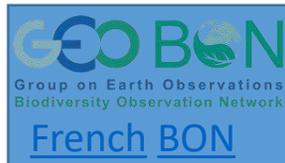


Pôle national de données de Biodiversité

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

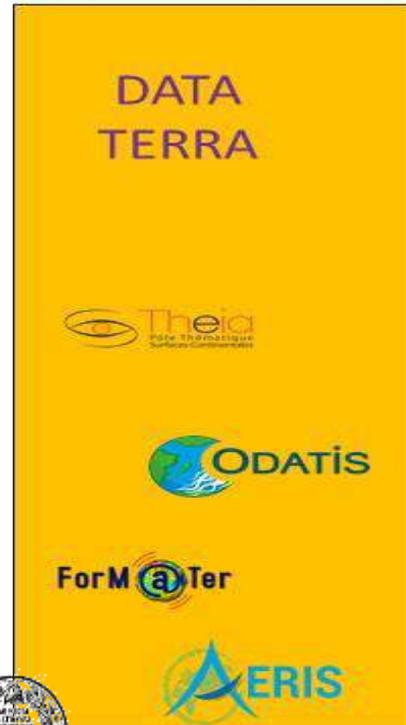


Yvan LE BRAS, Chef de projet (UMS PatriNat)
Sandrine Pavoine, MNHN (UMR CESCO)
Anne-Sophie ARCHAMBEAU, GBIF-France (UMS PatriNat)
Cécile CALLOU, Dir UMS BBEES (CNRS-MNHN)
Aurélié DELAUAUD (FRB)
Dominique JOLY, DAS CNRS (INEE)
Thomas Milon, Chef de projet "Système d'information sur la biodiversité« (UMS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

 #PNDB @Yvan2935



Pôles de données



TGIR/IR nationales

IN-SYLVA Fr	TREEFORCE
RECOLNAT	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

Miroir européen

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

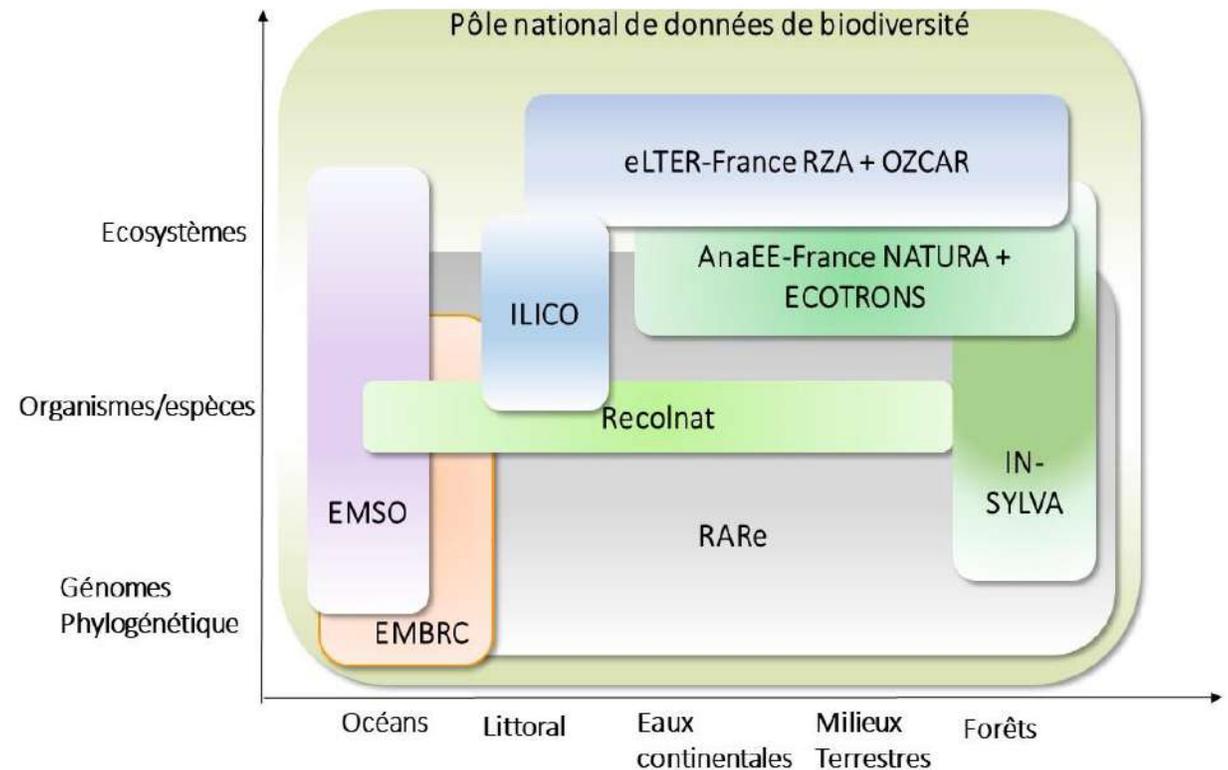


Figure 4: Paysage des infrastructures de recherche françaises dans le domaine de la biodiversité et des écosystèmes. L'encadré orange indique une infrastructure commune avec le domaine « biologie-santé ».

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

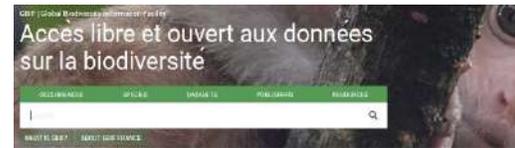
1. S'appuyer sur bases de métadonnées/données et systèmes d'information **existant** :
 - en les mettant en relation et en générant des synergies (notamment référentiels et ontologies)
 - en contribuant à renforcer leur visibilité et leur cohérence

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)

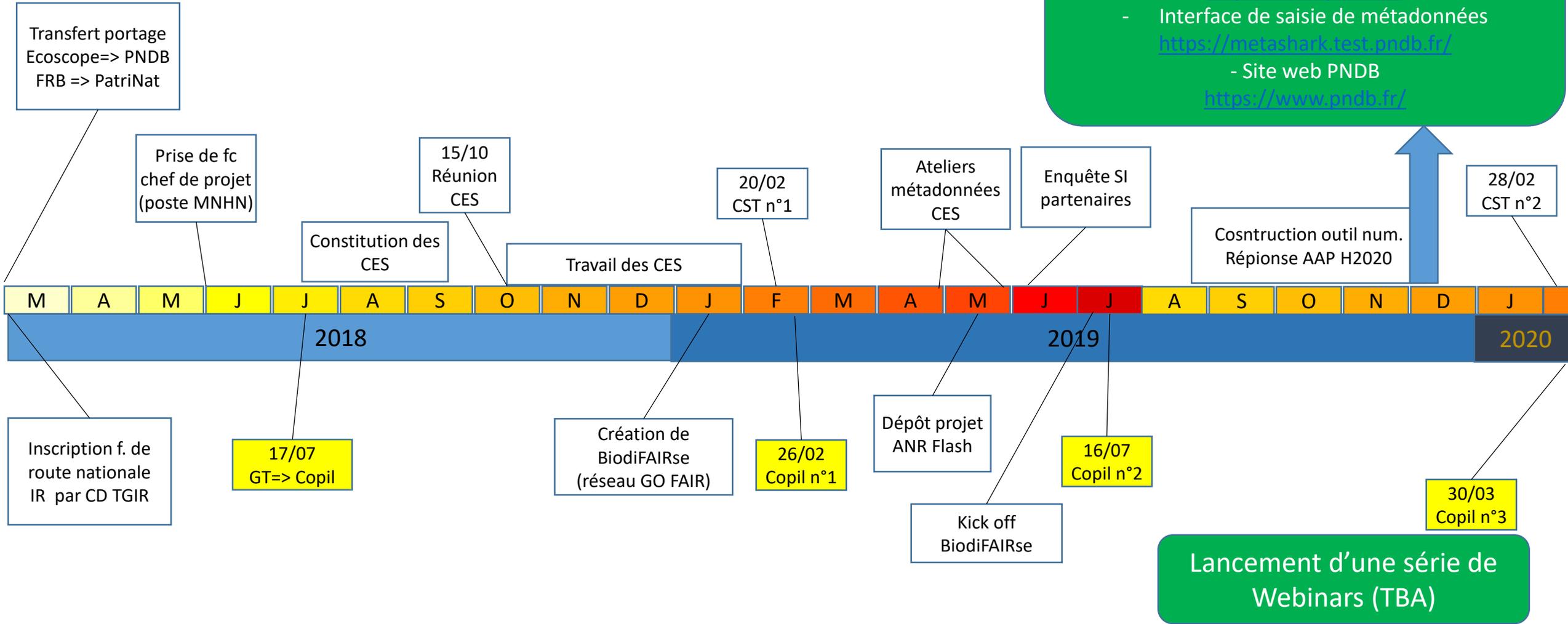
3 volets

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

Premiers livrables :

- Portail de données PNDB
<https://data.pndb.fr/>
- Interface de saisie de métadonnées
<https://metashark.test.pndb.fr/>
- Site web PNDB
<https://www.pndb.fr/>



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

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

PNDB Outils ▾ Animation ▾ Projets ▾ FAQ

Pôle National de Données de Biodiversité!

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

[Accéder aux données »](#)

[Tester MetaShARK »](#)

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/>

PNDB DONNÉES RÉSUMÉ A PROPOS aller à: DOLx or II Go CONNEXION

PÔLE NATIONAL DE DONNÉES DE BIODIVERSITÉ

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

Exemple: espèces invasives

brgm Géosciences pour une Terre durable
cirad LA RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT
cnrs
FRB FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ
Ifremer
INERIS maîtriser le risque pour un développement durable

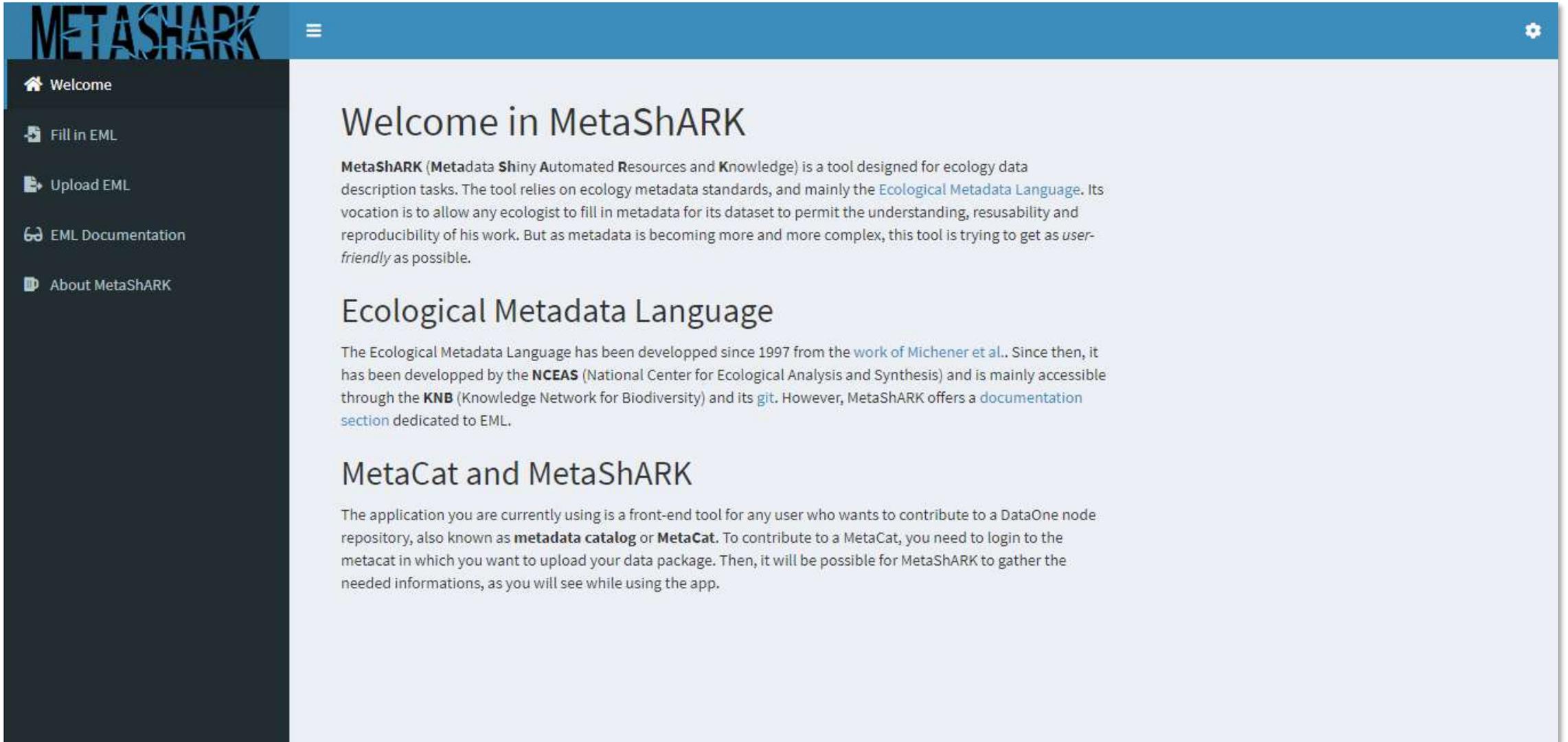
INRAE
IRD Institut de Recherche pour le Développement FRANCE
MUSEUM NATIONAL D'HISTOIRE NATURELLE
OFB OFFICE FRANÇAIS DE LA BIODIVERSITÉ
Météo France

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, 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/>



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 for settings on the right. Below the header is a dark grey sidebar with a list of navigation links: '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 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/>

The image displays four screenshots of the Galaxy / Ecology web interface. The top-left screenshot shows a world map with red data points and a sidebar with tool categories like 'GIS Data Handling' and 'STATISTICS AND VISUALISATION'. The top-right screenshot shows a 'History' panel with a search bar and a list of workflows, including 'treat biodiversity data & tuto GBIF data handling'. The bottom-left screenshot shows a 'Complete EBV workflow example from GBIF data' with a grid of workflow steps like 'Count', 'Filter', 'DGR2ogr', and 'Wallace'. The bottom-right screenshot shows a 'Your workflows' table listing various workflows with columns for Name, Tags, Owner, Steps, and Published status.

Name	Tags	Owner	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



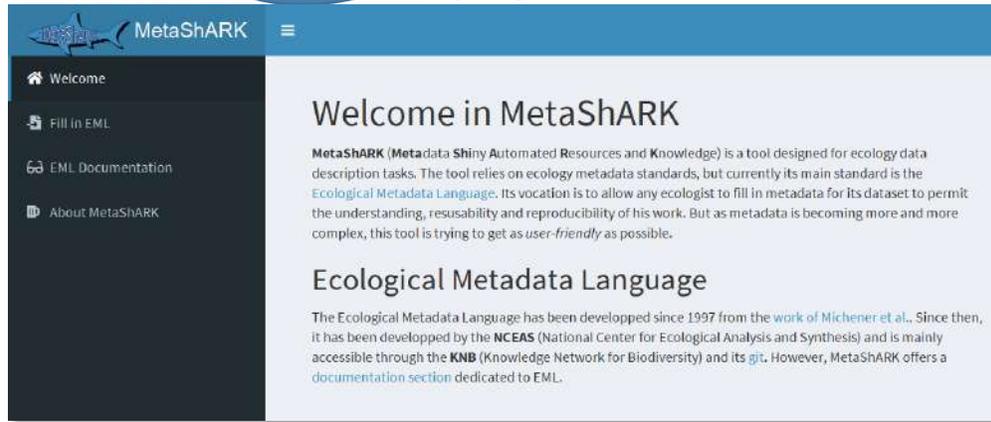
dataset1

dataset2

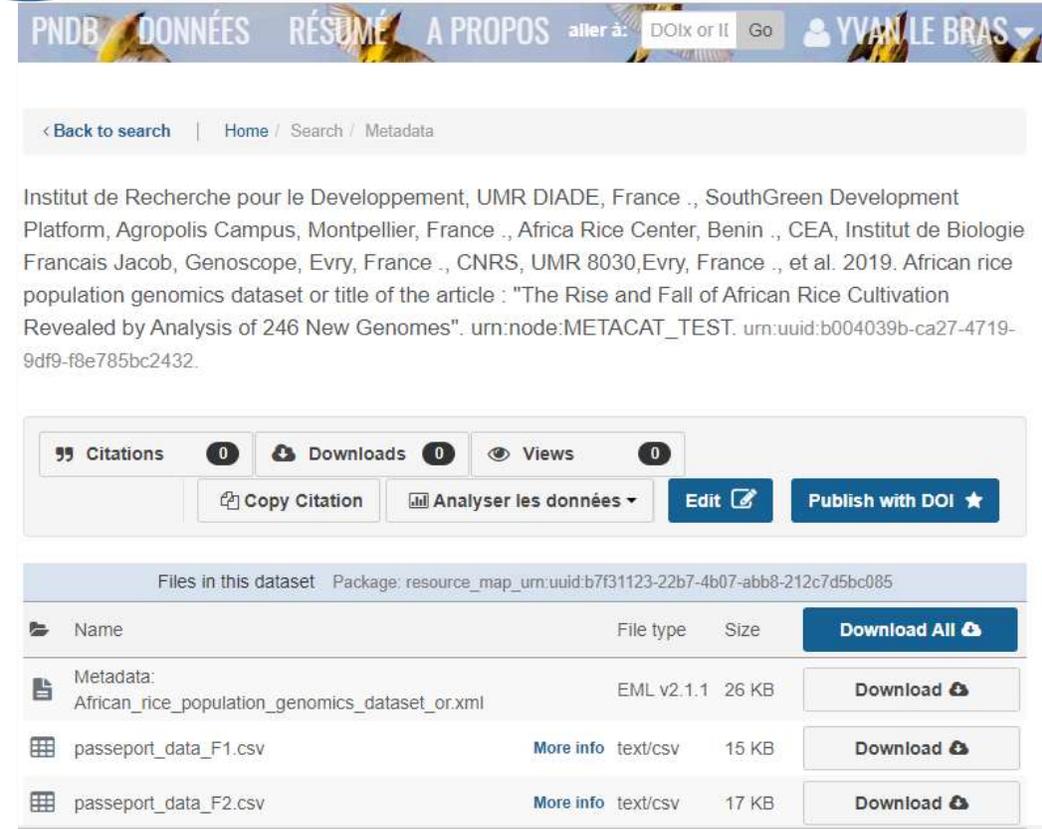
script1

article1

protocol1



EML

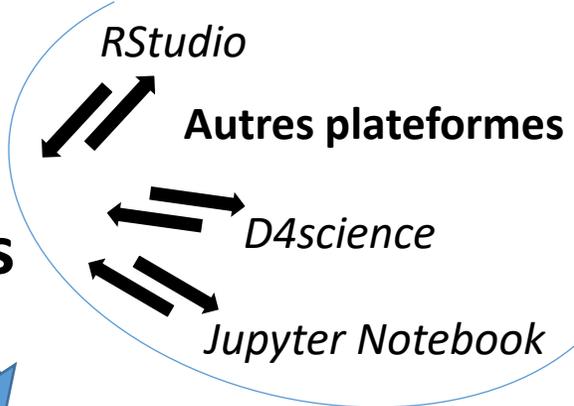


Portail de données métadonnées :

<https://openstack-192-168-100-101.genouest.org/metacatui>

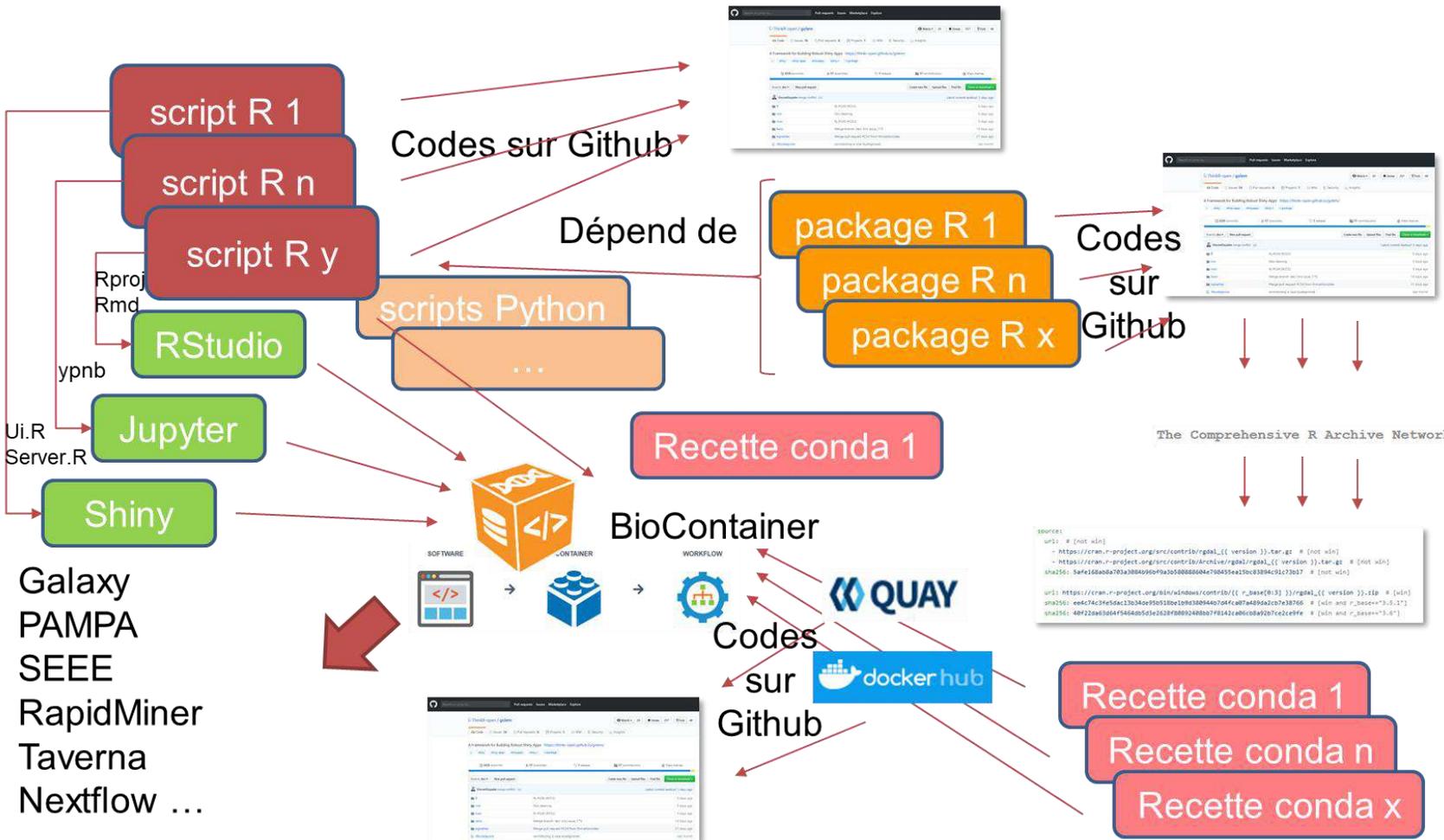
<http://data.test.pndb.fr/data> <https://data.pndb.fr/data>

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



Machines virtuelles (local ou cloud)
Galaxy

EcoinfoFAIR



Galaxy / Ecology

Workflows

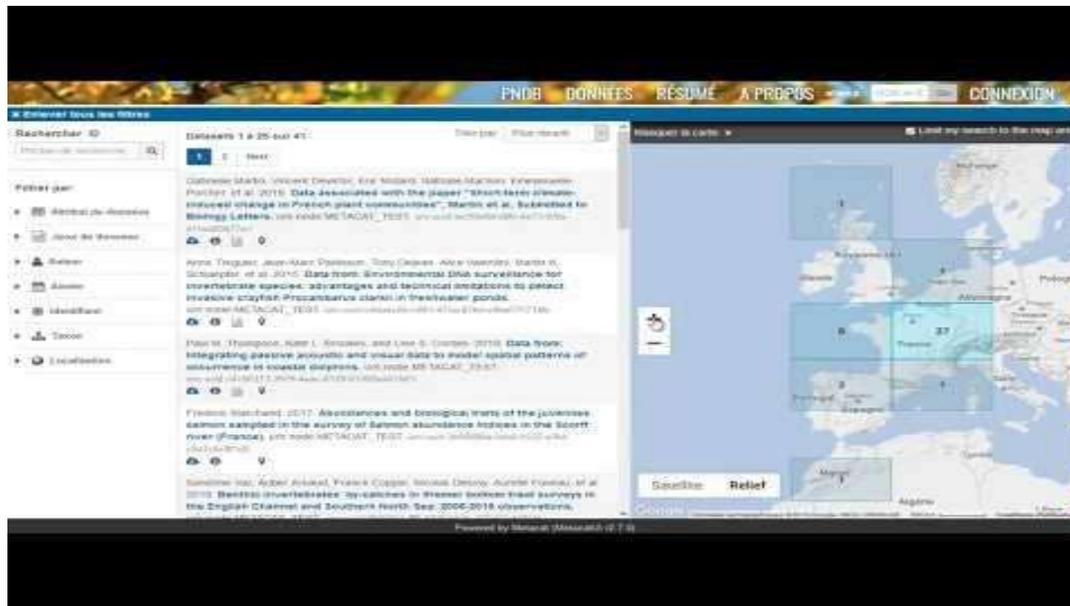
Histories

Complete EBV workflow example from GBIF data

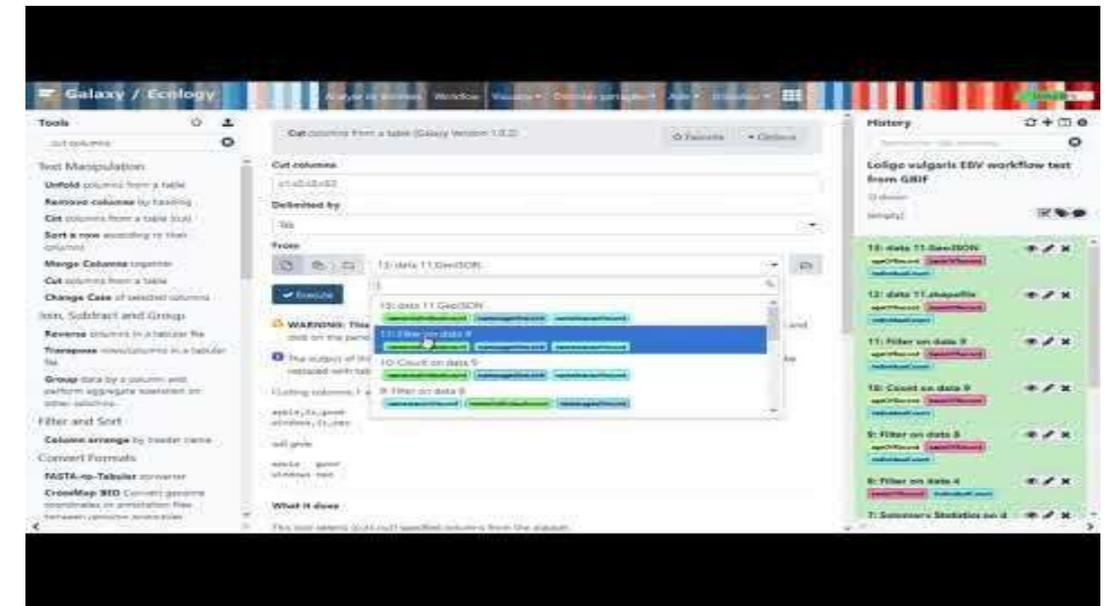
Démonstrations !!!!!!!



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>

Merci de votre attention



PNDB team

Coline Royaux - M2 EBE (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

<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](#)