

Département : REM

Auteur(s) : S. Ybert

Contributeur(s) : Marie-Anne Cambon ; Cécile Cathalot ; Valérie Harscoat ; Lénaïck Menot ;
Constance Pages ; Ewan Pelleter ; Cécile Pertuisot ; Olivier Rouxel

Référence : REM2023/011

PLAN DE GESTION DES DONNEES

Projet : Mission DMA et CCZ

Référence ANR : 22-MAFM-0001



Visibilité Archimer :

- Internet
- Intranet Ifremer
- Equipe :
- Groupe d'utilisateurs :
- Confidentiel

Diffusion : Equipe projet

: Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre de France 2030 portant la référence ANR 22 MAFM 0001

Plan de gestion des données - Mission DMA et CCZ

Version du document

1/17

Table des matières

1	Objet	5
2	Résumé du projet.....	5
3	Description des données et collecte ou réutilisation de données existantes	6
3.1	Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?	6
3.2	Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?	7
4	Documentation et qualité des données	9
4.1	Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?	9
4.2	Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?	9
5	Stockage et sauvegarde pendant le processus de recherche	9
5.1	Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?	9
5.2	Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?	10
6	Exigences légales et éthiques, codes de conduite	10
6.1	Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ? ..	10
6.2	Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?	10
6.3	Comment les éventuelles questions éthiques seront-elles prises en compte, et les codes déontologiques respectés ?	11
7	Partage des données et conservation à long terme	12
7.1	Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?	12
7.2	Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?	13
7.3	Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?	16
7.4	Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?	16
8	Responsabilité et ressources en matière de gestion des données	17
8.1	Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?	17

8.2 Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ? 17

{
.
{
{
{
{
{
{
{
.
.
.

1 Objet

L'objet du plan de gestion des données est d'expliciter la nature des données qui seront recueillies ou produites durant le projet ainsi que la manière dont elles seront documentées, analysées, archivées et publiées. Le plan de gestion de données a vocation à évoluer au cours du projet jusqu'à une version définitive qui sera publiée à la date de fin du projet.

2 Résumé du projet

Le projet « Mission DMA et CCZ » consiste à mener les explorations prévues dans le cadre des contrats d'exploration liant l'Ifremer, sous patronage de la France, à l'Autorité Internationale des Fonds Marins (AIFM) sur la période allant de mars 2022 à mars 2027. Ces explorations concernent deux zones géographiques : la dorsale médio-atlantique (DMA) pour le contrat d'exploration des sulfures polymétalliques ; la zone Clarion-Clipperton (CCZ) pour le contrat d'exploration des nodules polymétalliques. Technologiquement, ces missions représenteront les premières validations en environnement réel et par grandes profondeurs, du premier drone (AUV) français capable de descendre jusqu'à 6 000 mètres.

Pour la partie du projet liée au contrat d'exploration sulfures, les objectifs principaux sont de :

- Poursuivre l'exploration régionale, la localisation et l'étude géologique de sites actifs, inactifs et fossiles ;
- Poursuivre l'exploration locale du champ hydrothermal de TAG ;
- Réaliser une évaluation préliminaire de la nature des panaches naturels et des interactions eau-roche, étudier les interactions biogéochimiques associées ;
- Inventaire de la biodiversité et étude des échanges biologiques entre les différents sites (par larves, microorganismes, faune...) à travers les panaches, les courants de fonds...

Pour la partie du projet liée au contrat d'exploration des nodules, les objectifs principaux sont de :

- Améliorer la classification des ressources minérales dans certaines parties du contrat par la collecte de nouvelles données géologiques, géophysiques et optiques haute résolution.
- Améliorer nos connaissances sur les communautés biologiques et notamment : quantifier et modéliser l'impact des essais d'extraction (dans le cadre du projet JPIO MiningImpact II) ; caractériser les habitats et la faune benthique ; améliorer les outils de taxonomie intégrative ; définir une stratégie d'évaluation et de suivi de l'impact sur l'environnement.

Le projet est financé par l'ANR, référence 22-MAFM-0001, dans le cadre de l'objectif 10 « investir dans le champ des fonds marins » du plan France 2030.

Le coordinateur et unique participant du projet est l'Ifremer.

La durée du projet est de 60 mois sur la période allant du 25/03/2022 au 24/03/2027.

3 Description des données et collecte ou réutilisation de données existantes

3.1 Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Le projet prévoit la réalisation de quatre campagnes d'exploration, deux campagnes d'exploration sur chacun des contrats d'exploration AIFM, qui permettront de recueillir de nouvelles données et échantillons. La première des quatre campagnes s'est déroulée du 9 juillet au 24 août 2022, HERMINE2 : <https://doi.org/10.17600/18001851>.

Les principaux outils et méthodes qui seront utilisés pour recueillir de nouvelles données sont listés ci-dessous. Le détail des outils et opérations mis en œuvre sont par ailleurs explicités dans les rapports liés aux différentes campagnes (dossiers de demande de campagne, rapport de campagne etc.).

- Sondeurs multifaisceaux (sondeur de coque du navire, sondeur embarqué dans l'AUV) pour l'acquisition de données de bathymétrie et réflectivité ;
- Caméras (ex. embarquées sur l'AUV, sur un carottier) pour l'acquisition de photos et vidéos ;
- Sonars embarqués sur AUV ;
- Capteurs CTD et auxiliaires (ex. néphélomètre), etc. pour l'acquisition de données physico-chimiques dans la colonne d'eau ;
- Magnétomètres ;
- Gravimètre ;
- Carottiers multitubes, carottiers USNEL, etc. seront utilisés pour la collecte d'échantillons de sédiments, nodules et faune associée ;
- Drague à roche, systèmes de prélèvements par ROV et HOV ;
- Outils de collecte d'eau et de fluides tels que le préleveur PIF, CTD Rosette, la pompe à larve SALSA, aspirateurs à faune des engins sous-marins, Petites Boîtes (PBT) de prélèvements biologiques, bouteilles de prélèvements titane et bouteille étanche au gaz – BEG, etc.

Le projet utilisera également les données acquises lors de précédentes campagnes sur les zones des deux contrats d'exploration ainsi que celles produites à partir de ces données. Les précédentes campagnes menées par l'Ifremer dont les données seront utilisées sont listées ci-dessous avec leur DOI par lequel l'ensemble des données, publications et rapports associés à ces campagnes sont accessibles:

- WP1 – DMA :
 - BICOSE (2014) : <https://doi.org/10.17600/14000100>
 - LEVE_SMF (2016) : <https://doi.org/10.17600/16010800>
 - HERMINE (2017) : <https://doi.org/10.17600/17000200>
 - BICOSE 2 (2018) : <https://doi.org/10.17600/18000004>
- WP2 – CCZ :
 - NODINAUT (2004) : <https://doi.org/10.17600/4010060>
 - BIONOD (2012) : <https://doi.org/10.17600/12010030>
 - NODULE 2015 : <https://doi.org/10.17600/15001400>

Le projet utilisera par ailleurs des données acquises et/ou produites dans le cadre de la participation de l'Ifremer aux campagnes menées avec les partenaires des projets MiningImpact et MiningImpactII de la JPI Oceans :

- Lefaible, Nene (2022): Meiofauna counts from multi corer samples from a dredge-experiment in the German License area during SONNE cruise SO268/2, Clarion-Clipperton Zone. *PANGAEA*, <https://doi.pangaea.de/10.1594/PANGAEA.942174>
- Lefaible, Nene (2022): Meiofauna counts from multi corer samples from a disturbance experiment in the Belgian License area during Island Pride cruise IP21, Clarion-Clipperton zone. *PANGAEA*, <https://doi.pangaea.de/10.1594/PANGAEA.942177>
- Bonifacio, P., Neal, L., Omnes, E., Baptiste, F., Dahlgren, T.G., Menot, L., 2019. The polychaete fauna of the Clarion-Clipperton Fracture Zone from boxcore samples during SONNE cruise SO239. *PANGAEA*. <https://doi.org/10.1594/PANGAEA.902860>
- Bonifácio, P., Neal, L., Omnes, E., Menot, L., 2021. Polychaete fauna belonging to family Polynoidae of the Clarion-Clipperton Fracture Zone sampled during SONNE cruise SO239. <https://doi.org/10.1594/PANGAEA.926674>
- Vanreusel, A., Hilário, A., Ribeiro, P.A., Menot, L., Martínez Arbizu, P., 2017. Epifaunal composition and densities measured along ROV video transects during SONNE cruise SO239. <https://doi.org/10.1594/PANGAEA.879189>

Enfin, le projet utilisera d'autres données acquises lors de thèses et en cours de publication :

Djedjroh, 2022 : <https://www.theses.fr/s218582>

3.2 Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Une part importante des données (abondance de la faune, données d'environnement et isotopique, etc.) sont au format csv qui est utilisable facilement avec de nombreux outils. D'autres données sont dans des formats spécifiques : par exemple, les données de bathymétrie sont au format NetCDF ; les données de séquençage haut débit sont au format FASTQ¹.

3.2.1 Données numériques

Bathymétrie et réflectivité

Le format des jeux de données de bathymétrie est le format NetCDF qui est un format commun dans le domaine marin. Les Modèles Numériques de Terrain de bathymétrie et de réflectivité associée sont au format GeoTIFF.

Données de localisation des opérations

La localisation de l'ensemble des opérations réalisées dans les campagnes du projet sont intégrées dans un SIG.

¹ Le format FASTQ est un format de fichier texte permettant de stocker à la fois des séquences biologiques (uniquement des séquences nucléiques) et les scores de qualité associés. Source : [Wikipédia](https://fr.wikipedia.org/wiki/FASTQ)

Photos et vidéos

Les formats préférentiels pour les photos et vidéos sont respectivement jpeg et mp4. En cas de reconstruction 3D ou de mosaïques 2D, les formats pourront être tif, geotif, mnt ou obj.

Métadonnées des campagnes

Le format des métadonnées des campagnes (lien vers le catalogue des campagnes, voir chapitre 7.2) est le format CSR² (Cruise summary reports), qui est utilisé au niveau européen par les organisations impliquées dans le projet SeaDataNet et qui assure l'interopérabilité des métadonnées des campagnes entre les partenaires du projet. Les métadonnées sont accessibles en moyenne un an après la campagne.

Données sur la faune

Les données associées à la faune telles qu'abondance, densité, diversité, mesures morphométriques, biomasse, etc., sont au format csv. La base de données taxonomique utilisée est le [World Register of Marine Species](#) (WoRMS).

Données environnementales

Les données environnementales comprennent les données de température, salinité, pH, O₂, CH₄, H₂S, CO₂, concentration en métaux, courants, turbidité, taux d'activité métabolique, concentrations de cellule microbienne et teneur en matière organique collectés dans les environnements pertinents. Des données géochimiques, notamment sur la composition des roches, sédiments et fluides en éléments majeurs et traces seront également produites. Toutes ces données sont généralement au format csv.

Données de séquençage ADN

Des données de séquençage seront produites à partir d'échantillons de faune (allant du domaine de la microfaune à la macrofaune) à des fins d'identification, d'études de connectivité et d'approches génomiques. Les données seront en accès libre dans les quatre ans après leur soumission dans les bases de données ou lors de la publication, selon la première de ces éventualités. Les données de séquençage haut débit sont au format FASTQ qui est un format standard pour ce type de données.

Données isotopiques

Les données isotopiques incluent des données relatives au carbone, notamment sur les espèces volatiles (ex. CO₂, CH₄), à l'azote, au cuivre, fer, zinc, cadmium, germanium, sélénium, silicium, strontium et soufre. Ces données (existantes et qui seront produites à partir des échantillons collectés dans les quatre campagnes du projet) sont au format csv.

Documents

- Tableurs et bases de données aux formats xls et csv ;
- Rapports, publications, posters, compte-rendu etc. au format pdf et/ou doc, odt.

3.2.2 Echantillons physiques

- Roches et minéralisations (ex minéralisations hydrothermales, roches volcaniques) ;
- Sédiments ;
- Spécimen de faune (micro, méio, macro) ;

² Voir: <https://www.seadatanet.org/Standards/Metadata-formats/CSR>

- Fluides (ex. eau de mer ; fluide hydrothermal) ;
- Particules (ex en suspension dans l'eau de mer, pièges à particules).

4 Documentation et qualité des données

4.1 Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Des informations relatives aux métadonnées et à leur format sont fournies aux chapitres 3.2 et 7 en notant que ceux-ci sont pour partis indépendants du projet dans la mesure où ils sont définis par les plateformes et base de données utilisées par le projet.

4.2 Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Les données acquises font l'objet de contrôle qualité à différentes étapes du processus listées ci-dessous. Ces mesures de contrôle qualité impliquent les équipes techniques des navires, les équipes scientifiques ainsi que les personnels du SISMER.

- A bord lors de leur acquisition ;
- Au retour de campagne (contrôle technique sur la qualité instrumentale d'acquisition des mesures scientifiques opéré par l'armateur) ;
- Lors de leur intégration dans les différentes bases de données et portails (contrôle qualité des métadonnées opérés par le SISMER et contrôle qualité des données opérés par le SISMER et par les scientifiques).

Ces mesures s'inscrivent dans le cadre du processus qualité « Recueillir et mettre à disposition des données sur le milieu marin », partie intégrale du système qualité de l'Ifremer et piloté par le SISMER.

5 Stockage et sauvegarde pendant le processus de recherche

5.1 Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

À partir de leur collecte lors d'une des quatre campagnes du projet, les données et métadonnées associées sont très rapidement intégrées sur des disques réseaux des unités de recherche concernées par les données ainsi que dans les bases de données/portails tels que ceux gérés par le SISMER (voir chapitre 7). Les disques réseaux et l'infrastructure gérée par le SISMER font l'objet de sauvegardes quotidiennes protégeant ainsi les données tout au long du processus de recherche. Pour le partage de jeux de données entre chercheurs, ces derniers disposent également d'outils sécurisés tels que Filesender (un service de Renater, dont les serveurs sécurisés sont basés en France et qui est membre

du consortium des Réseaux nationaux de Formation et de Recherche – NRENs financé par l'Union européenne) ou le cloud de l'Ifremer, service fourni par la société Owncloud sur les serveurs propriétaires de l'Ifremer.

5.2 Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Les disques réseaux Ifremer ainsi que le cloud privé Owncloud sont à accès restreints et les données sont stockées sur l'infrastructure propre de l'Ifremer. L'accès à ces espaces est régi par la politique de sécurité informatique de l'Institut.

L'accès aux données collectées lors des campagnes est géré par les gestionnaires des bases de données, les chefs de mission et les chercheurs responsables de différentes opérations et jeux de données associés.

6 Exigences légales et éthiques, codes de conduite

6.1 Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Le projet ne prévoit pas de collecte ou d'utilisation de données à caractère personnel.

6.2 Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

En tant qu'unique titulaire du projet, l'Ifremer entend appliquer le droit commun de la propriété intellectuelle, en accord avec les stipulations de la convention de financement ANR et ses engagements en matière de science ouverte.

Données numériques et physiques collectées lors des campagnes (Chap. 3.2) : Ces données seront la propriété exclusive de l'Ifremer, répertoriées selon les règles stipulées au chapitre 4 et partagées selon les modalités détaillées au chapitre 7.

Inventions, savoir-faire, bases relatifs aux données et protégés par le Droit de la Propriété intellectuelle

{ Tout développement innovant de méthode ou outils scientifiques de traitement de données développé
{ par les équipes de l'Ifremer dans le cadre du projet sera la propriété exclusive de l'Ifremer qui décidera
{ seul de l'opportunité d'un enregistrement à titre de protection, notamment en matière de brevet et/ou
{ de logiciel, et en financera les procédures.

:

Dans le cas d'un dépôt de titre de propriété industrielle, l'Ifremer s'assurera d'exploiter ou faire exploiter l'invention par un partenaire afin d'assurer la dissémination des savoirs créés et financés à l'occasion du projet.

Dans le cas de bases de données constituées pour le traitement des données agrégées et collectées pour le projet, l'Ifremer assure leur diffusion la plus large possible dans le cadre de la science ouverte, en en proposant le téléchargement sur les services SEANOE par exemple.

Dans le cas d'un logiciel, et selon les éventuelles licences applicables dans le cas d'un logiciel composé ou dérivé d'autres logiciels, l'Ifremer s'efforcera de diffuser le code source sous une licence libre ou open source.

6.3 Comment les éventuelles questions éthiques seront-elles prises en compte, et les codes déontologiques respectés ?

Les campagnes d'exploration se déroulent dans les zones maritimes au-delà des juridictions nationales que sont la Haute Mer et la zone des fonds marins au-delà des limites du plateau continental (appelée « la Zone »). D'un point de vue règlementaire, la gestion de l'environnement relève :

- De l'AIFM pour la stratégie globale à l'échelle de la Zone, via la Convention des Nations Unies sur le droit de la mer, et les règlements pour l'exploration ;
- Des contractants à l'échelle de chaque contrat d'exploration selon les règles et recommandations établies par l'AIFM, en particulier les Recommandations à l'intention des contractants en vue de l'évaluation d'éventuels impacts sur l'environnement liés à l'exploration des minéraux marins dans la Zone ([ISBA 25 LTC 6 Rev.2](#)).

D'un point de vue opérationnel, les principaux outils mis en œuvre et/ou identifiés pour la gestion de l'environnement sont :

- La définition, à l'échelle de chaque contrat et lorsque cela est possible, potentiellement dès la phase d'exploration une fois les zones prospectives minières définies, de zones témoins de préservation et de zones témoins d'impact ;
- L'obligation, à toutes les échelles, de s'assurer que les activités n'aient pas d'effets nocifs importants sur les écosystèmes marins vulnérables, parmi lesquels figurent les écosystèmes des sources hydrothermales actives.

Dans ce contexte, l'Ifremer en tant que contractant propose de promouvoir des activités de recherche, menées au titre de son contrat d'exploration des sulfures polymétalliques (WP1 – DMA), en respectant scrupuleusement le code de conduite InterRidge (voir page suivante). Ces activités de recherche comprennent notamment :

1. des actions d'exploration pour localiser et étudier en priorité les minéralisations hydrothermales fossiles.
2. des actions visant à comprendre et à quantifier l'export de métaux dans l'océan par les panaches hydrothermaux naturels.
3. des études sur les sites actifs relais découvert dans le plan de travail du contrat d'exploration avec la caractérisation des échanges inter sites (larves, microorganismes, faune) via les panaches naturels, les courants de fonds pour établir les relations entre ces sites en termes de connectivité et de processus biologiques.

Rappel du code de conduite InterRidge:

- *Avoid, in the conduct of scientific research, activities that will have deleterious impacts on the sustainability of populations of hydrothermal vent organisms.*
- *Avoid, in the conduct of scientific research, activities that lead to long lasting and significant alteration and/or visual degradation of vent sites.*
- *Avoid collections that are not essential to the conduct of scientific research.*
- *Avoid, in the conduct of scientific research, transplanting biota or geological material between sites.*
- *Familiarize yourself with the status of current and planned research in an area and avoid activities that will compromise experiments or observations of other researchers. Assure that your own research activities and plans are known to the rest of the international research community through InterRidge and other public domain data bases.*
- *Facilitate the fullest possible use of all biological, chemical and geological samples collected through collaborations and cooperation amongst the global community of scientists.*

We also reaffirm our commitment to open international sharing of data, ideas and samples in order to avoid unnecessary re-sampling and impact on hydrothermal vents, and to further our global understanding of these habitats for the good of all people on Earth.

D'une manière globale, les activités planifiées au cours des quatre campagnes d'exploration prévues dans le projet pour collecter de nouvelles données sont toutes considérées par l'AIFM (ISBA_25_LTC_6_Rev.2) comme n'étant pas susceptibles de causer de graves dommages à l'environnement et n'appellant donc pas d'évaluation de leur impact sur l'environnement..

7 Partage des données et conservation à long terme

7.1 Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

Toutes les données produites durant le projet doivent être facilement accessibles. La majorité des outils utilisés pour stocker et archiver les données ont un portail d'accès sur lequel les données ouvertes sont facilement identifiables. Seanoe, l'Océanothèque d'Ifremer et Archimer contribuent aux principes d'accès ouvert et de sciences ouvertes pour un accès gratuit à l'ensemble des données scientifiques financées par des fonds publics au bénéfice de la recherche. Seanoe est développé dans le cadre du pôle de données et de services pour l'océan "ODATIS" qui promeut le principe de données « FAIR ». Les jeux de données de Seanoe peuvent facilement être utilisés par d'autres projets. Par exemple, les données publiées dans Seanoe sont automatiquement dupliquées dans le portail d'ingestion des données d'EMODnet afin que les centres de données du réseau soient informés de l'existence et de la disponibilité de ces données. Si ces données sont d'intérêts pour ces centres de données, ces derniers pourront mettre en œuvre les actions nécessaires à leur intégration dans leurs bases de données nationales ainsi que dans les plateformes de données européennes telles que SeaDataNet, EurOBIS, et les plateformes thématiques d'EMODnet.

Une part importante des données acquises/produites durant le projet ne sera rendue accessible qu'après valorisation et publication. Sur Seanoe, un embargo allant jusqu'à 2 ans peut être appliqué à

un jeu de données pour couvrir la période de préparation d'une publication, l'archivage et la description de jeux de données durant cette période sur Seanoe, Sextant et Archimer étant toujours possible durant cette période. Ces outils permettent en effet de gérer les droits d'accès afin par exemple d'en restreindre l'usage aux membres de l'équipe projet.

Les données de géophysique acquises en dehors des eaux sous juridiction françaises et stockées sur la base de données géosciences du SISMER sont visibles mais ne sont pas publiquement accessibles. Une requête doit être envoyée au gestionnaire de la base de données pour en demander l'accès et l'utilisation à des fins commerciales n'est pas autorisé.

Dans la mesure du possible, les données et autres produits seront publiés sous licence CC0 ou CC-BY. Seanoe, Sextant et l'Océanothèque d'Ifremer permettent aux producteurs de données de sélectionner la licence s'appliquant à leurs données. Sur la base de données géosciences du SISMER, tous les jeux de données de géophysique sont sous licence CC-BY, à l'exception des données acquises dans des eaux étrangères.

Bien que les vidéos archivées sur la base de données vidéos du SISMER ne soient pas actuellement sous licence, elles sont protégées par les droits d'auteurs. Une demande écrite doit être faite auprès de l'Ifremer avant de pouvoir copier, traduire, reproduire, vendre, publier ou diffuser une vidéo. Les séquences vidéo peuvent être directement téléchargées en basse ou moyenne définition mais une demande préalable doit être faite au chef de mission concerné pour pouvoir les télécharger en haute définition.

Dans le cadre de ces deux contrats d'exploration avec l'AIFM, l'Ifremer doit par ailleurs communiquer l'ensemble des données à l'AIFM par le biais de ses rapports annuels d'activités et/ou rapports quinquennaux. Les données doivent normalement être communiquées à l'AIFM dans un délai de quatre ans maximum après leur acquisition au cours d'une campagne d'exploration. Les données communiquées à l'AIFM sont ensuite intégrées dans la base de données DeepData, mise en place récemment par l'Autorité. Sauf exception, notamment pour permettre la publication d'articles scientifiques avant ouverture des données, les données relatives à l'environnement sont considérées comme non confidentielles et ouvertes et sont donc accessibles via DeepData. Les données relatives à l'évaluation des ressources minérales sur les zones des contrats sont considérées comme confidentielles par l'AIFM et leur accès n'est pas ouvert sur la base de données de l'Autorité.

Les publications, rapports, communications orales etc. feront mention du soutien apporté par l'ANR au titre de France 2030, par exemple avec la mention « Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre de France 2030 portant la référence ANR 22 MAFM 0001 ».

7.2 Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

La préservation des données et publications s'appuiera sur les plateformes, bases de données, outils et processus opérés et mis en œuvre par le Service des Systèmes d'Informations Scientifiques pour la MER (SISMER) de l'Ifremer. Le SISMER est reconnu comme Centre national de données océanographiques (NODC) par l'organisation internationale IODE et comme Centre de données de référence par l'organisation CoreTrustSeal ([certified Core Trustworthy Data Repositories](#)). Par ailleurs, Sextant contribue

au Système d'Information pour le Milieu Marin (SIMM), système fédérateur du Ministère de l'environnement, et au Pôle ODATIS, pôle de Données et Services pour l'océan de l'Infrastructure du système Terre, Data Terra. La diffusion et archivage des données géochimiques seront aussi conformes au DMP de l'IR RéGÉF (Réseau Géochimique et Expérimental Français) en cours de rédaction.

Les publications seront également déposées sur l'archive ouverte nationale HAL. Un système de transfert automatique des publications déposées sur ARCHIMÉRIE vers HAL est en cours de développement par l'Ifremer et devrait être opérationnel autour de mi 2023.

7.2.1 Base de données géosciences du SISMER

Les jeux de données de bathymétrie et de réflectivité associée acquis avec les sondeurs multifaisceau des navires et engins de la flotte océanographique française sont archivés de manière permanente sur des bandes magnétiques et référencés dans la base de données géosciences du SISMER. Elles sont accessibles sur la base de données en moyenne un an après la campagne, une fois la vérification et le traitement des données terminés. Les données sont archivées dans leur format natif ainsi qu'au format NetCDF. Les équipes Ifremer de l'unité mixte de recherche Geo-Océan utilisent le logiciel Globe pour traiter et valider les données de bathymétrie et produire des jeux de données validés, un jeu de données intermédiaire entre les données brutes et un modèle numérique de terrain (MNT).

Les jeux de données de bathymétrie et de réflectivité associée sont accessibles sur demande via le catalogue des campagnes océanographiques françaises³ sur la page de description des métadonnées de chaque campagne et sur le SeaDataNet *Common Data Index* (CDI)⁴. Ils sont également utilisés par le volet bathymétrie du projet EMODnet pour générer des MNT en accès ouvert.

7.2.2 BIGOOD - Biology and GeOlogy Ocean Database (Ifremer-SISMER)

La base de données BIGOOD est gérée par le SISMER et contient les données suivantes:

- Opérations de prélèvement : position (latitude, longitude), profondeur, équipement (drague, carottier, benne...);
- Descriptions des échantillons ;
- Stockage des échantillons : emplacement physique et mouvements (transferts, prêts) ;
- Analyses associées (par le biais de fichiers attachés) : chimie, minéralogie, granulométrie (Folk, laser), stratigraphie, lithologie, mesures par Multi-Sensor Core Logger.

Cette base de données est alimentée par les scientifiques au moyen d'une interface graphique connectée à la base de données BIGOOD disponible en laboratoire ou en lithothèque : le logiciel Archimède. Ce logiciel permet la gestion des métadonnées des échantillons, la localisation physique des échantillons, la génération d'étiquettes, le suivi des mouvements des échantillons (transferts, prêts...).

Cette base de données est étroitement liée au catalogue des campagnes océanographiques françaises (voir 7.2.1).

Un geoviewer permet d'afficher tous les prélèvements répertoriés dans la base de données BIGOOD et d'effectuer des filtres et sélections. Une aide en ligne est disponible pour aider à utiliser ce geoviewer (<https://echantillons.ifremer.fr/Echantillons>).

³ <https://campagnes.flotteoceanographique.fr/>

⁴ <https://cdi.seadatanet.org/search>

La base de données BIGOOD garantit la standardisation des données et l'utilisation de vocabulaire standard tel que OBIS pour la taxonomie et P01 (BODC) pour les paramètres. Les jeux de données sont exportés vers l'*Ocean Biodiversity Information System* (OBIS)⁵, dont les données sont intégrées au *Global Biodiversity Information Facility*⁶ (GBIF). Un identifiant unique est attribué à chaque échantillon. BIGOOD fait partie du système qualité certifié ISO 9001 de l'Ifremer et dispose donc d'un protocole de contrôle qualité. BIGOOD n'offre pas de service de diffusion des données, ce service étant assuré par un système tel que Seanoe.

7.2.3 SEANOE (Ifremer – SISMER)

SEANOE (<https://www.seanoe.org/>) est une solution de publication des données environnementales scientifiques marines créée par l'Ifremer en tant que NODC. Elle permet aux scientifiques de publier, en libre accès, un jeu de données et de le citer de manière fiable et pérenne dans un article avec un DOI. La préservation à long terme est assurée par les infrastructures de l'Ifremer. Les données sont décrites avec des métadonnées qui peuvent être converties au format Dublin-core OI-PMH (*Open archives initiative - protocol for metadata harvesting*), Data cite ou ISO 115. Chaque jeu de données est contrôlé avant d'être publié. SEANOE se réserve le droit de refuser une publication si le dépôt ne correspond pas à ses critères de qualités ou à sa thématique (les sciences marines).

7.2.4 Base de données « Vidéos Sciences Marines » (Ifremer – SISMER)

Les vidéos acquises par des engins sous-marins (ROV, HOV...) lors de campagnes sur des navires de la flotte océanographique française seront archivées dans la base de données vidéos du SISMER et disponibles sur le portail en ligne des vidéos de sciences marines (<http://video.ifremer.fr/index>). Les vidéos diffusées à travers ce portail sont organisées par campagne océanographique, par engin sous-marin, par plongée et par caméra. Ces données sont accompagnées par des métadonnées relatives aux plongées des engins sous-marins et aux conditions d'acquisition : positionnement et attitude de l'engin, données de température et salinité, position et paramétrage de la caméra. Les commentaires enregistrés en parallèle du tournage vidéo sont également accessibles lorsqu'ils existent. Les données sont déposées sur le portail après traitement et intégration en moyenne un an après la campagne.

7.2.5 Océanothèque (Ifremer – SISMER)

Les photos acquises par des engins sous-marins (ROV, HOV...) lors de campagnes sur des navires de la flotte océanographique française seront archivées sur l'Océanothèque de l'Ifremer (<https://image.ifremer.fr/search>), basée sur le principe d'accès ouvert et dont l'objectif est de faciliter l'accès à la documentation scientifique au plus grand nombre en le mettant à disposition à titre gratuit sur internet.

7.2.6 Centre de Ressources en Échantillons et Archives Marins (CREAM) – Ifremer

Les carottes de sédiments et les échantillons de roches et de minéralisations ainsi que certains échantillons biologiques sont archivés dans le Centre de Ressources en Échantillons et Archives Marins géré par l'unité mixte de recherche Geo-Ocean de l'Ifremer (Centre de Bretagne à Plouzané). Les échantillons se voient attribués un identifiant unique IGSN, maintenant relié au système Datacite, et sont recensés dans la base de données BIGOOD. Une fois toutes les études réalisées, les échantillons

⁵ <https://obis.org/>

⁶ <https://www.gbif.org/fr/>

de sédiments, roches et minéralisations prélevés dans les campagnes du projet seront archivés au sein de CREAM.

7.2.7 Archimer (Ifremer-SISMER)

Les documents tels que publications et rapports seront archivés sur l'archive institutionnelle de l'Ifremer Archimer (<https://archimer.ifremer.fr/search>) qui contient l'ensemble de la production des personnels Ifremer. Les documents collectés par Archimer comprennent également les publications citant les campagnes océanographiques françaises ainsi que les publications des UMR auxquelles l'Ifremer est associé.

La préservation à long terme est gérée par l'infrastructure Ifremer. Les documents sont ouverts et décrits par l'intermédiaire de métadonnées qui peuvent être converties au format Dublin-core OI-PMH (*Open archives initiative - protocol for metadata harvesting*) afin d'assurer leur interopérabilité avec d'autres outils. Archimer permet également de lier les documents aux jeux de données et/ou campagnes associés. Le dépôt de document sur Archimer est contrôlé et validé par le Service information scientifique et technique de l'Ifremer.

7.2.8 GenBank

GenBank est la base de données publique de séquences génétiques du NIH (National Institutes of Health) américain recensant les séquences ADN annotées. GenBank fait partie de la *International Nucleotide Sequence Database Collaboration*, qui comprend également la *DNA DataBank* du Japon (DDBJ) et la *European Nucleotide Archive* (ENA). GenBank sera utilisée pour archiver les séquences ADN des échantillons de faune.

7.2.9 Sequence Read Archive

La *Sequence Read Archive* (SRA) du NCBI (*National Centre for Biotechnology Information*), qui fait partie du NIH (*National Institutes of Health*) américain, sera utilisée pour archiver les séquences haut débit ADN en format FASTQ zippé. La préservation à long terme des séquences est assurée par le NIH et ses infrastructures. Un identifiant unique est attribué à chaque jeu de données qui peut ainsi être recherché sur le site de la SRA pour accès aux données brutes ainsi qu'aux métadonnées associées et tables attributives. Les données peuvent être rendues publiques immédiatement après validation de leur dépôt, à une date définie par le producteur de données ou après publication. La soumission de données à la SRA est contrôlée et validée par les équipes de la SRA (<https://www.ncbi.nlm.nih.gov/sra>).

7.3 Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Voir chapitre 7.2.

7.4 Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Chaque jeu de données publié dans Seanoe se verra attribué un DOI dans lequel l'auteur des données est clairement identifié. Ceci permet aux données d'être citées de manière précise et fiable dans le long

terme. Sextant et Archimer offre également la possibilité d'attribuer un DOI aux données/documents déposés.

8 Responsabilité et ressources en matière de gestion des données

8.1 Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

La gestion des données et échantillons collectés au cours des campagnes du projet relève de l'Ifremer, au travers des chefs de missions et de l'équipe scientifique du projet, assistés par le SISMER en tant que Centre National de Données Océanographiques désigné par la France et la Commission Océanographique Intergouvernementale de l'Unesco et en tant qu'un des partenaires du pôle de données et de services pour l'océan « ODATIS »⁷ de l'infrastructure de recherche data terra⁸. Dans le cadre du système qualité ISO 9001 de l'Ifremer, le SISMER pilote le processus qualité « Recueillir et mettre à disposition des données sur le milieu marin ». Il est également certifié depuis 2019 par la RDA en tant que « *Core Trustworthy Data Repositories* ».

8.2 Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Les coûts associés aux portails, plateformes et infrastructures présentés dans les chapitres précédents ne sont pas pris en charge par le projet. Par contre, le budget du projet tient compte des frais nécessaires à l'acquisition, au traitement des données (tant d'un point de vue temps personnel, frais de fonctionnement et équipements) ainsi que les coûts de publications.

⋮
⋮
⋮
⋮
⋮
⋮
⋮
⋮
⋮

⁷ <https://www.odatis-ocean.fr/en/>

⁸ <https://www.data-terra.org/>