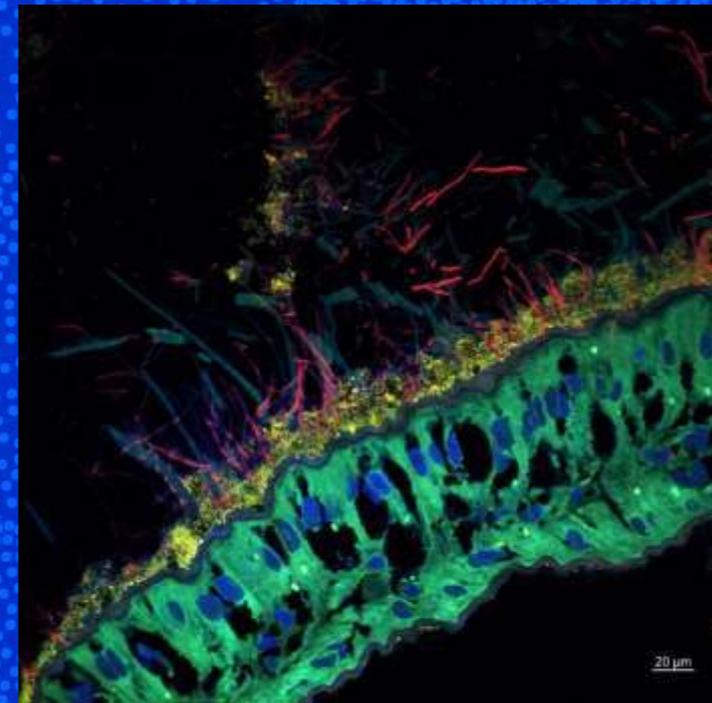




Traçabilité des données séquencées

Exemple des données du LMEE au sein de BEEP (Laboratoire de Microbiologie des Environnements Extrêmes)

Pauline Auffret IRSI-SEBIMER
Valérie Cueff-Gauchard BEEP/LMEE



16 octobre 2025

Les différents écosystèmes étudiés à BEEP

Sources hydrothermales marines profondes

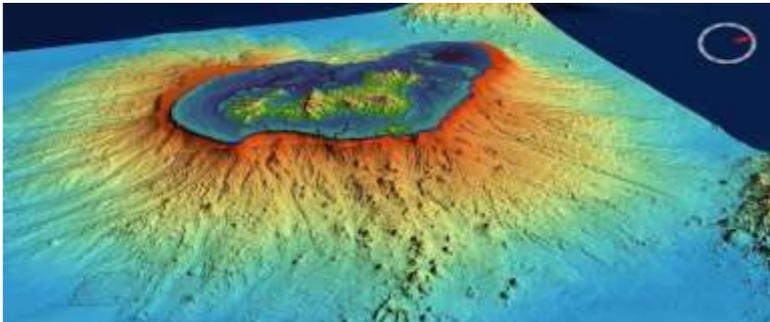


Plaines à nodules polymétalliques

Suintements froids – lacs de saumure



Volcans sous-marins



Coraux froids profonds



Sources thermales côtières / terrestres

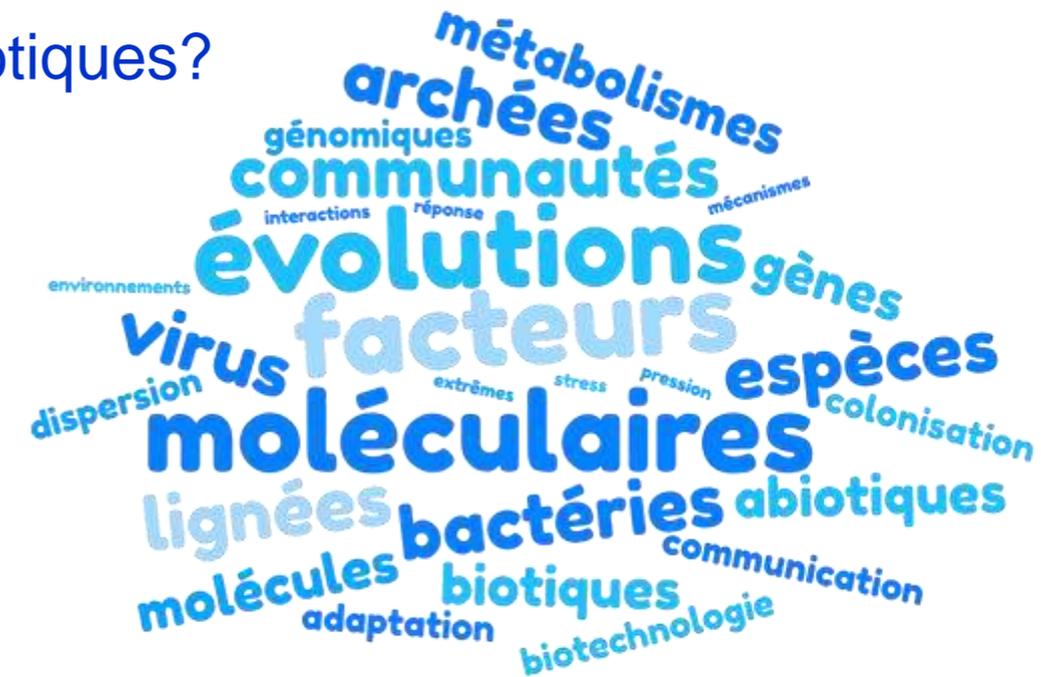


Questions scientifiques autour des microorganismes

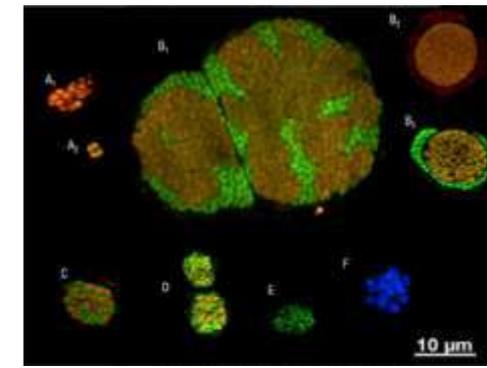
- Biodiversité : espèces et lignées présentes dans ces écosystèmes ? métabolismes mis en jeu ?
- Dynamique spatiale et temporelle des communautés microbiennes en relation avec les facteurs biotiques et abiotiques ? Fonctionnement de ces écosystèmes
- Interactions entre microorganismes?

entre microorganismes et compartiments abiotiques?

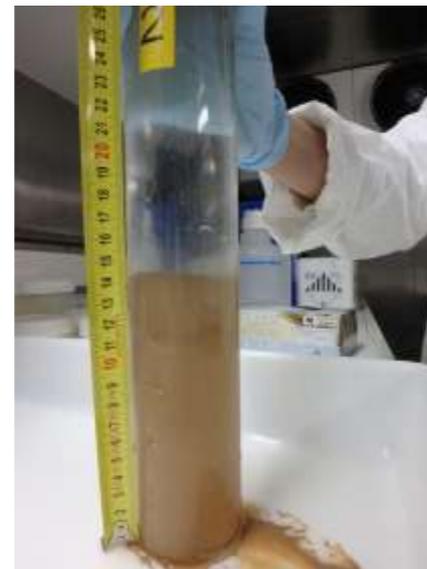
entre microorganismes et faune (symbiose)?



Les différents types d'échantillons étudiés en écologie microbienne



Carotte sédimentaire



Cheminée hydrothermale / roche / nodule



Eau de mer / fluide hydrothermal



Animaux



16/09/2025

Traçabilité des données séquencées

Les moyens de prélèvements

Submersibles
Victor 6000 / Nautilie



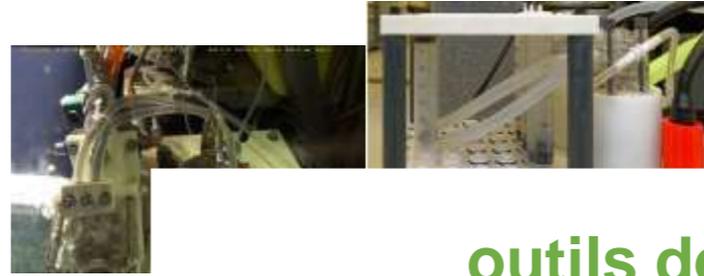
Depuis le navire



bathysonde
avec rosette
de bouteilles
Niskin

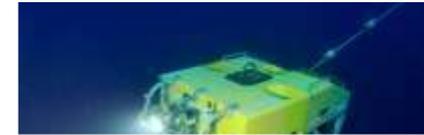


carottiers :
Kullenberg,
multitubes



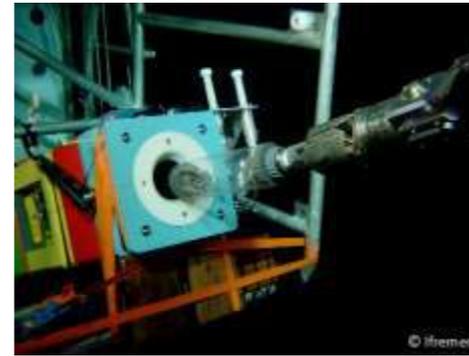
POC
PLU

outils dédiés à nos études



Remontée sans
dépressurisation

Fixation *in situ*



PERISCOP (Amex,
Sorbonne université)

Pour faune



IBIS (Carnot
DEEPECOS)

Pour fluide



FISH (Carnot
DEEPECOS)

Pour animaux

anche)



jaune



carc



beep

carottiers lame



16/10/2025

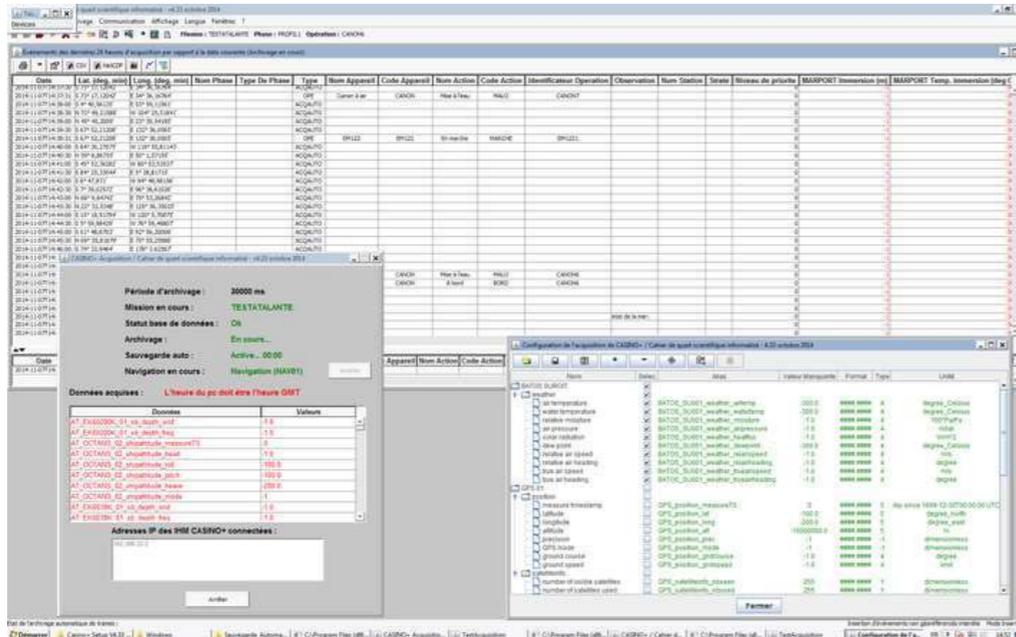
Traçabilité des données séquencées

Base de données des échantillonnages

Cf exposé Olivier Soubigou

Depuis le navire

JouARNI de bord : CASINO+



Submersibles
Nautilie / Victor 6000

MIMOSA3

Adelie : Adelie Import / Adelie Video

Après correction
de la navigation

Sealog : pour compléter
les données liées aux
échantillonnages

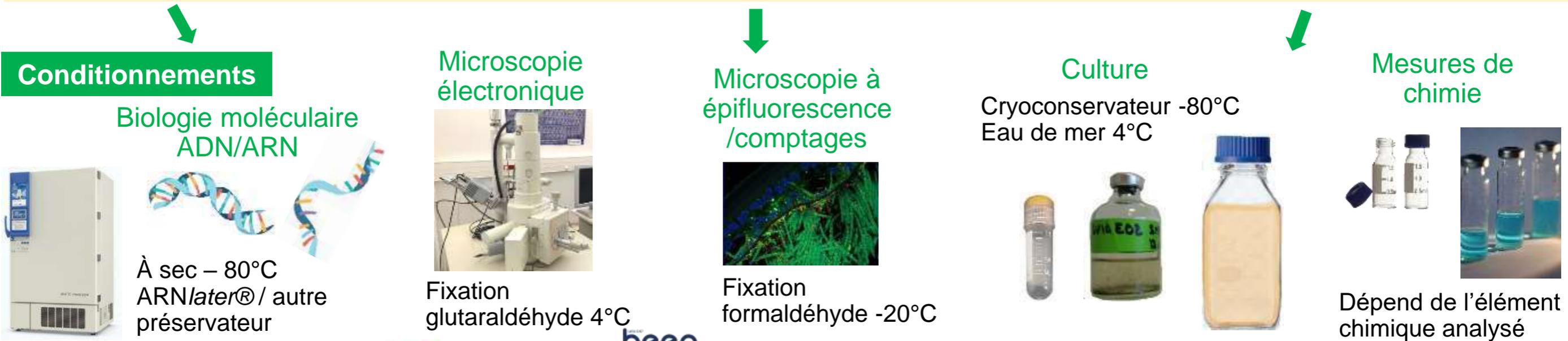
Base de données des campagnes



16/09/2025

Traçabilité des données séquencées

Traitement des échantillons en fonction de l'étude

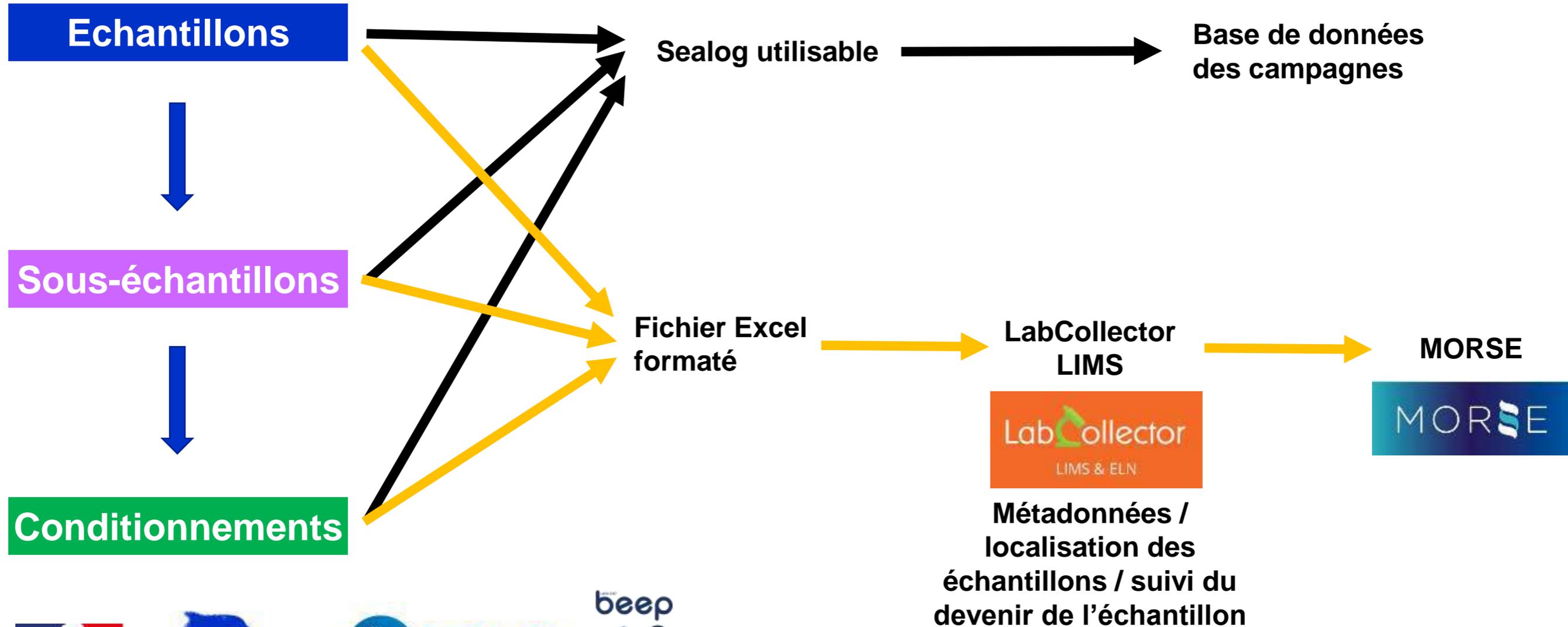


16/09/2025

Traçabilité des données séquencées

Base de données des échantillons

Cf exposé Lénaïck Menot et Brendan Hennebaut



16/09/2025

Traçabilité des données séquencées

Exemple d'expériences de laboratoire : isolement de souches en culture

Mise en collection

Echantillons

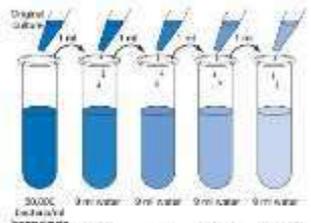
Enrichissements

Test de différentes conditions de culture



Isolements

Obtention d'une souche microbienne pure (bactérie/archée)



Souchothèque LMEE Ifremer

et/ou



Biologics Software

Avec interface WEB pour rechercher et commander souches

<https://ubocc.bio-aware.com/page/marine-search>

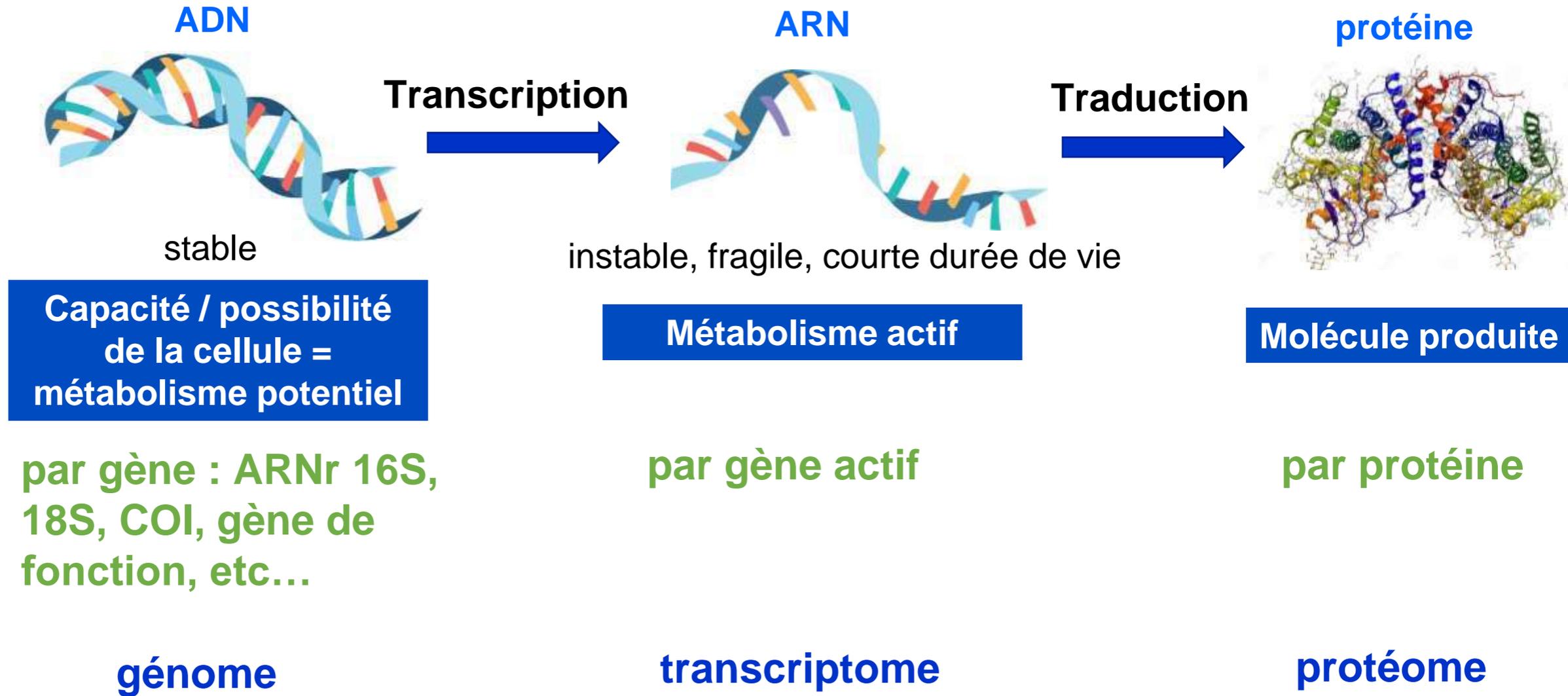


Traçabilité des données séquencées



16/09/2025

Autre approche expérimentale : biologie moléculaire



Approche ciblée

par gène : ARNr 16S, 18S, COI, gène de fonction, etc...

par gène actif

par protéine

Approche globale

génomome

transcriptome

protéome



16/09/2025

Traçabilité des données séquencées

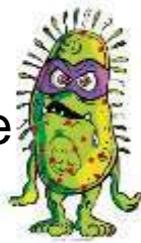
10

Approche ciblée sur un individu

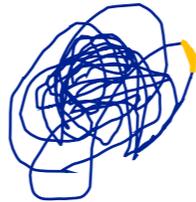
Barcoding

Séquençage de 1^{ère} génération en approche ciblée

1 individu d'espèce animale ou 1 culture pure = 1 espèce microbienne



Extraction d'ADN



1 ADN total

PCR

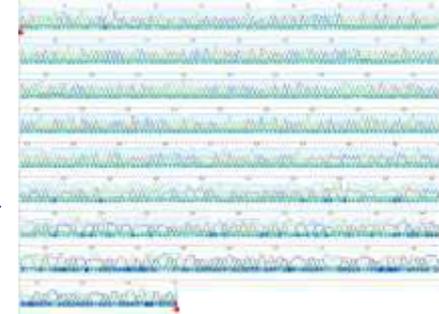


Amplification d'un gène marqueur 200-2000 pb



Millions de copies d'1 gène

Séquençage Sanger



Max 900 pb

→ gestion de ces données

Stockage des données brutes



Ordinateur portable de bureau

Nettoyage des données



Traitement des données



Gestion individuelle des données

Publication des séquences



par scientifiques



16/09/2025

Traçabilité des données séquencées

Approche ciblée sur un individu

Barcoding

Séq

1 individu d'espèce animale ou 1 culture pure = 1 espèce microbienne



Gestion individuelle sur ordinateurs → Traçabilité à améliorer !!!

- Gestion collective des données à mettre en place
 - Par laboratoire
 - Et/ou centralisé au niveau de l'Ifremer sur Datarmor



En réseau

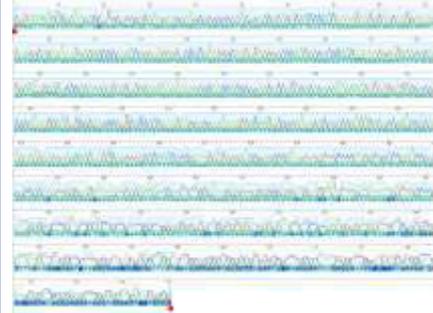


Par laboratoire
Permet de relier à
l'échantillon de départ



Datarmor

Autre
outil?



Max 900 pb

→ gestion de ce

Stockage des données brutes

- Soumission éventuelle par service SEBIMER ?



Ordinateur
portable
personnel

Traitement
des données

par scientifiques

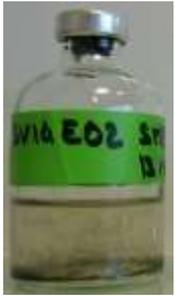


16/09/2025

Traçabilité des données séquencées

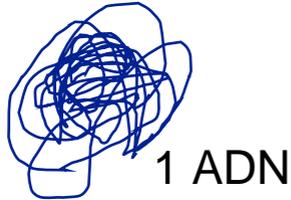
Séquençage NGS = séquençage haut débit

Génome



Culture, 1 espèce
1 individu

Extraction
ADN

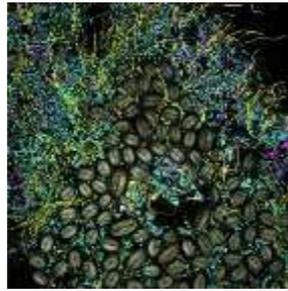


Séquençage ↓ Short-read / Long-Read

Tous les gènes

↓
Potentiel métabolique
de l'individu

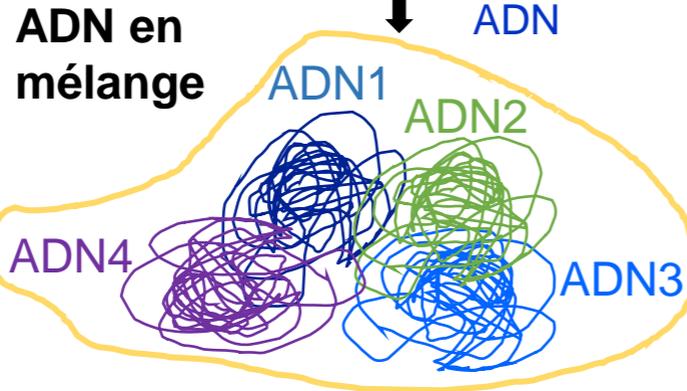
Métabarcoding



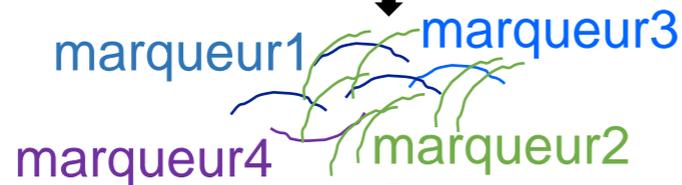
Communauté
Holobionte

1 échantillon

Extraction
ADN



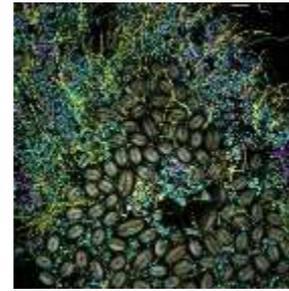
PCR



Séquençage

Identification des organismes

Métagénomique



Communauté
microbienne
Holobionte

Extraction
ADN

ADN en
mélange



Séquençage

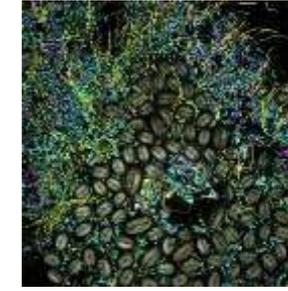
Tous les gènes

Bioinformatique ↓ Binning



Potentiel métabolique de la
communauté, / individu

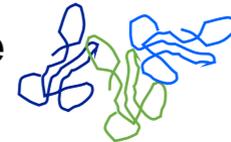
Métatranscriptomique



Communauté
microbienne
Holobionte

Extraction
ARN

ARN en
mélange



Séquençage

Tous les gènes

Bioinformatique ↓ Binning



Partenaires actifs;
métabolismes actifs à un
instant T



SeBiMER
Bioinformatique marine



Biologie et Écologie des
Écosystèmes marins Profonds

16/09/2025

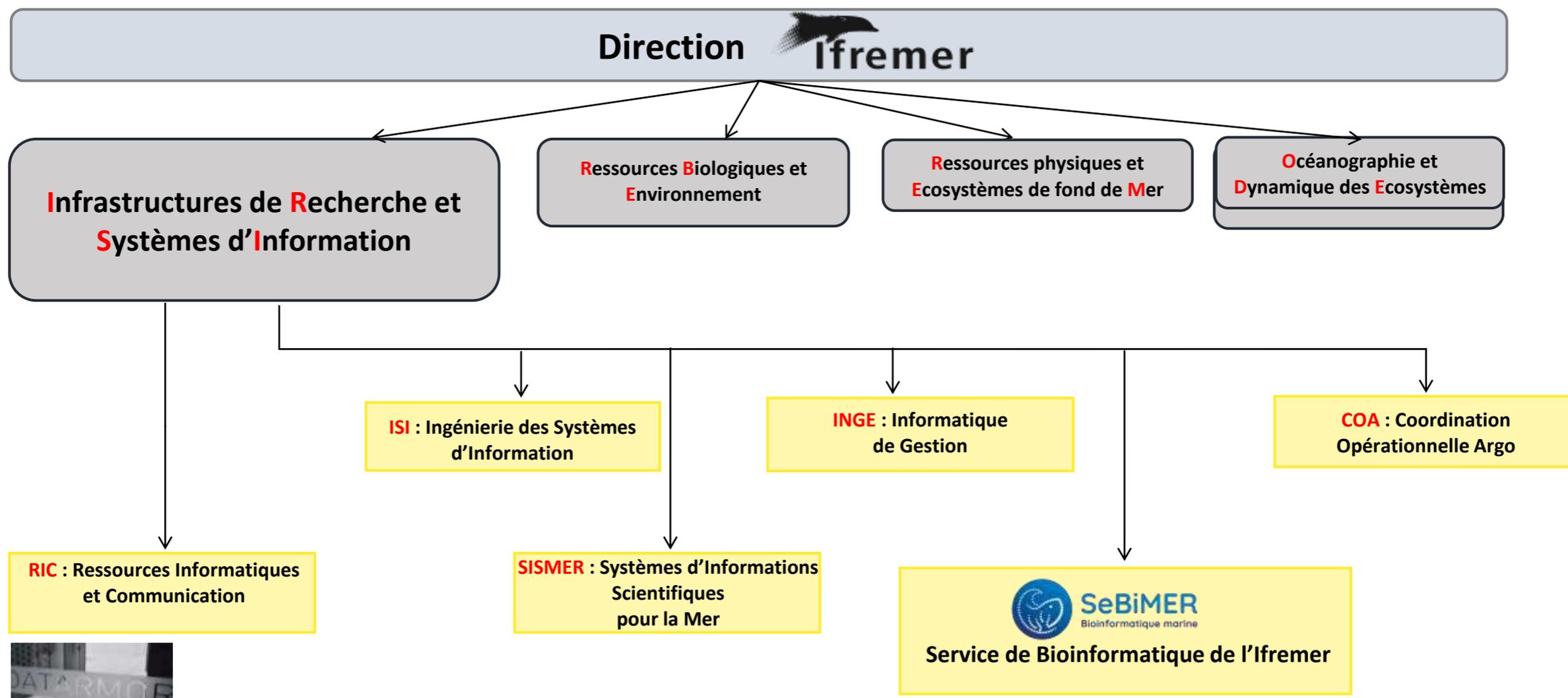
Traçabilité des données séquencées

01.

Présentation du SeBiMER

Service de Bioinformatique de l'Ifremer

SeBiMER : présentation



Bâtiment Blaise Pascal (209), 2^{ème} sous-sol



16/09/2025

Traçabilité des données séquencées

SeBiMER : activités



Patrick Durand, PhD



2017-now



Laura Leroi, engineer



2018-now



Cyril Noël, PhD



2019-now



Pauline Auffret, engineer



2022-now



Yaëlle Pihan, engineer



2024-2026



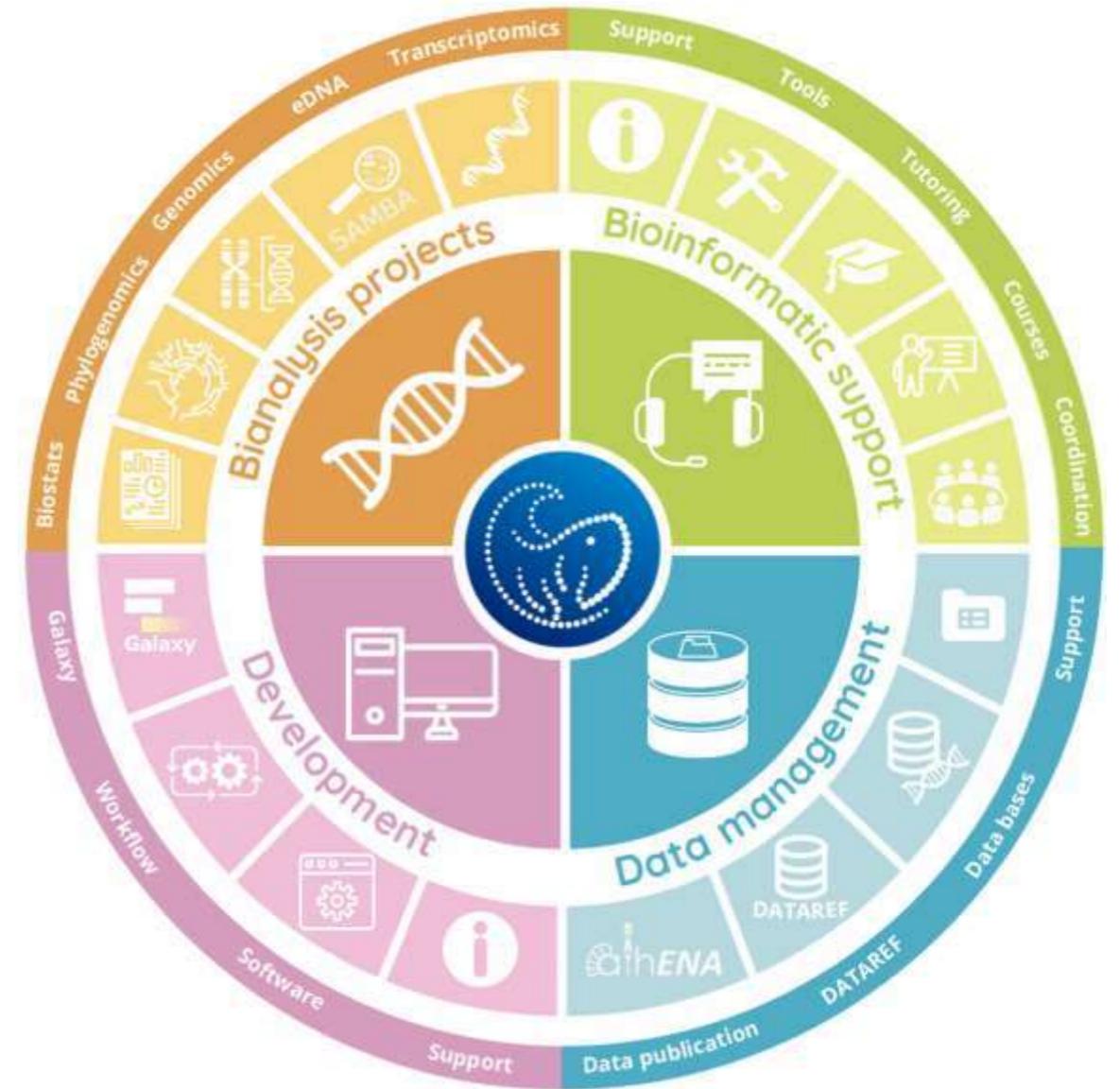
Antoine Daussin, engineer



2025-2026



Master students
in season



16/09/2025

Traçabilité des données séquencées

SeBiMER : thématiques principales

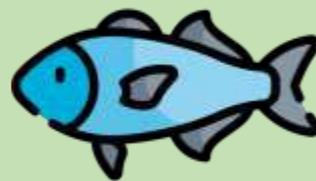
ADN environnemental / metabarcoding

- Surveillance de l'émergence de pathogènes susceptibles d'affecter les humains et les animaux
- Détection et suivi la biodiversité des grands fonds
- Impact des changements globaux sur la bioversité des écosystèmes
- Description des communautés bactériennes associées aux crevettes des cheminées hydrothermales



Transcriptomique

- Impact de l'acidification des océans sur la métabolisme du foie du bar
- Caractérisation des mécanismes moléculaires impactés par les vagues de chaleur chez les bécards géants et leurs symbiotes
- Identification des gènes de résistance à une maladie chez l'huître creuse



Assemblage de génomes / métagénomique

- Reconstruction de génomes de bactéries symbiotiques ou de micro-organismes extrémophiles à partir de métagénomes
- Assemblage de génomes complets d'espèces marines d'intérêt écologique ou économique (poissons, mollusques, crustacés).
- Étude de la diversité génétique et des variations intra-espèces chez l'huître plate



02.

Typologie des données

Des échantillons aux données : overview

Equipes de recherche RBE, REM, ODE

ADN environnemental / metabarcoding



Transcriptomique



Assemblage de génomes / métagénomique



Extraction d'ADN ou ARN
Conservation des échantillons
In situ ou au laboratoire

1. Echantillonnage

Préparation des échantillons pour le séquençage
amplification, ajout d'adaptateurs etc



© Oxford Nanopore Technologies

Séquençage *in situ*

MinION : séquenceur ADN portable, alimenté par USB, pour générer des lectures longues directement sur le terrain (débit et précision inférieurs aux plateformes de laboratoire)



© Illumina

Séquençage par une plateforme

Illumina (MiSeq, NovaSeq) pour le séquençage à haut débit de fragments courts, PacBio (Sequel II/IIe) pour obtenir des séquences longues, ou Oxford Nanopore (GridION, PromethION) pour générer des lectures très longues.

2. Séquençage

Données de séquençage

Transformation du signal brut (électrique ou optique) produit par le séquenceur



3. Data



SeBiMER
Bioinformatique marine



Biologie et Ecologie des
Écosystèmes marins Profonds

16/09/2025

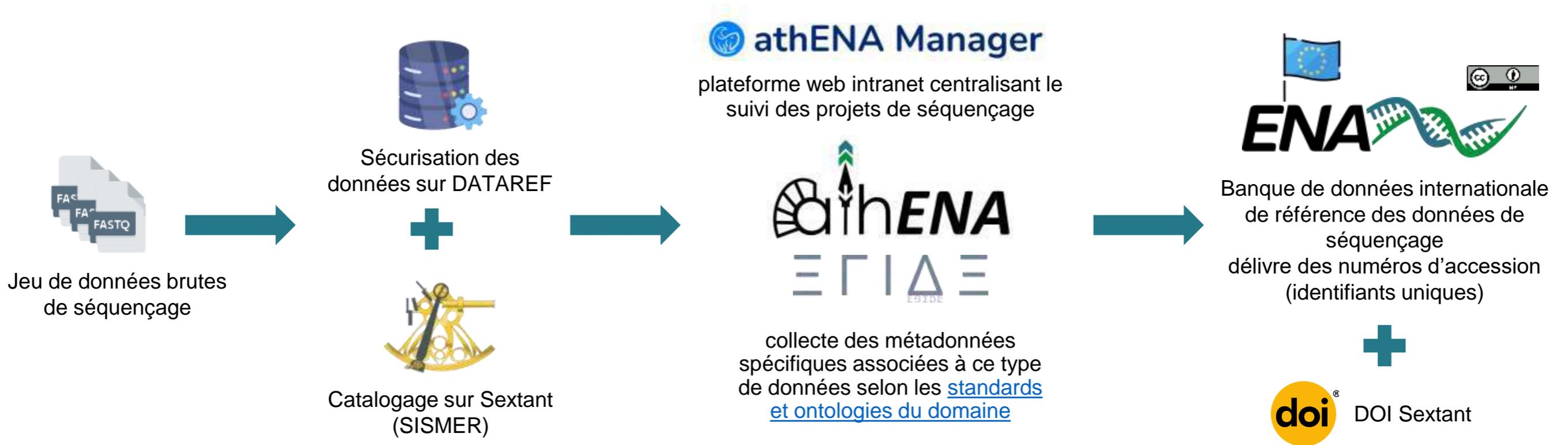
Traçabilité des données séquencées

03.

Gestion des (méta)données de séquençage

Flux global de gestion des données au SeBiMER

Au SeBiMER : téléchargement des données sur DATAREF, collecte des métadonnées avec Egide¹ et athENA², data brokering vers ENA³ (European Nucleotide Database)



1. <https://gitlab.ifremer.fr/bioinfo/workflows/athena>
2. <https://gitlab.ifremer.fr/bioinfo/software/egide>
3. <https://www.ebi.ac.uk/ena/browser/view>



Le template de métadonnées athENA

Le [template de métadonnées athENA](#) est basé sur le modèle de métadonnées de l'ENA, et est composé de 7 onglets à remplir

- 1 ligne par échantillon / par fichier de séquençage
- distinction champs obligatoires / facultatifs, liste déroulante disponible pour les vocabulaires contrôlés.

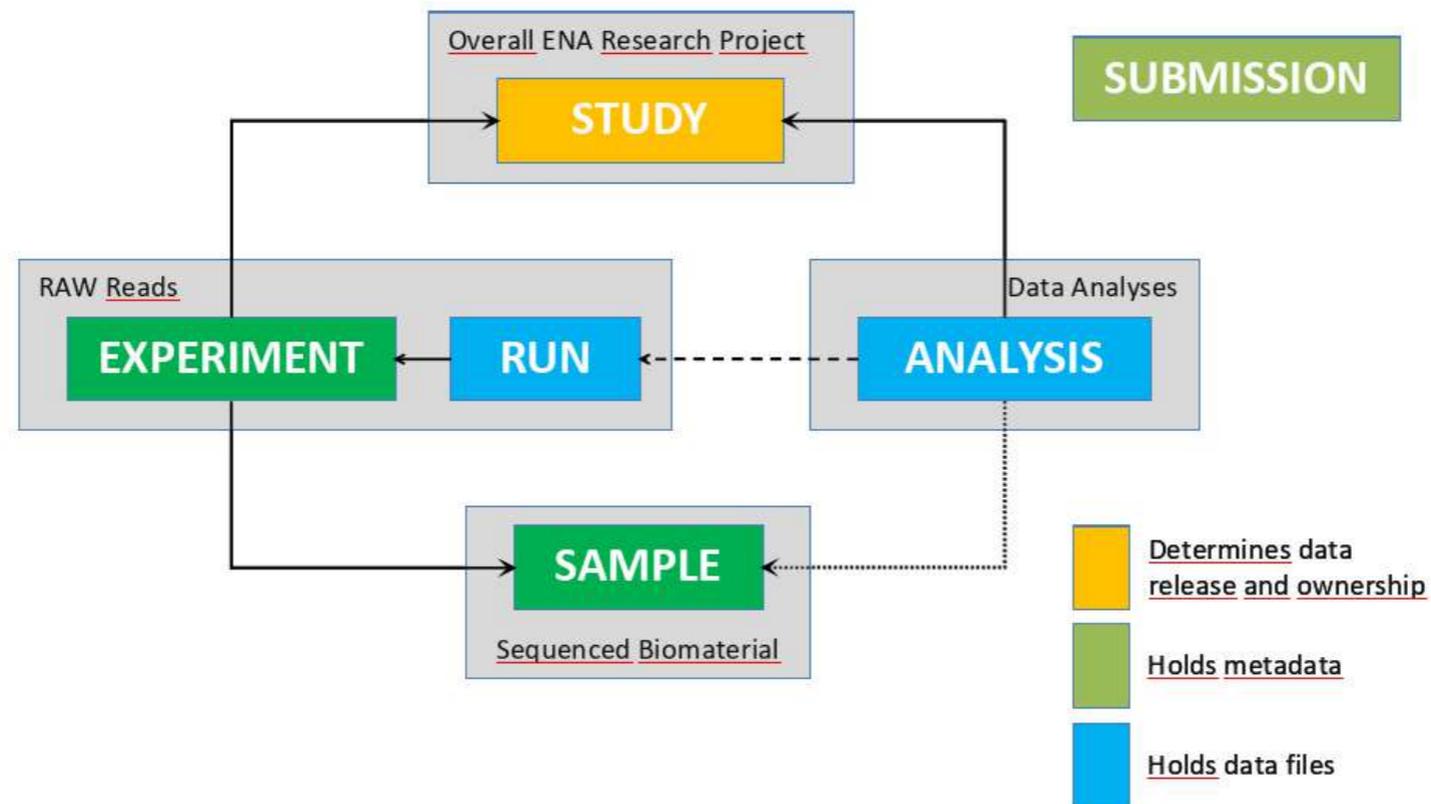
STUDY	➤ STUDY	mandatory
	➤ COLLABORATORS	mandatory
	➤ SAMPLES_general	mandatory
	➤ SAMPLES_meta	mandatory for metaB/metaG/metaT
SAMPLE	➤ SAMPLES_indiv	mandatory for NON-metaB/metaG/metaT
	➤ SAMPLES_chemphys	optional
	➤ SEQUENCING	mandatory
EXPERIMENT		
RUN		



16/09/2025

Le modèle de métadonnées ENA

- Interdépendance des objets du [modèle de métadonnées de séquençage de l'ENA](#) :

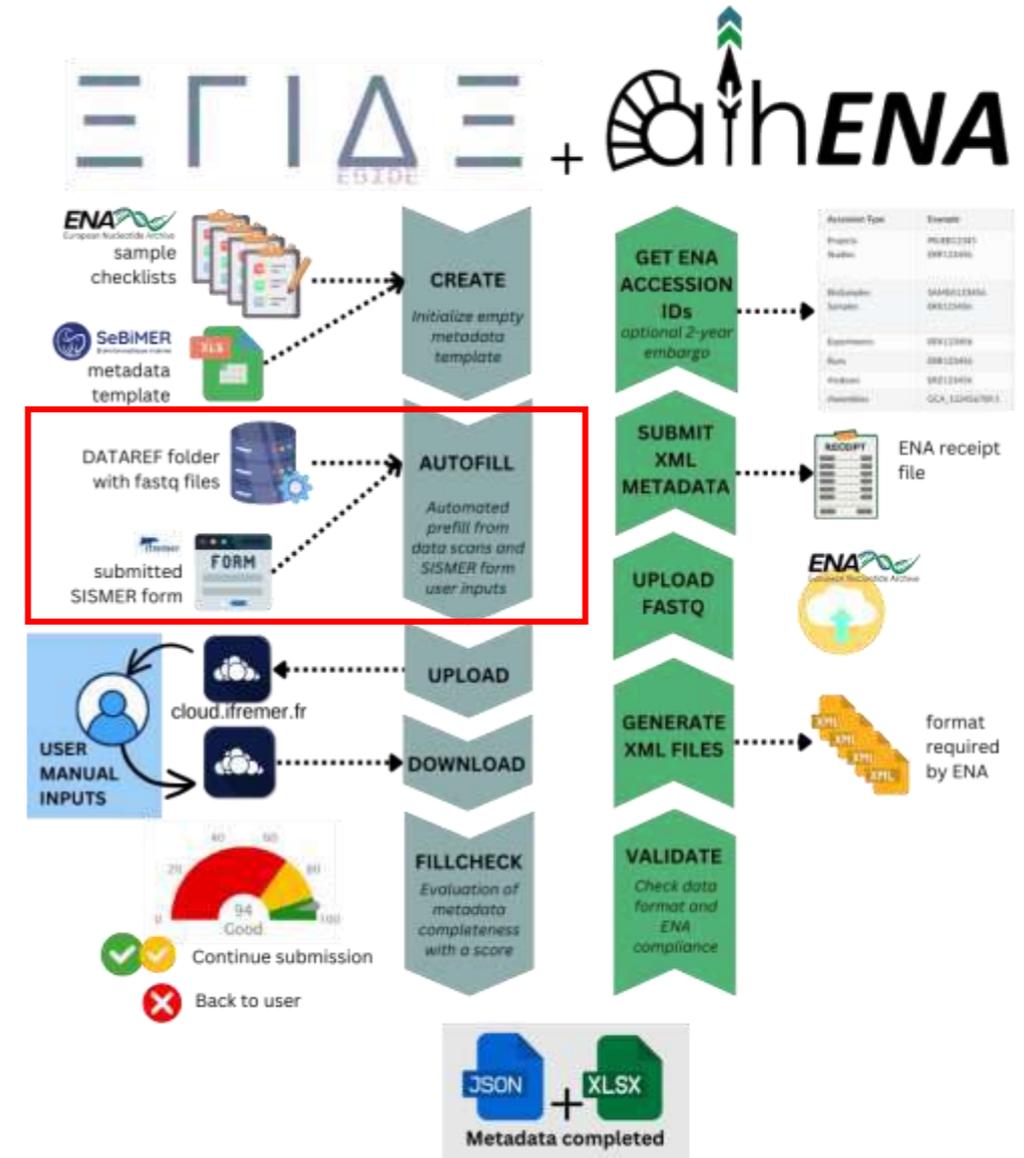


Collecte automatisée des métadonnées *via* Egide

Grâce au programme [Egide](#) (@David Goudenège) une partie des métadonnées est automatiquement importée dans le formulaire athENA avant de le donner aux utilisateur·ices :

- certaines informations consignées dans le **formulaire SISMER** vont peupler automatiquement les onglets **STUDY**, **COLLABORATORS** (Titre, identité des producteurs de la donnée) et **SAMPLES_general** (coordonnées GPS)
- en analysant les données elles-mêmes (fichiers de séquences .fastq.gz par exemple), Egide en déduit des métadonnées de l'onglet **SEQUENCING**

(Instrument et version, taille des séquences..)



Le Groupe de Travail METIS

- Les métadonnées de séquençage doivent contenir ou référencer les **métadonnées des échantillons biologiques d'origine** !
- Comme ces métadonnées n'ont pas pu être récupérées par Egide → c'est aux producteur.ices de la donnée de les saisir manuellement dans athENA.
- Or, retour des producteur.ices de la donnée : ces métadonnées ont déjà été saisies en grande partie dans d'autres SI Ifremer (ex. LabCollector, MORSE...) et/ou dans des fichiers Excel personnels



Création du GT METIS : Moissonnage et Traçabilité des Informations de Séquençage

Objectif : Optimiser la collecte de métadonnées de séquençage tout au long de leur cycle de vie à Ifremer en mettant en place l'**interopérabilité** des différents SI Ifremer dans le but de moissonner et d'harmoniser les **métadonnées de séquençage** existantes.

2024/2025

<https://gitlab.ifremer.fr/bioinfo/gt-metis>



SeBiMER
Bioinformatique marine



16/09/2025

Traçabilité des données séquencées

04.

Bilan et perspectives collectives

Bilan et perspectives collectives

- athENA : pipeline libre (<https://gitlab.ifremer.fr/bioinfo/workflows/athena>) mais consolidation requise (portabilité et données hors biologie marine)
- Extension d'athENA à d'autres types de données (séquençage SANGER, puces)
- Un outil à tester : [MADBOT](#) (Metadata And Data Brokering Online Tool)
- Nouveau GT à la suite de METIS : implémentation du module Egide d'interrogation de MORSE



SeBiMER
Bioinformatique marine



16/09/2025

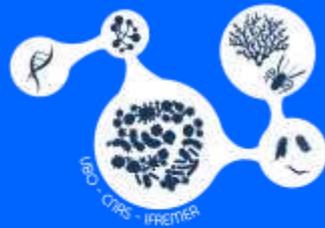
Merci !

<https://bioinfo.gitlab-pages.ifremer.fr/documentation/sebimer-intra/sebimer@ifremer.fr>



SeBiMER
Bioinformatique marine

UMR 6157
beep



Biologie et Ecologie des
Ecosystèmes marins Profonds

