

Plan de gestion des données

P704-08 – DeepSea'nnovation

AG. Vincent, J. Opderbecke (Ifremer)

Unité Systèmes Sous-Marins

PDG/DFO/SM/2022-0033 Indice A du 11/04/2022



Plan de gestion des données

P704-08 – DeepSea'nnovation

Diffusion : (facultatif)

Visibilité GED :

- Libre
- Restreinte SM
- Restreinte SMD
- Documentation opérationnelle
- Archiver : Internet

Référence : 2022-0033 Indice A du 11/04/2022

Résumé : Ce document constitue le plan de gestion des données pour le projet PIA3 – EQUIPEX DeepSea'nnovation n°ANR-21-ESRE-0042.

Mots-clés / Catégorie(s) / Classement GED :

Révisions

Indice	Objet	Rédigé par	Vérifié par	Validé par Responsable
A	Création du document	11/04/2022 A.G. Vincent (Ifremer) J. Opderbecke (Ifremer)	11/04/2022 Comex du projet	11/04/2022 J. Opderbecke (Ifremer)

Sommaire

1. Documents de référence	5
2. Introduction	5
3. Contexte.....	5
4. Description des données et collecte ou réutilisation de données existantes	6
4.1. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?	6
4.2. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?.....	7
5. Documentation et qualité des données	8
5.1. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?	8
5.2. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?	9
6. Stockage et sauvegarde pendant le processus de recherche	9
6.1. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?	9
6.2. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?.....	10
7. Exigences légales et éthiques, codes de conduite.....	10
7.1. 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
7.2. 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
7.3. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?	11
8. Partage des données et conservation à long terme.....	11
8.1. 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 ?.....	11
8.2. 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) ?.....	11
8.3. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?	12
8.4. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?	12

9. Responsabilités et ressources en matière de gestion des données.....	12
9.1. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?	12
9.2. 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) ?	13

1. Documents de référence

[1] Convention attributive d'aide n° ANR-21-ESRE-0042, ref. Ifremer 21/1002147

[2] Agence nationale de la recherche (ANR) : ANR - Modèle de PGD (français) - Modèle créé avec DMP OPIDoR. Version du 14/12/2020

2. Introduction

Ce document constitue le plan de gestion des données du projet DeepSea'nnovation (01/10/2021 – 21/06/2029). Il s'appuie sur le modèle fourni par l'ANR [2].

Il est dû, dans une version initiale, six mois après la signature de la convention attributive d'aide, soit au plus tard le 18 Avril 2022 [1].

Il doit être mis à jour tous les 2 ans à compter de la signature de la convention attributive d'aide. Une version révisée est donc prévue en Avril 2024, Avril 2026 et Avril 2028.

Une version mise à jour à la date de fin de projet sera fournie, en Juin 2029.

Les dispositifs décrits dans ce document sont décidés et mis en application au niveau du comité exécutif du projet.

3. Contexte

Le projet DeeSea'nnovation a été conçu pour développer des équipements et mettre en place les processus d'exploitation opérationnelle, ce qui comprend pour les équipements qui en produisent (capteurs, analyseurs, ...) la mise en place du cycle de la donnée depuis l'acquisition jusqu'à l'exploitation scientifique.

Un des objectifs du projet est donc de définir les processus pour automatiser la chaîne de traitement afin que les données acquises à bord des navires par les équipements issus du projet soient archivées et mises à disposition de la communauté scientifique : acquisition de la donnée, traitement de la donnée pour la calibration et validation, distribution et pérennisation de la donnée pour être le plus conforme possible aux principes FAIR.

L'ensemble du processus sera adapté pour chaque équipement.

Les différents équipements ou types d'équipement sont les suivants :

- Lot 1.1 Imagerie acoustique colonne d'eau (Sondeur et ADCP)
- Lot 1.2 Caméra bioluminescence
- Lot 1.3 Caméras zooplancton
- Lot 1.4 Lidar sous-marin
- Lot 1.5 Imageur géophysique CSEM
- Lot 2.1 Spectromètre de masse (mesure des gaz dissous)
- Lot 2.2 Chambre benthique
- Lot 2.3 Profileur benthique
- Lot 2.4 Sonde de gradient de température
- Lot 2.5 Rhizons
- Lot 3.1 Foreuse à roche
- Lot 3.2 Vibro-carottier
- Lot 3.3 Main de prélèvement

- Lot 3.4 Préleveur à larves et à plancton
- Lot 3.5 Préleveur ADN environnemental

Chaque lot est générateur de données de différents types. Certains types de données sont communs à tous les lots, alors que d'autres (données issues des équipements) sont spécifiques à chaque lot. Nous considérons dans ce document l'ensemble des données générées par les lots.

Les données concernées dans ce document sont donc :

1. les données numériques produites par les capteurs développés dans le cadre du projet, enrichies de leurs métadonnées (navigation, calibration, etc) ;
2. les données numériques liées aux échantillons prélevés par les équipements développés dans le cadre du projet lorsqu'elles ont un caractère suffisamment générique (par ex : données liées aux carottes) ;
3. la matière grise produite dans le cadre du projet : actes de colloques, rapports techniques à intérêt général (par ex : description des équipements, guides opérateur, rapports de qualification, ...), ... ;
4. les publications à comité de lecture.

Le projet comprend deux phases principales : phase de développement et phase de mise en exploitation. La gestion des données peut couvrir des données générées en phase de prototypage et développement, elle caractérisera la nature expérimentale et la qualité spécifique de ces données.

Compte-tenu de la variété des données concernées, ce document constitue un plan de gestion des données ainsi que des objets numériques issus du projet DeeSea'nnovation et ne se limite pas aux données *stricto sensu* telles qu'envisagées dans les plans de gestion des données classiques. En effet ce projet a pour objectif de développer de nouveaux instruments qui seront décrits via la matière grise produite ce qui nécessite d'étendre le plan de gestion de données aux objets numériques.

Les différents paragraphes suivants sont abordés d'un point de vue général. Ils seront précisés et déclinés par la suite pour chaque lot, i.e. pour chaque équipement générateur de données spécifiques.

4. Description des données et collecte ou réutilisation de données existantes

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

Des données préexistantes pourront être utilisées dans le cadre du projet, en tant que données exemple ou données de référence pour des équipements similaires (terrestres, observatoires, laboratoires, ...). Ces données seront utilisées en respectant les règles de confidentialité/utilisation propres à chaque jeu de données.

Les données produites dans le cadre du projet peuvent être classées selon la phase du projet où elles sont produites :

1. Données de prototype au cours du cycle de développement,
2. Données exemples suite aux recettes techniques et qualifications,
3. Données des missions en mer en phase d'exploitation.

En fonction de la phase du projet, les méthodologies ainsi que le processus de collecte et de mise à disposition pourront évoluer.

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

Comme précisé, ce plan de gestion des données considère :

- les données numériques produites par les capteurs ;
- les données numériques liées aux échantillons prélevés ;
- la matière grise (actes de colloques, rapports techniques à intérêt général, ...) ;
- les publications.

Si l'on exclue la matière grise et les publications, le tableau modèle ci-dessous recense les données produites ou collectées par les différents équipements. Il sera complété au fil de la définition technique des équipements, au cours de la phase de développement.

	Type de données	Métadonnées associées	Format	Volume
Lot 1.1	Imagerie acoustique colonne d'eau			
Lot 1.2	Caméra bioluminescence			
Lot 1.3	Caméra zooplancton			
Lot 1.4	Lidar sous-marin			
Lot 1.5	Imagerie géophysique CSEM			
Lot 2.1	Mesure des gaz dissous			
Lot 2.2	Chambre benthique			
Lot 2.3	Profileur benthique			
Lot 2.4	Sonde de gradient de température			
Lot 2.5	Rhizons			

Lot 3.1 Foreuse à roche				
Lot 3.2 Vibro-carottier				
Lot 3.3 Main de prélèvement				
Lot 3.4 Préleveur à larves et plancton				
Lot 3.5 Préleveur ADN environnemental				

5. Documentation et qualité des données

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

Une fiche de donnée et de leurs métadonnées sera associée à chaque équipement ; la trame sera donnée dans ce document. Elle précisera également les conventions de nommage des différents fichiers afin de retrouver sans ambiguïté les métadonnées associées aux données.

Les métadonnées génériques suivront les standards utilisés dans la communauté lorsqu'ils sont déjà définis (datation, navigation, paramètres d'environnement, ...) et on suivra les recommandations du pôle Odatis pour les conventions, formats et attributs : <https://www.odatis-ocean.fr/donnees-et-services/principes-de-gestion-des-donnees/formats-attributs-conventions>.

Dans le cas de nouveaux types de données ou métadonnées vouées à servir la communauté, le pôle Odatis pourra être consulté afin de proposer des standards via des groupes d'experts constitués pour une thématique donnée et ingérer dans les serveurs de vocabulaire nationaux et internationaux les nouveaux types de données, avec leurs définitions, leur protocole d'obtention et la méthodologie de traitement de la nouvelle données associés.

	Métadonnées	Format
Lot 1.1 Imagerie acoustique colonne d'eau		
Lot 1.2 Caméra bioluminescence		
Lot 1.3 Caméra zooplancton		

Lot 1.4 Lidar sous-marin		
Lot 1.5 Imagerie géophysique CSEM		
Lot 2.1 Mesure des gaz dissous		
Lot 2.2 Chambre benthique		
Lot 2.3 Profileur benthique		
Lot 2.4 Sonde de gradient de température		
Lot 2.5 Rhizons		
Lot 3.1 Foreuse à roche		
Lot 3.2 Vibro-carottier		
Lot 3.3 Main de prélèvement		
Lot 3.4 Préleveur à larves et plancton		
Lot 3.5 Préleveur ADN environnemental		

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

Pour chaque équipement, une procédure de calibration et un échéancier seront fournis : fiche de calibration, état de calibration, fiche de contrôle.

Une procédure de contrôle de la qualité des données sera fournie. Elle pourra s'appuyer sur les pratiques mises en place au SISMER pour les données acquises par les engins sous-marins lors des campagnes en mer.

Dans la phase d'exploitation, l'état d'entretien et de calibration et les historiques devront être consignés et actualisés pour chaque équipement dans une base de données de référence.

6. Stockage et sauvegarde pendant le processus de recherche

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

En fonction de la phase du projet (développement ou exploitation), le processus de stockage et de sauvegarde sera adapté.

Pendant la phase de développement, au niveau des lots :

- Un stockage local est assuré.
- Un répertoire projet est mis en place sur une plateforme cloud de l'Ifremer, avec une arborescence permettant à chaque lot d'organiser ses données dans un répertoire approprié.

Les données à intérêt pour la communauté sont stockées et sauvegardées sur le cloud du projet (cloud de l'Ifremer).

Le choix du cloud de l'Ifremer permet de s'appuyer sur les outils de l'institut et de s'assurer des bonnes pratiques en termes de stockage et de sauvegarde.

Pendant la phase d'exploitation, les données collectées pendant les missions en mer sont transmises aux centres de données et de service qui auront été identifiés (par exemple, le SISMER). La description de la procédure de stockage et sauvegarde des données pendant la phase d'exploitation fait l'objet du projet lui-même et sera détaillée au fur et à mesure de l'avancement du projet. La procédure d'intégration des données collectées sera mise en place en concertation avec le pôle Odatis et ses centres de données et de services.

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

Pendant la phase de développement, les données et métadonnées seront stockées sur le cloud du projet, hébergé par l'Ifremer ce qui offre les garanties nécessaires en termes de stockage et sauvegarde des données, ou par les équipes qui s'assureront d'une sauvegarde à intervalle régulier afin d'éviter toute perte de données pendant le processus de recherche.

Lors de la mise en exploitation des équipements, la gestion des données sera en conformité avec la procédure qui aura été définie (qui peut varier d'un équipement à l'autre). La gestion des données s'appuiera autant que faire se peut sur le pôle Odatis qui fédère au niveau national des activités de gestion de données et d'expertise scientifique en océanographie, ainsi que sur les centres de données et de services existants (SISMER, SEANOE, SEXTANT par exemple), ce qui permet de garantir les protections requises pour le stockage et la sauvegarde des données.

Les cas d'exception peuvent donner lieu à des principes d'accès avec des modalités de contrôle adaptés.

7. Exigences légales et éthiques, codes de conduite

7.1. 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 ?

Il n'y a pas de données à caractère personnel. Les cas particuliers, le cas échéant, seront traités à part.

7.2. 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 ?

Les données seront publiées sous Licence Creative Commons, à définir ultérieurement.

Les aspects de propriété intellectuelle sont traités dans l'accord de consortium.

Les cas particuliers seront traités à part.

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

Chacun des participants au projet est tenu de respecter les chartes d'éthique et de déontologie de son organisme d'affiliation dont les principes sont définis par la Charte nationale de déontologie des métiers de la recherche.

Si une mise en concurrence est nécessaire, la question est traitée au niveau du comité directeur du projet.

8. Partage des données et conservation à long terme

8.1. 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 ?

La matière grise et les publications seront déposées dans des archives ouvertes libres d'accès, HAL ou Archimer en fonction de l'institut de rattachement de l'auteur principal.

Pendant la phase de développement, les données issues des capteurs seront stockées et sauvegardées. Le partage à destination d'une communauté scientifique plus large sera géré au cas par cas.

Pendant la phase d'exploitation, les données issues des équipements développées dans le cadre du projet seront déposées dans des centres de données et de services du pôle Odatis, afin de permettre un partage pérenne (SISMER, SEANOE, SEXTANT, ...).

La distribution se fera via le catalogue Odatis et le SISMER.

Les cas particuliers seront traités à part.

8.2. 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) ?

En phase de développement, le choix des données à conserver relèvera de la responsabilité des responsables de lots, selon des principes décidés par le comité exécutif du projet. Durant cette première phase du projet, le processus d'exploitation sera défini pour chaque équipement.

En phase d'exploitation, les données seront préservées sur le long terme en optant pour des centres de données et de services fiables, fédérés au niveau national ayant une certification CoreTRUSTSeal. Nous nous appuyerons sur le pôle Odatis dont l'objectif est de mettre à disposition des données, des produits, des logiciels, des outils et /ou des services destinés principalement à la communauté scientifique française travaillant dans le domaine de la recherche océanographique. Le pôle pourra proposer les entrepôts adéquats en fonction du schéma de mise en exploitation retenu selon les spécificités des instruments développés dans le cadre du projet.

Cette partie sera donc déclinée pour chaque lot / instrument.

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

Dans la mesure du possible, les données et métadonnées associées seront proposées sous un format standard ou générique (CSV, NetCDF, ...), qui sera documenté.

Dans le cas de données spécifiques, des logiciels open source ou des fonctions logicielles (pilotes de lecture ou décodage, ...) sont nécessaires pour exploiter les données, ils seront mis à disposition.

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

Un numéro d'identification unique fourni par les plateformes de sauvegarde, ou un DOI, sera attribué à tous les objets numériques destinés à être sauvegardés et partagés.

Le tableau ci-dessous recense les différents types de données générées par le projet, l'entrepôt visé et quel identifiant sera attribué. A noter, toutes les campagnes en mer sont déjà répertoriées par le SISMER, avec un DOI.

Type de donnée	Entrepôt	Identifiant unique pérenne
Publications	HAL ou Archimer	DOI
Matière grise	HAL ou Archimer	Référence interne ou DOI
Jeux de données	SISMER (ou autre Centre de Services) ou pôle Odatis	DOI
Données numériques liées aux échantillons	Pôle Odatis	DOI
Campagnes en mer	SISMER	DOI

Afin d'assurer la valorisation du projet, un DOI sera attribué au projet et à chaque équipement, via des fiches mises à disposition sur une archive ouverte. Les jeux de données produits, les échantillons, les publications et la matière grise y feront référence.

On s'attachera à automatiser dans la mesure du possible le lien entre les objets numériques. A l'heure actuelle, seul le lien entre les campagnes en mer et les publications est automatisé.

9. Responsabilités et ressources en matière de gestion des données

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

Pendant la phase de développement, les responsables des lots sont responsables de la gestion des données produites par leurs équipes respectives.

Pendant la phase d'exploitation du projet et au-delà du projet, en fonction du schéma de mise en exploitation retenu, la responsabilité de la gestion des données incombera au chercheur principal et chef de mission des campagnes océanographiques ou à l'organisme chargé de la mise en œuvre de l'équipement (Genavir par exemple).

9.2. 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) ?

Le projet s'appuie sur les centres de données et de services (SISMER, SEXTANT, SEANOE, ...) du pôle national de données et de services pour l'océan Odatis. Ce pôle et ses centres de données assurent les ressources en termes de budget et de temps alloués nécessaires à leurs missions, en particulier, selon le principe FAIR.