

Compte rendu de l'atelier technique **ODATIS - PEPR Grands Fonds Marins**

Gestion des données issues d'échantillons en sciences de
la vie et de la terre

Yacine MOUFID
Clémence COTTEN

16 et 17 septembre 2025
Salon de l'Océan, Ifremer Plouzané

Sommaire

1. Introduction.....	5
1.1. Première journée – présentation et retours d’expériences	5
1.2. Deuxième journée – groupes de discussions par thématiques	5
1.3. Profils des participants.	6
2. Synthèse des présentations – mardi 16 septembre	9
2.1. Introduction à l’Atelier par Erwann Quimbert, Pierre-Marie Sarradin et Yacine Moufid	9
2.1.1. Le pôle Océan ODATIS	9
2.1.2. Le PEPR Grands Fonds Marins	9
2.1.3. Le volet Données et échantillons du PEPR	10
2.1.4. Présentation de l’atelier technique en cours	10
2.2. Christine David-Beausire – Recommandation de la flotte océanographique française à propos de la gestion de données	10
2.3. Olivier Soubigou – Outils d’acquisition à bord des navires de la flotte océanographique française.....	12
2.3.1. Suite logicielle TECHSAS.....	12
2.3.2. Sealog.....	12
2.3.3. Ocean Data Tool.....	13
2.3.4. Formations	13
2.4. Xavier Saint-Laurent – Modernisation du ROV Victor et projet DeepSea’Nnovation	13
2.4.1. Le projet DEEPSEA’Nnovation	14
2.4.2. Modernisation du ROV Victor	14
2.4.3. Gestion des données	14
2.5. Lénack Menot – Gestion des données d’échantillons biologiques au Laboratoire Environnement Profond (Ifremer)	15
2.5.1. Données collectées	15
2.5.2. Chaîne de gestion des données	15
2.5.3. Intégration et diffusion	16
2.5.4. Limites et difficultés	16
2.5.5. Discussion.....	16
2.6. François Dusoulier – L’accès aux données des collections naturalistes dans le cadre de l’infrastructure Récolnat	17
2.6.1. Objectifs et périmètre	17
2.6.2. Chiffres et outils	17
2.6.3. Gestion taxonomique	18
2.6.4. Enjeux actuels.....	18
2.6.5. Limites.....	18
2.7. Brendan Hennebaut – MORSE, entre enjeux réglementaires et traçabilité des échantillons.....	19
2.7.1. Objectifs	19

2.7.2.	Fonctionnement	19
2.7.3.	Métadonnées et volumes	19
2.7.4.	Enjeux	20
2.8.	Pauline Auffret et Valérie Cueff-Gauchard – Traçabilité des données séquencées, exemple des données du Laboratoire de Microbiologie des Environnements Extrêmes.....	20
2.8.1.	Contexte.....	20
2.8.2.	Prélèvements	20
2.8.3.	Gestion des échantillons	21
2.8.4.	Exemples d'analyses.....	21
2.8.5.	Données de séquençage haut débit	21
2.8.6.	Chaîne bio-informatique (SeBiMER)	21
2.9.	Annie Lebreton – Challenges liés à l'acquisition et à la gestion des données au sein du PEPR ATLASEa	22
2.9.1.	Contexte et organisation	22
2.9.2.	Collecte et sélection	22
2.9.3.	Diffusion des données	22
2.9.4.	Défis.....	22
2.9.5.	Retours	23
2.10.	Valérie Chavagnac – Gestion des données d'échantillons de fluide de l'observatoire EMOS-Açores	23
2.10.1.	Contexte.....	23
2.10.2.	Dispositifs en place	24
2.10.3.	Échantillonnage.....	24
2.10.4.	Traitements	24
2.10.5.	Contraintes.....	24
2.11.	Agathe Laës-Huon – Gestion des données d'échantillonnage et analyse de fluide à l'Ifremer, Exemple des missions MoMAR 2010-2020 25	
2.11.1.	Sources et dispositifs	25
2.11.2.	Données.....	25
2.11.3.	Difficultés	25
2.12.	Ewan Pelleter et Marine Vernet – Gestion des données géologiques à l'UMR Géo-Océan.....	26
2.12.1.	Intégration et archivage	26
2.12.2.	Accès	26
2.12.3.	Limites actuelles.....	26
2.12.4.	Solutions envisagées	27
2.12.5.	Perspectives.....	27
2.13.	Isabelle Billy – Gestion des données de carottage, du terrain au laboratoire : Comment F.A.I.R ?.....	27
2.13.1.	Objectifs :	27
2.13.2.	Cycle de vie des données	28
2.13.3.	Gestion au laboratoire	28
2.13.4.	Conditions de succès	28
2.14.	Pierre Sans-Jofre – Gestion des données géologiques au MNHN..	29

2.14.1. Ensemble « Terre et Univers »	29
2.14.2. Base JACIM	29
2.15. Eva Moreno – Gestion des carottes sédimentaires au MNHN	30
2.16. Christophe Scheffer et Pierre-Yves Arnould : Ecosystèmes et stratégie de la gestion des échantillons et des données analytiques de l'INSU : Première phase dédiée au domaine Terre-Solide 31	
3. Synthèse des groupes de discussion – mercredi 17 septembre.....	31
3.1. LIMS et aspects réglementaires	32
3.1.1. Participants	32
3.1.2. Introduction à la thématique	32
3.1.3. Synthèse des échanges	32
3.2. Traçabilité de l'acquisition à la diffusion	33
3.2.1. Participants	33
3.2.2. Introduction à la thématique	34
3.2.3. Synthèse des échanges	34
3.3. Acquisition des échantillons en mer	35
3.3.1. Participants	35
3.3.2. Introduction à la thématique	36
3.3.3. Synthèse des échanges (Adélie Delacour et Bernard Dennielou).....	36
3.4. Gestion des données issues des échantillons en biologie	40
3.4.1. Participants	40
3.4.2. Introduction à la thématique	40
3.4.3. Synthèse des échanges (Brendan Hennebaut et Bérengère Husson).....	41
3.5. Gestion des données issues des échantillons en géologie	42
3.5.1. Participants	42
3.5.2. Introduction à la thématique	42
3.5.3. Synthèse des échanges (Marine Vernet).....	42
3.6. Gestion des données issues des échantillons en chimie.....	43
3.6.1. Participants	43
3.6.2. Introduction à la thématique	43
3.6.3. Synthèse des échanges	43
4. Points clés retenus dans les présentations et les discussions lors des deux journées	44
4.1.1. Points bloquants et limites rencontrées	44
4.1.2. Pistes d'améliorations évoquées	45
4.1.3. Conclusion	46

1. Introduction

L'atelier technique ODATIS, organisé dans le cadre du PEPR Grands Fonds Marins, a réuni les 16 et 17 septembre 2025 plus de 70 participants sur le site de l'Ifremer à Plouzané et en visioconférence. Il avait pour objectif d'aborder un enjeu central de la recherche en océanographie profonde : la gestion des données issues des échantillons collectés en mer.

Dans un contexte marqué par l'essor des données produites, la diversité des pratiques entre disciplines et l'évolution rapide des réglementations internationales (APA, BBNJ, CITES, etc.), cet atelier visait à dresser un état des lieux des outils existants, à partager des retours d'expériences et à identifier collectivement des pistes d'amélioration.

Cet événement a permis d'aborder les problématiques liées à l'acquisition, la traçabilité, l'archivage et la réutilisation des données et échantillons collectés lors des campagnes océanographiques, dans un contexte de forte hétérogénéité des pratiques et des outils entre disciplines et instituts.

1.1. Première journée – présentation et retours d'expériences

Au cours de la première journée, 21 intervenants ont présenté leurs travaux et leurs retours d'expérience couvrant ainsi l'ensemble du cycle de vie des données et des échantillons issus des grands fonds : de leur acquisition à bord, à leur conservation et leur traçabilité, jusqu'à leur analyse et leur diffusion dans des bases et référentiels spécialisés. L'intérêt majeur de cette journée a résidé dans la diversité des profils réunis : gestionnaires de données, chercheurs et responsables d'outils ont chacun apporté leur éclairage, permettant de comprendre simultanément les enjeux techniques, scientifiques et organisationnels. Cette confrontation des points de vue a contribué à offrir une vision d'ensemble des pratiques et des défis liés à la gestion des données issues d'échantillons océanographiques, et à souligner la nécessité d'approches coordonnées et complémentaires.

Au-delà de la diversité des approches, les échanges ont permis de mettre en lumière la richesse des initiatives déjà engagées pour améliorer la traçabilité, l'ouverture et la gestion FAIR des données, qu'il s'agisse d'outils logiciels, de plateformes partagées ou de projets collaboratifs. Mais ils ont aussi révélé la répétition de blocages communs : hétérogénéité des systèmes utilisés, lourdeur de la saisie, manque de moyens humains dédiés, difficultés d'interopérabilité ou encore complexité des procédures réglementaires.

Cette convergence des constats, observée à travers des disciplines et des instituts pourtant très différents, souligne l'importance d'avancer vers une structuration plus cohérente et des solutions partagées à l'échelle nationale et internationale.

1.2. Deuxième journée – groupes de discussions par thématiques

La seconde journée de l'atelier a été consacrée à des sessions de discussions organisées par thématiques, portant notamment sur les LIMS, l'IGSN, la traçabilité et la gestion des échantillons biologiques, géologiques et chimiques. Ces échanges, qui ont réuni chercheurs et gestionnaires de données issus de structures et disciplines variées, ont permis de confronter les pratiques, de partager des retours d'expérience et de comparer les difficultés rencontrées. Ils ont favorisé l'identification de bonnes pratiques et mis en lumière plusieurs points

de blocage communs, tout en faisant émerger des pistes concrètes pour améliorer les outils, les recommandations et les workflows.

À l'issue de ces discussions, trois thématiques se sont distinguées par la volonté de leurs participants de prolonger les travaux sous la forme de groupes de travail pérennes. Ces groupes se concentreront respectivement sur :

- Le développement d'outils embarqués pour améliorer l'acquisition et la transmission des données à bord.
- Un travail de centralisation des recommandations et outils en matière de traçabilité des échantillons et des données.
- L'accompagnement d'équipes de chimistes pour faciliter la publication de leurs jeux données dans les banques et catalogues officiels.

1.3. Profils des participants.

Sur les deux journées, 73 participants se sont inscrits via un formulaire qui nous a permis d'établir un profil général de l'assemblée.

Les questions du formulaire ont porté sur le domaine de travail des inscrits ainsi que sur leur niveau d'implication dans le processus de vie des données

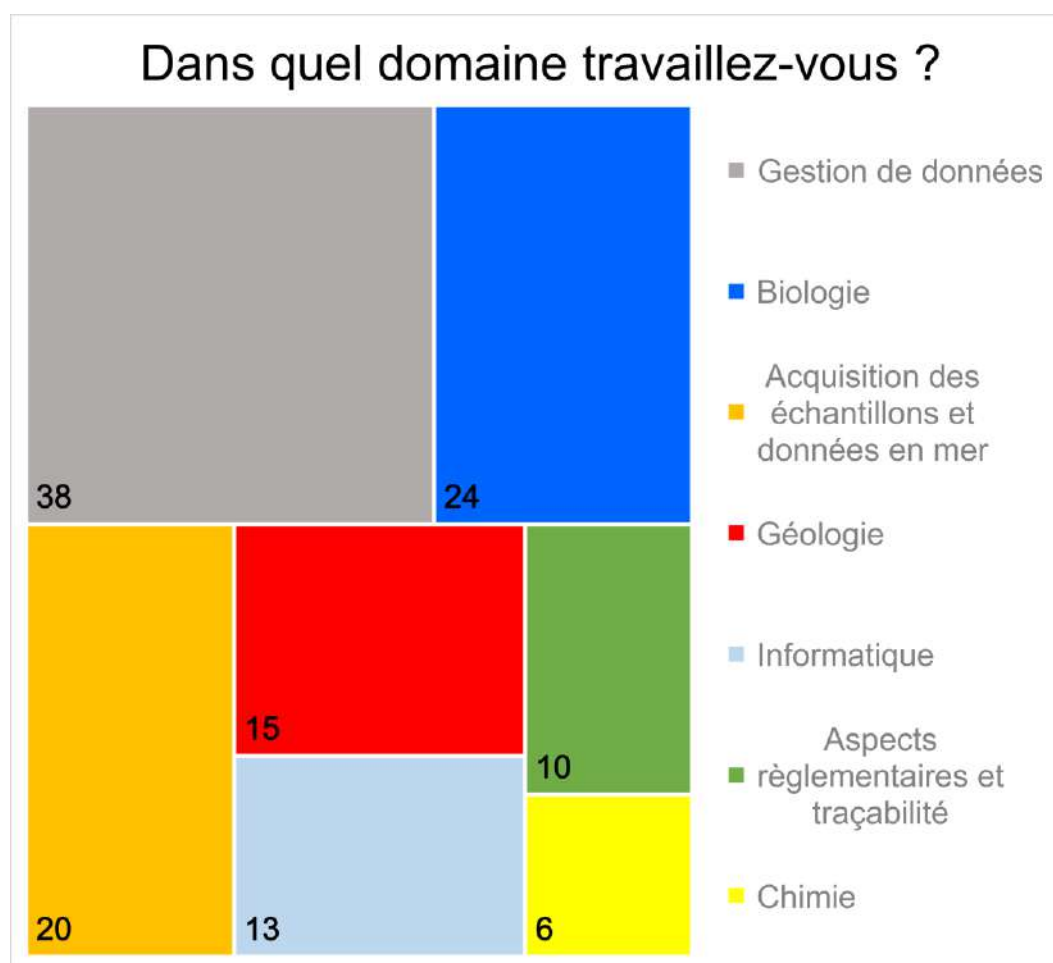


Figure 1 : Réponses à la question : Dans quel domaine travaillez-vous ? (Plusieurs réponses possibles)

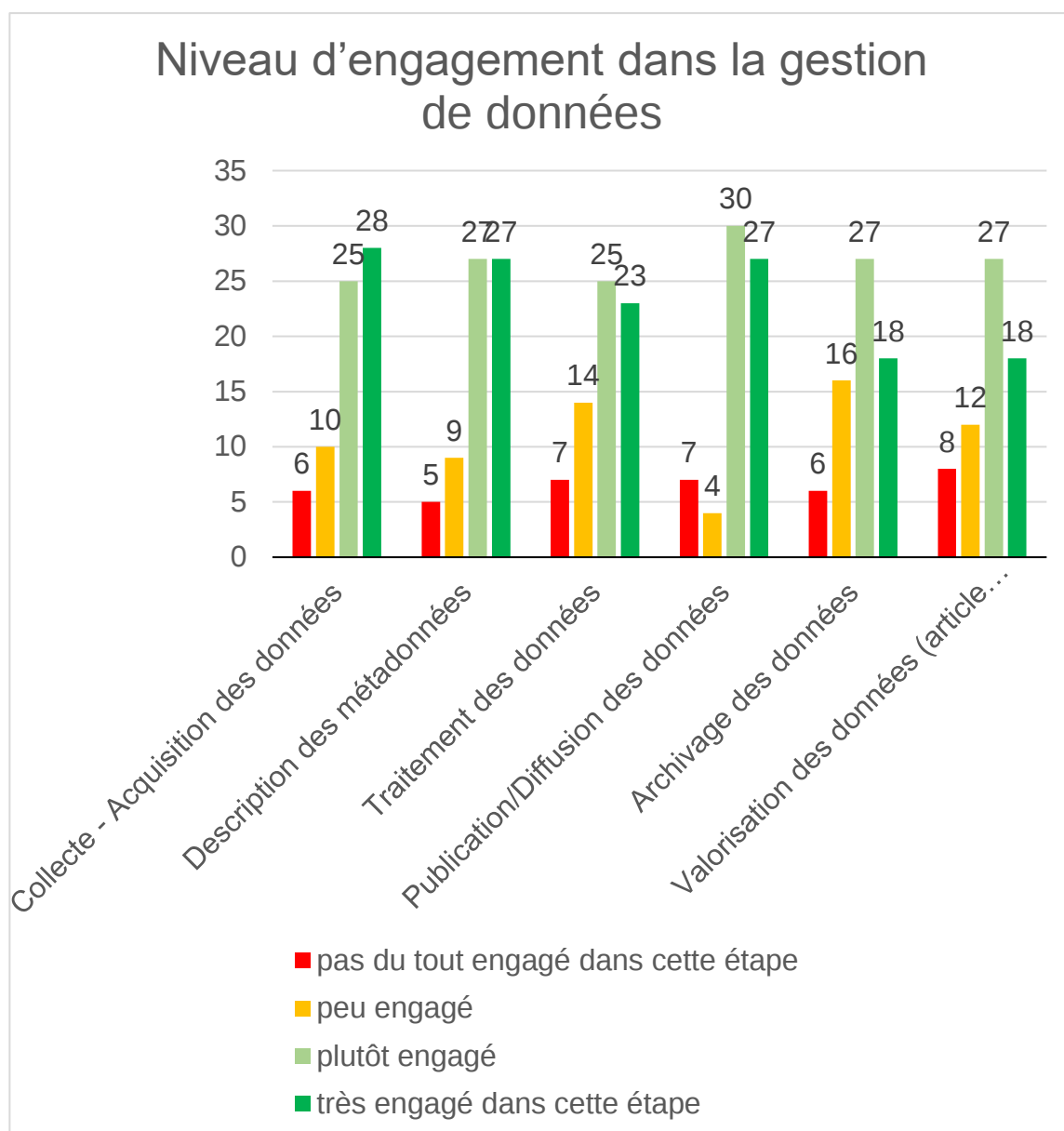


Figure 2 : Réponses à la question : Quelle est votre implication dans les différentes étapes du cycle de vie des données ?

Il apparaît ici que les trois disciplines scientifiques visées par cet atelier ont été représentées mais de manière fortement inégale.

En effet, 24 participants ont déclaré travailler dans le domaine de la biologie, 15 dans le domaine de la géologie et seulement 4 dans la chimie.

En revanche, on constate qu'une majorité des participants ont déclaré travailler dans le domaine de la gestion de données.

En ce qui concerne le niveau d'engagement dans les étapes de vie des données, une nette majorité des participants se déclarent plutôt engagés ou très engagés

dans les 6 étapes proposées (Collecte, description, traitement, publication/diffusion, archivage et valorisation).

2. Synthèse des présentations – mardi 16 septembre

Les supports de présentation sont disponibles sur la page dédiée à l'atelier sur le site web ODATIS : <https://www.odatis-ocean.fr/activites/ateliers-techniques/atelier-gfm-sept-2025>

2.1. Introduction à l'Atelier par Erwann Quimbert, Pierre-Marie Sarradin et Yacine Moufid

2.1.1. Le pôle Océan ODATIS

Présentation générale d'ODATIS et de Data Terra par Erwann Quimbert : description des centres de données (CDS), des centres d'expertise (CES) et de l'offre de services.

ODATIS constitue le pôle Océan de l'infrastructure de recherche Data Terra ayant pour objectif de proposer un accès généralisé à l'ensemble des produits et données de l'observation du système Terre.

Data Terra est soutenue par 34 organismes de recherche et ODATIS par 8 d'entre eux :

- CNRS
- CNES
- Ifremer
- IRD
- SHOM
- Universités Marines
- CEREMA

La cible principale d'ODATIS est la communauté scientifique mais le pôle travaille régulièrement avec des industriels ou des collectivités publiques.

ODATIS est structuré en Centres de Données et de Services (CDS) et en Consortium d'Expertise Scientifique (CES) et propose, outre l'accès à des données issues de l'observation de l'océan, une grande offre de services techniques et d'accompagnement.

Pour plus d'information, le lien du site web ODATIS : <https://www.odatis-ocean.fr/>

2.1.2. Le PEPR Grands Fonds Marins

Présentation du contexte et de l'équipe opérationnelle du PEPR par Pierre-Marie Sarradin.

Co-pilotage par :

- L'IRD
- CNRS

- Ifremer

2.1.3. Le volet Données et échantillons du PEPR

Présentation de la place centrale accordée à la gestion des données et des échantillons au sein de ce PEPR. Nécessité absolue de partage et de réutilisation des échantillons et données relatifs aux grands fonds compte tenu de la difficulté d'acquisition.

Besoin de faire évoluer les pratiques vers les recommandations FAIR.

Plusieurs moyens d'actions tels que les ateliers techniques.

2.1.4. Présentation de l'atelier technique en cours

Présentation du programme des deux journées de l'atelier ainsi que du profil des participants à partir des résultats issus du formulaire d'inscription.

2.2. Christine David-Beausire – Recommandation de la flotte océanographique française à propos de la gestion de données

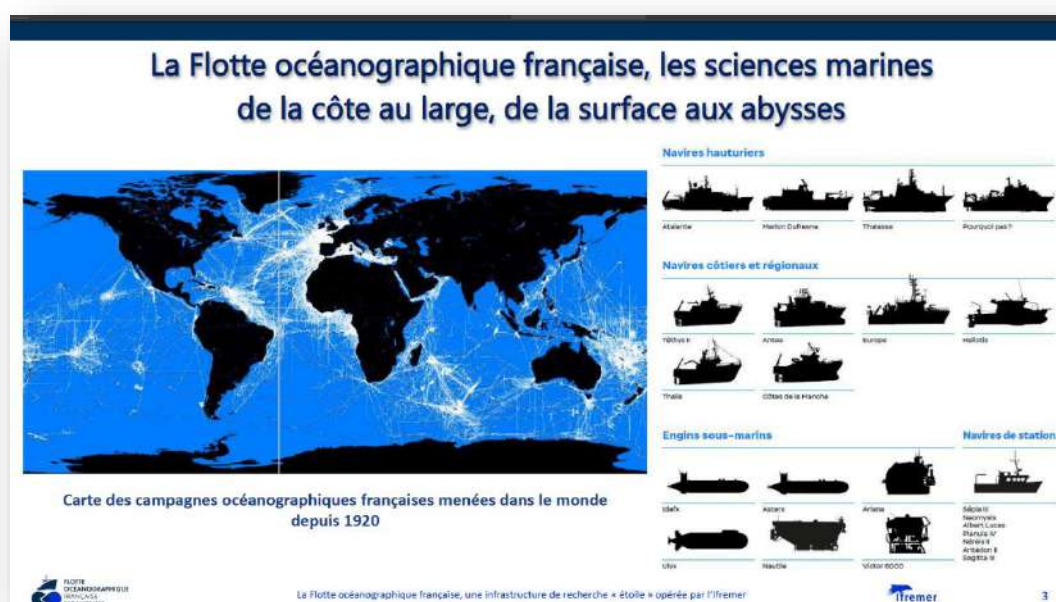


Figure 3 : Extrait de la présentation, Christine David-Beausire ([Lien vers la présentation](#))

Présentation de la FOF, infrastructure de recherche « étoile » directement rattachée au ministère de la Recherche. Elle regroupe différents navires et engins, et se distingue par son importance stratégique et son financement direct par le ministère (contrairement aux autres IR financées par les organismes gestionnaires). La FOF est confiée à un seul opérateur : l'Ifremer.

La flotte est principalement utilisée pour la recherche académique et la formation universitaire, mais elle contribue aussi à l'appui aux politiques publiques, à des

partenariats (marine nationale, archéologie sous-marine) et à des affrètements privés ou pour des organismes étrangers. En moyenne, elle assure 3 500 jours d'activité par an, dont plus des deux tiers pour la science et l'enseignement.

La FOF fournit aux équipes les moyens opérationnels pour accéder aux données et aux échantillons, via un appel à projets annuel. Les données occupent une place centrale dans le processus de recherche, désormais institutionnalisée au niveau national (Plan pour la science ouverte).

- Le cahier des charges des infrastructures de recherche impose :
- Le respect des principes FAIR,
- La mise en place d'un plan de gestion des données couvrant tout le cycle de vie,
- La favorisation de l'ouverture et du partage des données,
- L'interopérabilité et la souveraineté des solutions choisies,
- L'utilisation de licences compatibles.

Dans la convention avec Genavir, ce dernier est responsable de l'acquisition et de la qualification des données numériques (navigation, temps réel, capteurs, etc.). Tout est formalisé dans des processus qualité décrivant l'ensemble de la chaîne, de l'acquisition à la livraison au SISMER.

Des axes d'amélioration restent toutefois à travailler, identifiés lors d'un atelier FOF-ODATIS en juin 2023 :

- Intégration des navires de station,
- Prise en compte des échantillons et prélèvements (traçabilité et référencement),
- Définition d'un plan de gestion des données pour les capteurs de la FOF et pour les équipements scientifiques embarqués.

L'objectif est d'obtenir une vision coordonnée de l'ensemble et de relancer la dynamique engagée.

2.3. Olivier Soubigou – Outils d'acquisition à bord des navires de la flotte océanographique française

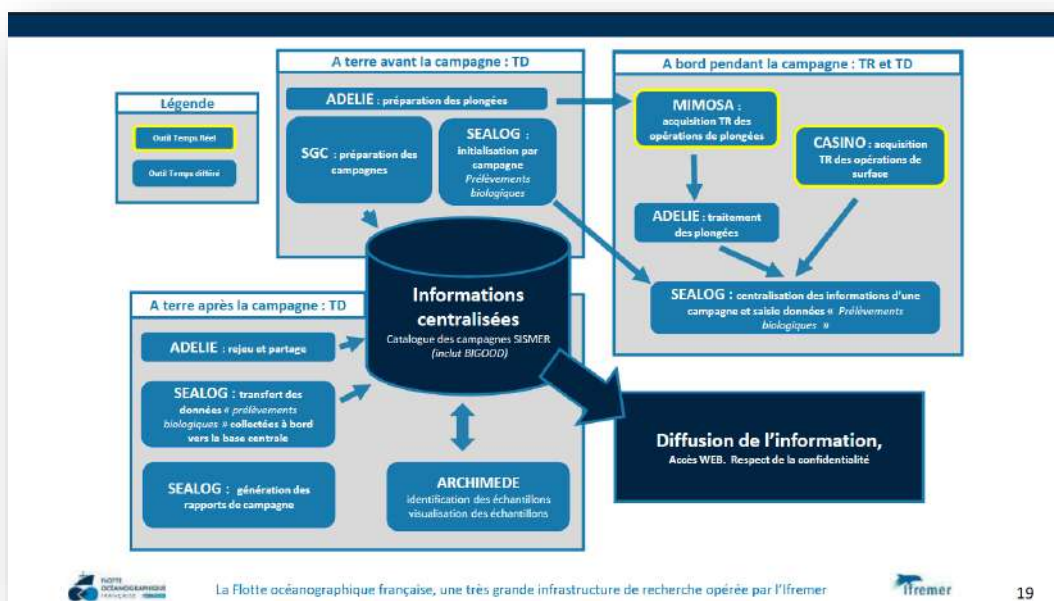


Figure 4 : Extrait de la présentation, Olivier Soubigou ([Lien vers la présentation](#))

Présentation des outils disponibles à bord pour l'acquisition et le suivi des données.

2.3.1. Suite logicielle TECHSAS

- Module d'acquisition : interface avec les capteurs, assurant l'archivage des données.
- Journal de bord : composante en cours de refonte (« CASINO »). Des entretiens avec les scientifiques ont été réalisés et les spécifications sont en cours de rédaction.
- Brique de visualisation : permet de suivre en temps réel ce qui se passe sur le navire via un tableau de bord personnalisable. Des postes dédiés sont disponibles dans les zones de travail, mais l'outil est également accessible sur tous les ordinateurs connectés au réseau de bord.
- Brique de planification : module récent permettant de planifier la route, de calculer le temps nécessaire entre deux stations, et d'intégrer l'impact carbone via la vitesse de transit.

2.3.2. Sealog

Outil dont la forme actuelle est en fin de vie. Il permet de créer une base de mission avant la campagne (informations de préparation), puis de gérer :

- Les opérations de surface (provenant de CASINO),
- Les opérations de plongée (données des engins sous-marins),
- Les mouillages.

Il produit un rapport chronologique des opérations, auquel peuvent être ajoutés les prélèvements et mesures associées. C'est un travail fastidieux, mais il garantit la bancarisation des données dans la base des campagnes gérée par le SISMER.

2.3.3. Ocean Data Tool

Projet développé par un freelance pour la NOAA, proposant un outil plus simple et moderne pour l'acquisition et la gestion des opérations.

- Préparation de mission (journal de bord en amont).
- Suivi de campagne avec une interface web conviviale permettant de générer les opérations dans un rapport chronologique.
- Phase d'exploitation avec extraction et filtrage des données.

Cet outil couvre une partie des fonctions de TECHSAS et de Sealog, avec un fonctionnement plus simple que CASINO.

2.3.4. Formations

Les chefs de mission peuvent demander des formations spécifiques à l'utilisation de ces outils.

2.4. Xavier Saint-Laurent – Modernisation du ROV Victor et projet DeepSea'Nnovation

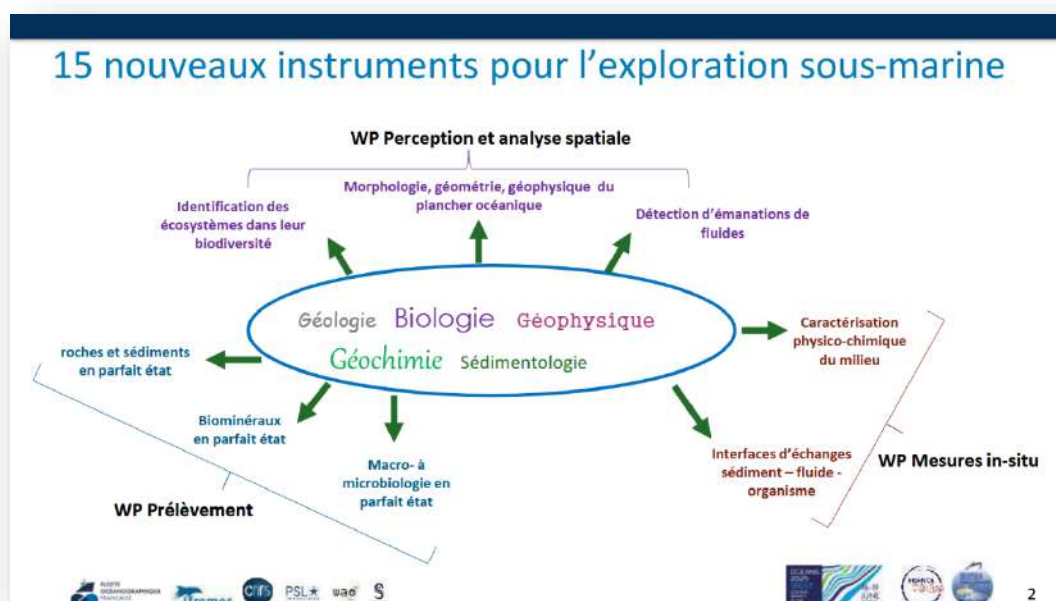


Figure 5 : Extrait de la présentation, Xavier Saint-Laurent ([Lien vers la présentation](#))

Le projet DEEPSEA'Nnovation a débuté en 2021 et se poursuivra jusqu'en 2029. Il vise à développer ou acquérir 15 nouveaux équipements pour les engins de la Flotte Océanographique Française (FOF). Ce programme accompagne également la modernisation du ROV Victor.

2.4.1. Le projet DEEPSEA'Nnovation

Ce projet est structuré en trois grands « work packages » (WP) :

1. Prélèvements
2. Perception spatiale
3. Mesures *in-situ*

Chaque WP comprend cinq instruments, couvrant un large spectre disciplinaire.

Concernant les prélèvements, des outils innovants sont développés pour prélever sans détruire les échantillons, ou pour réaliser des collectes qui étaient jusque-là impossibles.

- Une main de prélèvement a déjà été testée sur la mission CHEREEF afin de manipuler des coraux et autres organismes fragiles.
- Une foreuse à roches et un vibro-carottier sont en cours de développement.
- Un échantillonneur de larves et plancton est prêt à être testé.
- Des préleveurs d'ADN, pour petits et grands volumes, sont également en préparation.

Ces instruments généreront des données de natures très diverses, ce qui impose de leur associer des métadonnées précises afin d'assurer leur traçabilité et leur intérêt scientifique.

2.4.2. Modernisation du ROV Victor

Le ROV Victor est en cours de refonte (environ 80 % de ses composants sont renouvelés : caméras, bras, etc.). L'objectif est de tirer le meilleur parti de chaque plongée.

Trois types de données sont distingués :

- Celles produites par l'engin,
- Celles issues du navire (positionnement, navigation),
- Celles liées aux prélèvements (les moins bien maîtrisées, mais nécessitant un suivi rigoureux par métadonnées).

Un logiciel, MIMOSA v3, permettra de suivre toute la plongée et servira de cahier de quart numérique pour les opérateurs.

Un système de base de données baptisé DataPortal 2 est en cours de développement. Accessible sur le web, il centralisera les logiciels d'acquisition des données de Victor. La base PostgreSQL pourrait être déployée directement sur l'engin, permettant un accès quasi en temps réel aux données, y compris à bord.

2.4.3. Gestion des données

Le projet devra mettre en place un plan de gestion des données (PGD).

Des questions d'interopérabilité avec les autres logiciels embarqués restent à résoudre, afin d'éviter les ressaisies et pertes d'information.

2.5. L  na  ck Menot – Gestion des donn  es d’  chantillons biologiques au Laboratoire Environnement Profond (Ifremer)

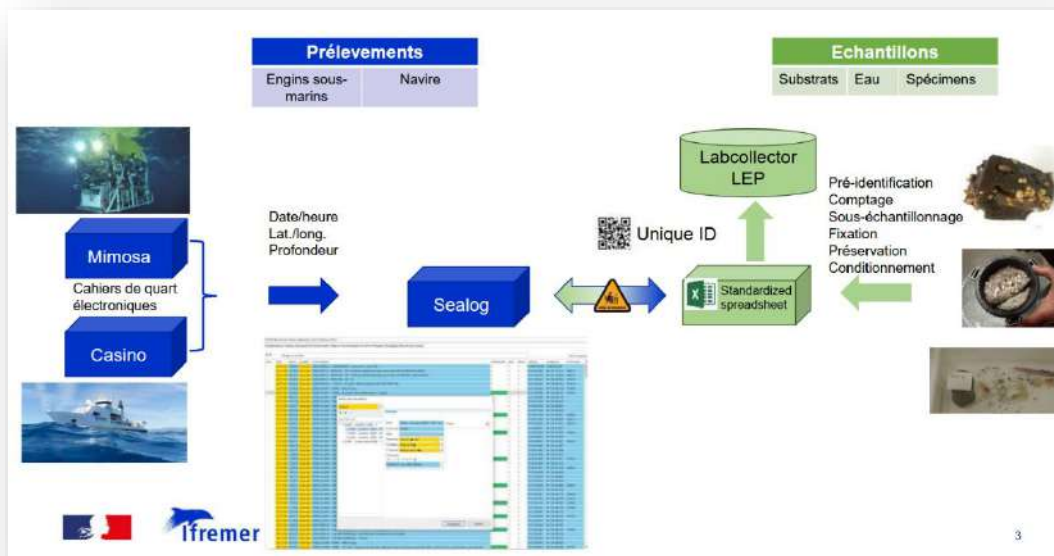


Figure 6 : Extrait de la pr  sentation, L  na  ck Menot ([Lien vers la pr  sentation](#))

2.5.1. Donn  es collect  es

Le c  ur des donn  es concerne les organismes benthiques. Pour chaque organisme collect  , diff  rentes informations sont relev  es :

- Nom de l'esp  ce,
- Nombre d'individus,
- Taille, sexe, poids, stade de d  veloppement,
- Sous-  chantillons pour analyses (ADN, isotopes pour   tudier le r  gime alimentaire, lipides, etc.).

Les   chantillons proviennent de substrats durs, meubles, d'eau, ou d'organismes collect  s directement sur le fond marin. Certains sont associ  s    des pr  levements effectu  s depuis la surface ou par des engins sous-marins.

L'objectif principal est d'assurer une tra  cabilit   compl  te, depuis le pr  l  vement jusqu'   la donn  e produite.

2.5.2. Cha  ne de gestion des donn  es

- CASINO pour les op  rations de surface.
- MIMOSA pour les op  rations de plong  e.
- Les donn  es de CASINO et MIMOSA sont int  gr  es dans Sealog, qui sert de registre des op  rations.
- Les listes d'  chantillons sont g  r  es    bord avec LabCollector (LIMS). Les informations de pr  -identification, comptage, fixation, etc., sont saisies dans Excel puis import  es dans LabCollector.

- Chaque échantillon se voit désormais attribuer un numéro unique, intégré dans LabCollector et, lorsque c'est possible, dans Sealog, afin de maintenir le lien entre les systèmes.

2.5.3. Intégration et diffusion

- Les données issues de Sealog sont intégrées par le SISMER dans BIGOOD (puis vers le catalogue des campagnes) et également dans MORSE. L'objectif avec le numéro unique est de pouvoir aussi intégrer les données de LabCollector dans BIGOOD.
- Elles dépendent du référentiel WoRMS, utilisé comme base taxonomique.
- Elles alimentent également GBIF et OBIS, mais pour l'instant ce processus est manuel et devrait être automatisé à terme.

2.5.4. Limites et difficultés

- La saisie est chronophage et fastidieuse, sans personnel dédié à bord.
- Les pratiques diffèrent selon les équipes et les objectifs scientifiques, ce qui entre en contradiction avec la rigidité des bases de données et de leurs référentiels.
- Résultat : des référentiels flous, sources d'erreurs et de manque d'homogénéité.
- Les données d'imagerie prennent une place croissante (les informations biologiques sont de plus en plus extraites des images plutôt que des prélèvements), mais aucune solution FAIR n'est encore disponible. Le CES imagerie benthique d'ODATIS devrait apporter des solutions.

2.5.5. Discussion

Un identifiant unique, généré par Brendan Hennebaut (QR code), est utilisé par le laboratoire Environnement profond. Mais toutes les équipes ne le connaissent pas ou ne se l'approprient pas.

Les changements de noms d'espèces ne sont pas toujours signalés dans les bases, ce qui complique l'identification.

En général, un fragment de l'échantillon est conservé et le devenir de l'échantillon (utilisation complète, transfert) est indiqué dans LabCollector.

Des questions demeurent sur les référentiels : peut-on se baser sur les listes Sea Data Net ? Des mappings sont prévus, avec l'idée de rendre les systèmes plus arborescents et flexibles.

2.6. François Dusoulie – L'accès aux données des collections naturalistes dans le cadre de l'infrastructure Récolnat

 **Récolnat** 

L'accès aux données des collections naturalistes
Atelier ODATIS, PEPR GFM, Ifremer (Plouzané), 16 septembre 2025

Chiffres clés

Corpus de **10,9 millions** dont 10 millions d'herbiers !
Ca. **400 000** types et figurés, **360** collections de **92** institutions
Ca. **15 000** utilisateurs/an utilisant le portail des collections
Ca. **1 200** publications/an utilisant l'IR
ca. **450** personnes impliquées



Figure 7 : Extrait de la présentation, François Dusoulie ([Lien vers la présentation](#))

L'Infrastructure de Recherche Récolnat (Réseau national des collections naturalistes) vise à rendre accessibles les données issues des collections naturalistes françaises.

2.6.1. Objectifs et périmètre

- Centraliser et rendre accessibles les spécimens physiques et leurs données associées.
- Constituer des réservoirs de données permettant de remonter le temps grâce aux collections (ex. pollens fixés sur des spécimens).
- Ne pas inclure les formes domestiquées, qui relèvent de l'INRAE.

Récolnat est une infrastructure du Système Terre et Environnement, tous domaines confondus. Elle rassemble les collections et favorise la concertation sur les pratiques taxonomiques.

2.6.2. Chiffres et outils

- 10 millions de spécimens intégrés à Récolnat, sur environ 130 millions estimés en France.
- Données accessibles via un portail web permettant la consultation et l'ajout de spécimens.
- Utilisation de Darwin Core (DwC) pour garantir l'interopérabilité.
- Mise en place du standard MIDS (informations minimales associées aux spécimens numérisés), même si la complétude reste très variable.

- Gestion de spécimens numériques « étendus » : possibilité de lier les spécimens à d'autres sources de données.

2.6.3. Gestion taxonomique

- Conservation du nom verbatim (nom tel qu'indiqué sur l'étiquette, même avec faute de frappe).
- Intégration parallèle du nom normalisé via WoRMS ou TaxRef.
- Conservation de l'historique des déterminations (re-identifications par des taxonomistes).
- Possibilité d'indiquer lorsqu'un taxon n'est plus à jour, par exemple lorsqu'un taxon unique est divisé en plusieurs.

2.6.4. Enjeux actuels

Construction d'un miroir européen via le projet DiSSCo (European Research Infrastructure Consortium), qui rassemblera 23 pays sous un portail unique (« DiSSCover »).

Chantiers en cours :

- Intégration massive de données (notamment géologiques, en lien avec l'IR REGEF),
- Élaboration de descripteurs communs,
- Création d'un corpus 3D,
- Réduction de la « dette de description » des collections,
- Accompagnement humain pour les gestionnaires de collections.
- Nouvelle phase d'intérêt scientifique prévue pour 2026–2030.

2.6.5. Limites

- Le standard *Darwin Core* manque de descripteurs spécifiques pour la géologie, hormis certains communs avec la paléontologie.
- Trop de descripteurs peut décourager leur remplissage.
- Possibilité de signaler les bases géologiques dans Récolnat pour indiquer l'existence d'échantillons.
- Dans la logique du spécimen étendu, un échantillon peut être relié directement à la campagne de collecte.

2.7. Brendan Hennebaut – MORSE, entre enjeux réglementaires et traçabilité des échantillons

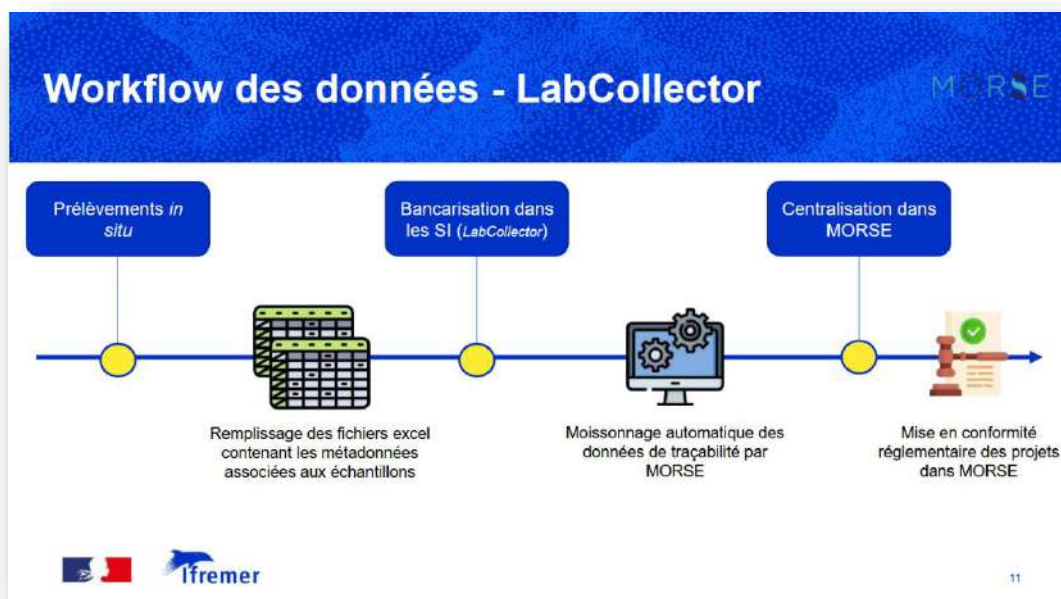


Figure 8 : Extrait de la présentation, Brendan Hennebaut ([Lien vers la présentation](#))

2.7.1. Objectifs

Le projet MORSE vise à assurer la traçabilité des échantillons biologiques de l'Ifremer et à répondre aux réglementations (APA, BBNJ). Initialement limité aux cas soumis à l'APA, il a été étendu à l'ensemble des réglementations sur les échantillons biologiques.

2.7.2. Fonctionnement

Mise en place d'un système d'information capable de tracer tous les échantillons biologiques stockés à l'Ifremer.

Moissonnage des informations depuis :

- Les instances LIMS de LabCollector,
- La base projet SAP,
- Et à terme, d'autres SI (base des campagnes, Quadriga, SIH, etc.).

Chaîne de gestion : prélèvement → fichiers Excel → LabCollector → moissonnage par MORSE → mise en conformité réglementaire.

2.7.3. Métadonnées et volumes

Traçabilité assurée via des métadonnées : localisation, stockage, taxonomie, documents réglementaires associés.

114 000 échantillons de référence sont déjà recensés, représentant environ 457 000 échantillons physiques (ex. plusieurs tubes par lot).

Grande diversité : ADN, tissus, spécimens entiers, etc.

2.7.4. Enjeux

- Améliorer l'interopérabilité avec les SI internes.
- Mieux relier les échantillons aux données de séquençage (GT METIS).
- Étendre MORSE à d'autres organismes de recherche.
- Créer des liens avec les structures de données externes (Récolnat, GBIF, etc.), compatibles avec le standard Darwin Core.
- Prendre en compte l'évolution des réglementations RGPD
- Les noms des opérateurs de collecte peuvent être anonymisés (par ex. remplacés par l'organisme), comme le pratique Récolnat.

La refonte du système Quadrige intègrera cette anonymisation, avec la possibilité de conserver les informations personnelles en back-office.

2.8. Pauline Auffret et Valérie Cueff-Gauchard – Traçabilité des données séquencées, exemple des données du Laboratoire de Microbiologie des Environnements Extrêmes



Figure 9 : Extrait de la présentation, Pauline Auffret ([Lien vers la présentation](#))

2.8.1. Contexte

L'UMR Biologie et Ecologie des Ecosystèmes Profonds (BEEP) travaille sur divers écosystèmes, notamment en écologie microbienne : identification des espèces présentes, étude des métabolismes, analyses sur substrats variés (cheminées hydrothermales, sédiments, espèces animales, fluides, etc.).

2.8.2. Prélèvements

Réalisés à partir du navire (bouteilles Niskin, carottiers, etc.) ou via les sous-marins de la Flotte Océanographique Française (préleveurs d'eau, bouteilles titane, carottiers, boîtes étanches, aspirateur à faune).

Outils spécifiques développés :

- PERISCOPE (maintien de la pression),
- IBIS (prélèvement de fluide hydrothermal sans dépressurisation),
- Systèmes de fixation in situ des animaux.

2.8.3. Gestion des échantillons

- Traitements différenciés selon le type de substrat.
- Sous-échantillons conditionnés selon les analyses prévues.
- Utilisation de LabCollector pour centraliser la localisation et le devenir des échantillons.

2.8.4. Exemples d'analyses

- Isolement de souches en culture : archivage soit à l'Ifremer (via LabCollector), soit à l'UBO (logiciel BioloMICS).
- Analyses moléculaires (ADN, ARN, protéines) :
- Approches ciblées (gène marqueur).
- Approches globales (métagénomique, metabarcoding).
- Les données brutes sont stockées et traitées localement, puis publiées dans des bases comme GenBank (mais sans traçabilité idéale).
- Besoin d'une gestion centralisée (Geneious en réseau, module dédié dans LabCollector, DataRef...).

2.8.5. Données de séquençage haut débit

- Très volumineuses, stockées sur Datarmor.
- Données majoritairement au format FASTQ compressé (standard mais pauvre en métadonnées de contexte).

2.8.6. Chaîne bio-informatique (SeBiMER)

- Couvre le cycle de la donnée : ADN environnemental, transcriptomique, assemblage de génomes, métagénomique.
- Données traitées et cataloguées dans Sextant, décrites dans Athena (spécifique génomique), puis publiées dans ENA (banque européenne).
- Le programme EGIDE cherche à automatiser la récupération des métadonnées (via DataRef, fichiers FASTQ, etc.).
- Le GT METIS vise à améliorer la liaison avec MORSE.
- Athena est un pipeline libre, mais qui doit être consolidé pour couvrir l'ensemble des types de données de séquençage.

2.9. Annie Lebreton – Challenges liés à l'acquisition et à la gestion des données au sein du PEPR ATLASEa

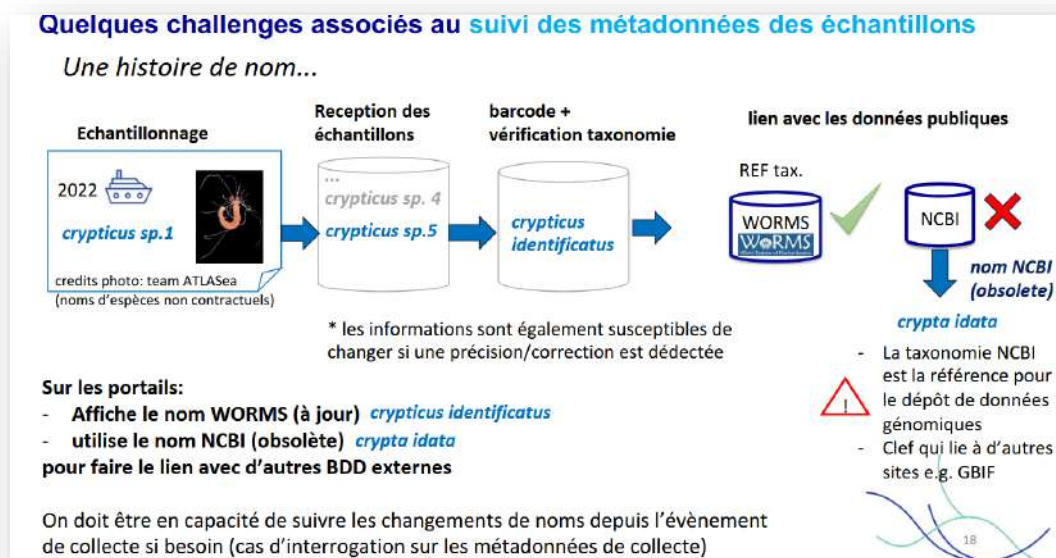


Figure 10 : Extrait de la présentation, Annie Lebreton ([Lien vers la présentation](#))

2.9.1. Contexte et organisation

Le PEPR ATLASEa regroupe quatre projets ciblés, centrés sur la collecte d'échantillons et la production de données génomiques.

2.9.2. Collecte et sélection

- Collecte d'échantillons lors de missions variées, mise en collection au Muséum, prises de photos.
- Sélection d'espèces non encore séquencées (suivi via l'outil GoaT).
- Alimentation régulière de GoaT et dépôts dans l'ENA dès que possible.

2.9.3. Diffusion des données

Portails :

- Science MNHN (expéditions, spécimens)
- Tracker (outil interne développé par SeBiMER)
- Portail public de suivi, Marine Genome Portal (analyse des génomes, production ATLASEa + bases publiques).

Chaque projet dispose d'une base encore en maturation, bien que la production de données ait débuté.

2.9.4. Défis

- Gestion des mises à jour issues de NCBI et ENA.

- Collecte des métadonnées sur le terrain : temps limité, nombreux cas particuliers, outils devant rester adaptables.
- Problèmes d'alignement taxonomique (parfois WoRMS ≠ NCBI).
- Nécessité de synchronisation et de fusion des bases, ainsi que de former les équipes.
- Développement d'un outil pour compiler et valider les données en mission.

2.9.5. Retours

- Importance de tout tracer et documenter, car une donnée manquante peut rendre un échantillon inexploitable.
- Pérennité assurée par l'interopérabilité entre bases génomiques (Europe/USA/Japon).
- Projet Catalogue of Life (GBIF) pour construire une taxonomie unifiée (branches issues de NCBI et de WoRMS).

2.10. Valérie Chavagnac – Gestion des données d'échantillons de fluide de l'observatoire EMOS-Açores



Figure 11 : Extrait de la présentation, Valérie Chavagnac ([Lien vers la présentation](#))

2.10.1. Contexte

Le site hydrothermal Lucky Strike accueille depuis 2010 un observatoire, fruit d'une coopération entre l'Ifremer, le CNRS et l'Université des Açores. L'objectif est de mieux comprendre :

- Les interactions entre fluides hydrothermaux et écosystèmes profonds,
- Les liens entre processus géologiques, biologiques et physiques,
- L'évolution de ces interactions à différentes échelles de temps.

2.10.2. Dispositifs en place

Deux stations SEAMON :

- Est : axée sur la biochimie et la physique,
- Ouest : centrée sur la sismique et la déformation du plancher océanique.

Transmission des données jusqu'en 2023 via une bouée d'archivage.

Évolutions et ajustements réguliers des stations.

2.10.3. Échantillonnage

- Réalisé lors des campagnes de relevage.
- Matériels adaptés aux fortes températures et à la corrosion.
- Désormais possible de réaliser des prélèvements automatiques et séquencés.

Pour contextualiser les échantillons, recours aux logiciels des sous-marins (MIMOSA, Sealog, etc.).

2.10.4. Traitements

- À bord : extraction des gaz et fluides des bouteilles, mesures (volume, pH, filtration, etc.).
- À terre : analyses avancées en salle blanche (géochimie, isotopes).

Données disponibles dans les rapports de campagnes ou sur Seanoe.

2.10.5. Contraintes

- Coordination avec la FOF indispensable pour optimiser les protocoles.
- Contraintes administratives liées à l'usage de produits chimiques.
- Peu de spécialistes des fluides hydrothermaux → stockage dispersé.
- Nécessité d'ouvrir l'accès aux échantillons, compte tenu du coût élevé de leur collecte.
- Réflexion sur la mise en place d'une base nationale des échantillons de fluides, inspirée de la cybercarothèque, en lien avec l'IR REGEF.

2.11. Agathe Laës-Huon – Gestion des données d'échantillonnage et analyse de fluide à l'Ifremer, Exemple des missions MoMAR 2010-2020

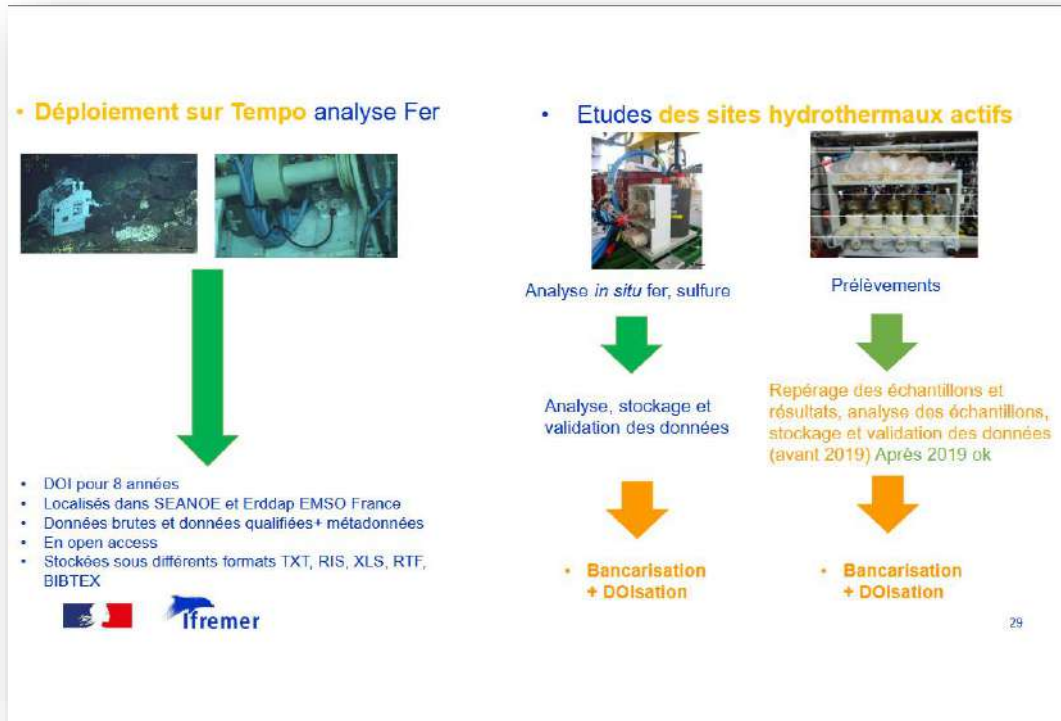


Figure 12 : Extrait de la présentation, Agathe Laës-Huon ([Lien vers la présentation](#))

2.11.1. Sources et dispositifs

- Missions Momarsat et module écologique TEMPO.
- Sites actifs, notamment la moulière observée au pied de la cheminée « Tour Eiffel ».
- Instruments utilisés :
 - CHEMINI : analyse du fer et des sulfures (in situ et en grands fonds).
 - PEPITO (23 poches de prélèvement).
 - PIF (30 seringues de prélèvement).
 - Mesures : oxygène dissous, électrodes soufre, pH, température.

2.11.2. Données

- Données TEMPO disponibles sur Seanoe et ERDDAP EMSO.
- Métadonnées et données archivées localement sur disque Ifremer (pas encore de bancarisation officielle).
- Depuis 2021 : tableaux de prélèvements gérés dans LabCollector.

2.11.3. Difficultés

- Traçabilité compliquée entre cahier de quart, Sealog et LabCollector.

- Questions sur la propriété des données : analyses stockées de façon hétérogène, peu de dépôts dans des bases officielles.
- Travaux en cours sur la « parentalité » des données (lien entre données brutes et analyses).
- Besoin d'un outil générique et d'un référent en laboratoire/SISMER pour harmoniser les pratiques.

2.12. Ewan Pelleter et Marine Vernet – Gestion des données géologiques à l'UMR Géo-Océan

1. Pas de workflow **continu** d'outils pour saisir, garder, consulter, vérifier, exporter les informations sur les échantillons

2. Pas d'**exhaustivité des données en base**, pas de captation systématique dès le bord, pas de vocabulaires communs entre les outils

3. Pas de système de génération d'étiquettes utilisable à bord
Archimède permet un étiquetage post campagne

4. Export des métadonnées « échantillons »
rapport de fin de mission : Sealog uniquement

Figure 13 : Extrait de la présentation, Ewan Pelleter et Marine Vernet ([Lien vers la présentation](#))

2.12.1. Intégration et archivage

- BIGOOD intègre les échantillons biologiques, géologiques et de fluides.
- Les échantillons géologiques sont étiquetés et archivés au CREAM.
- Les informations sont bancarisées via Archimède, puis intégrées à BIGOOD.
- Tous les échantillons géologiques stockés au CREAM reçoivent un PID IGSN.

2.12.2. Accès

Informations disponibles via : catalogue des campagnes, portail des échantillons, cybercarothèque, ODATIS, SeaDataNet (pour les analyses).

2.12.3. Limites actuelles

- Workflow incomplet : saisies multiples, risques d'erreurs.
- Certaines informations bancarisées automatiquement (SISMER), d'autres non (opérations, engins).

- Multiplication des saisies (MIMOSA, CASINO, ADELI) = redondances et erreurs.
- Contrairement à la NOAA, les prélèvements ne sont pas systématiquement tagués.
- Conversion Sealog → BIGOOD également source d'erreurs.

2.12.4. Solutions envisagées

- Vérifications à terre via Archimède (Sigistes à bord = génération de tableaux SIG puis import).
- Import en masse possible.
- Archimède offre visualisation, saisie, mises à jour (module cartographique).

Limites : logiciel lourd, pas encore de rapprochement avec Sealog (ex. mouillages), nécessite d'avoir une personne en charge du SIG à bord.

2.12.5. Perspectives

Progresser vers des logiciels embarqués permettant de taguer en amont les prélèvements avec des référentiels validés.

Idee d'un outil unique intermédiaire (Sealog enrichi ou Archimède enrichi), utilisable en mer et à terre.

2.13. Isabelle Billy – Gestion des données de carottage, du terrain au laboratoire : Comment F.A.I.R ?

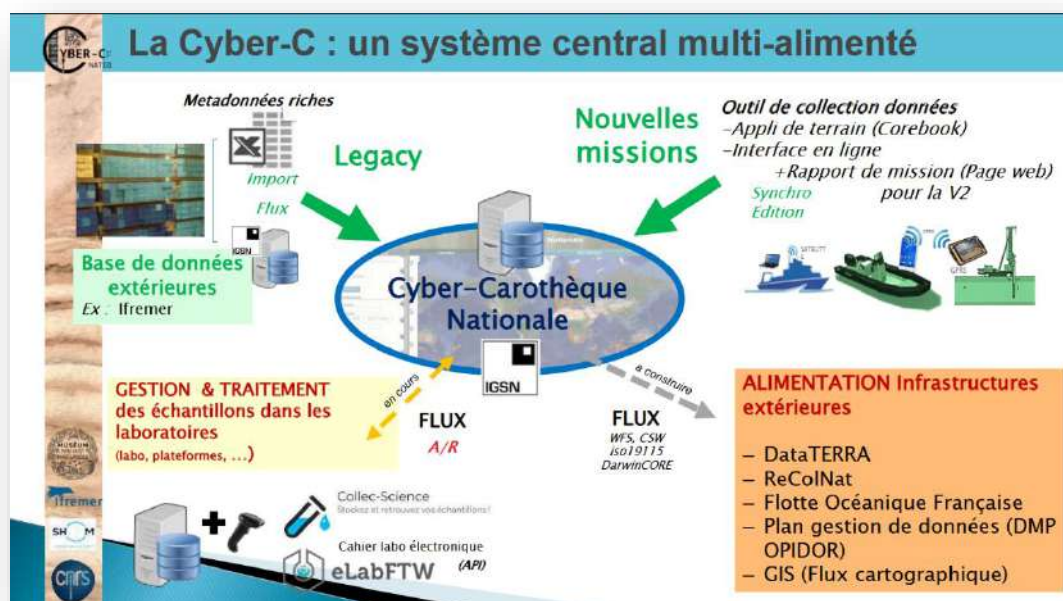


Figure 14 : Extrait de la présentation, Isabelle Billy ([Lien vers la présentation](#))

La cybercarothèque est une base de données et un portail web pour bancariser les carottes sédimentaires.

2.13.1. Objectifs :

- Mise en base et attribution d'IGSN,

- Gestion et traçabilité des carottes,
- Mise à disposition via un portail.

2.13.2. Cycle de vie des données

Prélèvements → base centrale → diffusion.

Alimentation par :

- Inventaires historiques (« legacy » via Excel),
- Moissonnage (ex. BIGOOD),
- Saisies directes à bord (appli de terrain CoreBook, à adapter pour les missions marines).

2.13.3. Gestion au laboratoire

- Utilisation des IGSN pour assurer les liens avec la base centrale (ex. Collec-Science).
- Flux futurs vers Récolnat, la FOF, etc.
- Application mobile avec : GPS, gestion des tronçons, travail hors ligne (synchronisation au retour), vocabulaire contrôlé (listes déroulantes).
- Génération de rapports PDF et Word.

2.13.4. Conditions de succès

- Adhésion et implication du chef de mission.
- Premier maillon de la chaîne de sauvegarde des données.

2.14. Pierre Sans-Jofre – Gestion des données géologiques au MNHN

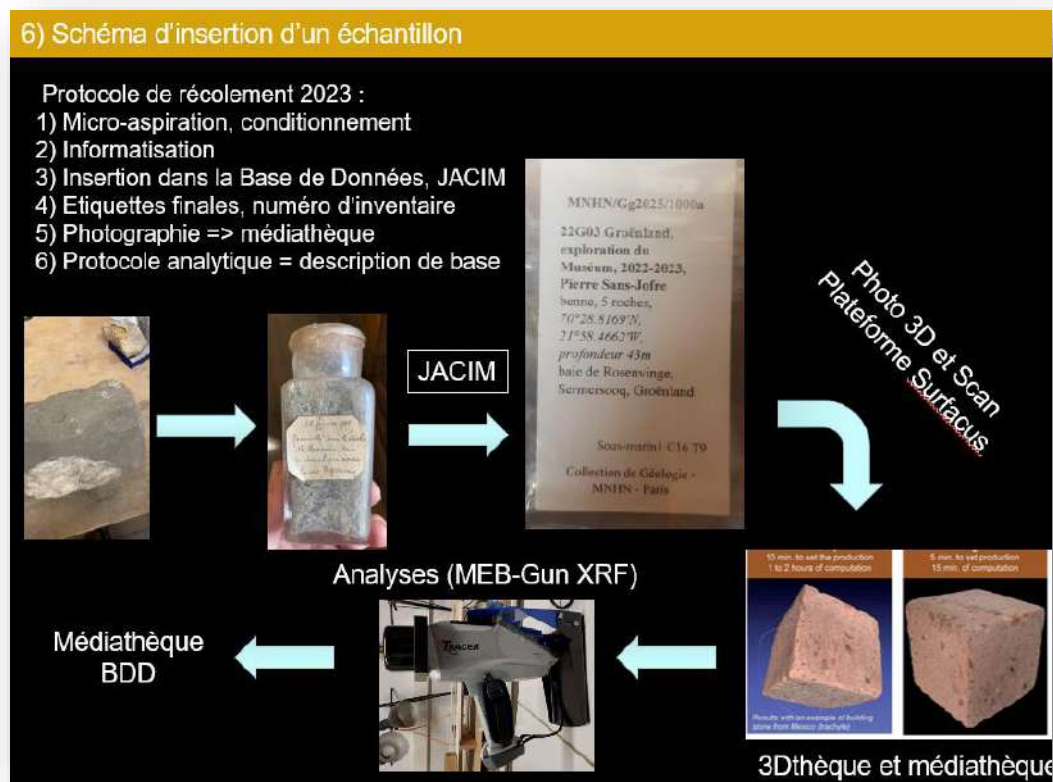


Figure 15 : Extrait de la présentation, Pierre Sans-Jofre ([Lien vers la présentation](#))

Présentation de l'organisation et la gestion des collections géologiques au Muséum national d'Histoire naturelle.

2.14.1. Ensemble « Terre et Univers »

Comprend cinq collections principales, dont celle consacrée aux sédiments et roches océaniques, pilotée par Eva Moreno.

Ces collections rassemblent des échantillons de natures diverses, stockés dans des conditions variables selon leur type et leur sensibilité.

2.14.2. Base JACIM

Utilisée pour la saisie et la consultation des données. Elle permet également l'import de fichiers Excel. Cependant, des limites importantes subsistent : faible interopérabilité avec les autres systèmes, ergonomie perfectible et accessibilité limitée pour les utilisateurs externes. Une nouvelle version de JACIM est en cours de développement pour répondre à ces enjeux.

L'accès aux collections peut également se faire via le portail MUSE, qui centralise les données naturalistes du Muséum. L'objectif est de faciliter la visibilité et l'exploitation des échantillons par la communauté scientifique.

2.15.Eva Moreno – Gestion des carottes sédimentaires au MNHN

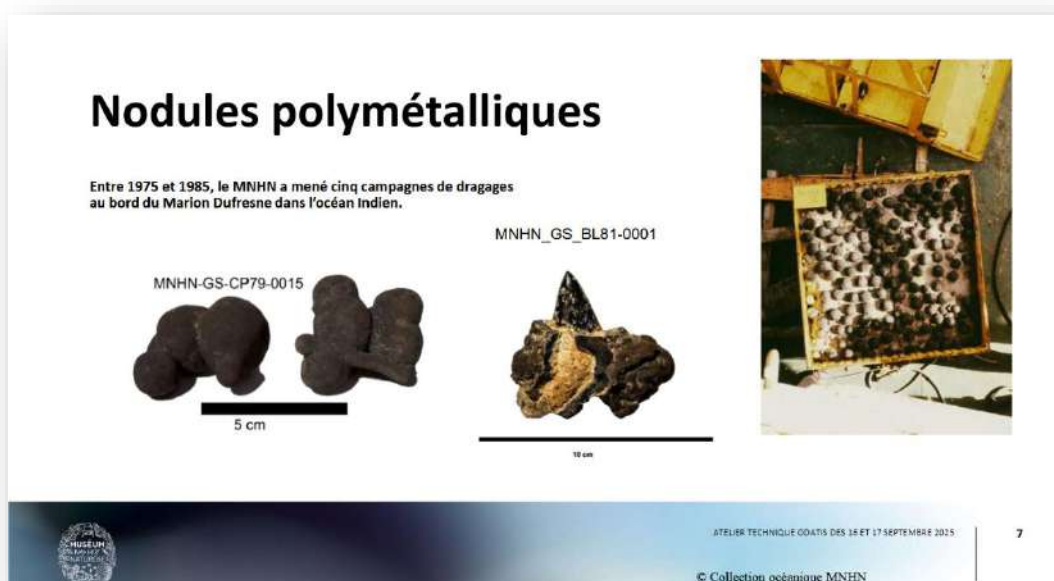


Figure 16 : Extrait de la présentation, Eva Moreno ([Lien vers la présentation](#))

Eva Moreno est responsable des collections de carottes sédimentaires et d'autres échantillons marins collectés depuis 1972. Ces collections comprennent des carottes, nodules, dragages et bénéficient d'une forte représentation pour l'océan Indien, grâce notamment aux campagnes menées par le Marion Dufresne et dans le cadre des TAAF. Elles s'accompagnent également d'une collection d'images et de films associés aux missions de prélèvement.

Historiquement, les données étaient gérées dans la base GEOCORES, avant leur transfert vers la base du Muséum. Aujourd'hui, la gestion repose sur plusieurs outils :

- JACIM pour la saisie et la consultation,
- COLHELPER pour la gestion des prêts,
- MUSE pour l'accès en ligne et la diffusion publique.

Malgré ces dispositifs, la gestion des demandes et des prêts reste encore perfectible, avec des processus parfois lourds ou hétérogènes. L'enjeu est d'améliorer l'accessibilité, la traçabilité et la valorisation de ces collections uniques, afin de renforcer leur intégration dans les grandes infrastructures nationales et internationales.

2.16.Christophe Scheffer et Pierre-Yves Arnould : Ecosystèmes et stratégie de la gestion des échantillons et des données analytiques de l'INSU : Première phase dédiée au domaine Terre-Solide

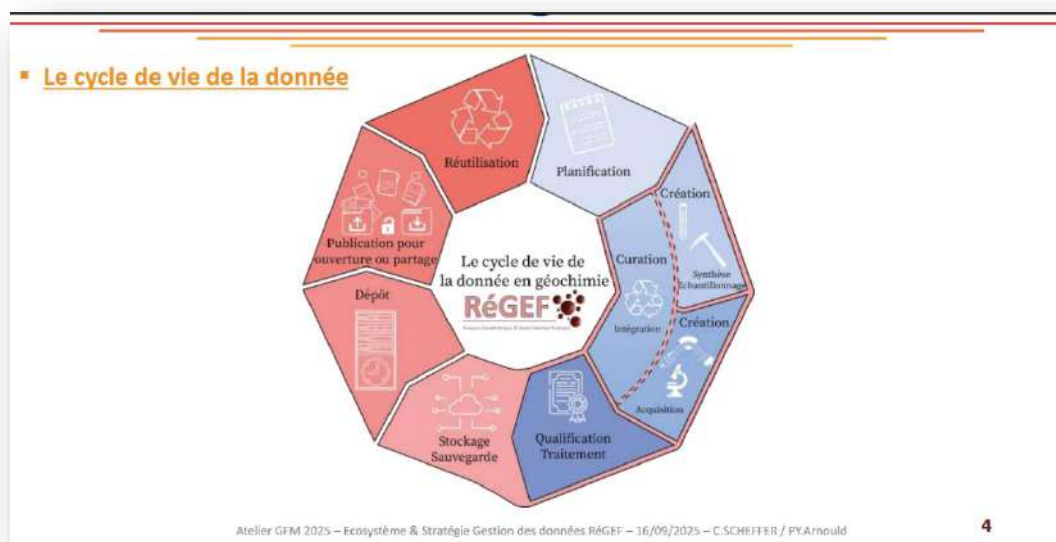


Figure 17 : Extrait de la présentation, Christophe Scheffer et Pierre-Yves Arnould ([Lien vers la présentation](#))

L'IR REGEF s'appuie sur 12 réseaux par disciplines et instruments, couvrant l'ensemble du cycle de vie des données en géochimie.

Depuis juin 2023, elle a été missionnée par l'INSU pour participer à un groupe de travail « Échantillons et données », chargé de construire et définir les bonnes pratiques. Ce groupe est divisé en cinq sous-groupes, dont trois impliquent directement REGEF :

- Métadonnées
- Qualité des données
- Valorisation des données

Les travaux s'appuient sur le système américain IGSN (SESAR) pour définir des descripteurs standards. L'objectif est de proposer un dispositif utilisable par toute la communauté française.

Le groupe réunit une quinzaine de personnes issues de plusieurs instituts. Une réunion est prévue le 10 février 2026 à Paris afin d'harmoniser les vocabulaires et descripteurs, avec la volonté d'être précis sans être trop complexes.

L'interface RESANA facilite la mise en place de groupes de travail trans-instituts. Un service national de numérotation universelle des échantillons est en cours de définition. Le cahier des charges prévoit la possibilité d'attribuer des IGSN directement sur le terrain.

3. Synthèse des groupes de discussion – mercredi 17 septembre

3.1. LIMS et aspects réglementaires

3.1.1. Participants

- Pauline Chauvet (CNRS)
- Valérie Cueff-Gauchard (Ifremer)
- Bertrand David (CNRS)
- Jérôme Detoc (Ifremer)
- Catherine Dreanno (Ifremer)
- Brendan Hennebaut (Ifremer)
- Charlie Widehem (Ifremer)

3.1.2. Introduction à la thématique

Les systèmes de gestion informatisée de laboratoire (LIMS) prennent une place de plus en plus centrale dans la recherche scientifique, en particulier lorsqu'il s'agit de gérer des données et des échantillons. Ils constituent aujourd'hui l'un des outils les plus pertinents pour garantir une traçabilité complète, depuis la collecte des échantillons jusqu'au partage entre collègues de laboratoires des données.

La traçabilité, depuis l'acquisition des échantillons jusqu'à la diffusion des données, est également un élément central pour garantir à la fois la qualité scientifique et la conformité réglementaire. Dans le contexte de l'exploration et de l'étude des grands fonds marins, il est essentiel de s'assurer que la recherche, la production et l'utilisation des échantillons respectent pleinement les cadres légaux et environnementaux, tels que les accords sur l'accès et le partage des avantages (APA) ou encore les discussions en cours autour du traité BBNJ.

Toutefois, l'un des principaux défis liés à l'utilisation des LIMS réside dans la cohérence et l'interopérabilité des systèmes employés. Chaque institut ou laboratoire a tendance à utiliser son propre outil, ce qui génère une fragmentation et complique la circulation, la comparaison et l'intégration des données. Cette situation freine le partage et la mise en valeur des informations.

L'introduction de cette thématique invitait donc à réfléchir aux possibilités d'harmonisation et, à terme, d'homogénéisation des solutions LIMS. Une telle démarche représenterait une avancée considérable dans la gestion des données, en renforçant leur interopérabilité et leur pérennité, et en facilitant les synergies entre communautés scientifiques.

3.1.3. Synthèse des échanges

Les discussions ont mis en évidence la nécessité d'assurer une traçabilité complète des données et des échantillons, depuis leur acquisition jusqu'à leur diffusion. Les enjeux majeurs concernent la préservation des données, leur centralisation et le respect des cadres réglementaires. À Ifremer, la gestion réglementaire liée à l'APA est aujourd'hui assurée par MORSE, avec un suivi par mission ou projet. Les champs obligatoires nécessaires sont déjà intégrés dans le module « échantillon » de LabCollector, et les évolutions attendues (CODECO, BBNJ, CITES) seront également prises en compte par MORSE, de manière transparente pour l'utilisateur.

Une problématique récurrente a été soulignée : la perte d'informations lors du départ d'étudiants, de post-doctorants ou de salariés (cahiers de laboratoire

emportés, accidents, retraites), les données restant alors dans des bases personnelles difficilement accessibles. Aujourd'hui, l'utilisation du cahier de laboratoire électronique (ELN) intégré à LabCollector permet d'assurer un suivi partagé et sécurisé. L'outil se distingue par sa modularité, offrant la possibilité de concevoir de nouveaux modules dédiés à différents types de résultats, tels que les souches isolées, extraits d'ADN/ARN, séquences ou encore données issues d'analyses chimiques.

Cependant, plusieurs limites ont été identifiées. L'aspect financier constitue un frein : l'ajout de nouveaux utilisateurs ou le développement de modules spécifiques représente un coût non négligeable. Par ailleurs, le déploiement de LabCollector reste très fragmenté au sein des organismes (Ifremer, UBO, CNRS, IRD), faute de formation adaptée et de coordination. L'outil est souvent utilisé de manière individuelle, sans réel accompagnement structuré. De plus, la majorité des laboratoires ne disposent pas de gestionnaire de données dédié (Lab Manager), ce qui complique la prise en main et la pérennisation des pratiques.

Un autre problème majeur réside dans le fait que différents LIMS sont utilisés selon les laboratoires et instituts. Cette diversité d'outils accentue la fragmentation des pratiques et complique la mise en commun et l'exploitation des informations. L'homogénéisation des systèmes représenterait une avancée significative pour harmoniser la gestion des données et renforcer la coopération entre équipes. À défaut, il apparaît indispensable de travailler à une meilleure interopérabilité entre les différents LIMS existants, afin de permettre une circulation fluide des données et d'éviter les silos d'information.

En conclusion, LabCollector apparaît comme un outil pertinent et modulable pour renforcer la traçabilité et la valorisation des données scientifiques. Son efficacité dépend toutefois de la mise en place de formations adaptées, d'une meilleure coordination inter-organismes, d'un investissement financier durable et, plus largement, de l'harmonisation ou de l'interopérabilité des systèmes de gestion afin d'assurer une véritable continuité dans la chaîne des données.

3.2. Traçabilité de l'acquisition à la diffusion

3.2.1. Participants

- Pierre-Yves Arnould (CNRS)
- Pauline Auffret (Ifremer)
- Claire Daguin Thiébaud (CNRS)
- Clara Emery (IFB)
- Mark Hoebeke (CNRS)
- Annie Lebreton (CNRS)
- Yacine Moufid (Ifremer)
- Elodie Pétorin (CNRS)
- Claudia Ratti (MNHN)
- Aurélien Schmitt (OSU STAMAR)
- Marine Vernet (Ifremer)

3.2.2. Introduction à la thématique

La traçabilité, depuis l'acquisition des échantillons jusqu'à la diffusion des données, est un élément central pour garantir à la fois la qualité scientifique et la conformité réglementaire.

Au-delà de ces aspects réglementaires, la traçabilité constitue un levier fondamental pour répondre aux critères FAIR (Facile à trouver, Accessible, Interopérable et Réutilisable). Elle permet de renforcer la solidité, la fiabilité et la pérennité des données, afin de garantir leur réutilisation dans d'autres contextes scientifiques et interdisciplinaires.

Cette thématique met ainsi en avant l'importance de développer des pratiques standardisées et robustes, capables d'assurer une continuité entre la collecte sur le terrain, le stockage, le traitement et la diffusion des données. La réflexion autour d'un identifiant unique pour chaque échantillon constitue une des meilleures solutions. Elle faciliterait le suivi tout au long du cycle de vie de l'échantillon et renforcerait l'interopérabilité entre bases de données et institutions.

3.2.3. Synthèse des échanges

Les échanges ont souligné que la traçabilité constitue une condition indispensable pour répondre aux principes FAIR et garantir la solidité, la fiabilité et la réutilisation des données scientifiques. Elle doit couvrir l'ensemble du cycle de vie de l'échantillon, depuis l'acquisition sur le terrain jusqu'à la diffusion dans des catalogues ou entrepôts officiels.

Plusieurs obstacles freinent aujourd'hui la mise en œuvre d'une traçabilité robuste :

- Freins techniques : complexité des outils, difficulté d'appropriation, manque de solutions simples et « clés en main ».
- Manque de ressources humaines : absence de temps et de moyens dédiés, faible implication des responsables de projets dans la démarche Open Science.
- Manque de reconnaissance : la gestion des données reste souvent sous-valorisée, peu visible et donc insuffisamment intégrée aux priorités scientifiques.

Problèmes pratiques :

- Risque de perte d'informations faute d'obligation ou de procédures claires,
- Réticence de certains chercheurs à partager leurs données (embargo, peur d'une réutilisation non maîtrisée),
- Capacités de stockage insuffisantes,
- Hétérogénéité des systèmes de description et des pratiques entre équipes.

Pistes de solutions

Les participants ont convergé sur plusieurs leviers d'amélioration :

- Former et accompagner dès le doctorat : proposer des formations pratiques (« cafés data », modules intégrés aux écoles doctorales) pour inculquer une culture de la donnée dès le début du parcours scientifique.
- Simplifier les outils : développer ou promouvoir des solutions faciles d'usage et adaptées aux besoins opérationnels du terrain.

- Encadrer la gestion de données : mettre en place des obligations minimales de dépôt et de description, à l'image de procédures déjà éprouvées (ex. JACIM).
- Recenser et valoriser les ressources existantes :
 - Catalogues d'entrepôts de confiance (Life Data Gouv, OPIDoR, etc.),
 - Outils FAIR (RDMkit, RO-Crate pour regrouper ressources et métadonnées),
 - Standards (XML, BCprov, etc.).
- Encourager la mutualisation : partager les retours d'expérience réussis (par ex. SEBIMER pour les données de séquençage, TARA pour les observatoires marins génomiques) afin de créer une base commune de bonnes pratiques.

Projet de groupe de travail

À l'issue de la discussion, il a été proposé de structurer un groupe de travail dédié, avec pour objectifs :

- Émettre des recommandations sur les pratiques et outils garantissant la traçabilité des données et métadonnées,
- Réaliser un état des lieux des ressources existantes (entrepôts, standards, formats, outils) et des actions déjà en cours dans d'autres infrastructures de recherche,
- Évaluer l'adéquation des outils avec les besoins des communautés, à l'aide de cas d'utilisation,
- Diffuser largement les résultats sous forme de rapport, de page web ou via les réseaux de correspondants ODATIS.

Une première étape concrète a été définie : identifier les acteurs travaillant déjà sur la description générique des échantillons et recenser les pratiques existantes pour établir une base comparative. Un sondage doit également être lancé afin de planifier les prochaines séances de travail.

3.3. Acquisition des échantillons en mer

3.3.1. Participants

- Anne-Sophie Alix (Ifremer)
- Florian Besson (Ifremer)
- Isabelle Billy (Univ. Bordeaux)
- Christophe Brandilly (Ifremer)
- Cédric Cotté (MNHN)
- Adélie Delacour (Univ. Saint Etienne)
- Bernard Dennielou (Ifremer)
- Nicolas Gabarron (Ifremer)
- Morgane Hubert (Ifremer)
- Bérengère Husson (CNRS)
- Xavier Saint-Laurent (Ifremer)
- Lénaïck Menot (Ifremer)
- Eva Moreno (MNHN)
- Ewan Pelleter (Ifremer)

- Delphine Pierre (Ifremer)
- Mathilde Pitel (Ifremer)
- Christophe Scheffer (Ir REGEF)
- Olivier Soubigou (Ifremer)
- Laurent Toffin (Ifremer)

3.3.2. Introduction à la thématique

L'acquisition des échantillons en mer constitue une étape clé qui conditionne la qualité et la valorisation des données scientifiques. Un groupe de travail intitulé « Bancarisation des opérations de prélèvement et des échantillons à bord des navires » avait déjà souligné l'importance de mettre en place des outils et procédures pour assurer la traçabilité dès la collecte.

Aujourd'hui, malgré l'existence de logiciels embarqués (Casino, Mimosa, Sealog, Adélie), souvent vieillissants, l'absence d'un flux homogène et structuré reliant les opérations de prélèvement aux bases de données institutionnelles est un obstacle. Cette lacune limite la traçabilité, pourtant indispensable pour répondre aux obligations réglementaires (INSPIRE, Nagoya) et aux standards scientifiques internationaux.

L'enjeu est donc de réfléchir à des solutions modernes permettant une captation systématique et interopérable des informations dès le bord, en s'appuyant notamment sur les nouvelles technologies afin de renforcer la pérennité, la réutilisation et la valorisation optimale des échantillons collectés dans le cadre de l'exploration des grands fonds.

3.3.3. Synthèse des échanges (Adélie Delacour et Bernard Dennielou)

1. Introduction

Les présentations données sur cette thématique lors du 1er jour ont montré qu'il existe des failles/lacunes entre les logiciels dédiés à l'acquisition en mer des métadonnées sur les opérations de prélèvements et les échantillons récoltés (Casino, Mimosa, Adélie, Sealog), et les outils informatiques utilisés dans les différents laboratoires (Archimède, Corebook, Cybercarothèque Nationale, Labcollector) pour la bancarisation de ces métadonnées.

La bancarisation relève de la responsabilité des équipes scientifiques embarquées et repose sur ce qui existe déjà dans les laboratoires. Cependant certains laboratoires ne possèdent pas de logiciel spécifique et les scientifiques utilisent alors des fichiers Excel non formalisés pour consigner les échantillons et leurs données relatives limitant ainsi l'accessibilité de ces informations à la communauté scientifique. Ainsi, beaucoup des métadonnées décrivant les opérations de prélèvement et les échantillons récoltés, bien que conservées, ne sont pas bancarisées et demeurent inaccessibles.

Pour les métadonnées bancarisées il existe des portails de consultation et d'interrogation :

- (Cybercarothèque CNRS/Universités <https://cybercarotheque.fr/>),
- Géoportail échantillons de l'Ifremer sur Sextant (<https://echantillons.ifremer.fr/>),
- Catalogue des campagnes de la FOF (<https://campagnes.flotteoceanographique.fr/>),

- Portail de bases de données naturalistes du MNHN (<https://science.mnhn.fr>).
- La base de données BasExp regroupant les informations concernant les expéditions du Muséum national d'Histoire naturelle,
- InfoTerre du BRGM (<https://infoterre.brgm.fr/viewer/>).

Il est nécessaire de disposer d'un workflow continu pour saisir, archiver et consulter les données et les métadonnées liées afin de conserver la traçabilité de la donnée et que les opérateurs ne répètent pas la saisie dans plusieurs logiciels (ce qui conduit à une perte de temps pour le scientifique et peut induire des erreurs dans les informations saisies). Il existe ensuite trois bases de données qui peuvent recevoir des métadonnées sur les échantillons marins :

- BIGOOD, hébergée par le SISMER qui couvre tout type d'échantillon prélevé en mer (sans distinction de discipline),
- La Cybercarothèque hébergée par l'OASU à Bordeaux pour les carottes de sédiment et de glace.
- Au Muséum national d'Histoire naturelle, huit de douze ensembles de collections naturalistes possèdent des spécimens en lien avec la mer : anthropologie culturelle, botanique, collections animales vivantes, invertébrés marins, roches et sédiment océaniques, paléontologie, ressources biologiques-cellules vivantes et cryoconservées, vertébrés.

D'autres bases de données permettent de recueillir des données acquises en mer comme la Banque Nationale de Physique Chimie où peuvent être déposées les données issues des bathysondes (cf. CR du groupe de discussion « Bancarisation des données de Chimie »)

Objectif : avoir un flux continu de l'opération jusqu'à la base de données pour atteindre un inventaire complet (toutes disciplines confondues) des opérations de prélèvements et des échantillons récoltés avec toutes les métadonnées liées et standardisées.

2. Identification des principaux enjeux relatifs à chaque thématique

Les principaux enjeux concernent :

- La traçabilité des opérations et des échantillons pour en permettre une « FAIRisation » aboutie.
- La bancarisation des données et des métadonnées produites lors de la collecte d'un échantillon. Pour certaines communautés scientifiques, il manque des outils dans les logiciels pour remplir les phases d'acquisition de données.
- La mise en ligne et l'accessibilité des métadonnées sur les opérations de prélèvement et les échantillons

Les principales problématiques identifiées sont :

- La possible/certaine perte d'informations et de données entre les différents logiciels utilisés lors des campagnes océanographiques lors des recopies/retranscriptions numériques ou manuelles, les conversions

(coordonnées, unités...), et bien sur l'absence de bancarisation et l'impossibilité de revenir sur les opérations et échantillonnages passés ;

- Les fonctions et l'ergonomie des logiciels utilisés sur les navires de la FOF comme cahier de quart et pour consigner les opérations, les prélèvements et/ou les mesures (Casino, Sealog, Mimosa, Adélie) et aux liens/connexions possibles entre ces différents logiciels.

Ces logiciels ne sont par ailleurs pas toujours utilisés par certaines communautés de scientifiques (par manque de temps et/ou de formation).

Le logiciel intérateur/agrégateur Sealog qui a le mérite d'exister ne répond pas complètement aux besoins exprimés de bancarisation et n'est utilisé que par quelques équipes scientifiques de l'Ifremer.

Le logiciel Corebook, développé par le CNRS dans le cadre de l'Equipex CLIMCOR dans la mouvance de la Cybercarothèque permet de saisir des informations sur les carottes de sédiment mais pas pour d'autres types d'échantillons et il ne communique pas avec les logiciels de la FOF

- Discussions autour des solutions potentielles, des bonnes pratiques et des points de blocage

Il manque des fonctions dans les logiciels de captation des métadonnées (Casino, Mimosa, Adélie) et il manque un logiciel intermédiaire « agrégateur », abouti et intégré à la FOF, entre ces logiciels pour l'intégration des métadonnées dans les bases de données.

Ce rôle d'agrégation n'est que partiellement et imparfaitement rendu par le logiciel Sealog. Le Corebook, en tant qu'outil de captation des informations de carottage sur le terrain, ne concerne que les carottes de sédiment et n'est pas intégré dans les flux des navires de la FOF.

Les obstacles majeurs identifiés à ces enjeux concernent :

- Le temps de développement et le financement nécessaires à la création/refonte d'un outil/logiciel polyvalent.
- La diversité des communautés scientifiques qui utilisent ces logiciels, mais qui n'ont ni la même façon de travailler, ni les mêmes intérêts en termes de collecte de données et métadonnées,
- Le manque de vocabulaire scientifique et de description commun et donc de lexique ou référentiel homogène et commun de la captation des métadonnées jusqu'à leur bancarisation.

Les solutions potentielles évoquées lors de la discussion sont

- Faire évoluer les logiciels d'acquisition des métadonnées (Mimosa, Adélie, Casino, corebook) pour pouvoir taguer/renseigner les opérations conduisant à collecter des échantillons ou les mesures in situ de façon homogène.
- Avoir un logiciel commun, interactif et pratique, qui reprendra les fonctions nécessaires présentes dans certains logiciels (Sealog, Casino, Archimède, Corebook) et apportera des fonctions nouvelles à définir.

3. Production de propositions concrètes et des actions à court et moyen terme

Avant de discuter de la création/évolution d'un logiciel, il est important de constituer un groupe de travail pour procéder à :

- L'état des lieux des solutions existantes sur les navires de la FOF et les laboratoires scientifiques, puis
- La définition des besoins (champs pour la métadonnée par discipline), choix à faire, puis
- La rédaction du cahier des charges pour les logiciels à chaque étape (acquisition, validation/contrôle qualité et bancarisation), puis
- Le développement du/des logiciel(s)

L'état des lieux et la définition du besoin doivent être faits par les différents utilisateurs de la FOF afin de couvrir tous les domaines et tous les laboratoires. Il a été proposé d'établir des sous-groupes de réflexion en fonction de la spécificité des opérations de prélèvement mais transverses aux domaines d'étude (biologie, géologie, chimie, océanographie, halieutique) :

- Groupe 1 : échantillonnage et mesure sur le fond depuis le navire (carottages, dragues à roches, multitube):
- Groupe 2 : échantillonnage, mesure et observation sur le fond réalisés par submersible (Victor, Nautilie, Ariane, AUV) :
- Groupe 3 : échantillonnage et mesure dans colonne d'eau (chalut, acoustique, CTD, rosettes, lignes de mouillage instrumentées, ROV, AUV)

La question d'intégrer les scientifiques travaillant en domaine halieutique a été soulevée pour avoir une vision représentative de l'ensemble des utilisateurs de la FOF.

Une fois cette réflexion engagée sur les besoins, manques et solutions spécifiques à chaque communauté, il a été proposé d'organiser une discussion afin d'harmoniser les besoins, les notions de référentiels communs et les recommandations dans l'objectif de proposer un cahier des charges pour le(s) logiciel(s).

Un certain nombre de recommandations complémentaires ont été formulées par les personnes présentes à la réunion :

- Il est indispensable d'impliquer les équipes scientifiques de différents instituts dans la refonte de Casino et les évolutions de Mimosa3 et Adélie (et que ce soit indépendant du type d'instrument).
- Il serait nécessaire également d'informer les futurs porteurs de projets de campagne sur l'importance/obligation ? de la bancarisation (FAIRisation) des échantillons => besoin de documents d'information et suivi du projet pour la gestion des données.

Autres points abordés et qui demandent une discussion :

- Quelles sont les possibilités de financement via le PEPR ?
- Quelle est la définition d'un échantillon (vs opération vs prélèvement) pour chaque communauté ?
- Système d'identification unique d'un échantillon (IGSN) : est-ce qu'il faut le faire lors de la campagne ou bien après la campagne une fois que tous les échantillons ont été collectés ?
- L'édition d'étiquettes pour les échantillons à bord (incluant l'identifiant unique)
- Le souhait que les logiciels impliqués dans le flux de donnée, spécifiquement l'outil agrégateur, offrent un service aux scientifiques pour la production de documents de synthèses sur les opérations et les échantillons et pour la production du rapport de mission. C'est à cette condition que les équipes scientifiques adopteront de nouveaux outils. Les logiciels devront être utiles, utilisables et utilisés.

La future étape consiste en la définition des dates des futures discussions et à l'élaboration d'un rétroplanning permettant d'avancer rapidement et efficacement sur le cahier des charges.

3.4. Gestion des données issues des échantillons en biologie

3.4.1. Participants

- Cédric Cotté (MNHN)
- Valérie Cueff-Gauchard (Ifremer)
- Pauline Auffret (Ifremer)
- Pauline Chauvet (CNRS)
- Claire Daguin-Thiébaud (CNRS)
- Catherine Dreanno (Ifremer)
- Clara Emery (IFB)
- Brendan Hennebaut (Ifremer)
- Bérengère Husson (CNRS)
- Annie Lebreton (CNRS)
- Lénaïck Menot (Ifremer)
- Elodie Pétorin (CNRS)
- Aurélien Schmitt (OSU STAMAR)

3.4.2. Introduction à la thématique

Ce groupe thématique avait pour objectif de rassembler des chercheurs et ingénieurs travaillant sur les échantillons biologiques collectés en grands fonds, mais appartenant à des laboratoires et instituts différents. Il visait à créer un espace d'échange autour des bonnes pratiques existantes, tout comme des difficultés rencontrées dans la gestion des données associées. L'ambition était de favoriser une meilleure cohérence et d'identifier collectivement des pistes d'amélioration pour renforcer la valeur scientifique des données biologiques.

3.4.3. Synthèse des échanges (Brendan Hennebaut et Bérengère Husson)

Présentation individuelle de chaque personne présente à cet atelier (Ifremer, MNHN, Station Biologique de Roscoff, Université de la Rochelle, IRD, etc.)

L'atelier a débuté par une présentation des participants, représentant notamment l'Ifremer, le MNHN, la Station biologique de Roscoff, l'Université de La Rochelle, l'IRD et d'autres institutions.

- Gestion des données des SNO BENTHOBS & PHYTOBS

Une première session d'échanges a porté sur la gestion des données associées aux SNO BENTHOBS et PHYTOBS. Les discussions ont concerné l'utilisation de référentiels existants (BODC, Marine Regions, etc.), ainsi que les pratiques relatives à la phase d'analyse des données.

- Centralisation des données de laboratoire

La deuxième session est partie d'une question soulevée par une personnes de l'Université de La Rochelle sur la gestion centralisée des données d'un laboratoire. La proposition a été faite d'utiliser un LIMS, tel que LabCollector, qui permet de gérer de manière intégrée les échantillons, les produits chimiques et l'ensemble des données associées.

- Aspects réglementaires et données de séquençage

La troisième session a porté sur les aspects réglementaires, notamment liés à l'APA, ainsi que sur la gestion des données de séquençage.

Pauline Auffret est intervenue sur ce sujet.

Lénaïck Menot a soulevé la question de la protection des données de séquençage déposées sur les plateformes telles que ENA, ouvrant une discussion autour de la réglementation DSI.

- Retour d'expérience IRD

Cédric Cotté (IRD) a partagé son expérience, soulignant que la gestion des données de campagne au sein de l'IRD est très individualisée, chaque campagne gérant ses propres données sans réelle mutualisation.

- Conclusion

En conclusion, Bérengère Husson est revenue sur les possibilités opérationnelles et financières offertes par le PEPR Grands Fonds, en invitant à réfléchir aux apports que ce programme pourrait fournir en matière de gestion et de valorisation des données.

3.5. Gestion des données issues des échantillons en géologie

3.5.1. Participants

- Anne-Sophie Alix (Ifremer)
- Pierre-Yves Arnould (CNRS)
- Florian Besson (Ifremer)
- Isabelle Billy (Univ. Bordeaux)
- Adélie Delacour (Univ. Saint Etienne)
- Bernard Dennielou (Ifremer)
- Nicolas Gabarron (Ifremer)
- Christophe Scheffer (Ir REGEF)

3.5.2. Introduction à la thématique

L'objectif de ce groupe thématique était de permettre la rencontre et le dialogue entre spécialistes de la géologie marine impliqués dans l'étude et la gestion des échantillons issus des grands fonds. En partageant leurs expériences, les participants ont pu confronter leurs méthodes, mettre en lumière les bonnes pratiques déjà mises en place, mais aussi pointer les blocages ou difficultés persistantes. Cet échange visait à identifier des solutions pour améliorer la gestion et la valorisation des données géologiques.

3.5.3. Synthèse des échanges (Marine Vernet)

Constat : dans la plupart des organismes, les données d'analyse réalisées sur les échantillons géologiques sont stockées sur les serveurs des laboratoires. Pas de bancarisation ni de diffusion centralisée de ces données.

Une partie des données est déposée dans des entrepôts (i.e. SEANOE, Zenodo, autres) lors de la publication d'un article scientifique, mais ce n'est pas automatique.

Possible rôle des instances nationales de type Data Terra (Odatis, Formater), INSU? pour centraliser la diffusion des données d'analyses réalisées sur échantillons géologiques, en moissonnant les catalogues existants.

Possible référencement des données d'analyses réalisées sur échantillons collectés via la FOF sur le catalogue des campagnes géré par l'Ifremer ?

Question concernant la bancarisation :

Définir la granularité de la donnée à bancariser (donnée d'intérêt) : analyse brute, analyse pré-qualifiée, donnée associée à une publication. Coût de stockage conséquent pour les données brutes (cf. retex Regef).

Données historiques : fortes demandes sociétales pour accéder à ces données, mais manque d'outils/bases pour répondre à cette demande. Difficultés à accéder aux informations nécessaires pour décrire ces données (créateur plus en poste, informations dispersées...) compliquant la bancarisation. De plus, manque de financement ciblé pour traiter ce type de données, cela se fait le plus souvent sur base de volontariat.

Pistes :

Commencer par les métadonnées décrivant les données d'analyse, et notamment (à minima), celles qui permettent d'identifier qui a réalisé quelle analyse sur quel échantillon, et qui est le point de contact pour obtenir ces données d'analyses.

Commencer par les données produites actuellement, dans l'attente d'éléments/financement permettant de traiter les données historiques.

Une partie des analyses réalisées à bord lors des campagnes en mer est issue d'équipements standards, et pourrait de ce fait être décrite, bancarisée et diffusée.

3.6. Gestion des données issues des échantillons en chimie

3.6.1. Participants

- Christophe Brandily (Ifremer)
- Morgane Hubert (Ifremer)
- Agathe Laës-Huon (Ifremer)
- Yacine Moufid (Ifremer)
- Steven Piel (Ifremer)

3.6.2. Introduction à la thématique

Ce groupe thématique a réuni les acteurs travaillant sur les échantillons et données chimiques liés aux grands fonds marins. L'objectif est de favoriser les échanges entre équipes de différentes institutions afin de partager expériences et pratiques, qu'elles soient positives ou au contraire marquées par des obstacles récurrents. En confrontant ces retours d'expérience, les participants pourront dégager des recommandations communes et ouvrir la voie à une gestion plus harmonisée et efficace des données chimiques.

3.6.3. Synthèse des échanges

Nous avons rencontré trois chimistes du Laboratoire Environnement Profond (LEP) et du Laboratoire Détection, Capteurs et Mesures (LDCM) afin d'échanger sur la manière dont leurs équipes gèrent les données issues des observations in situ et celles provenant de leurs analyses en laboratoire.

Au cours des discussions, il est apparu que ces données, bien que collectées et archivées localement, ne sont pas transmises au SISMER. Or, ce service est dédié à la bancarisation des jeux de données dans la BPC (Banque de données de Physico-Chimie) et dans leur publication vers les catalogues européens officiels, garantissant ainsi leur visibilité et leur diffusion à la communauté scientifique internationale et au reste de la société.

Face à ce constat, Steven Piel a présenté le cheminement complet des données physico-chimiques :

- De leur acquisition à bord ou lors d'expérimentations,
- À leur structuration et archivage,

- Jusqu'à leur mise à disposition publique via les catalogues officiels.

Cette présentation a permis de clarifier les enjeux de traçabilité et de valorisation des données, ainsi que les étapes clés nécessaires à leur bonne intégration dans les systèmes existants.

s

À l'issue de l'échange, nous avons convenu de mettre en place un accompagnement concret auprès des équipes du LEP et du LDCM. Celui-ci visera à :

- Intégrer les archives de données existantes dans les circuits officiels de bancarisation et de diffusion.
- Assurer la prise en charge systématique des futures données produites, afin qu'elles suivent directement le cheminement standardisé jusqu'à leur publication.

Cette entrevue marque ainsi une première étape vers une meilleure harmonisation des pratiques, avec pour objectif de renforcer la qualité, la pérennité et la visibilité internationale des données produites par les deux laboratoires.

4. Points clés retenus dans les présentations et les discussions lors des deux journées

Ce dernier chapitre propose une synthèse des échanges ayant eu lieu lors des deux journées de l'atelier technique, présentée sous forme de points clés. Il met en évidence les points de blocage rencontrés par les équipes ainsi que les pistes d'amélioration envisagées. Cette synthèse offre une vision transversale des discussions et fait office de bilan de l'atelier.

4.1.1. Points bloquants et limites rencontrées

- Hétérogénéité forte des pratiques et outils : coexistence de systèmes multiples (LabCollector, Sealog, Excel, Archimède, JACIM...), entraînant doublons, ressaisies, pertes d'information et manque d'interopérabilité.
- Charge de travail élevée et manque de ressources humaines : saisies manuelles chronophages, masse de données et d'échantillons très importante, absence fréquente de gestionnaires de données dédiés, fragmentation des pratiques au sein des équipes.
- Interopérabilité incomplète : incompatibilités entre référentiels (WoRMS vs NCBI), manque de descripteurs pour la géologie dans Darwin Core, faible dialogue entre bases nationales et internationales.
- Traçabilité lacunaire : rupture fréquente entre prélèvement, analyses et dépôt final ; difficultés pour relier les données de séquençage, imagerie ou fluides à leur échantillon source.
- Outils vieillissants ou peu ergonomiques : Sealog en fin de vie, CASINO jugé lourd, Archimède complexe, JACIM limité en ergonomie et accessibilité.
- Questions de gouvernance : incertitudes sur la parentalité et la propriété des données, perte d'informations lors des départs de chercheurs ou étudiants, manque de règles claires et partagées.

- Valorisation insuffisante : certaines données restent stockées localement (serveurs de laboratoire) et n'atteignent pas les dépôts officiels (SISMER, Seanoë, ENA, EmodNet, etc.).

4.1.2. Pistes d'améliorations évoquées

Simplification et harmonisation des outils

- Garantir la continuité du flux de données entre les outils embarqués (Casino, Mimosa, Adélie, Sealog) et ceux utilisés à terre (Archimède, Corebook, LabCollector, Cybercarothèque, etc.), afin d'assurer un workflow homogène de l'acquisition à la bancarisation.
- Développer un outil intermédiaire "agrégateur" intégré à la Flotte Océanographique Française, capable de centraliser et redistribuer automatiquement les métadonnées issues des différents logiciels embarqués.
- Limiter la multiplication des logiciels et des interfaces sources de redondance et d'erreurs.
- Développer soit un outil unique intégré, soit une interopérabilité forte via API permettant à chaque système de dialoguer sans ressaisie.
- Favoriser des outils plus légers et ergonomiques, utilisables à bord comme à terre, avec des interfaces adaptées aux contraintes des missions.

Renforcement de la traçabilité dès la collecte

- Généraliser l'attribution d'identifiants uniques appliqués directement lors du prélèvement.
- Intégrer ces identifiants dans tous les flux pour assurer un suivi complet.
- Développer des applications mobiles et hors-ligne (type CoreBook) permettant de saisir les informations directement sur le terrain, même sans connexion.

Interopérabilité et standardisation

- Élaborer un lexique et des référentiels communs pour uniformiser la description des opérations, prélèvements et échantillons, toutes disciplines confondues.
- Poursuivre l'adoption de standards internationaux : Darwin Core, MIDS, WoRMS, TaxRef, Catalogue of Life.
- Développer des passerelles automatiques entre bases nationales et internationales.
- Compléter les standards existants, par exemple en géologie, où certains descripteurs manquent encore.

Allègement de la charge de saisie et automatisation

- Déployer des outils de moissonnage automatique (EGIDE, DataRef) pour récupérer les métadonnées sans ressaisie.
- Mettre en place des modules de contrôle qualité pour réduire les erreurs et garantir la cohérence entre systèmes.
- Intégrer aux outils d'acquisition des fonctions automatiques de génération de documents de synthèse (fiches d'échantillons, rapports de mission), pour en faire de véritables services opérationnels pour les scientifiques.

Gouvernance et accompagnement humain

- Désigner des référents « données » dans les laboratoires et équipes embarquées pour garantir la continuité et l'homogénéité des pratiques ou

augmenter les moyens humains et pécuniers en matière de gestion de données.

- Former les chercheurs, doctorants et ingénieurs aux outils et aux standards de gestion des données, et ce dès les études.
- Clarifier la question de la parentalité et de la propriété des données, afin d'éviter les pertes ou les flous lors des départs de personnes.

Ouverture, valorisation et mutualisation

- Poursuivre l'effort de dépôt systématique des données dans les bases nationales et internationales reconnues.
- Développer des interfaces publiques conviviales (portails web, moteurs de recherche fédérés) pour accroître la visibilité et l'impact des données collectées.
- Créer ou améliorer les catalogues dédiés, afin de faciliter le référencement et l'échange des échantillons entre scientifiques, laboratoires et instituts.
- Valoriser les jeux de données bien documentés, en les intégrant aux démarches de science ouverte et aux collaborations internationales (DiSSCo, GBIF, OBIS, Catalogue of Life).

4.1.3. Conclusion

En conclusion, il ressort des présentations et des discussions un constat partagé : la gestion des données et des échantillons, en particulier pour les grands fonds marins, constitue un enjeu stratégique majeur qui souffre aujourd'hui d'un manque de moyens humains et financiers. Alors que la masse de données à traiter ne cesse de croître, que les préconisations en matière de gestion FAIR se renforcent et que les réglementations sur les échantillons se multiplient (APA, BBNJ, CITES, RGPD), les équipes font face à une charge de travail considérable avec des ressources limitées.

Plusieurs interventions ont également souligné que de nombreuses difficultés identifiées relèvent moins d'enjeux techniques que de problématiques humaines et organisationnelles : manque de coordination entre structures, absence de gouvernance claire, et insuffisance de profils dédiés à la gestion des données. Renforcer la dimension collective, la communication et la structuration des rôles et l'acculturation aux principes FAIR apparaît donc essentiel pour améliorer durablement les pratiques.

De nombreux outils, pratiques et initiatives existent déjà et témoignent d'une dynamique réelle, mais ils gagneraient à être consolidés et mieux soutenus. Un appui renforcé en termes de moyens permettrait de faciliter la traçabilité, le partage et la réutilisation des échantillons et des données, et de répondre plus efficacement non seulement aux attentes de la communauté scientifique, mais aussi aux besoins plus larges de la société en matière de connaissance, de protection et de valorisation durable des océans.

À l'issue de cet atelier, trois groupes de travail ont été mis en place. Ils contribueront à améliorer la bancarisation de certaines données, à renforcer la traçabilité et à mettre en œuvre des flux automatisés entre l'acquisition à bord et les bases de données.