon index are systematically computed.

✓ Execute

Calculate community metrics from abundance data



Compute GLM on community data Compute a GLM of your choice on community data (Galaxy Version 0.0.1)

Input metrics file

No tabular dataset available.

Metrics data file, with location, year, and metrics informations that can be used as interest variable.

Unitobs informations file

No tabular dataset available.

Unitobs file, with all informations available about unitobs.

Response variables

☑ Select/Unselect all

× Year × Site × Habitat

Choose the response variables you want to include in your analysis.

Random effect ?

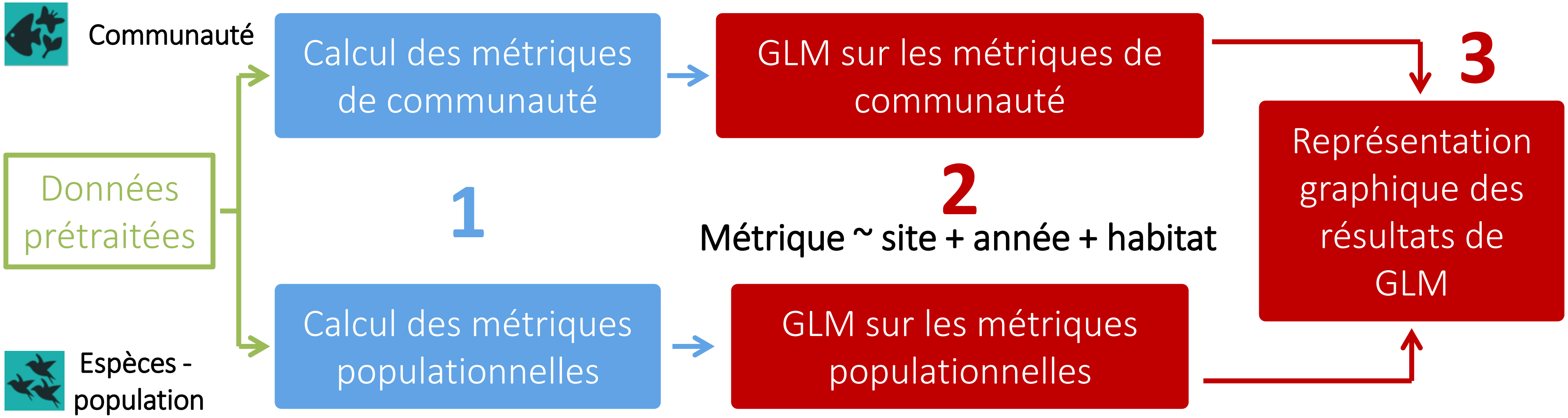
□ Select/Unselect all

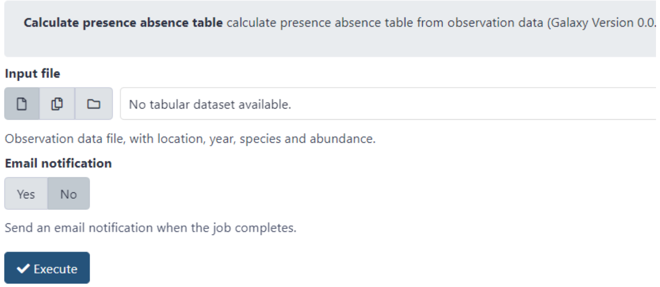
× Site

Allocate a random effect on site or year makes your model more reliable as random events on a peculiar site or year can affect populations, it takes account of pseudoreplication. However, avoid applying it on a less than 10 levels variable (less than 10 different sites and/or year).

✓ Execute

Compute GLM on community data with selected interest variables


Communauté
Calcul des métriques de communauté
GLM sur les métriques de communauté
3
Représentation graphique des résultats de GLM
Données prétraitées
1
2
Métrique ~ site + année + habitat
Calcul des métriques populationnelles
GLM sur les métriques populationnelles
Espèces - population



Calculate presence absence table calculate presence absence table from observation data (Galaxy Version 0.0.

Input file

No tabular dataset available.

Observation data file, with location, year, species and abundance.

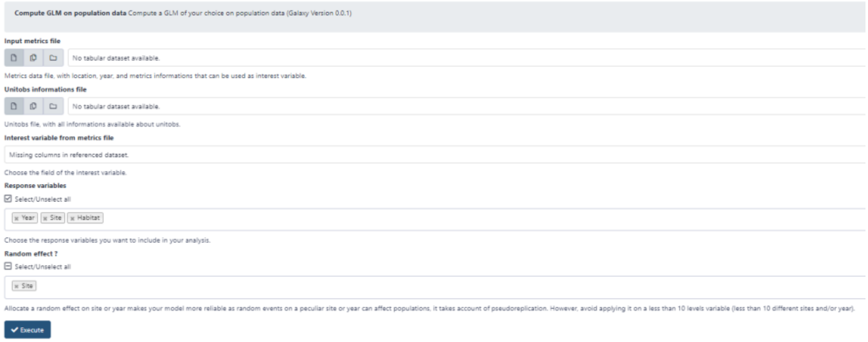
Email notification

Yes No

Send an email notification when the job completes.

✓ Execute

Calculate presence absence table from abundance data



Compute GLM on population data Compute a GLM of your choice on population data (Galaxy Version 0.0.1)

Input metrics file

No tabular dataset available.

Metrics data file, with location, year, and metrics informations that can be used as interest variable.

Unitobs informations file

No tabular dataset available.

Unitobs file, with all informations available about unitobs.

Interest variable from metrics file

Missing columns in referenced dataset.

Choose the field of the interest variable.

Response variables

☑ Select/Unselect all

× Year × Site × Habitat

Choose the response variables you want to include in your analysis.

Random effect ?

□ Select/Unselect all

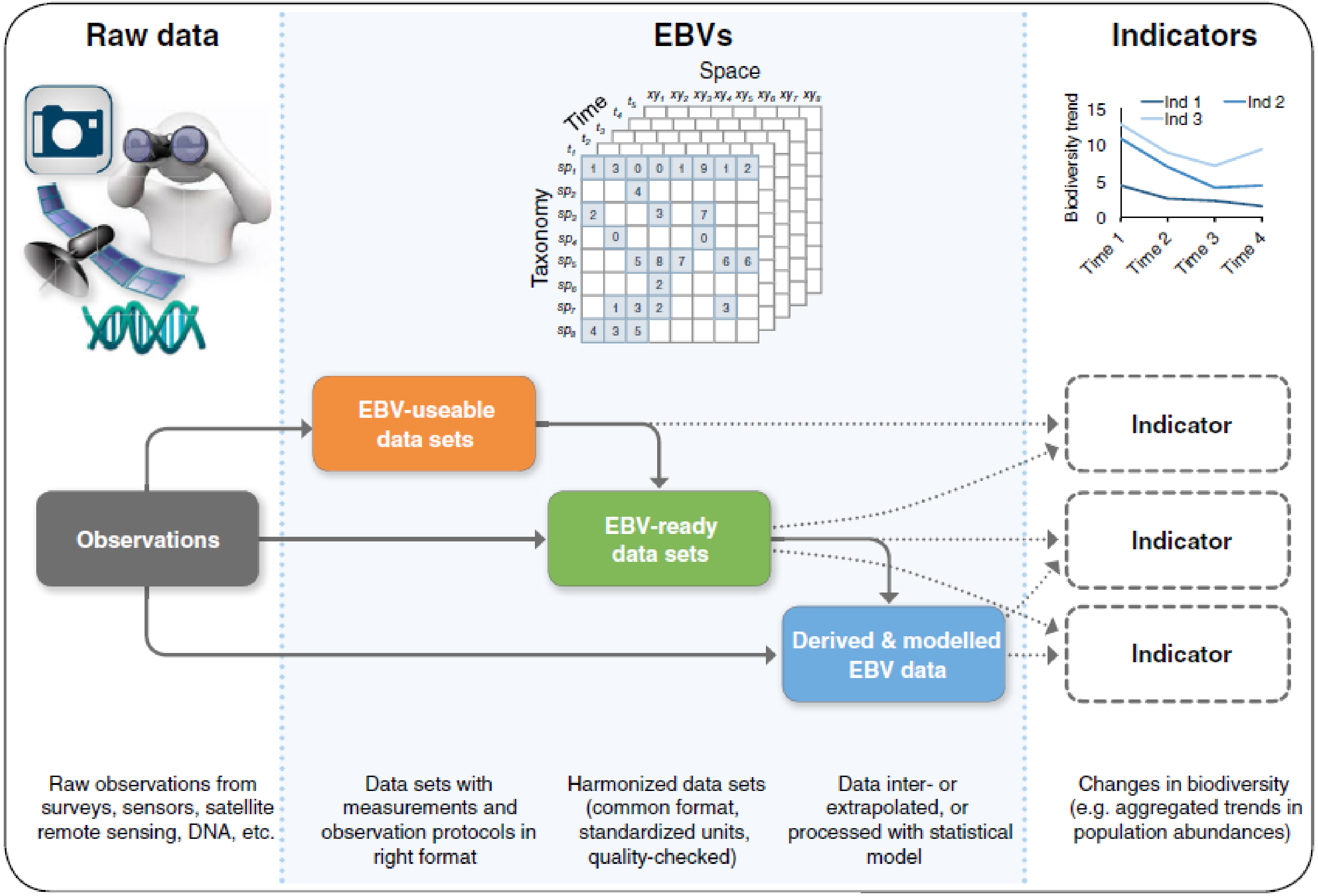
× Site

Allocate a random effect on site or year makes your model more reliable as random events on a peculiar site or year can affect populations, it takes account of pseudoreplication. However, avoid applying it on a less than 10 levels variable (less than 10 different sites and/or year).

✓ Execute

Compute GLM on population data with selected interest variables

Les Variables Essentielles de Biodiversité (EBV)



	xy_1	xy_2	xy_3	xy_4	xy_5	xy_6	xy_7	xy_8
sp_1	1	3	0	0	1	9	1	2
sp_2			4					
sp_3	2			3		7		
sp_4		0				0		
sp_5			5	8	7		6	6
sp_6				2				
sp_7		1	3	2			3	
sp_8	4	3	5					

Observations

EBV-useable data sets

EBV-ready data sets

Derived & modelled EBV data

Indicator

Indicator

Indicator

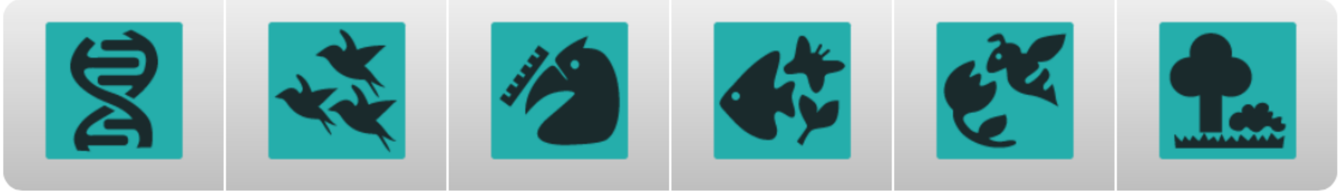
Raw observations from surveys, sensors, satellite remote sensing, DNA, etc.

Data sets with measurements and observation protocols in right format

Harmonized data sets (common format, standardized units, quality-checked)

Data inter- or extrapolated, or processed with statistical model

Changes in biodiversity (e.g. aggregated trends in population abundances)



Genetic composition

Species population

Species traits

Community composition

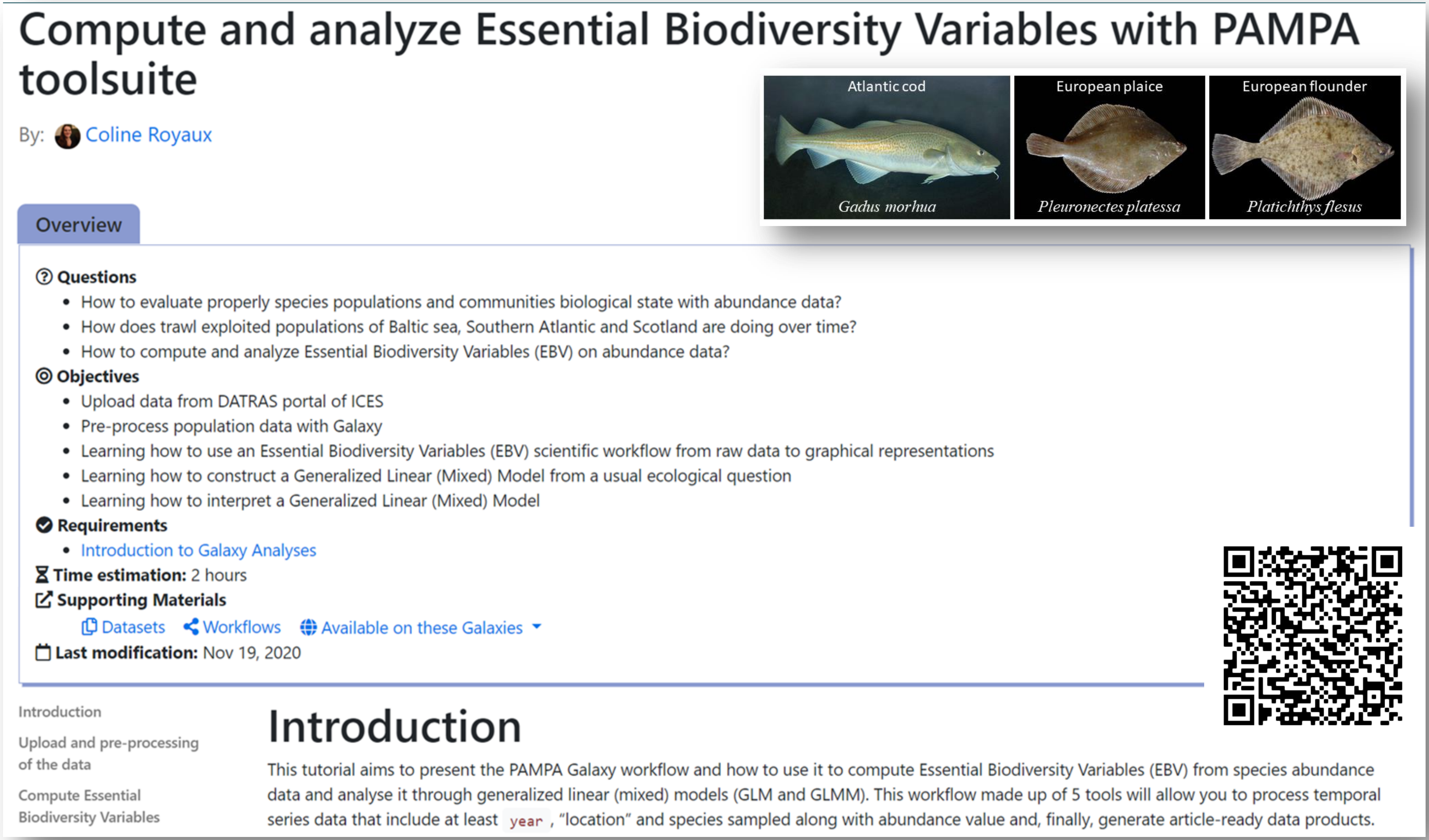
Ecosystem function

Ecosystem structure

Kissling *et al.* 2017

Tutoriel sur données de pêche (CPUE)

https://bit.ly/2VjDTaQ



Compute and analyze Essential Biodiversity Variables with PAMPA toolsuite

By: Coline Royaux

Overview

Questions

- How to evaluate properly species populations and communities biological state with abundance data?
- How does trawl exploited populations of Baltic sea, Southern Atlantic and Scotland are doing over time?
- How to compute and analyze Essential Biodiversity Variables (EBV) on abundance data?

Objectives

- Upload data from DATRAS portal of ICES
- Pre-process population data with Galaxy
- Learning how to use an Essential Biodiversity Variables (EBV) scientific workflow from raw data to graphical representations
- Learning how to construct a Generalized Linear (Mixed) Model from a usual ecological question
- Learning how to interpret a Generalized Linear (Mixed) Model

Requirements

- Introduction to Galaxy Analyses

Time estimation: 2 hours

Supporting Materials

Datasets Workflows Available on these Galaxies

Last modification: Nov 19, 2020

Introduction
Upload and pre-processing of the data
Compute Essential Biodiversity Variables

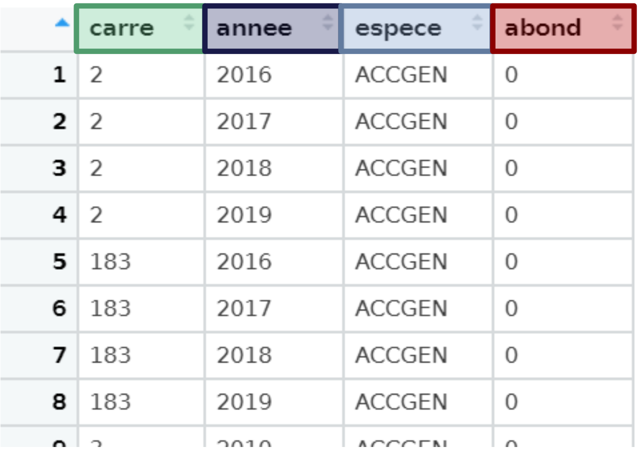
Introduction

This tutorial aims to present the PAMPA Galaxy workflow and how to use it to compute Essential Biodiversity Variables (EBV) from species abundance data and analyse it through generalized linear (mixed) models (GLM and GLMM). This workflow made up of 5 tools will allow you to process temporal series data that include at least `year`, "location" and species sampled along with abundance value and, finally, generate article-ready data products.

Outils Galaxy-E **généralisable**

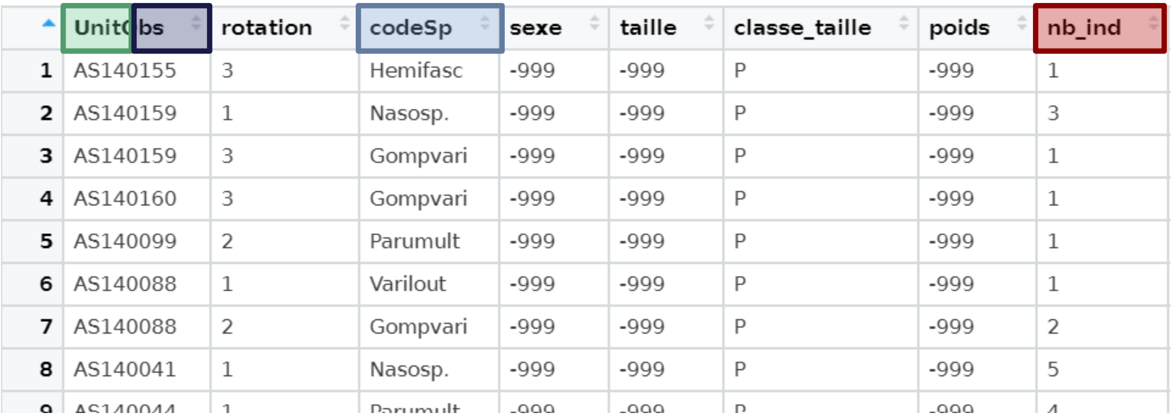
Variables primaires smilaires

Données Suivi Temporel des Oiseaux Communs



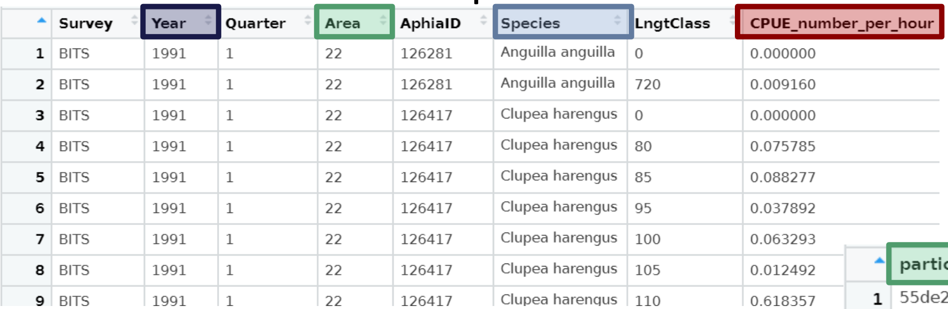
	carre	annee	espece	abond
1	2	2016	ACCGEN	0
2	2	2017	ACCGEN	0
3	2	2018	ACCGEN	0
4	2	2019	ACCGEN	0
5	183	2016	ACCGEN	0
6	183	2017	ACCGEN	0
7	183	2018	ACCGEN	0
8	183	2019	ACCGEN	0
9	[illegible]	[illegible]	ACCGEN	0

Données de caméras sous-marines



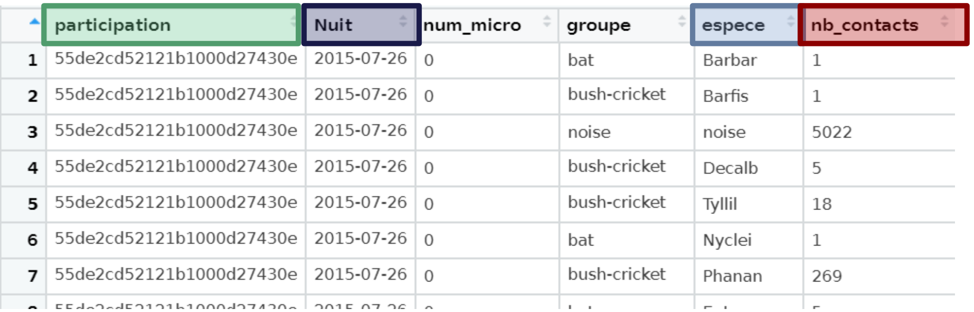
	Unit(bs	rotation	codeSp	sexe	taille	classe_taille	poids	nb_ind
1	AS140155	3	Hemifasc	-999	-999	P	-999	1
2	AS140159	1	Nasosp.	-999	-999	P	-999	3
3	AS140159	3	Gompvari	-999	-999	P	-999	1
4	AS140160	3	Gompvari	-999	-999	P	-999	1
5	AS140099	2	Parumult	-999	-999	P	-999	1
6	AS140088	1	Varilout	-999	-999	P	-999	1
7	AS140088	2	Gompvari	-999	-999	P	-999	2
8	AS140041	1	Nasosp.	-999	-999	P	-999	5
9	AS140044	1	Parumult	-999	-999	P	-999	4

Données de pêche



	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

Données Vigie - Chiro



	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
8	55de2cd52121b1000d27430e	2015-07-26	0	[illegible]	[illegible]	[illegible]

🟩 Site

🟪 Année

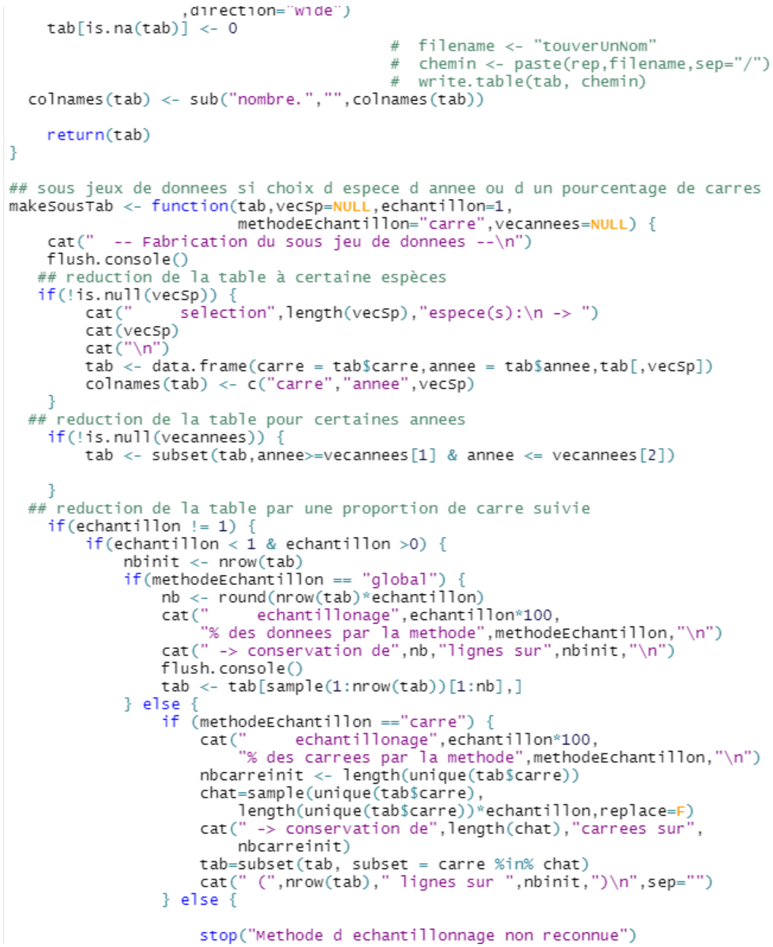
🟦 Espèce

🟥 Occurence

Atomisation

Actuellement, en écologie...

Un script R pour chaque fichier de données

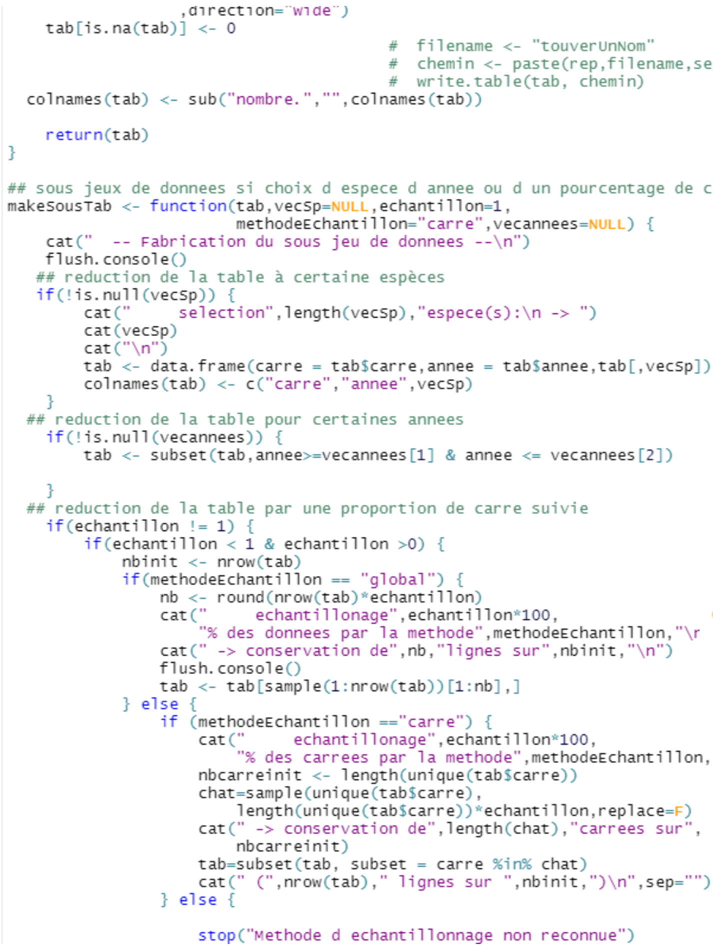


```r
                  ,direction="wide")
    tab[is.na(tab)] <- 0
                                        #  filename <- "touverUnNom"
                                        #  chemin <- paste(rep,filename,sep="/")
                                        #  write.table(tab, chemin)
  colnames(tab)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

Atomisation

Actuellement, en écologie...

Un script R pour chaque fichier de données

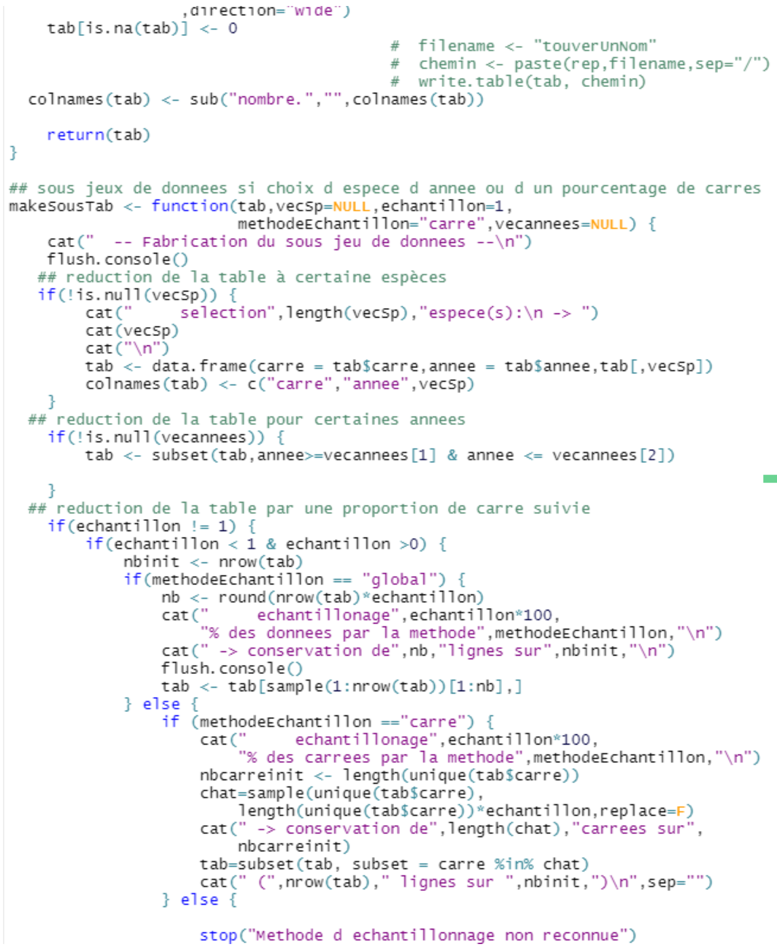


```r
                  ,direction="wide")
    tab[is.na(tab)] <- 0
                                        #  filename <- "touverUnNom"
                                        #  chemin <- paste(rep,filename,sep="/")
                                        #  write.table(tab, chemin)
  colnames(tab) <- 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,"\r
                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 echantillonnage non reconnue")
```

Atomisation

Actuellement, en écologie...

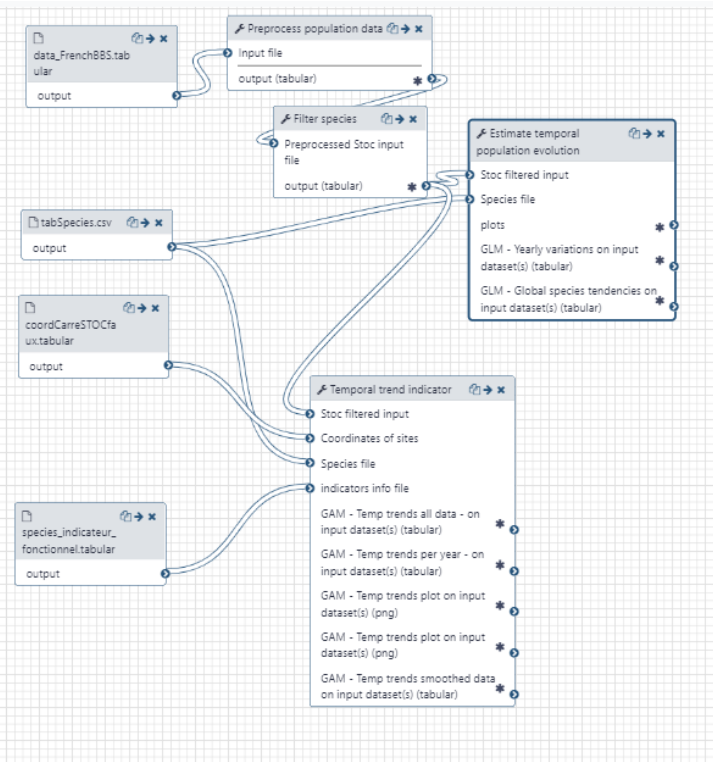
Un script R pour chaque fichier de données



```
                   ,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 echantillonnage non reconnue")
```

Avec Galaxy...

Plusieurs scripts R atomisés pour analyser différents fichiers de données

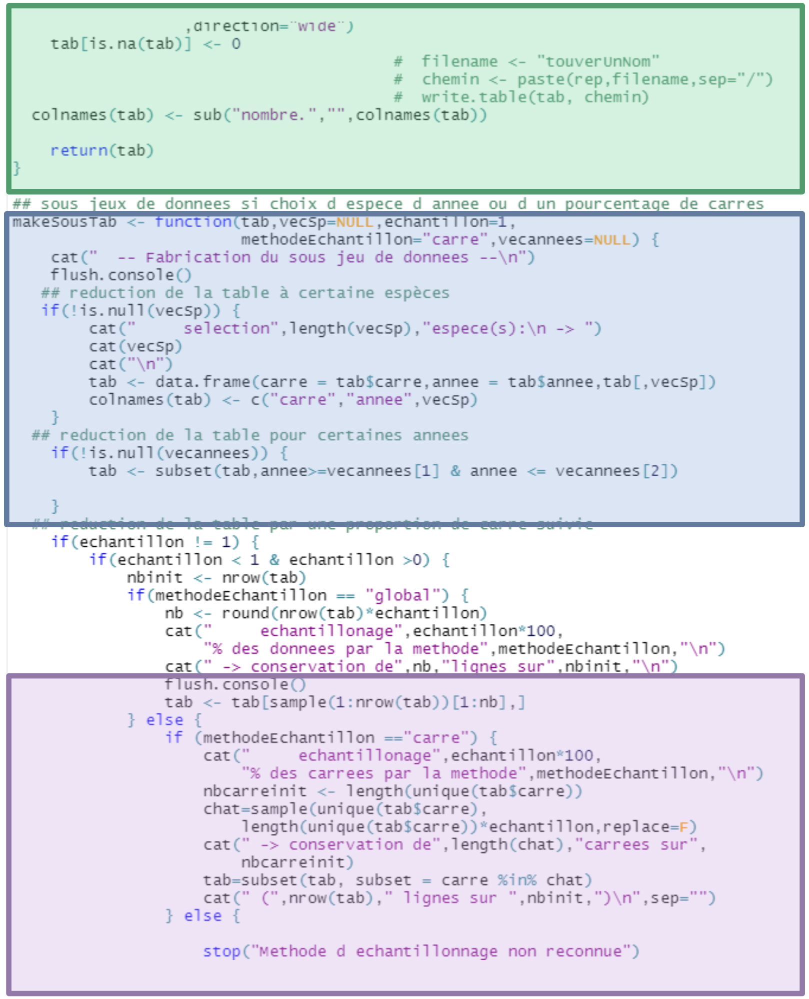


data_FrenchBBS.tabular
output
Preprocess population data
Input file
output (tabular)
Filter species
Preprocessed Stoc input file
output (tabular)
Estimate temporal population evolution
Stoc filtered input
Species file
plots
GLM - Yearly variations on input dataset(s) (tabular)
GLM - Global species tendencies on input dataset(s) (tabular)
tabSpecies.csv
output
coordCarreSTOCfaux.tabular
output
Temporal trend indicator
Stoc filtered input
Coordinates of sites
Species file
indicators info file
GAM - Temp trends all data - on input dataset(s) (tabular)
GAM - Temp trends per year - on input dataset(s) (tabular)
GAM - Temp trends plot on input dataset(s) (png)
GAM - Temp trends plot on input dataset(s) (png)
GAM - Temp trends smoothed data on input dataset(s) (tabular)
species_indicateur_fonctionnel.tabular
output

Atomisation

Actuellement, en écologie...

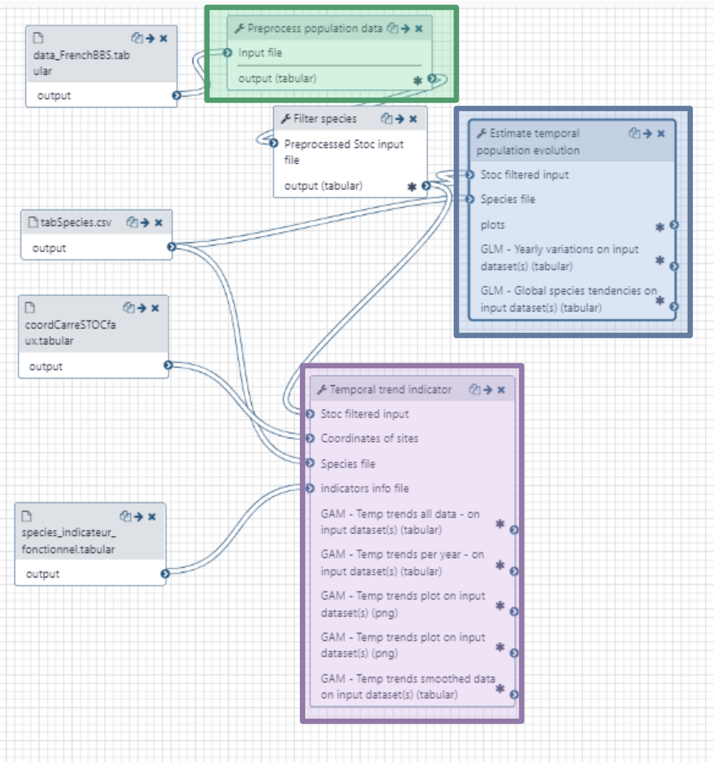
Un script R pour chaque fichier de données



```r
                    ,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 echantillonnage non reconnue")
```

Avec Galaxy...

Plusieurs scripts R atomisés pour analyser différents fichiers de données


data_FrenchBBS.tabular
output
Preprocess population data
Input file
output (tabular)
Filter species
Preprocessed Stoc input file
output (tabular)
Estimate temporal population evolution
Stoc filtered input
Species file
plots
GLM - Yearly variations on input dataset(s) (tabular)
GLM - Global species tendencies on input dataset(s) (tabular)
tabSpecies.csv
output
coordCarreSTOCfaux.tabular
output
Temporal trend indicator
Stoc filtered input
Coordinates of sites
Species file
indicators info file
GAM - Temp trends all data - on input dataset(s) (tabular)
GAM - Temp trends per year - on input dataset(s) (tabular)
GAM - Temp trends plot on input dataset(s) (png)
GAM - Temp trends plot on input dataset(s) (png)
GAM - Temp trends smoothed data on input dataset(s) (tabular)
species_indicateur_fonctionnel.tabular
output