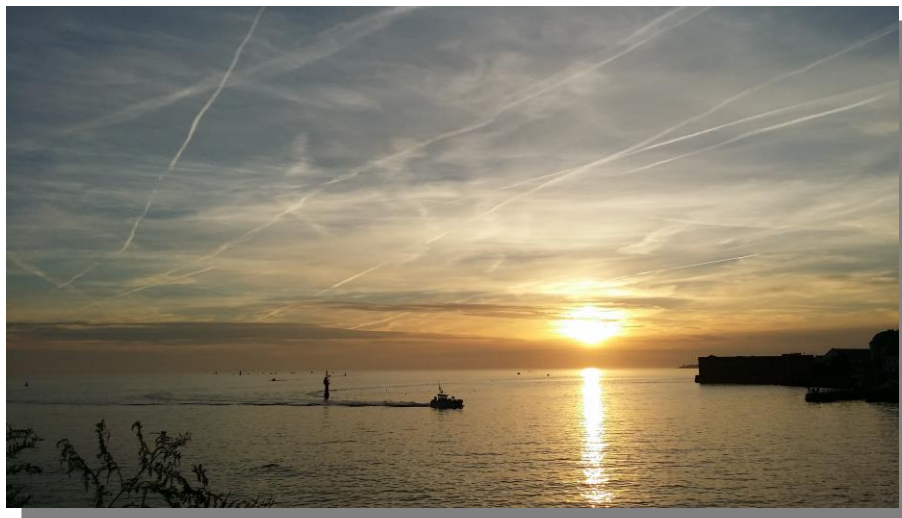

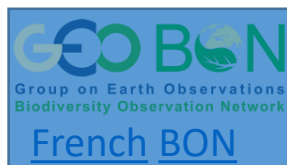


# Pôle national de données de Biodiversité

de la gestion de données à son analyse, présentation des outils et services PNDB



 #PNDB @Yvan2935



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)  
Aurélié DELAVAUD (FRB)  
Dominique JOLY, DAS CNRS (INEE)  
Laurent PONCET, Dir. UMS PatriNat, en charge du Centre de données (MNHN)  
Jean-Denis VIGNE, DGD-Recherche, expertise, valorisation, enseignement MNHN



Pôles de données



TGIR/IR nationales

Miroir européen

**PNDB**

**DATA TERRA**

IN-SYLVA Fr	TREEFORCE
<b>RECOLNAT</b>	DISSCO
ANAEE-Fr ECOTRONS	ANAEE
ANAEE-Fr NATURA	
eLTER-Fr Zones Ateliers	eLTER
eLTER-Fr OZCAR	
ILICO	JERICO
EURO-ARGO France	EURO-ARGO
EMSO France	EMSO
RESIF/EPOS	EPOS
ACTRIS France	ACTRIS
IAGOS France	IAGOS
ICOS France	ICOS
SAFIRE	EUFAR
CLIMERI	IS-ENES

Infrastructures du/portées par le Muséum dans le cadre de l'Infrastructure de recherche Terre-Environnement du Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation

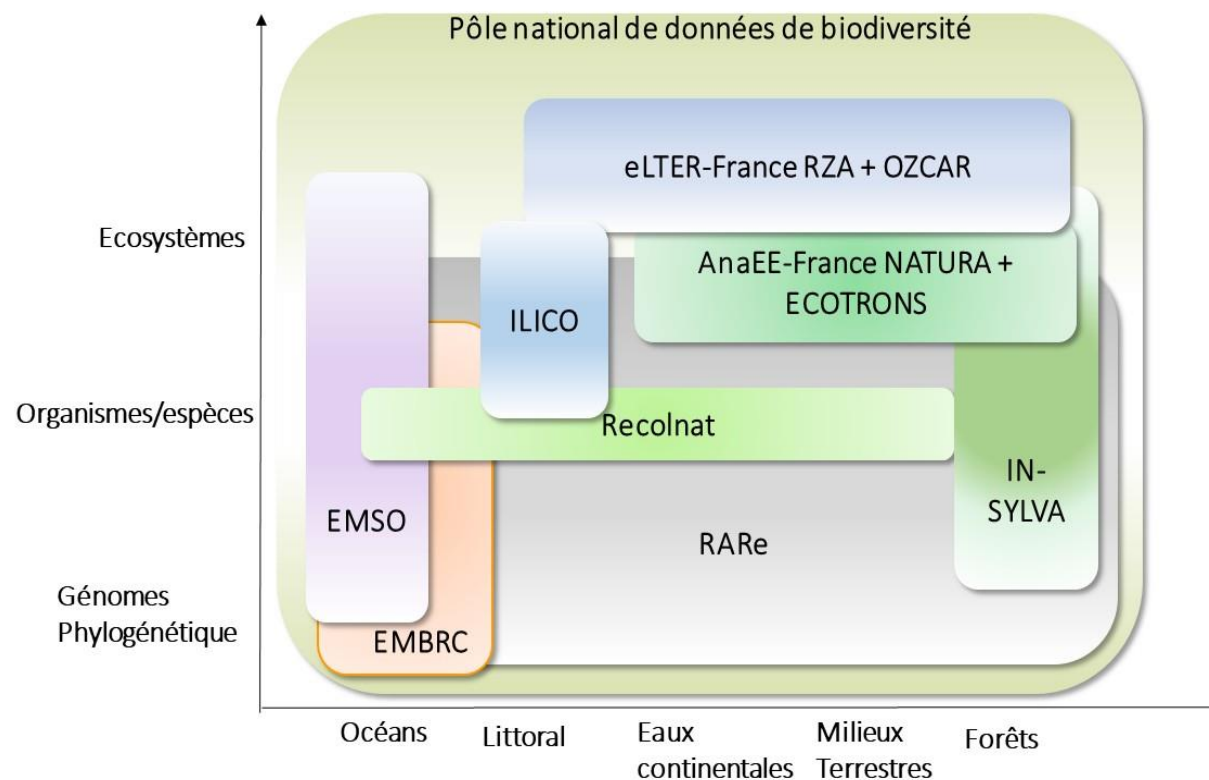


Figure 4: Paysage des infrastructures de recherche françaises dans le domaine de la biodiversité et des écosystèmes. L'encadré orange indique une infrastructure commune avec le domaine « biologie-santé ».

# Le PNDB, e-Infrastructure nationale de recherche

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

- **Consortium maître d'ouvrage :**

(11 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

## Objectifs du PNDB

1. Contribuer aux recherches concernant l'état et la dynamique de la biodiversité, en tenant compte :
  - De tous niveaux d'organisation du vivant, du génome aux communautés,
  - Des facteurs abiotiques (IR Syst. Terre-Env. : surf. continentale, océans, atmosphère, terre interne),
  - Des pratiques et comportements des sociétés humaines (socio-écosystèmes),
  - Des échelles de temps emboîtées (collections d'histoire naturelle, bioarchéologie, paléontologie).
2. Répondre aux besoins des communautés scientifiques au jour le jour:
  - Accéder aux métadonnées et, autant que possible, aux données, adaptées à la question posée (y compris données « expertise »),
  - Faciliter la mise à disposition, la validation et l'analyse des métadonnées/données en dépassant leurs hétérogénéité, dispersion et provenance,
  - Accroître la caractérisation et la fiabilité des données en améliorant la documentation et les formats des métadonnées (équilibre entre « *big data* » et « *quality data* »),
  - Favoriser la réutilisation et le couplage des données dans une approche intégrée.
3. Faire progresser la consolidation de toutes les connaissances disponibles sur la biodiversité, au service de la compréhension et de la décision, notamment en promouvant :
  - La bancarisation des données de recherche (information, formation, accompagnement),
  - Les approches intégrées (réutilisation, couplage),
  - L'animation des communautés scientifiques concernées, en lien avec les dispositifs internationaux.

# Stratégie du PNDB

Le PNDB ambitieuse  
d'apporter une valeur ajoutée  
par rapport à l'existant

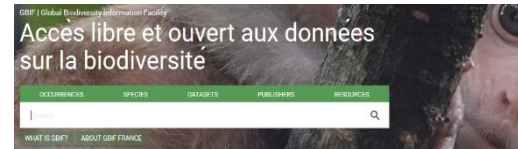
1. S'appuyer sur bases de métadonnées/données et systèmes d'information **existant** :
  - en les mettant en relation et en générant des synergies (notamment référentiels et ontologies)
  - en contribuant à renforcer leur visibilité et leur cohérence

les uns par rapport aux autres

GBIF, SINP-INPN, INDOR, SIB-SIMM, SISMER, INFOTERRE...

ReColNat, RARE, AnaEE, RZA...

iBOL2...



2. Contribuer à **étendre l'existant** avec d'autres BdD de recherche ou d'expérimentation :
  - BARcoding, « omiques », communautés, socio-écosystèmes, pressions, temps long...

3. Potentialiser les données à travers la mise à disposition d'**outils** (y compris existant) :
  - de traitement/croisement des jeux de données et d'accès aux plateformes de calcul intensif,
  - de modélisation (stimulant, en retour, les interactions et la conceptualisation)
  - d'aide à la bancarisation (e.g. ajout de métadonnées ; format et correspondance des métadonnées)

4. Générer une dynamique **d'animation** et **d'échanges** entre chercheurs/communautés :
  - en s'appuyant sur l'existant et le valorisant (e-learning GBIF, CES, CESAB...)
  - par une stratégie de communication, d'information, de rencontres/colloques
  - en sensibilisant toutes les communautés aux enjeux de l'*open data* (FAIR)

5. Favoriser la **cohérence** avec les efforts nationaux, européens et internationaux

# Cahier des charges du PNDB

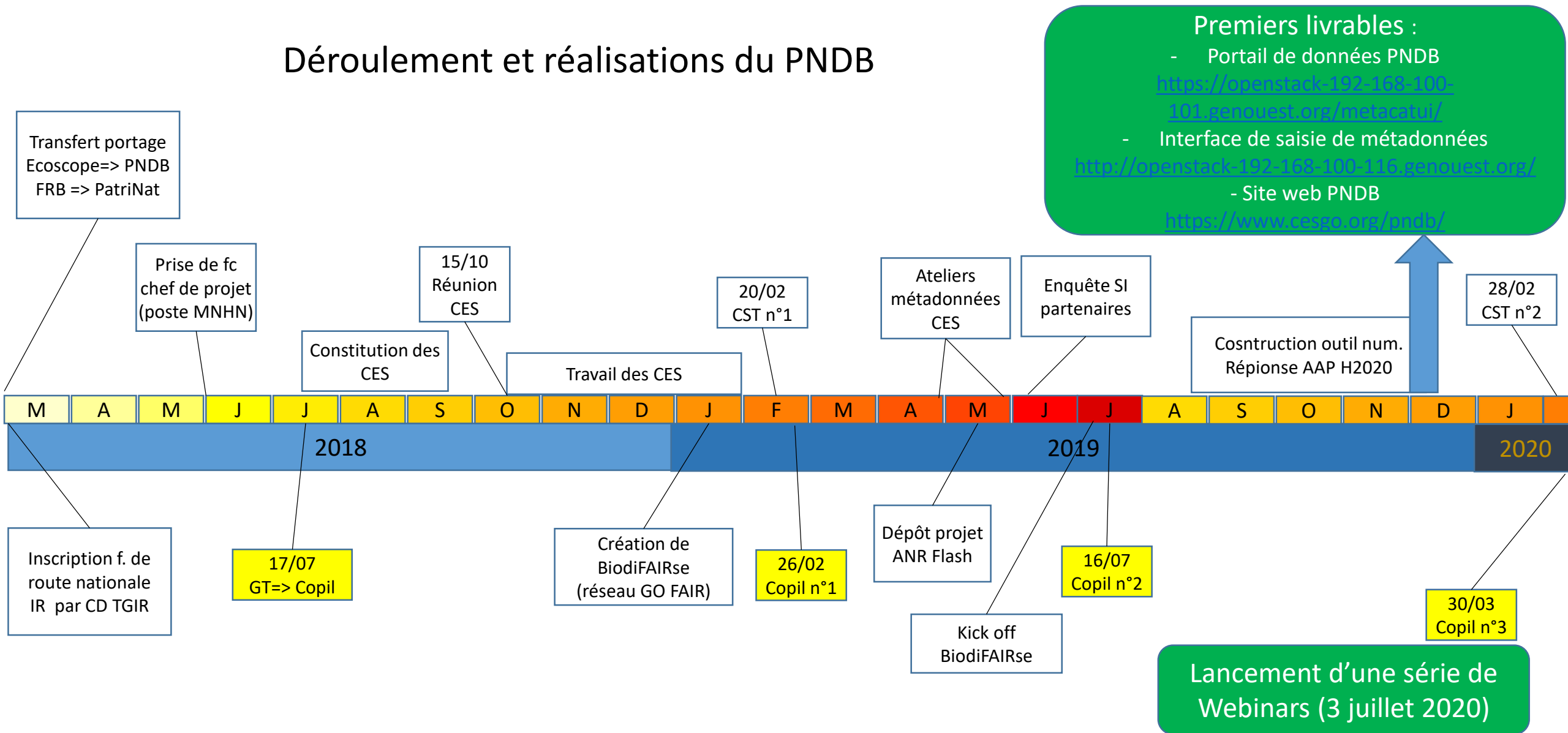
## 10 lignes de force

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)

# Déroulement et réalisations du PNDB



Visitez la version préliminaire du site web du PNDB pour suivre le projet

<https://www.cesgo.org/pndb/>

**PNDB**

- Outils
- Animation
- Projets

☰  🔍

## Pôle national de données de Biodiversité (PNDB)

*Un pôle de données, e-infrastructure, au service des scientifiques produisant, gérant et analysant des données de biodiversité*

AGENCE FRANÇAISE  
POUR LA BIODIVERSITÉ  
ÉTABLISSEMENT PUBLIC DE L'ÉTAT

Géosciences pour une Terre durable  
**brgm**



INRA  
SCIENCE & IMPACT



**INERIS**  
maîtriser le risque  
pour un développement durable

**cirad**  
LA RECHERCHE AGRONOMIQUE  
POUR LE DÉVELOPPEMENT

**FRB**  
FONDATION  
POUR LA RECHERCHE  
SUR LA BIODIVERSITÉ

**Ifremer**

**IRD**  
Institut de Recherche  
pour le Développement  
FRANCE



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, Interopérable, Réutilisable), et consistent à :



Visitez la pré-maquette du portail de données du PNDB et jouez avec les premiers jeux de données-métadonnées  
<https://openstack-192-168-100-101.genouest.org/metacatui/>



The screenshot shows the top navigation bar with links: PNDB, DONNÉES, RÉSUMÉ, A PROPOS, a search field with 'DOix or II' and 'Go', and 'CONNEXION'. The main heading is 'PÔLE NATIONAL DE DONNÉES DE BIODIVERSITÉ'. Below it is a search bar containing 'Exemple: espèces invasives' and a magnifying glass icon. A row of partner logos follows: brgm (Géosciences pour une Terre durable), cirad (LA RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT), cnrs (Centre National de la Recherche Scientifique), FRB (FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ), Ifremer, INERIS (maîtriser le risque pour un développement durable), INRAE, IRD (Institut de Recherche pour le Développement FRANCE), MUSÉUM NATIONAL HISTOIRE NATURELLE, OFB (OFFICE FRANÇAIS DE LA BIODIVERSITÉ), and a red circular logo with 'M' and 'RECHERCHE'.

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

Exemple: espèces invasives

Logos des partenaires: brgm, cirad, cnrs, FRB, Ifremer, INERIS, INRAE, IRD, MUSÉUM NATIONAL HISTOIRE NATURELLE, OFB, and a red circular logo.

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

<http://openstack-192-168-100-116.genouest.org/>

The screenshot shows the MetaShARK web application interface. At the top, there is a blue header with the 'METASHARK' logo on the left, a hamburger menu icon in the center, and a gear icon on the right. Below the header is a dark grey sidebar with a list of navigation items: 'Welcome', 'Fill in EML', 'Upload EML', 'EML Documentation', and 'About MetaShARK'. The main content area has a light blue background and contains the following sections:

## Welcome in MetaShARK

**MetaShARK** (**M**etadata **S**hiny **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 developed since 1997 from the [work of Michener et al.](#). Since then, it has been developed 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/>

The image displays several screenshots of the Galaxy / Ecology web interface. The top-left screenshot shows a world map with red data points, likely representing species occurrences. The top-right screenshot shows a 'History' panel with a workflow titled 'treat biodiversity data & tuto GBIF data handling'. The middle-left screenshot shows a workflow editor for 'Complete EBV workflow example from GBIF data', featuring steps like 'Count', 'Filter', 'OGR2ogr', 'Wallace', and 'RStudio'. The middle-right screenshot shows a 'Your workflows' table listing various workflows. The bottom-right screenshot shows a 'History' panel for an 'EBV workflow example from GBIF data' with steps like '11: data 9.GeoJSON', '10: data 9.shapefile', etc.

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

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

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

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



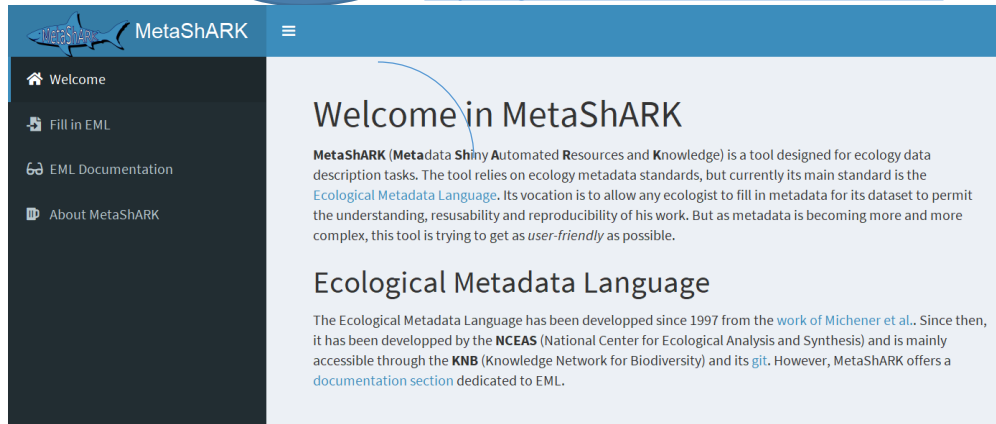
dataset1

dataset2

script1

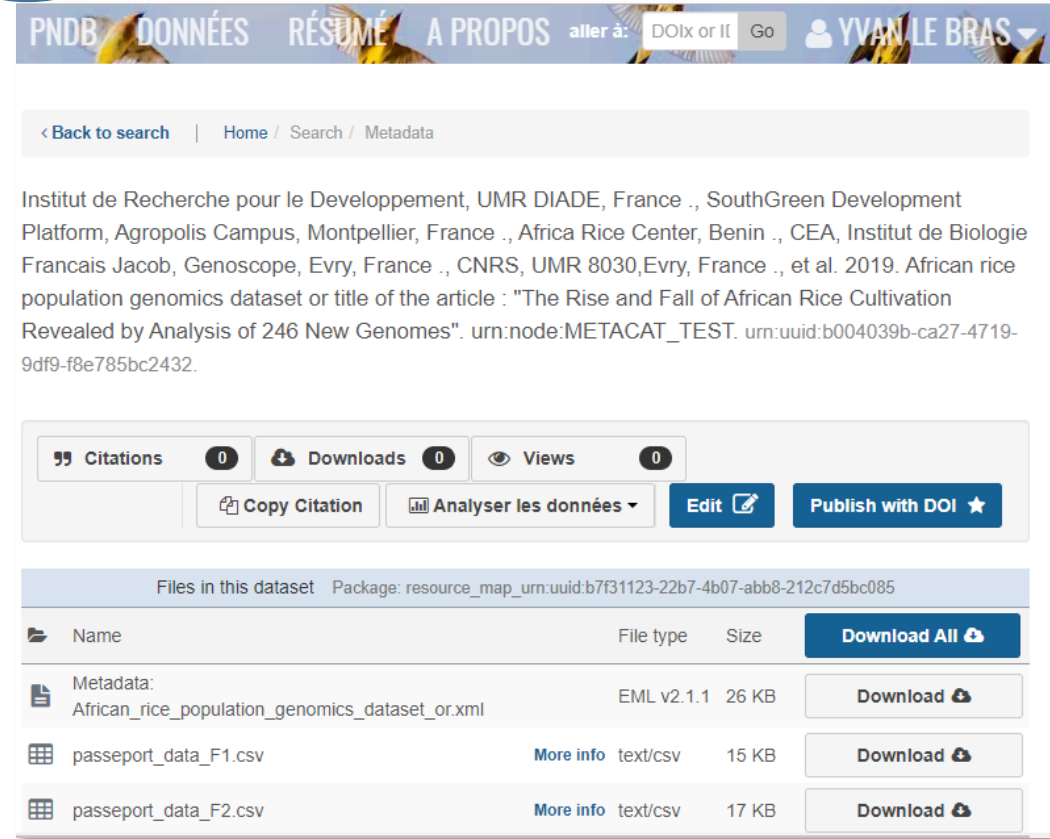
article1

protocol1



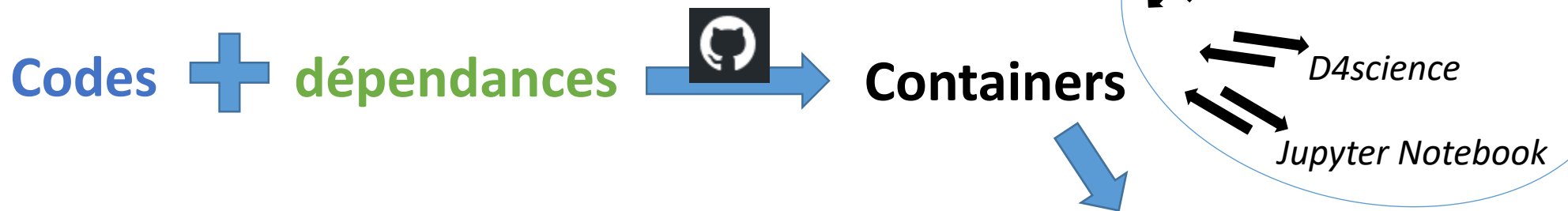
 **EML Assembly Line**  
<https://ediorg.github.io/EMLassembleline/index.html>

**EML**

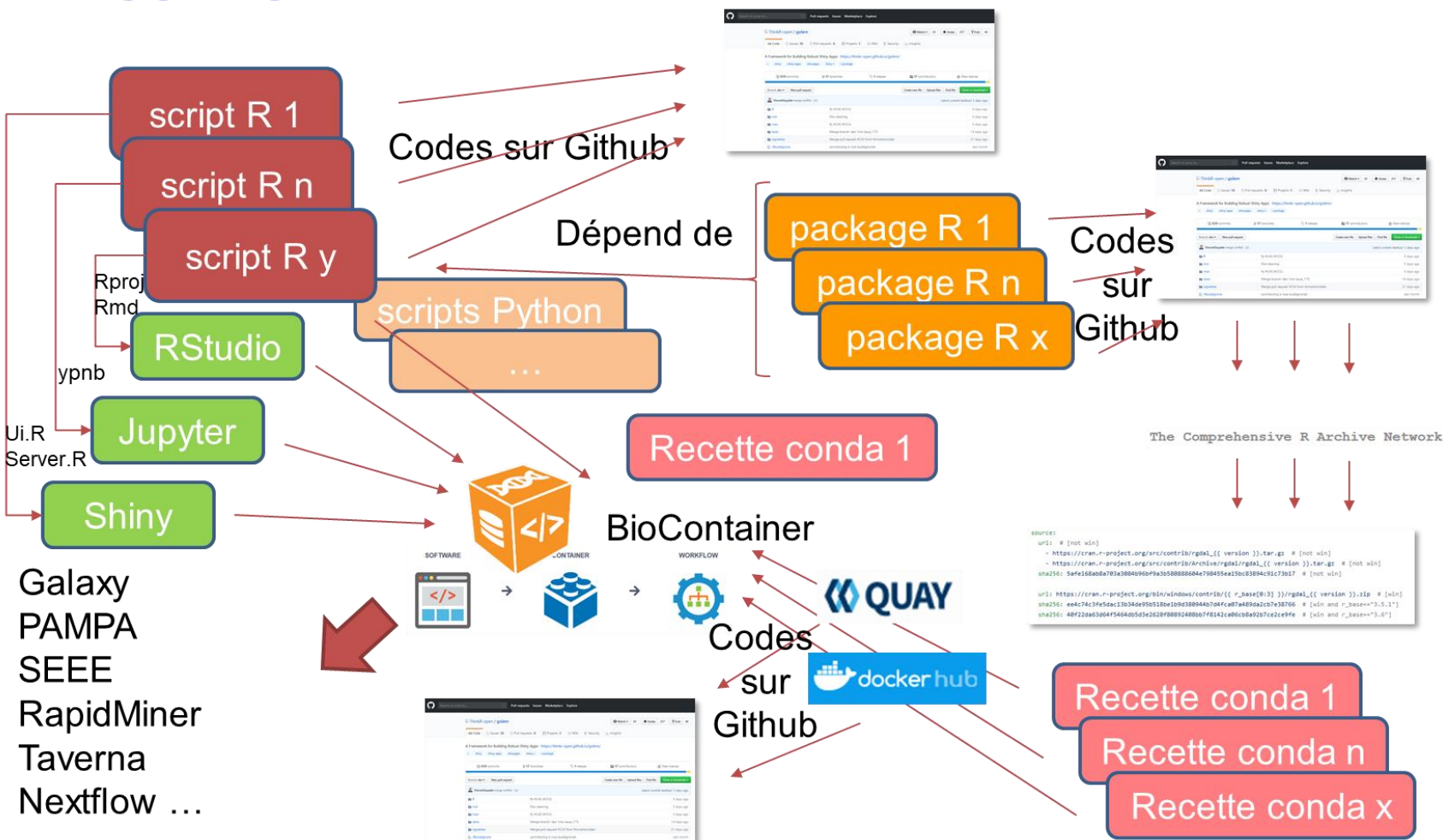


Portail de données métadonnées :  
<https://openstack-192-168-100-101.genouest.org/metacatui>

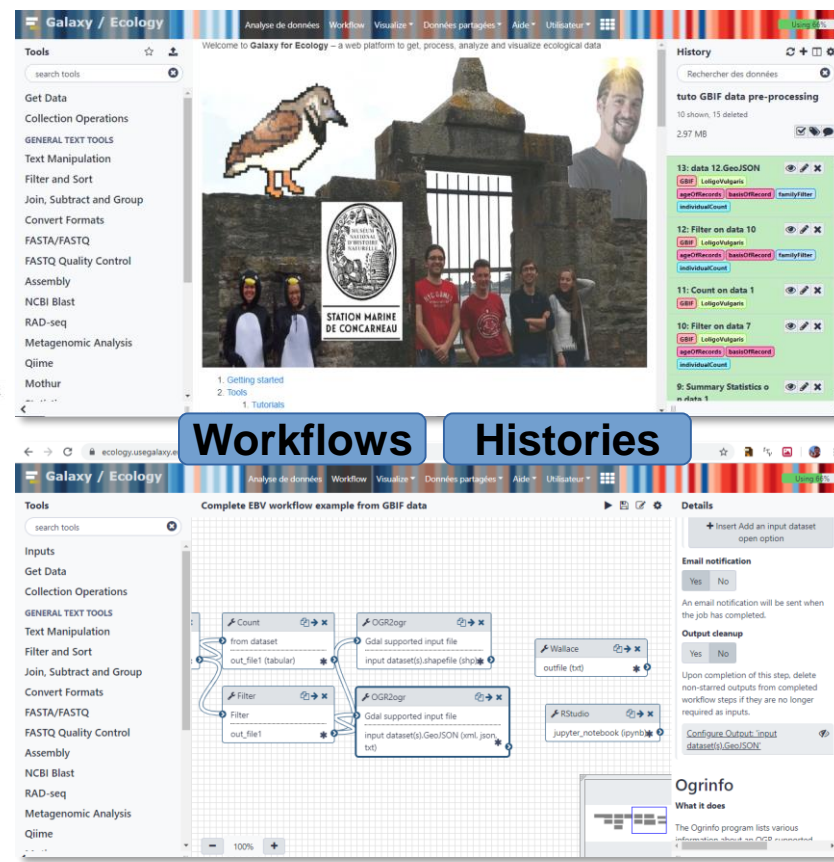
# Le paysage analyse via Github, Conda, Containers, Cloud et Galaxy



## EcoinfoFAIR



## Machines virtuelles (local ou cloud) Galaxy



# Merci de votre attention



## PNDB team

Coline Royaux - M2 EBE (Galaxy)



Elie Arnaud – Ingénieur R Shiny /  
métadonnées



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



Yvan Le Bras

Démo portail de données et métadonnées

<https://www.youtube.com/watch?v=STwsYDHet2A&feature=youtu.be>



Démo Galaxy Europe

<https://www.youtube.com/watch?v=HelAHggX6D4&t=268s>