



Atelier IR*FOF/Odatis « Données de la FOF »

Compte-rendu de l'atelier IR*FOF/Odatis autour des données de la FOF - 15 juin 2023



Titre court	
Atelier IR*FOF/Odatis « Données de la FOF »	
Titre long	
Compte-rendu de l'atelier IR*FOF/Odatis autour des données de la FOF - 15 juin 2023	
Auteurs	
Marine Vernet, Caroline Mercier, Cécile Pertuisot, Erwann Quimbert, Sabine Schmidt, Goulwen Peletier, Christine David-Beausire, Bernard Dennielou, Anne-Gaëlle Vincent, Isabelle Billy, Fabien Arnaud, Sylvie Van Iseghem, Olivier Norvez	
Dissémination	Copyright
Publique	©ODATIS

Historique

Version	Auteurs	Date	Commentaires
0.1	Marine Vernet, Caroline Mercier	15/06/2023	Notes
0.2	Marine Vernet	10/07/2023	Mise au propre
0.3	Cécile Pertuisot, Erwann Quimbert, Caroline Mercier, Sabine Schmidt	12/07/2023	Relecture
0.4	Goulwen Peletier, Christine David-Beausire, Bernard Dennielou, Anne-Gaëlle Vincent, Isabelle Billy, Fabien Arnaud, Sylvie Van Iseghem, Olivier Norvez	25/07/2023	Relecture
0.5	Marine Vernet	14/09/2023	Version finale



Table des matières

1. Organisation	4
2. Présentations	7
3. Discussions ouvertes.....	8
3.1. Procédures réglementaires.....	8
3.1.1. Réglementation d'acquisition et de diffusion des données issues de campagnes en mer.....	8
3.1.2. Moratoire sur les données.....	9
3.2. Services supports aux campagnes et d'aide à la traçabilité.....	9
3.2.1. Outil agrégateur pour les données issues de prélèvements.....	9
3.2.2. Outil d'aide à la préparation de campagne.....	10
3.3. Workflow de données et PGD.....	11
3.3.1. Instruments & acquisition.....	11
3.3.2. Workflow de données des navires de stations.....	12
3.3.3. Découpage du Plan de Gestion de Données IR*FOF.....	12
4. Relevé d'actions et de décisions	13

Liste des acronymes

AAP	Appel à projets
ANR	Agence Nationale de la Recherche
CSR	Cruise Summary Report
DMP	Data Management Plan
EBV	Essential Biodiversity Variables
ECV	Essential Climate Variables
EOV	Essential Ocean Variables
ERC	European Research Council
FOF	Flotte Océanographique Française
INSU	Institut national des sciences de l'Univers
IPEV	Institut polaire français Paul-Émile Victor
IR	Infrastructure de Recherche
IR*	Infrastructure de recherche "étoile"
ISI	Service Ingénierie des Systèmes d'Information
OASU	Observatoire Aquitain des Sciences de l'Univers
PGD	Plan de Gestion des Données
PID	Persistent identifier
PNDB	Pôle National de Données de Biodiversité
RDA	Research Data Alliance
SOSA	Sensor, Observation, Sample, and Actuator
SGC	Système de Gestion des Campagnes
SNO	Service National d'Observation
WMO	World Meteorological Organization
ZEE	Zone Économique Exclusive



1. Organisation

Date : Jeudi 15 juin 2023

Lieu : Salle Pascal Le Mauguen, IPEV, Plouzané & Visioconférence Webex

Ordre du jour :

1. Introduction et tour de table

2. Présentations :

- IR* FOF et ses processus de données
- Workflow de données issues de campagnes en mer IR*FOF
- Workflow de données issues de campagnes en mer relevant de l'AAP INSU Lefe-Cyber
- Workflow de données issues de campagnes de type échantillons
- Cybercarothèque nationale et ses processus de données
- Présentation du PNDB et des services support à la gestion et utilisation des données de biodiversité
- Workflow de données issues des pêcheries australes
- Présentation du pôle Océan Odatis et de ses services dédiés au support aux campagnes
- Présentation des outils supports aux campagnes INSITUDE et SPASSO

3. Discussions ouvertes :

- Aspects réglementaires
- Services support aux campagnes et outils d'aide à la traçabilité et bancarisation
- Workflow de données et Plan de Gestion de Données

4. Conclusions et plan d'action



15 juin 2023 - Brest **Atelier Technique #15**

Données de la Flotte Océanographique Française

- > Workflow des données depuis la préparation de la campagne jusqu'à la bancarisation et la diffusion
- > Aspects réglementaires, documentaires et procédures
- > Plan d'action pour élaborer un plan de gestion de données commun
- > Sujets d'intérêt autour des questions de réglementation
- > Sujets d'intérêt autour des services supports aux campagnes

  

Participants :

NOM	Prénom	Organisme
ABEL-MICHAUX	Karine	GENAVIR
ANDRE	François	OMP
ARNAUD	Fabien	CNRS
BILLY	Isabelle	OASU/U.Bordeaux
BORREMANS	Catherine	IFREMER
CALVAT	Pascal	CNRS - OASU
CARIOU	Valérie	SHOM
CARVAL	Thierry	IFREMER
COTTEN	Clémence	IFREMER
CROSTA	Xavier	OASU/U.Bordeaux
DANIOUX	Nowlenn	IFREMER
DAVID-BEAUSIRE	Christine	IFREMER (IR* FOF)
DENNIELOU	Bernard	IFREMER (Geo-Ocean)
DIBARBOURE	Gérald	CNES
DOUET	Vincent	IPSL
DUTRION	Thibaut	IFREMER
FOUCHER	Eric	IFREMER
GASCO	Nicolas	MNHN - UMS Borea
GERMINEAUD	Cyril	CNES
HARSCOAT	Valérie	IFREMER
HOEBEKE	Mark	SU
KHVOROSTYANOV	Dimitry	CNRS - IPSL
LE HINGRAT	Françoise	IFREMER
LEROY	Sylvie	Sorbonne Université
MERCIER	Caroline	CNES
NORVEZ	Olivier	MNHN & FRB (PNDB)
PAYAN	Sébastien	SU
PELLICHERO	Violaine	SHOM

PELTIER	Goulwen	IFREMER (IR* FOF)
PERTUISOT	Cécile	IFREMER
PEUCH	Alexis	IFREMER (IR* FOF)
PIGNOL	Cécile	Univ- Savoie
PIOLLE	Jean-François	IFREMER
PONCELET	Cyrille	IFREMER (IR* FOF)
QUIMBERT	Erwann	IFREMER
ROUSSELET	Louise	LOCEAN/IPSL
SCHMECHTIG	Catherine	SU
SCHMIDT	Sabine	CNRS - OASU
SUDRE	Joël	CNRS - UMS CPST
TOSELLO	Vanessa	IFREMER
VAN ISEGHEM	Sylvie	IFREMER
VERNET	Marine	IFREMER
VINCENT	Anne-Gaëlle	IFREMER (IR* FOF)



2. Présentations

Différentes présentations ont été réalisées durant la matinée autour des points d'intérêts suivants :

- Présentation de l'IR*FOF et du processus des données de l'IR* FOF
- Workflow de données issues de campagnes en mer au SISMER, pour les données en route et les données issues de prélèvements biologiques et géologiques
- Workflow de données issues des campagnes en mer LEFE-CYBER, CoreBook/Cybercarothèque
- Présentation du Pôle National de Biodiversité et du workflow de données de biodiversité acquises lors de campagne,
- Présentation du pôle Océan OdatiS et de ses services,
- Services supports aux campagnes (logiciels SPASSO, INSITUDE).

Toutes les présentations sont disponibles en ligne sur la page du site OdatiS dédiée à l'atelier : <https://www.odatis-ocean.fr/activites/ateliers-techniques/atelier-technique-juin-2023-donnees-fof>

Ces présentations ont permis de présenter les acteurs, leurs services ainsi que les workflows existants de données issues de campagnes entrant dans le périmètre de l'IR*FOF, que ce soit pour les données collectées par les équipements de la FOF et celles collectées avec des équipements et analyses ultérieures propres aux unités de recherche, et les spécificités des différents types de données.

Elles ont également fait état d'un certain nombre d'outils existants, avec leurs fonctionnalités propres, souvent thématiques ou institutionnelles pour les outils de traçabilité des instruments.

Elles ont enfin fait remonter un certain nombre de besoins, qui ont servi de base à trois discussions libres synthétisées ci-dessous.



3. Discussions ouvertes

3.1. Procédures réglementaires

3.1.1. Réglementation d'acquisition et de diffusion des données issues de campagnes en mer

Une campagne océanographique traversant des ZEE étrangères doit recevoir l'**autorisation d'acquisition** de données dans ces ZEE auprès des états auxquelles elles sont rattachées. Cette condition ne concerne pas les données de navigation.

Cette demande d'autorisation d'acquisition est portée pour l'IR* FOF par la DFO auprès du ministère des Affaires Étrangères, qui l'adresse aux ambassades des pays concernés. La note verbale diplomatique reçue en retour est souvent teintée de conditions (limitation sur certains jeux de données, conditions d'acquisition particulières...). **Les données de routine ne sont pas incluses dans la demande d'autorisation.**

Cependant, bien que l'acquisition ait été autorisée par l'autorité du pays, la diffusion des données peut être restreinte.

Afin d'étudier l'application de la loi Open Data de 2016, un groupe de travail Politique de diffusion des données a été créé en 2018 à l'Ifremer. Il a abouti sur un guide pratique, en cours de validation. Mais quelques points restent à éclaircir, notamment concernant les licences de diffusion, mais également concernant la **diffusion de données « sensibles » ou à intérêt économique**. C'est également le cas des données acquises en routine, comme celle de Coriolis par exemple.

La convention de Montego Bay, traité international de 1982 est signé et appliqué par les états. Cette convention prime sur l'ensemble des autres politiques de données, y compris celles du WMO et de la Commission Océanographique Internationale, même si celles-ci prônent l'ouverture des données « core data » (données associées aux variables essentielles, EBV, EOVS & ECV) par défaut. La convention de Montego Bay stipule que **les données environnementales appartiennent au pays côtier**. L'accord de diffusion des données doit être demandé au pays qui les possède.

Cela pose problème car le volume de données concerné est important. Ces données sont donc bancarisées mais jamais diffusées faute d'obtention d'autorisation de l'état propriétaire.

Une piste serait de joindre la demande **d'autorisation de diffusion des données environnementales à la demande d'autorisation d'acquisition** validée par le pays propriétaire de la ZEE. Il faudrait également y **inclure l'autorisation d'acquisition des données de routine**.

Le groupe de travail sur la Politique de diffusion des données de l'Ifremer essaie également d'identifier d'autres pays en situation similaires (navires traversant les ZEE de plusieurs pays) afin d'étudier leurs pratiques. Pour les données hydrographiques, une piste est identifiée auprès du service international hydrographique (IHO) pour essayer de simplifier les processus.

Enfin, il est précisé que la non-acquisition n'est pas forcément possible selon les capteurs.



3.1.2. Moratoire sur les données

Le moratoire débute au moment où la donnée est qualifiée et scientifiquement utilisable. La définition laisse voie à interprétation. Selon le type de données, le temps d'acquisition et d'analyse peut varier, et donc le moratoire doit être adapté. Cela doit être précisé par les scientifiques dans le Plan de Gestion de Données.

Les moratoires peuvent différer entre les différents organismes, et en fonction du type de données. En général, le moratoire est justifié par le temps nécessaire à l'exploitation des données par les scientifiques.

Une **expertise juridique peut être sollicitée** si nécessaire auprès de l'**unité AMURE** (Marie Guizemanes, Thomas Leclerc). Un autre contact au sein de l'**université de Montpellier** pouvant être sollicité est celui d'Agnès Robin, juriste spécialiste en droit des données de recherche.

3.2. Services supports aux campagnes et d'aide à la traçabilité

3.2.1. Outil agrégateur pour les données issues de prélèvements

Concernant les **données d'échantillons, issues de prélèvements**, il existe une **pluralité d'outils, thématiques ou non, couvrant chacun différentes fonctionnalités** : préparation de la campagne, saisie des informations à bord (description des opérations réalisées, identification des données acquises), collecte des données à bord et transmission des données pendant ou après la campagne.

La pluralité d'outils, pas forcément interopérables, induit des **difficultés de collecte des données et des informations à leur sujet**, ce qui **complique la bancarisation** de l'ensemble des données issues des campagnes.

Compte-tenu des besoins spécifiques des différents acteurs et disciplines de recherche, un outil unique n'est pas envisageable. Cependant, afin de simplifier le processus, et permettre un workflow complet, **un outil de type agrégateur**, pouvant être utilisé en ligne par toute équipe scientifique, avant, pendant et après la mission, serait une solution. Cela pourrait prendre la forme d'une **interface usager unique, connectant un ensemble d'outils et briques existantes** permettant la préparation de la campagne, la **description des opérations et données acquises** au cours de la campagne, ainsi que leur **validation par les équipes scientifiques** et leur **transmission auprès des services** dédiés.

Cela simplifierait également la saisie du rapport de campagne pour les scientifiques.

Les **outils thématiques Sealog (outil FOF, avec collaboration SISMER pour le suivi du besoin utilisateur et l'intégration en base de données centrale) et CoreBook (outil CNRS, pour les prélèvements de type « carottes »)** pourraient servir de **base de travail**, à transférer dans le périmètre plus large de la FOF.

Sealog permet l'agrégation de métadonnées collectées par différents logiciels (Mimosa, Casino) et l'ajout de métadonnées, mais est chronophage et nécessite du travail de mapping pour une



bancarisation effective. CoreBook est accessible par le chef de mission en amont pour préparer et en aval pour finaliser la mission, et permet également la traçabilité des données collectées grâce à un service d'étiquetage des carottes pouvant être utilisé à bord. Cependant l'outil ne couvre que le carottage. Il est indépendant du système informatique des navires, il n'est pas connecté avec les données et informations venant des capteurs des navires et récupérables par Casino et Sealog. (Voir présentations de la [CyberCarothèque](#) et du [Workflow de données issues de prélèvements](#)).

La réflexion pourrait être portée non pas par type de métiers mais par type d'opérations :

- Opérations de prélèvement sur le fond de mer à partir du navire (carottage, dragage)
- Prélèvements biologiques et géologiques faits par submersibles
- Prélèvements faits dans la colonne d'eau

Il est primordial, **pour avoir une adhésion des chefs de missions et scientifiques, d'avoir un outil facile d'utilisation leur apportant une plus-value.** Ainsi, cet outil doit également apporter une aide à la production du rapport de mission. Il est également précisé qu'un effort de communication et de formation des scientifiques en amont serait nécessaire. Une attention devra être portée à l'interface de saisie mais également à celle de restitution pour favoriser l'utilisation de l'outil. Enfin, afin d'atteindre ces objectifs, il est nécessaire d'inscrire cet effort dans une démarche de co-construction avec les différents acteurs et utilisateurs.

3.2.2. Outil d'aide à la préparation de campagne

La base de données MADIDA référence les navires, systèmes sous-marins, équipements mobiles confiés en gestion à GENAVIR par Ifremer, opérateur de l'IR* FOF, et leurs utilisations. Il existe également un outil de recensement et de gestions des équipements à l'Ifremer (base P6) et un autre à la DT INSU. L'ouverture de ce type de base pourrait être envisagée pour les équipements des laboratoires sans base de données propre et dont la récupération des données est compliquée. Pour les outils de carottage, par exemple, la cyber-carothèque collecte les informations des outils gérés par les laboratoires. Ce catalogue pourrait être interrogé par l'outil de préparation de missions.

Cela permettrait, à travers un outil proposé aux scientifiques avant la campagne, de sélectionner à partir des bases d'instruments et catalogues existants (MADIDA, outil « P6 » Ifremer, etc ...), les instruments à utiliser (ex : CTD), les paramètres et les opérations.

Ensuite, pendant la campagne, il serait possible de saisir, de façon plus précise, les opérations réellement réalisées et les instruments réellement utilisés (ex : quel type de CTD). Ce qui faciliterait également la production du CSR dû par le chef de mission après la campagne.

L'idéal serait que cet **outil soit intégré à l'interface usager** suggérée ci-dessus.

Actions identifiées - 3.2 Services supports aux campagnes et d'aide à la traçabilité

- Inventorier l'ensemble des outils qui existent à la fois pour la description de la mission, l'acquisition et agrégation des données, le support à bord, et également le post-traitement.

- Identifier les services facilement utilisables. Cette demande pourrait peut-être être incluse dans la demande de campagne, en permettant aux chefs de mission de cocher les supports/outils dont ils auraient besoin.
- Identifier les bases de données cibles et leurs prérequis.
- Identifier les besoins des différentes disciplines et l'ergonomie souhaitée.
- Identifier les acteurs et le temps de développement souhaité.

3.3. Workflow de données et PGD

3.3.1. Instruments & acquisition

Une distinction doit être faite entre les données collectées à partir des équipements communs de l'IR* FOF, et celles acquises à partir d'équipements propres aux laboratoires de recherche, dont la description et la bancarisation sont plus hétérogènes.

Les **différentes bases de données** détaillées dans le 3.2.2 (MADIDA pour les équipements/instruments de l'IR*FOF et leurs utilisations, base P6 pour les capteurs Ifremer, outil de traçabilité des capteurs et instruments proposé par la DT INSU pour les unités de recherche) **contiennent des informations sur les instruments et capteurs nécessaires à la constitution d'un plan de gestion de données.**

Il faudrait donc qu'elles soient **aisément disponibles, identifiées et décrites de façon harmonisée.**

Par exemple, il serait intéressant d'avoir **un PID pour chaque instrument/capteur**, afin d'avoir un identifiant unique et une description de ce capteur. Une **initiative « PID instrument »** est en cours, coordonnée **par un groupe de travail RDA** ayant publié un schéma de métadonnée de description, en lien avec l'ontologie **SOSA**.

Un second travail serait d'**améliorer l'interopérabilité de ces différentes bases.**

Cela permettrait, en préparation de la campagne, de pré-saisir de façon standardisée dans l'outil dédié (décrit dans le 3.2) les instruments et capteurs embarqués en vue d'utilisation. Le **niveau de granularité et les métadonnées nécessaires restent à définir** en fonction des besoins.

Actions identifiées - 3.3.1 Instruments & acquisition

- Identifier où en est l'initiative « PID instruments » de la RDA.
- Travail des responsables des bases de données instruments identifiées (MADIDA, P6, base d'une unité DT INSU) pour étudier l'implémentation de PID de façon harmonisée, en lien avec quelqu'un de l'IR Data Terra connaissant l'ontologie SOSA.

3.3.2. Workflow de données des navires de stations

Les **navires de stations**, armés par la DT INSU, sont néanmoins **dans le périmètre de l'IR* FOF** et les données acquises à bord doivent rentrer le processus des données de l'IR*FOF.

Pour les données acquises sur ces navires avec les équipements des laboratoires, cela peut s'avérer compliqué au vu du nombre d'opérations réalisées. Toutefois il y a des sorties récurrentes comme celles menées dans le cadre de SNOs qui pourraient faire l'objet de DOIisation.

Pour les **données acquises par les capteurs installés à bord de ces navires**, comme les TSG et sondes de température, la **bancarisation n'est pas effectuée** ou en tout cas **pas connue de l'IR* FOF**.

Le SISMER ne collecte que les données qualifiées par GENAVIR, qui permet également le découpage des données par mission, ce qui n'est pas le cas pour les données issues des capteurs des navires de station. Une première étape pourrait être de **DOIifier les campagnes réalisées**, cependant les navires de stations ne passent pas par le SGC mais par un outil mis en œuvre par le CNRS. La DOIification pourrait être réalisée par le SISMER, mais il faudrait définir un nouveau mode opératoire avec la DT INSU. Les points de contacts au SISMER sont Nolwenn Danioux (SGC) et Fred Merceur (ISI).

Actions identifiées - 3.3.2 Workflow des données de navires de stations

- Vérifier auprès des SNO et IR (comme l'IR ILICO) le besoin de DOIification des campagnes,
- Identifier le processus actuel et les différents acteurs.

3.3.3. Découpage du Plan de Gestion de Données IR*FOF

Les campagnes réalisées sur financement ANR/ERC ont leur propres PGD. Pour autant l'IR*FOF doit être capable de dire comment vont être gérées les données acquises par ces campagnes. Cela pose également des problèmes liés aux conditions d'accès et de diffusion qui peuvent différer selon les campagnes et les organismes d'affiliation des chefs de mission.

Le plus simple serait que **l'IR* FOF définisse, pour les données acquises sur équipements de l'IR* FOF, une position de principe**, à laquelle **les scientifiques impliqués dans une campagne puissent faire référence** pour le PGD de leur campagne.

La plateforme **DMP OPIDOR** serait à recommander car elle permet de faire référence à un autre PGD.

Actions identifiées - 3.3.3 Découpage du Plan de Gestion de Données IR*FOF

- Transmission de modèles de PGD pertinents pour la FOF par ODATIS et Data Terra (référents possibles Cécile Pertuisot (SISMER/Pôle Données des Campagnes), Erwann Quimbert (SISMER/ODATIS), Joël Sudre (Data Terra))

4. Relevé d'actions et de décisions

Création de groupes de travail dédiés :

GT instruments dans la phase d'acquisition :

- Coordinateur : Joël Sudre (CNRS/Data Terra)
- Participation Thierry Carval (Ifremer/Coriolis), Karine Abel-Michau (Ifremer/GENAVIR), Cyril Poncelet (Ifremer/NSE), Fabien Arnaud (Parc national d'instruments Carottage Continental France et Cyber-Carothèque nationale)
- Objectifs :
 - Harmonisation des identifiants et de la description des instruments et interopérabilité des bases existantes.
 - Réflexion sur le niveau de granularité à appliquer et les champs descriptifs nécessaires

GT Utilisateurs outil agrégateur pour les données issues de prélèvement :

- Coordinateur : Bernard Dennielou (Ifremer/GEO-OCEAN)
- Participants : à définir, certaines personnes déjà identifiées initialement par Bernard Dennielou. Il est nécessaire d'inclure des représentants de la Cyber-carothèque ainsi que de l'IR*FOF (NSE et SM pour les logiciels).
- Objectifs :
 - Identification des besoins utilisateurs pour la création d'un outil agrégateur
 - Identification des outils existants et simples d'utilisation, en lien avec les besoins & bases de données cibles

GT Plan de gestion de données de l'IR*FOF :

- Coordinateur : Cécile Pertuisot (Ifremer/SISMER), Erwann Quimbert (Ifremer/Odatis)
- Participants : Joël Sudre (Data Terra), Christine David-Beausire (DFO), à définir
- Définition d'un PGD de l'IR*FOF pour les données acquises par les équipements de la FOF
- Contacts pour modèles de PGD et support : Joël Sudre, Cécile Pertuisot, Erwann Quimbert

Échéances :

- Les GT doivent être formalisés et débuter en septembre 2023. Cette formalisation peut prendre la forme d'un document précisant: le nom du GT, le contexte de sa création (besoin, motivation), les objectifs, les actions envisagées et la liste des personnes identifiées pour travailler dans ce GT.
- Une seconde réunion IR*FOF/Odatis avec des restitutions des groupes de travail est prévue en juin 2024

