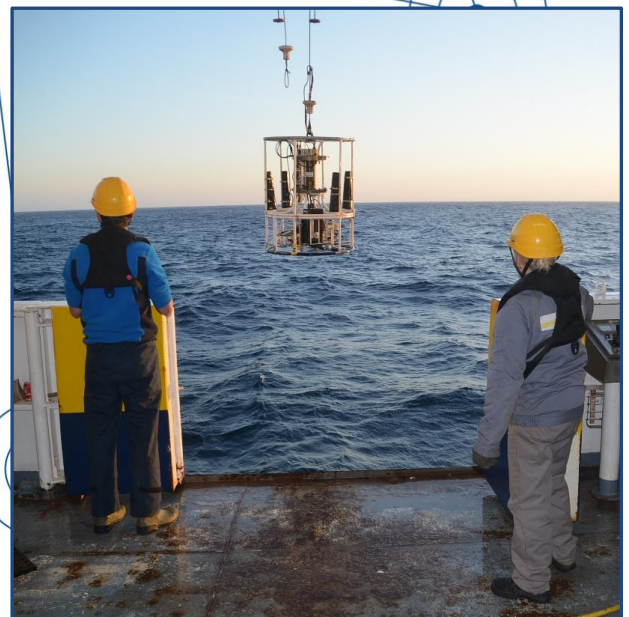
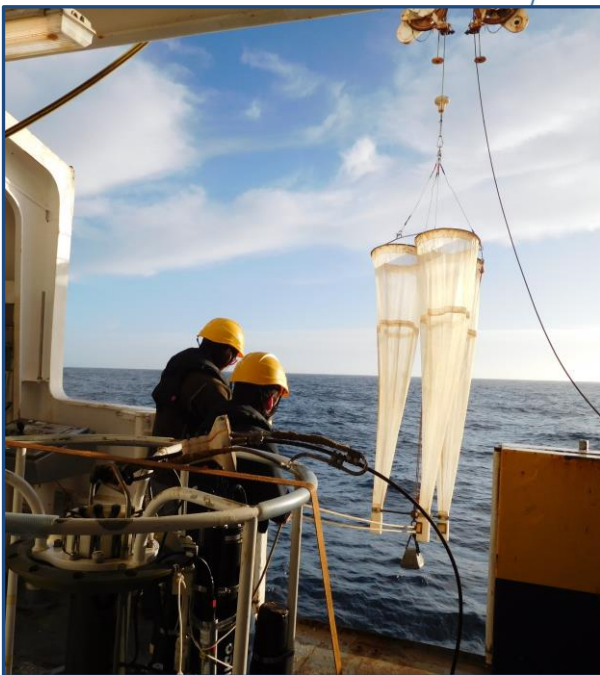


BANCARISATION DES DONNEES HYDROLOGIQUES DES CAMPAGNES HALIEUTIQUES

Proposition d'harmonisation et de restitution des données au titre de la DCSMM



BANCARISATION DES DONNEES HYDROLOGIQUES DES CAMPAGNES HALIEUTIQUES

Proposition d'harmonisation et de restitution des données au titre de la DCSMM

Fiche documentaire

| | |
|--|---|
| Titre du rapport : BANCARISATION DES DONNEES HYDROLOGIQUES DES CAMPAGNES HALIEUTIQUES : Proposition d'harmonisation et de restitution des données au titre de la DCSMM | |
| Référence interne : ODE/VIGIES Diffusion : <input checked="" type="checkbox"/> libre (internet) <input type="checkbox"/> restreinte (intranet) - date de levée d'embargo : AAA/MM/JJ <input type="checkbox"/> interdite (confidentielle) - date de levée de confidentialité : AAA/MM/JJ | Date de publication : 2018/02/28 Version : 5.2.0 Référence de l'illustration de couverture EVHOE 2015, J.P. Leaute - Ifremer Langue(s) : FR |
| Résumé/ Abstract : Cette étude fait état de la collecte des données hydrologiques actuellement effectuée à bord des six campagnes halieutiques DCF au titre de la DCSMM, et propose les solutions pour l'harmonisation de la bancarisation en vue d'une restitution uniformisée auprès des pilotes de la DCSMM et du grand public. | |
| Mots-clés/ Key words : DCSMM, Campagnes halieutiques, DCF, Collecte de données, Hydrologie, Bancarisation | |
| Comment citer ce document : Mabileau G., Baudrier J., 2018. Bancarisation des données hydrologiques des campagnes halieutiques - proposition d'harmonisation et de restitution des données au titre de la DCSMM. Rapport Ifremer, ODE/VIGIES, 52 p. | |
| Nom / référence du contrat : <input type="checkbox"/> Rapport intermédiaire <input checked="" type="checkbox"/> Rapport définitif | |
| Projets dans lesquels ce rapport s'inscrit (programme européen, campagne, etc.) : Directive Cadre Stratégie pour le Milieu Marin ; Programmes de Surveillance ; Optimisation des campagnes halieutiques DCF pour la collecte de données au large | |
| Auteur(s) / adresse mail Guillaume Mabileau / guillaume.mabileau@ifremer.fr Jérôme Baudrier / jerome.baudrier@ifremer.fr | Service, laboratoire Ifremer ODE/VIGIES Ifremer ODE/VIGIES |
| Encadrement(s) : Jérôme Baudrier / Jean-Paul Lecomte | |
| Destinataires : Coordination nationale de la DCSMM ; Pilotes scientifiques de la DCSMM ; Coordonnateurs de programmes thématiques DCSMM ; Responsable de campagnes halieutiques ; Responsables hydrologiques ; Equipes SISMER et Quadrige ² ; SIH... | |
| Contributeurs : E. Antajan, A. Auber, V. Badts, J.H. Bourdeix, P. Bourriau, D. Devreker, G. Durand, E. Gauthier, J. Gatti, T. Hattab, A. Huguet, M. Huret, A. Jadaud, A. Lamoureux, M. Le Moigne, A. Lefebvre, C. Loots, G. Maudire, L. Metral, C. Pertuisot, L. Petit De La Villeon, J.B. Romagnan, C. Saraux, M. Travers, V. Trenkel, S. Vaz, Y. Verin (Ifremer), F. Artigas (CNRS), A. Budria (AFB). | |

Contenu

| | |
|--|----|
| Figures et tableaux | 5 |
| Acronymes | 7 |
| Introduction..... | 9 |
| 1 Etat actuel de la bancarisation | 10 |
| 1.1 Campagnes en Manche-Mer du Nord..... | 10 |
| 1.1.1 Données de la CTD | 10 |
| 1.1.2 Données de la FerryBox..... | 11 |
| 1.1.3 Données des bouteilles Niskin | 12 |
| 1.1.4 Données des prélèvements zooplancton | 14 |
| 1.2 Campagnes en Atlantique..... | 14 |
| 1.2.1 Données de la CTD | 15 |
| 1.2.2 Données de la FerryBox..... | 16 |
| 1.2.3 Données des bouteilles Niskin | 16 |
| 1.2.4 Données des prélèvements zooplancton | 17 |
| 1.3 Campagnes en Méditerranée..... | 18 |
| 1.3.1 Données de la CTD | 18 |
| 1.3.2 Données de la Pocket FerryBox | 19 |
| 1.3.3 Données des bouteilles Niskin | 19 |
| 1.3.4 Données des prélèvements zooplancton | 20 |
| 1.4 Moratoires..... | 21 |
| 2 Restitution des données | 22 |
| 2.1 Besoins et suggestions des pilotes scientifiques | 22 |
| 2.1.1 Plateformes et type de fichiers..... | 22 |
| 2.1.2 Format des données | 22 |
| 2.1.3 Métadonnées et prétraitements | 23 |
| 2.2 Le Système d'Information sur le Milieu Marin (SIMM)..... | 23 |
| 3 Plateformes de bancarisation identifiées..... | 24 |
| 3.1 Banque nationale de physique-chimie (SISMER) | 24 |
| 3.1.1 Présentation..... | 24 |
| 3.1.2 Organisation des données | 24 |
| 3.1.3 Processus de qualification | 25 |
| 3.1.4 Diffusion..... | 26 |
| 3.2 Quadrige ² (VIGIES)..... | 28 |
| 3.2.1 Présentation..... | 28 |
| 3.2.2 Organisation des données | 28 |
| 3.2.3 Processus de qualification | 29 |

| | | |
|---------|---|----|
| 3.2.4 | Diffusion..... | 29 |
| 3.3 | Portail Coriolis..... | 32 |
| 3.4 | Autres plateformes de travail | 32 |
| 3.4.1 | Ecotaxa (OOV) | 32 |
| 3.4.2 | Datarmor (SISMER) | 34 |
| 4 | Discussion..... | 35 |
| 4.1 | Vers une bancarisation multiplateforme..... | 35 |
| 4.2 | Assurer la fusion des bases de données..... | 36 |
| 4.3 | Bancarisation des données physico-chimiques | 37 |
| 4.3.1 | FerryBox et Pocket FerryBox | 37 |
| 4.3.2 | Sonde CTD | 37 |
| 4.3.3 | Bouteilles Niskin | 39 |
| 4.4 | Bancarisation des données biologiques : Alternative Quadrige ² | 40 |
| 4.4.1 | Données de phytoplancton..... | 40 |
| 4.4.1.1 | Comptages..... | 40 |
| 4.4.1.2 | Intégration / saisie | 41 |
| 4.4.1.3 | Problématique de l'imagerie..... | 42 |
| 4.4.2 | Données de zooplancton | 42 |
| 4.4.2.1 | Comptages et biomasses..... | 42 |
| 4.4.2.2 | Intégration / saisie | 43 |
| 4.4.2.3 | Problématique de l'imagerie..... | 43 |
| 4.4.3 | Validation des critères de choix..... | 44 |
| 4.5 | Bancarisation des données biologiques : Autres alternatives | 46 |
| 4.5.1 | Alternative 2 | 46 |
| 4.5.2 | Alternative 3 | 48 |
| 5 | Conclusion | 50 |
| 6 | Références..... | 51 |

Figures et tableaux

| | | |
|-------------|--|----|
| Tableau 1 : | Paramètres collectés par la CTD sur les campagnes IBTS et CGFS | 10 |
| Tableau 2 : | Paramètres des FerryBox sur les campagnes IBTS et CGFS | 11 |
| Tableau 3 : | Paramètres des prélèvements de bouteilles Niskin sur les campagnes IBTS et CGFS..... | 12 |
| Tableau 4 : | Paramètres des prélèvements par filets sur les campagnes IBTS et CGFS..... | 14 |
| Tableau 5 : | Paramètres collectés par la CTD sur les campagnes PELGAS et EVHOE..... | 15 |
| Tableau 6 : | Paramètres des FerryBox sur les campagnes PELGAS et EVHOE | 16 |
| Tableau 7 : | Paramètres des prélèvements des bouteilles Niskin sur la campagne PELGAS..... | 16 |
| Tableau 8 : | Paramètres des prélèvements zooplancton sur les campagnes PELGAS et EVHOE... | 17 |
| Tableau 9 : | Paramètres collectés par la CTD sur les campagnes MEDITS et PELMED | 18 |

| | | |
|--------------|---|----|
| Tableau 10 : | Paramètres de la Pocket FerryBox sur les campagnes MEDITS et PELMED | 19 |
| Tableau 11 : | Paramètres des prélèvements des bouteilles Niskin sur la campagne PELMED | 19 |
| Tableau 12 : | Paramètres des prélèvements par filets sur les campagnes MEDITS et PELMED | 20 |
| Tableau 13 : | Contrôles qualité réalisés sur les métadonnées (SISMER) | 26 |
| Tableau 14 : | Contrôles qualité réalisés sur les données (SISMER) | 26 |
| Tableau 15 : | Compétences actuelles des banques en fonction du type de données | 36 |
| Tableau 16 : | Données physico-chimiques des bouteilles - Format du tableur à fournir au SISMER | 39 |
| Tableau 17 : | Alternative Quadrige ² - Liste des avantages et des inconvénients | 44 |
| Tableau 18 : | Alternative 2 - Liste des avantages et des inconvénients | 47 |
| Tableau 19 : | Alternative 3 - Liste des avantages et des inconvénients | 48 |
| | | |
| Figure 1 : | Extrait de la structure de la BDD Access des campagnes IBTS et CGFS | 13 |
| Figure 2 : | Portail d'accès aux données Nautilus (SISMER) | 27 |
| Figure 3 : | Portail d'accès aux données SeaDataNet | 27 |
| Figure 4 : | Diffusion des données bancarisées sur Quadrige ² | 30 |
| Figure 5 : | Interface actuelle Surval (visualisation) | 31 |
| Figure 6 : | Interface type Sextant (catalogue de données de Surval) | 31 |
| Figure 7 : | Data center de Coriolis | 32 |
| Figure 8 : | Ecotaxa - Interface de consultation de la base de données d'imagerie en ligne | 33 |
| Figure 9 : | Fusion et restitution des données par simple requête sur le portail SIMM | 36 |
| Figure 10 : | Structure du fichier cnv de la CTD envoyé au SISMER | 38 |
| Figure 11 : | Alternative Quadrige ² - processus de la collecte à la restitution par type de données | 40 |
| Figure 12 : | Processus de bancarisation des données de zooplancton | 43 |
| Figure 13 : | Alternative 2 - processus de la collecte à la restitution par type de données | 46 |
| Figure 14 : | Alternative 3 - processus de la collecte à la restitution par type de données | 48 |

Acronymes

- AFB** : Agence française pour la biodiversité
- AOA** : Algae Online Analyser
- BDD** : Base de données
- CDOM** : Matière organique dissoute colorée
- CGFS** : Channel GroundFish Survey (campagne)
- CNRS** : Centre national de la recherche scientifique
- CSR** : Cruise Summary Report
- CTD** : Conductivity Temperature Depth
- CUFES** : Continuous Underway Fish Egg Sampler
- DCF** : Data Collection Framework
- DCSMM** : Directive-Cadre Stratégie pour le Milieu Marin
- ERMS** : European Register of Marine Species
- EVHOE** : EValuation Halieutique de l'Ouest de l'Europe (campagne)
- IBTS** : International Bottom Trawl Survey (campagne)
- IDM** : Unité Informatique et Données Marines
- IGA** : Impact des Grands Aménagements
- IPEV** : Institut polaire français Paul-Emile Victor
- IRD** : Institut de recherche pour le développement
- ISO** : International Organization for Standardization
- ITIS** : Integrated Taxonomic Information System
- IUEM** : Institut Universitaire Européen de la Mer
- LRH** : Laboratoire Ressources Halieutiques
- MEDITS** : MEDiterranean Trawl Survey (campagne)
- MES** : Matière en suspension
- MIP** : Matière inorganique particulaire
- MOP** : Matière organique particulaire
- OOV** : Observatoire Océanologique de Villefranche-sur-Mer
- PELGAS** : PELagique GAScogne (campagne)
- PELMED** : PELagiques MEDiterranée (campagne)
- PFB** : Pocket FerryBox
- PSFM** : Prélèvement – Support – Fraction – Méthode (quadruplet)
- Q2** : Quadrige²
- REBENT** : Réseau de surveillance benthique
- REMI** : Réseau de contrôle microbiologique

REPHY : Réseau d'Observation et de Surveillance du Phytoplancton et des Phycotoxines

ROCCH : Réseau d'observation de la contamination chimique

SANDRE : Service d'Administration Nationale des Données et Référentiels sur l'Eau

SI : Système d'information

SIE : Système d'information sur l'eau

SIH : Système d'Informations Halieutiques

SIMM : Système d'information sur le milieu marin

SISMER : Systèmes d'Informations Scientifiques pour la Mer

SRM : Sous-Région Marine

TU : Temps universel (temps civil du méridien de Greenwich)

VIGIES : Valorisation de l'Information pour la Gestion Intégrée Et la Surveillance

WoRMS : World Register of Marine Species

Introduction

En France, l'élaboration du Programme de Surveillance (PdS) de la Directive Cadre "Stratégie pour le Milieu Marin" (DCSMM - European Commission, 2008) (1) s'est déroulée entre 2012 et 2014. A cette occasion, les campagnes halieutiques ont été identifiées comme support potentiel de suivis complémentaires à la surveillance actuelle, essentiellement ciblée vers les expertises halieutiques. L'optimisation des campagnes pour la DCSMM a été jugée prioritaire car elle permettait de collecter à des coûts modérés des données au large. Une étude de faisabilité a donc été lancée dès octobre 2013 pour juger de la compatibilité des suivis proposés avec les opérations liées à l'évaluation des stocks commerciaux. Elle a débouché sur le déploiement d'essais sur le terrain entre septembre 2014 et août 2015 (2), puis un rapport de synthèse a été diffusé aux principaux partenaires en novembre 2015 (3).

Le déploiement opérationnel des suivis DCSMM est effectif depuis début 2016, suite à l'adoption des PdS (4-7) en juin 2015 par les préfets coordonnateurs de chacune des sous-régions marines. L'action d'optimisation des campagnes halieutiques inscrite dans le programme de surveillance « Espèces commerciales » permet l'acquisition en routine de données au large pour de nombreuses thématiques (réseau trophique, contaminants, bruit, déchets, biodiversité, hydrologie...) (8).

Les données d'hydrologie collectées lors de ces campagnes sont entre autre destinées à être utilisées par les pilotes de la DCSMM pour le calcul des indicateurs du Bon Etat Ecologique des descripteurs D1 « Habitat pélagique », D5 « Eutrophisation », et D7 « Changements hydrographiques » (9). Leur nature est diverse puisque les paramètres physico-chimiques de la colonne d'eau sont concernés, ainsi que les communautés du phyto- et du zooplancton. La collecte de ces informations concerne ainsi six campagnes déployées au titre du règlement DCF (10) : IBTS (11), CGFS (12), PELGAS (13), EVHOE (14), PELMED (15), MEDITS (16). Pour chacune d'entre elles, la bancarisation des données d'hydrologie est pour le moment disparate en termes d'outils, de méthodes et de formats de stockage. L'objectif est donc d'harmoniser cette bancarisation afin de pouvoir mettre les données à disposition des pilotes scientifiques de la DCSMM en fonction de leurs besoins. Les données collectées étant publiques, il est également nécessaire de réfléchir à leur accessibilité auprès de la communauté scientifique et du grand public.

1 Etat actuel de la bancarisation

Actuellement, chacune des six campagnes halieutiques collecte et bancarise ses données de manière indépendante sur diverses bases de données et dans des formats différents. Un état des lieux avec les responsables de campagnes et les personnes responsables de la collecte des données hydrologiques à bord a permis d'identifier les différences et les similitudes entre chacune des méthodes de stockage et de bancarisation afin de réfléchir à leur harmonisation.

Il est à noter que tous les suivis décrits dans ce rapport ne concernent pas le programme de surveillance DCSMM. Cependant, pour les besoins de la présente étude, il était plus pertinent de réaliser l'inventaire de toutes les données hydrologiques collectées au cours des différentes campagnes. En effet, les pistes de réflexion pour l'archivage de ces informations nécessite de connaître tous les paramètres concernés (formats, types de données, etc.).

Par ailleurs, ces paramètres concernent des données physico-chimiques et de plancton, et ils peuvent ainsi être regroupés en quatre catégories selon le type d'appareil de collecte : les mesures hydrologiques collectées par bathysonde (CTD) ; les mesures automatisées par Ferrybox et Pocket Ferrybox (PFB) ; les données issues des prélèvements de bouteilles Niskin ; et les données issues des prélèvements par filets (WP2, MIK, etc.). Ces appareils sont pour la plupart communs aux six campagnes, mais la manière de bancariser les données et les traitements effectués dessus peuvent être différents.

Il existe déjà des points communs du fait que ce sont les mêmes équipes qui travaillent sur les campagnes IBTS et CGFS (Manche-Mer du Nord), EVHOE et PELGAS (Atlantique), ainsi que MEDITS et PELMED (Méditerranée).

1.1 Campagnes en Manche-Mer du Nord

Les campagnes IBTS et CGFS concernent la sous-région marine de la Manche-Mer du Nord à bord du navire *Thalassa*. Bien que certaines données hydrologiques collectées soient déjà bancarisées, elles ne sont pas toutes réunies sur une seule et même plateforme. En effet, en plus d'être stockées sur une base de données Access en local, les mesures hydrologiques CTD sont actuellement envoyées directement au SISMER qui les bancarise sur la banque physique-chimie, tandis que d'autres données issues de prélèvements par filet et bouteilles Niskin sont uniquement stockées sur la base locale.

1.1.1 Données de la CTD

Tableau 1 : Paramètres collectés par la CTD sur les campagnes IBTS et CGFS

| Paramètre | Appareil (collecte et analyse) |
|--------------|-------------------------------------|
| Température | CTD / SBE 19+V2 |
| Salinité | CTD / SBE 19+V2 |
| Densité | A partir de température et salinité |
| Fluorescence | ECO-FL-NTU (Wetlabs) sur CTD |
| Turbidité | ECO-FL-NTU (Wetlabs) sur CTD |
| Radiancé | PAR (Biospherical Inst.) sur CTD |
| Oxygène | SBE 43 sur CTD |
| pH | SBE 18 pH sur CTD |

Les données brutes récoltées par CTD sont initialement sorties sur fichier hex, puis convertis au format csv avant l'envoi au SISMER qui les bancarise sur la banque physique-chimie. Cette réception auprès du SISMER est accompagnée d'un processus de qualification des données basé

sur : 1) la vérification des métadonnées ; et 2) les contrôles automatique et visuel des données (minimum/maximum régionaux, climatologies) (voir section 3.1.3).

Des traitements sur les données collectées sont par ailleurs opérés en amont de l'envoi du fichier *cnv* au SISMER. Ces traitements sont réalisés avec le programme SBE data processing et suivent un protocole préétabli.

Ils consistent à :

- 1) l'alignement temporel des données par rapport aux délais de mesure des capteurs, selon les recommandations de SBE (module *Alignctd*)
- 2) le calcul de paramètres dérivés tels que la profondeur, la salinité, la densité, et la saturation en oxygène (module *Derive*)
- 3) l'ajout dans les métadonnées du fichier *cnv* des lignes renseignant sur le nom de la campagne, la position (coordonnées), l'heure et le nom de la station
- 4) la suppression manuelle des lignes inutiles et aberrantes afin de ne garder que les données de la descente
- 5) le calcul des mesures moyennées au demi-décibar (0,5 mètres) (module *Binavg*).

Le fichier *cnv* ainsi traité contient par ailleurs toutes les métadonnées nécessaires en en-tête, afin de tracer l'origine des mesures (date, coordonnées, outils), les traitements réalisés, ainsi que le nom et l'unité des paramètres qu'il contient.

Une fois archivées au SISMER, si ces données ne sont sous aucun moratoire, elles sont alors consultables/téléchargeables à partir de la banque de physique-chimie sur le portail d'accès Nautilus¹ (en passe d'être remplacé par un nouveau portail), et sont également accessibles via l'infrastructure européenne SeaDataNet². Les fichiers de sortie se présentent sous trois formats possibles : *medatlas* et *odv* (ascii), ou *NetCDF*.

A noter que le fichier *cnv* envoyé au SISMER et contenant les données est lisible sur un simple éditeur de texte (Bloc note, etc.).

1.1.2 Données de la FerryBox

Tableau 2 : Paramètres des FerryBox sur les campagnes IBTS et CGFS

| Paramètre | Appareil (collecte et analyse) |
|--|--------------------------------|
| Fluorescence / turbidité / oxygène | FerryBox |
| Température / Salinité / pH | FerryBox |
| Matière organique dissoute et colorée (CDOM) | FerryBox |
| 4 groupes spectraux | Analyseur algal sur FerryBox |

Les données sont suivies par FerryBox (17,18) à poste sur le navire *Thalassa* depuis CGFS 2017 (PFB auparavant), sont collectées automatiquement et seront bancarisées sur la plateforme Coriolis qui permettra l'accès aux données³ sous format *netcdf*.

Ces mesures concernent les données de température, turbidité, oxygène, salinité, pH et matière organique dissoute et colorée (CDOM) collectées toutes les minutes à une profondeur d'environ 5 mètres sous la surface, ainsi que les mesures spectrales de l'analyseur algal (4 paramètres

¹ Portail Nautilus : <http://donnees-campagnes.flotteoceanographique.fr>

² Portail SeaDataNet : http://seadatanet.maris2.nl/v_cdi_v3/search.asp

³ Portail Coriolis : <http://www.coriolis.eu.org/Data-Products/Data-Delivery>

exprimés en μg de Chlorophylle-a/L) (19,20). Cette collecte de données par FerryBox à bord de la *Thalassa* est identique pour toutes les campagnes, et les contenus de ces bases seront similaires aux campagnes de la façade Méditerranée qui recueille les données par Pocket FerryBox installée sur *l'Europe* (section 0).

L'archivage / bancarisation de ces données hautes fréquences sera mis en place par le SISMER. Il est à déterminer si les données seront reçues par Coriolis en flux réel, ou flux différé après traitement. Un processus de qualification est en outre envisagé.

1.1.3 Données des bouteilles Niskin

Les données issues des bouteilles Niskin se divisent en deux types bien distincts : les mesures physico-chimiques, et les données biologiques issues des prélèvements de phytoplancton.

Tableau 3 : Paramètres des prélèvements de bouteilles Niskin sur les campagnes IBTS et CGFS

| Paramètre | Appareil (collecte et analyse) |
|---|--|
| <i>Données physico-chimiques</i> | |
| Sels nutritifs (NH_4^+ , NO_3^- , PO_4^- , SiOH) | Bouteille Niskin en surface Analyseur Alliance (USTL-LOG) |
| Matières en suspension (MES+MO+MI) | Bouteille Niskin en surface Four + balance de précision |
| Pigments chlorophylliens | Bouteille Niskin en surface Spectromètre |
| <i>Données biologiques (taxonomiques)</i> | |
| Abondance phytoplancton | Bouteille Niskin en surface Microscope inversé |

En terme de bancarisation, les données en surface issues des bouteilles Niskin sont pour le moment stockées en local sur des bases Access dédiées aux campagnes IBTS et CGFS. Actuellement, seules des données physico-chimiques de la campagne IBTS 2004 sont bancarisées au SISMER sur la banque physique-chimie, mais une bancarisation généralisée est prévue prochainement. Celle-ci pourra suivre le même format que pour PELGAS (voir section 1.2.3).

Les bases de données Access sont structurées de manière à ce que chaque donnée issue des bouteilles Niskin soit stockée dans les tables dédiées, qui sont elles-mêmes reliées à une table « Station » par un identifiant unique, et permet ainsi de retrouver les informations associées (Figure 1).

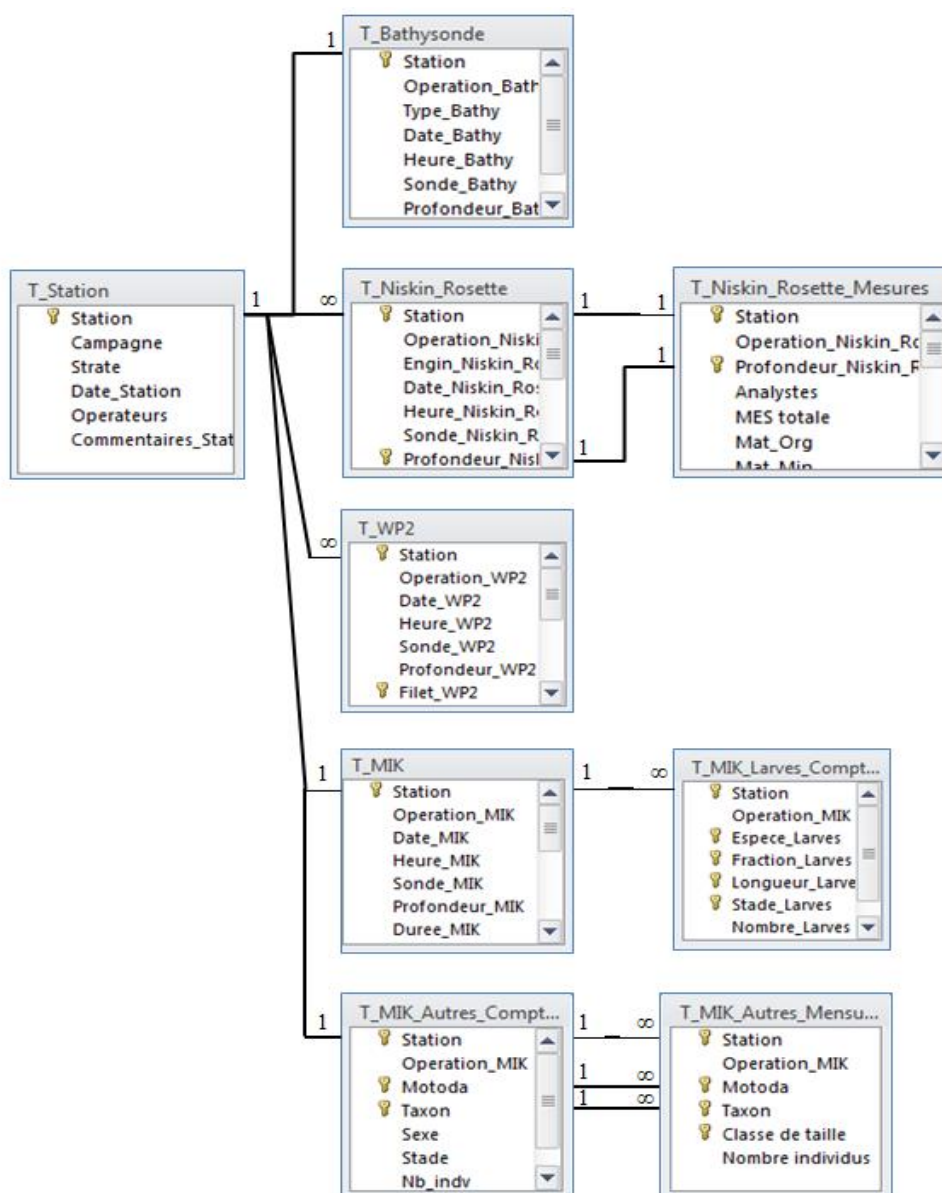


Figure 1 : Extrait de la structure de la BDD Access des campagnes IBTS et CGFS

On retrouve les mesures des paramètres physico-chimiques dans la table « T_Niskin_Rosette_Mesures » qui inclut en particulier les données relatives aux concentrations :

- Sels nutritifs dont Ammonium (NH₄⁺), Nitrate (NO₃⁻), Phosphate (PO₄⁻), et Silicate (SiOH)
- Matière en suspension totale avec le détail en matière organique et matière minérale
- Chlorophylle-a et phéopigments (concentrations moyennes sur deux répliques).

Ces mesures sont stockées selon l'identifiant de la station et la profondeur des bouteilles Niskin qui composent à eux deux la clé primaire de la table. Un champ « Observations » vient également compléter les informations sur les mesures lorsqu'elles sont non-renseignées ou qu'un événement particulier a été remarqué et noté à bord.

Les données biologiques relatives à l'abondance de phytoplancton sont, elles, stockées dans les tables « T_Phyto » et « T_Phyto_comptage ». La première compile les métadonnées de la méthode de dénombrement par microscope inversé (méthode Utermöhl) dont le nom de l'appareil, la surface de la cuve, le volume sédimenté, etc. La seconde stocke les comptages de cellules par classe de taille et par taxon, ainsi que l'abondance dérivée de ces mesures. Sur chacune de ces

tables de résultats on retrouve le code de la station et numéro de l'opération Niskin correspondant afin de pouvoir remonter aux informations sur l'origine des données.

1.1.4 Données des prélèvements zooplancton

Tableau 4 : Paramètres des prélèvements par filets sur les campagnes IBTS et CGFS

| Paramètre | Appareil (collecte et analyse) |
|--|---|
| Abondance Mésozooplancton | Filet WP2 (200-500µm) ZooScan |
| Abondance des œufs pélagiques de poisson | CUFES (300 µm) ZooCam |
| Abondance larves de poisson et macroplancton associé | Filet MIK (500µm) Binoculaire (larves) ZooScan (macrozooplancton associé) |

Les données issues des prélèvements par filets sont également stockées dans les bases de données Access dédiées aux campagnes IBTS et CGFS (Figure 1). De la même manière que pour les bouteilles Niskin, les mesures sont bancarisées dans les tables dédiées à chaque appareil de collecte, qui sont elles-mêmes reliées à la table principale « Station » par un identifiant unique.

Les appareils de collecte concernés sont le filet WP2, pour l'abondance de mésozooplancton (analyse par ZooScan), le CUFES pour les œufs pélagiques de poisson (prélèvements en surface analysés par ZooCam), et le filet MIK pour l'abondance de larves de poissons et macroplancton associé (ce filet n'est déployé que sur IBTS). On retrouve dans la table dédiée au WP2 les données associées au lieu de prélèvement (date, heure, profondeur, identifiant de la station), ainsi que les informations sur le prélèvement (taille de la maille du filet, nombre de tours du volucompteur, volume filtré, type de conservation, commentaires). Les analyses sont par ailleurs effectuées par ZooScan.

Pour les tables du filet MIK, outre les informations sur le lieu du prélèvement (date, heure, profondeur, durée, identifiant de la station), les données nécessaires au calcul de l'abondance de larves (taxon, fraction, longueur, stade, nombre d'individus, commentaires) sont accessibles.

A noter que certaines données analysées ne sont pas encore bancarisées dans ces bases de données Access, en particulier les comptages de zooplancton (WP2), et les comptages d'œuf (filet Mikey-M net / CUFES).

1.2 Campagnes en Atlantique

Les campagnes EVHOE et PELGAS concernent les sous-régions marines des mers Celtiques et du golfe de Gascogne à bord du navire *Thalassa*. Bien que la collecte des données sur ces deux campagnes soit similaire du point de vue des appareils utilisés, des méthodes, et des outils de stockage, la campagne PELGAS collecte certaines données qui ne le sont pas sur EVHOE, en particulier les prélèvements de bouteilles Niskin et certains paramètres de suivi du zooplancton.

1.2.1 Données de la CTD

Tableau 5 : Paramètres collectés par la CTD sur les campagnes PELGAS et EVHOE

| Paramètre | Appareil (collecte et analyse) |
|--------------|-------------------------------------|
| Température | CTD / SBE 19+V2 |
| Salinité | CTD / SBE 19+V2 |
| Densité | A partir de température et salinité |
| Fluorescence | WetLabs WETStar sur CTD |
| Turbidité | OBS-3+ sur CTD |
| Oxygène | SBE 43 sur CTD |
| pH | SBE 18 pH sur CTD |

Comme pour les autres campagnes les données hydrologiques récoltées par CTD sont automatiquement sorties sur fichiers hex qui sont ensuite convertis au format csv en vue de la bancarisation sur la banque physique-chimie par le SISMER. Cette conversion est accompagnée d'un traitement des données avec le programme SBE data processing suivant un protocole préétabli légèrement différent de celui des campagnes en Manche-mer du Nord. Il consiste en :

- 1) l'alignement temporel des données par rapport aux délais de mesure des capteurs, selon les recommandations de SBE (module *Alignctd*)
- 2) un lissage complémentaire sur les données (module *Loopedit*)
- 3) le calcul de paramètres dérivés tels que la profondeur (m), la salinité, la densité, et la saturation en oxygène (module *Derive*)
- 4) le calcul des mesures moyennées au décibar (tous les mètres) (module *Binavg*)

Ces traitements sont en grande partie similaires à ceux réalisés sur les données CTD en Manche-Mer de Nord (section 1.1.1), hormis l'ajout du module de lissage et le moyennage au décibar au lieu du demi-décibar.

Le fichier csv ainsi traité contient par ailleurs toutes les métadonnées nécessaires en en-tête, afin de tracer l'origine des mesures (date, coordonnées, outils), les traitements réalisés, ainsi que le nom et l'unité des paramètres qu'il contient. Ces métadonnées intègrent en outre les informations importantes au sujet des coefficients de calibration de chaque capteur, incluant les numéros de série (*SerialNumber*), les dates de calibration (*CalibrationDate*), ainsi que le chemin d'accès au fichier xmlcon dont sont issues ces informations.

Une fois archivées au SISMER, si ces données ne sont sous aucun moratoire, elles sont alors consultables/téléchargeables à partir de la banque de physique-chimie sur le portail d'accès Nautilus¹, et sont également accessibles via l'infrastructure européenne SeaDataNet². Les fichiers de sortie se présentent sous trois formats possibles : medatlas et odv (ascii), ou NetCDF.

¹ Portail Nautilus : <http://donnees-campagnes.flotteoceanographique.fr>

² Portail seaDataNet : http://seadatanet.maris2.nl/v_cdi_v3/search.asp

1.2.2 Données de la FerryBox

Tableau 6 : Paramètres des FerryBox sur les campagnes PELGAS et EVHOE

| Paramètre | Appareil (collecte et analyse) |
|--|--------------------------------|
| Fluorescence / turbidité / oxygène | FerryBox |
| Température / Salinité / pH | FerryBox |
| Matière organique dissoute et colorée (CDOM) | FerryBox |
| 4 groupes spectraux | Analyseur algal sur FerryBox |

La collecte de données FerryBox en façade Atlantique est sur le point de commencer sur la *Thalassa*. Le protocole en termes de traitement, qualification et bancarisation sera le même que pour les façades Manche-Mer du Nord et Méditerranée (voir sections 1.1.2 et 0). La bancarisation de ces données hautes fréquences sera faite sur Coriolis¹ au format netcdf.

1.2.3 Données des bouteilles Niskin

Tableau 7 : Paramètres des prélèvements des bouteilles Niskin sur la campagne PELGAS

| Paramètre | Appareil (collecte et analyse) |
|---|---|
| Sels nutritifs (PO ₄ ⁻ , NO ₃ ⁻ , NO ₂ ⁻ , NH ₄ ⁺ , SiOH) | Bouteilles Niskin avec prélèvements à 3 niveaux (surf, max fluo, fond) |
| Matières en suspension (MOP et MIP) | Bouteilles Niskin avec prélèvements à 3 niveaux (surf, max fluo, fond) et filtrations |
| Chlorophylle par classe de taille | Bouteilles Niskin avec prélèvements à 3 niveaux (surf, max fluo, fond) et filtrations |
| Phéopigments par classe de taille | Bouteilles Niskin avec prélèvements à 3 niveaux (surf, max fluo, fond) et filtrations |

Des données de prélèvements par bouteilles Niskin sont réalisées sur PELGAS et sont envoyées chaque année sous format Excel au SISMER qui les bancarise sur la banque de physique-chimie après en avoir contrôlé la qualité. Le fichier est organisé de manière à pouvoir identifier chaque mesure en ligne selon le nom, la date, et les coordonnées de la station, ainsi que l'heure et la profondeur du prélèvement à 3 niveaux (surface, max fluo, fond).

Les données détaillent ensuite les concentrations pour chaque niveau dont les unités sont renseignées en en-tête de colonne. Les mesures concernent :

- Sels nutritifs (µM) dont Phosphate (PO₄⁻), Nitrate (NO₃⁻), Nitrite (NO₂⁻), Ammonium (NH₄⁺) et Silicate (SiOH)
- Matière en suspension (mg/l), avec le détail des matières organique et inorganique particulières (MOP et MIP)
- Chlorophylle-a (µg/L), avec le détail par classe de taille (> 20µm ; 3-20µm ; 3µm)
- Phéopigment (µg/L), avec le détail par classe de taille (> 20µm ; 3-20µm ; 3µm).

Une fois archivées au SISMER, si ces données ne sont sous aucun moratoire, elles sont alors consultables/téléchargeables à partir de la banque de physique-chimie sur le portail d'accès

¹ Portail Coriolis : <http://www.coriolis.eu.org/Data-Products/Data-Delivery>

Nautilus¹, et sont également accessibles via l'infrastructure européenne SeaDataNet². Les fichiers de sortie se présentent sous trois formats possibles : madatlas et odv (ascii), ou NetCDF.

Les bouteilles Niskin ne concernent pour le moment que la campagne PELGAS, mais il est possible que cela soit réalisé également sur EVHOE à l'avenir, pour harmonisation avec les autres campagnes en réponse à la surveillance DCSMM.

1.2.4 Données des prélèvements zooplancton

Tableau 8 : Paramètres des prélèvements zooplancton sur les campagnes PELGAS et EVHOE

| Paramètre | Appareil (collecte et analyse) |
|--|---------------------------------------|
| <i>PELGAS et EVHOE</i> | |
| Abondance Mésozooplancton | Filet WP2 (200µm) ZooCam |
| Abondance des œufs pélagiques de poisson | CUFES (300 µm) ZooCam |
| <i>PELGAS uniquement</i> | |
| Biomasse Mésozooplancton | Filet WP2 (200µm) |
| Larves (opportuniste) | Filet Carré / Multinet (Hydrobios) |
| Abondance particules | LOPC (Laser Optical Plankton Counter) |

Les données de zooplancton sont bancarisées à divers endroits.

Les paramètres d'abondance mésozooplancton (prélèvement par filet WP2) et œufs pélagiques de poissons (pompés à 5m de profondeur par le CUFES) sont issus des données d'imagerie analysée par ZooCam (analyses au microscope jusqu'en 2014) (21,22). Ces données sont bancarisées sur la plateforme EcoTaxa de Villefranche-sur-Mer³ (voir section 3.4.1) qui stocke actuellement l'intégralité des données de PELGAS et EVHOE depuis 2016.

Pour chaque campagne, la base permet ainsi la consultation des résultats, et leur export automatisé sous forme de tableaux de comptage (summary), ainsi que le détail des données caractéristiques par individu observé. On retrouve en outre toutes les métadonnées nécessaires à l'identification de l'observation (campagne, station, coordonnées, date, heure et profondeur du prélèvement).

L'export de ces données au format standard tsv (séparateur par tabulation) permet la lecture et la conversion des fichiers sur n'importe quel éditeur de texte ou Excel.

Les données de biomasse mésozooplancton sont quant à elles stockées sur fichiers csv en local (à l'Ifremer). Ils comprennent les données issues des prélèvements par filets WP2 et concernent les biomasses (poids secs) pour cinq classes de tailles : >5000µm ; 2000-5000µm ; 1000-2000µm ; 500-1000µm ; et 200-500µm.

A noter que des prélèvements opportunistes par filets carré et filets MIK ont également eu lieu, ils sont uniques à PELGAS et non-systématiques. Les données sont stockées sur EcoTaxa et/ou en local. Le LOPC mesurant l'abondance de particules est lui aussi exclusif à PELGAS, mais sa maintenance devenant compliquée, cet instrument est susceptible de disparaître.

¹ Portail Nautilus : <http://donnees-campagnes.flotteoceanographique.fr>

² Portail SeadataNet : http://seadatanet.maris2.nl/v_cdi_v3/search.asp

³ EcoTaxa : <http://ecotaxa.obs-vlfr.fr>

1.3 Campagnes en Méditerranée

Les campagnes MEDITS et PELMED concernent la sous-région marine de la mer Méditerranée Occidentale à bord du navire *L'Europe* et recueillent entre autres des données hydrologiques issues des mesures de la CTD, la Pocket FerryBox, de certains filets à plancton, en particulier le WP2, ainsi que des bouteilles Niskin (PELMED uniquement). La collecte de ces données en Méditerranée est relativement récente et les informations ne sont donc pas encore valorisées ou bancarisées sur une seule et même base structurée.

1.3.1 Données de la CTD

Tableau 9 : Paramètres collectés par la CTD sur les campagnes MEDITS et PELMED

| Paramètre | Appareil (collecte / analyse) |
|--------------|-------------------------------------|
| Température | CTD / SBE 19+V2 |
| Salinité | CTD / SBE 19+V2 |
| Densité | A partir de température et salinité |
| Turbidité | OBS, D & A 3+ sur CTD |
| Oxygène | SBE 43 sur CTD |
| pH | SBE 18 pH sur CTD |
| Radiance | PAR (Biospherical Inst.) sur CTD |
| Fluorescence | WetLabs WETStar sur CTD |

Comme pour les autres campagnes, les données brutes de sortie sont stockées sur un fichier hex qui est ensuite converti à un format *cnv*, avant l'envoi au SISMER qui les bancarise sur la banque physique-chimie après en avoir contrôlé la qualité (pour PELMED actuellement). Cette conversion est accompagnée d'une calibration, mais aucun traitement supplémentaire n'est opéré sur les données (alignement, moyennage, lissage), comme cela peut être le cas sur d'autres campagnes (voir sections 1.1.1 et 0). Ces tâches sont opérées grâce au programme SBE data processing. Pour MEDITS, des données de CTD ne sont déclarées dans les CSR (Cruise Summary Report) des campagnes que depuis 2015 et sous forme d'essais pour la DCSMM. Les données sont donc uniquement sauvegardées au SISMER, sans bancarisation sur la banque physique-chimie.

Un fichier *cnv* pour chaque station contient ainsi les données compilées de la CTD, précédées des métadonnées du jeu renseignant sur l'origine des mesures (date, coordonnées de la station, configuration de l'appareil, nom des paramètres et unités de mesure). Ces métadonnées intègrent en outre les informations importantes au sujet des coefficients de calibration de chaque capteur, incluant les numéros de série (*SerialNum*), les dates de calibration (*CalDate*), ainsi que le chemin d'accès au fichier *xmlcon* dont sont issues ces informations. Les données brutes sont listées à la suite de ces métadonnées. Le fichier *cnv*, au format *ascii*, est par ailleurs lisible sur n'importe quel éditeur de texte (type Bloc note).

Pour les données de PELMED, une fois archivées au SISMER, si ces données ne sont sous aucun moratoire, elles sont alors consultables/téléchargeables à partir de la banque de physique-chimie sur le portail d'accès Nautilus¹, et sont également accessibles via l'infrastructure européenne SeaDataNet². Les fichiers de sortie se présentent sous trois formats possibles : *medatlas* et *odv* (*ascii*), ou *NetCDF*.

¹ Portail Nautilus : <http://donnees-campagnes.flotteoceanographique.fr>

² Portail SeadataNet : http://seadatanet.maris2.nl/v_cdi_v3/search.asp

1.3.2 Données de la Pocket FerryBox

Tableau 10 : Paramètres de la Pocket FerryBox sur les campagnes MEDITS et PELMED

| Paramètre | Appareil (collecte / analyse) |
|--|-------------------------------------|
| Fluorescence / turbidité / oxygène | Pocket FerryBox |
| Température / Salinité / pH | Pocket FerryBox |
| Matière organique dissoute et colorée (CDOM) | Pocket FerryBox |
| 4 groupes spectraux | Analyseur algal sur Pocket FerryBox |

Les données brutes des mesures de la Pocket FerryBox (17,18) sont stockées sur deux fichiers de sortie distincts :

Le premier au format standard csv stocke la majeure partie des données de la Pocket FerryBox dont l'heure et les coordonnées de la mesure, ainsi que les paramètres incluant : température, turbidité, oxygène, salinité, pH et CDOM mesurés à une profondeur d'environ 5 mètres sous la surface toutes les minutes aux unités de mesures standards.

Le second, issu des mesures de l'analyseur algal (19,20), est un fichier spécifique au format AOA (Algae Online Analyser) compilant les données spectrales (4 paramètres exprimés en μg de Chlorophylle-a/L) identifiées par l'heure de la mesure, permettant ainsi de relier ces données à celles du fichier csv de la Pocket FerryBox. Le fichier AOA est uniquement compatible au programme dédié de l'analyseur, mais un export dans un format txt standard est possible.

Actuellement, ces données ne sont ni valorisées, ni bancarisées sur une plateforme spécifique. Les fichiers bruts sont stockés individuellement en local. Un transfert vers le SISMER est prévu afin que les données soient bancarisées sur la plateforme Coriolis. Cette bancarisation sera identique en termes de format et de données aux autres campagnes en Atlantique et Manche-Mer du Nord (voir sections 1.1.2 et 1.2.2).

1.3.3 Données des bouteilles Niskin

Tableau 11 : Paramètres des prélèvements des bouteilles Niskin sur la campagne PELMED

| Paramètre | Appareil (collecte / analyse) |
|---|--|
| <i>Données physico-chimiques</i> | |
| Matière en suspension (MES) | Bouteilles Niskin en surface et au maximum de fluorescence (systématique sur PELMED) et filtrations |
| Chlorophylle par classe de taille | Bouteilles Niskin en surface et au maximum de fluorescence (systématique sur PELMED) et filtrations |
| <i>Données biologiques taxonomiques</i> | |
| Abondance phytoplancton | Bouteilles Niskin en surface et au maximum de fluorescence (systématique sur PELMED) et microscopie inversée |

Seule la campagne PELMED recueille aujourd'hui des données issues de prélèvements par bouteille Niskin à deux niveaux en Méditerranée. Cela concerne les paramètres relatifs à la matière en suspension, la Chlorophylle-a, ainsi que l'abondance de phytoplancton.

Les mesures physico-chimiques sont réalisées pour chaque station en surface et au maximum de fluorescence (PELMED). La collecte ayant débutée en 2016, les échantillons n'ont pas encore été traités et les résultats ne sont pas encore stockés sur tableur, mais ils concernent :

- Concentration de Chlorophylle-a ($\mu\text{g/L}$), avec le détail par classe de taille ($> 20\mu\text{m}$; $3-20\mu\text{m}$; $3\mu\text{m}$)
- Concentration de Matière en suspension (mg/l), avec le détail des matières organique et inorganique particulaires (MOP et MIP).

Pour les données biologiques, l'analyse de l'abondance de phytoplancton par microscopie inversée est confiée à un sous-traitant rattaché à l'IUEM de Brest qui étudie la composition spécifique du phytoplancton. Les résultats numériques, exprimés en nombre de cellules/L, sont stockés sur un fichier Excel sous forme de tableau croisé par nom de taxon (en lignes) et par identifiant de station (en colonnes). Les caractéristiques des stations sont par ailleurs stockées sur un fichier différent où l'on retrouve notamment la date, l'heure, et les coordonnées.

Ces données sont actuellement stockées sur fichier csv en local dans des répertoires dédiés aux côtés des données de prélèvements zooplancton par filets (voir section 1.3.4).

A noter que les prélèvements de bouteille Niskin pourraient également permettre les suivis de sels nutritifs (nitrate, nitrite, ammonium, silicate, phosphate) et phéopigments sur PELMED, comme cela est déjà fait sur les campagnes IBTS, CGFS, et PELGAS.

1.3.4 Données des prélèvements zooplancton

Tableau 12 : Paramètres des prélèvements par filets sur les campagnes MEDITS et PELMED

| Paramètre | Appareil (collecte / analyse) |
|--|--|
| <i>MEDITS et PELMED</i> | |
| Abondance Mésozooplancton | Filet WP2 ($200\mu\text{m}$) Microscope |
| <i>PELMED uniquement :</i> | |
| Abondance larves de poisson et macroplancton associé | Filet MIK ($500\mu\text{m}$) Binoculaire (Larves) |

Les données issues des prélèvements par filets à plancton concernent en particulier le filet WP2 ($200\mu\text{m}$) pour les mesures d'abondance de mésozooplancton.

La campagne PELMED stocke d'ores et déjà les données WP2 analysées de 2014 et 2015 en local, mais l'analyse des données 2016 et 2017 par les spécialistes planctonologistes (zooplancton) est toujours en cours (sous-traitance). Ce recueil n'a débuté qu'en 2016 pour la campagne MEDITS. A noter que, faute d'appareil et de place à bord, les analyses des prélèvements par ZooCam/ZooScan ne sont pas possibles pour ces deux campagnes, et les analyses qualitatives et quantitatives sont donc pour l'instant menées à la binoculaire.

De plus, d'autres prélèvements par filets MIK ($500\mu\text{m}$) et Bongo (200 et $500\mu\text{m}$) sont également réalisés sur la campagne PELMED, mais leur collecte est très récente. Ces données concernent en particulier les paramètres d'abondance de zooplancton, de larves de poisson et macroplancton associé.

En ce qui concerne la bancarisation, les concentrations taxonomiques par station sont actuellement stockées en local sur des fichiers csv sous le même format que les données de phytoplancton (voir section 1.3.3). Ces fichiers sont répertoriés par année et par appareil de collecte, et sont donc consultables individuellement. Un fichier liste par ailleurs les différentes stations réalisées et leurs caractéristiques, et permet donc d'associer les données des prélèvements aux caractéristiques de la station en question en se référant à son identifiant.

1.4 Moratoires

La question des moratoires sur les données a été posée au groupe de travail afin d'éclaircir la vision de tous. S'agissant de données DCSMM acquises sur financements publics il n'y a pas de raison d'appliquer un quelconque moratoire. Il faut s'en référer directement aux conventions qui ont été signées en amont et qui précisent les conditions de diffusion des données. Il faut noter en revanche que les délais avant de rendre des données publiques sont différents entre des données brutes, des données résultant d'analyses au laboratoire et des données traitées (produits de type indicateurs ou cartes...).

2 Restitution des données

Ces données collectées à bord des navires de campagnes halieutiques sont en particulier destinées à alimenter les calculs des indicateurs par les pilotes scientifiques de la DCSMM. Afin d'assurer une restitution optimale auprès de ces équipes, il est donc important de cerner les besoins en termes de mise à disposition. De plus, s'agissant de données publiques, la restitution auprès de la communauté scientifique et du grand public doit également être prise en compte dans la réflexion.

Les descripteurs de la DCSMM et les critères concernés par ces données sont :

- **D1 « Habitat pélagique »** (D1 PH1 : changes of plankton functional types (life form) index Ratio ; D1 PH2 : plankton biomass and/or abundance ; D1 PH3 : changes in biodiversity index).
- **D5 « Eutrophisation »** (D5C1 : concentration en nutriments dans la colonne d'eau ; D5C2 : concentration en chlorophylle-a dans la colonne d'eau ; D5C3 : occurrence d'algues nuisibles ; D5C4 : transparence de la colonne d'eau ; D5C5 : concentration en oxygène dissous au fond).
- **D7 « Changements hydrographiques »**, en particulier pour les données de température et salinité (D7C1 : étendue spatiale et répartition de la modification permanente des conditions hydrographiques sur les fonds marins et dans la colonne d'eau).

2.1 Besoins et suggestions des pilotes scientifiques

2.1.1 Plateformes et type de fichiers

Les besoins ne concernent pas des plateformes ou des formats de fichiers en particulier. L'important est principalement de permettre une consultation et une **récupération en routine** des données, permettant l'exploitation des jeux sur des logiciels standards.

De plus, dans le cas d'une bancarisation multiplateformes, il est indispensable que les jeux récupérés **puissent être ensuite fusionnés** simplement (par exemple via les identifiants uniques de station), afin de permettre le croisement et l'exploitation de toutes les données physico-chimiques et biologiques dans leur ensemble. En effet, le fait de pouvoir relier toutes les informations entre elles, quel que soit le type de données, est un point primordial.

2.1.2 Format des données

S'agissant du développement d'indicateurs et de méthodes de calculs à visée pérenne, il est important de garantir des formats de données qui, une fois décidés, **resteront figés** (en termes de structure, ordre de colonnes, unités de paramètres, etc.). Ceci afin d'assurer une exploitation en routine, qui évitera par exemple de devoir modifier les scripts de calcul rendus inutilisables, des conversions inhabituelles, ou des erreurs de calculs et d'interprétation. Cela permettrait en outre d'avoir des analyses standardisées et facilement reproductibles.

Plus précisément, il serait utile que les données suivent des types de **formats harmonisés** selon des normes adoptées au niveau national (e.g. formats du SANDRE) et/ou international (e.g. norme ISO). Notamment, les formats du type : date, heure, longitude/latitude (degrés décimaux ou minutes, secondes ?), etc., devraient être standardisés.

2.1.3 Métadonnées et prétraitements

La mise à disposition de **métadonnées complètes** est également un point essentiel afin d'assurer une parfaite compréhension de l'origine des données (nom, date, heure, coordonnées, profondeur), des paramètres restitués (nom, unité, méthode d'analyse) et des éventuels traitements ayant été réalisés en amont. Plus particulièrement, les données CTD étant prétraitées avant leur bancarisation, il est indispensable que les données mises à disposition des pilotes aient subi des traitements suivant des **protocoles identiques** pour les six campagnes, afin de rendre les données exploitables dans leur ensemble.

Par ailleurs, la mise à disposition des données issues des campagnes ne se limite pas uniquement aux besoins de la DCSMM. C'est pour cette raison qu'il serait important de restituer les données au niveau le plus brut possible, afin de permettre des calculs adaptés à différentes problématiques qui nécessiteront l'utilisation de ces données à l'avenir. C'est dans cette optique qu'il est important d'avoir une **résolution la plus fine possible**, et de restituer chacun des paramètres stockés, y compris les paramètres supports servant aux différents calculs intermédiaires.

2.2 Le Système d'Information sur le Milieu Marin (SIMM)

La bancarisation des données hydrologiques doit également être pensée en fonction de besoins plus larges, notamment dans l'optique du développement du Système d'Information sur le Milieu Marin (SIMM) à la charge de l'AFB et du Ministère de la Transition Ecologique et Solidaire. Ce SI fédérateur est construit en particulier pour les besoins de la DCSMM. A l'instar du Système d'Information sur l'Eau¹ (SIE), le contenu du SIMM sera **ouvert au grand public** (open-data) et **centralisera l'accès aux données** des banques utiles à la DCSMM et aux gestionnaires d'aires marines protégées (Quadrige², Sextant, Coriolis, SIH, etc.). Il sera accessible sur le portail milieumarinfrance.fr confié à l'Ifremer et en cours de développement au sein du service IDM (Informatique et données marines). Une première mise en exploitation de ce portail est prévue d'ici la **fin 2018**.

Dans ce cadre, les données et leurs métadonnées seront harmonisées et devront suivre des référentiels communs. De plus, lorsqu'il sera développé et que les différentes banques pour l'alimenter seront identifiées, le SIMM sera alors en mesure d'aller récupérer les données nécessaires dans ces bases pour les implémenter sur le portail dédié. Dans le cas des données hydrologiques des campagnes halieutiques, il est donc primordial que les systèmes de bancarisation des paramètres soient compatibles avec la tâche de récupération du SIMM, en termes d'accessibilité, d'export, et de format. Ainsi, dans le cas où les données seraient bancarisées sur divers SI, le portail SIMM pourrait ainsi faire le lien et uniformiser les différentes sources afin de rendre la restitution de n'importe quels paramètres possibles en une seule extraction selon les requêtes faites par les utilisateurs.

¹ Système d'Information sur l'Eau (SIE) : <http://www.eaufrance.fr/>

3 Plateformes de bancarisation identifiées

3.1 Banque nationale de physique-chimie (SISMER)

3.1.1 Présentation

La banque nationale de données de physique-chimie marine est un système d'information développé et géré par le service SISMER de l'Ifremer et elle est alimentée en temps différé par les données des campagnes françaises (mises en œuvre par l'IFREMER, le CNRS, l'IRD, l'IPEV, les Universités,...) ou de certains programmes internationaux. Ces données se présentent principalement sous forme de profils verticaux (bathysondes, bouteilles), de séries temporelles (mouillages de courantomètres, capteurs T&S,...) ou de trajectoires (ADCP de coque). La base de données, mise à jour en continu, intègre ainsi des données physico-chimiques et quelques données de comptage phytoplanctonique, mais elle ne permet pas encore de bancariser les données de zooplancton.

3.1.2 Organisation des données

La Banque nationale de physique-chimie gère les données des campagnes relatives à l'hydrologie et à la biogéochimie (CTD et bouteilles Niskin) en utilisant les **standards européens de SeaDataNet¹**. Ces standards concernent en particulier la liste des campagnes, des navires, des régions marines, des instruments, des types de données, des paramètres, etc. D'autres données de référence sont gérées en interne telles que la liste des centres de données, des responsables scientifiques, etc.

Afin de décrire comment les données ont été acquises et traitées, un certain nombre de métadonnées ou "données à propos des données" est, en effet, à renseigner avant la réception au SISMER, sans quoi aucun archivage pérenne n'est possible. Ces métadonnées permettent ainsi de compiler toutes les informations nécessaires et indispensables au bon usage et à la réutilisation des données sur le long terme.

Pour les données de **profils verticaux** (*i.e.* CTD et bouteilles Niskin), pour chaque station effectuée, les **informations indispensables** à la bancarisation concernant :

- la date, l'heure (TU)
- les coordonnées (latitude/longitude)
- la description des paramètres et leur unité
- le paramètre de référence verticale : pression (dbar) ou profondeur (m)

Il est également nécessaire de renseigner les méthodes d'analyse chimique, ainsi que les types de capteurs utilisés. Ces métadonnées pour les données CTD apparaissent en principe en en-tête des fichiers cnv générés par la SBE lors des traitements et de la conversion. Ce fichier complet est requis avant l'envoi au SISMER pour qu'il les archive/bancarise (détails en section 4.3.2). Pour les bouteilles Niskin, des fichiers du type csv peuvent être envoyés (détails en section 4.3.3).

L'**uniformisation de la structure des fichiers** commune à toutes les campagnes est primordiale afin d'assurer une intégration optimale des données dans la banque du SISMER. Cette intégration est réalisée à l'aide de l'outil NEMO (développé en interne au SISMER et utilisé au niveau européen pour les projets SeaDataNet et EMODnet Chimie) permettant de formater les

¹ <https://www.seadatanet.org/Standards>

fichiers reçus en formats ASCII : ODV, MEDATLAS ou en format binaire NetCDF (formats définis pour le projet européen SeaDataNet). NEMO ne prend pas en entrée des fichiers de type Excel, Word, Open Office. Seuls les fichiers texte (ascii) peuvent être pris en compte. En outre, le format des métadonnées et des données doit être **totalelement homogène** ; c'est-à-dire qu'il ne doit pas changer d'une station à l'autre.

Pour une bonne pratique de gestion de données et en vue d'interopérabilité des données, le SISMER bancarise les données moyennées au décibar (données CTD) selon les standards IODE et propose un nombre de décimales par défaut à conserver pour l'archivage.

Le SISMER bancarise déjà, entre autre, les données physico-chimiques suivantes : la température, la salinité, l'oxygène dissous, la turbidité, la fluorescence, la chlorophylle-a, les phéopigments, la matière en suspension et les sels nutritifs.

Concernant les données de phytoplancton, le SISMER ne bancarise pour le moment que quelques données issues de 46 campagnes. Il s'appuie pour cela sur le vocabulaire de SeaDataNet utilisant les codifications taxonomique d'ITIS et du WORMS pour des paramètres standards tels que l'abondance et le comptage d'entités biologiques spécifiques.

Pour plus d'information, le portail des données marines¹ du SISMER décrit plus en détail la manière dont sont gérées les données de la réception à la bancarisation.

3.1.3 Processus de qualification

Le SISMER effectue des contrôles qualité en temps différé sur les métadonnées et données reçues. Ces contrôles qualité ont pour but de fiabiliser les données délivrées aux utilisateurs mais également de permettre des échanges de données cohérentes et comparables au niveau européen et international.

Cela consiste à : 1) s'assurer que les métadonnées sont complètes, et cohérentes par rapport aux normes européennes/internationales et autres métadonnées ; 2) mettre un indicateur qualité à chaque valeur numérique du jeu de données selon une échelle standard allant de 0 à 9². Cet indicateur de qualité (ou flag) est le résultat de contrôles automatiques puis de contrôles visuels pour validation par un opérateur (suite à des échanges avec le responsable de la donnée dans les cas douteux). Les contrôles qualité réalisés sur les métadonnées et les données sont résumés respectivement dans les tableaux 13 et 14.

La qualification est possible pour de nouveaux paramètres DCSMM du type oxygène, pH, PAR... Ces contrôles qualité sont exactement les mêmes pour toutes les campagnes et sont simplement dépendants des données envoyées au SISMER, dont le contenu et la quantité sont finalement variables selon les campagnes.

¹ Portail des données marines : <https://data.ifremer.fr/Tout-savoir-sur-les-donnees>

² Echelle de flag : http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=L20

Tableau 13 : Contrôles qualité réalisés sur les métadonnées (SISMER)

| Erreur | Résultat/Flag | Commentaire |
|-------------------------------|-----------------------|--|
| Double | Elimination | Le jeux de données contenant le plus de paramètres ou le plus précis est gardé |
| Vitesse du navire trop élevée | Date flag 4 ou 5 | Correction impossible ou possible |
| | Latitude flag 4 ou 5 | Correction impossible ou possible |
| | Longitude flag 4 ou 5 | Correction impossible ou possible |
| Station à terre | Latitude flag 4 ou 5 | Correction impossible ou possible |
| | Longitude flag 4 ou 5 | |
| Profondeur hors statistiques | Flag 3, 4 ou 5 | Profondeur douteuse, fausse or modifiée |

Tableau 14 : Contrôles qualité réalisés sur les données (SISMER)

| Erreur | Résultat/Flag | Commentaire |
|---|---------------|--|
| Profil avec seulement le paramètre de référence | Elimination | Tout le profil/série temporelle est effacé |
| Valeur hors statistiques | Flag 2 | LEVITUS, REYNAUD, MEDAR/MEDATLAS |
| Pic, Gradient trop élevé | Flag 3 | |
| Mesure sous le fond | Flag 3 | |
| Valeur hors limites régionales | Flag 4 | Un minimum et un maximum défini par région |
| Pression non croissante | Flag 4 | Pour la pression |
| Profil Constant | Flag 4 | Profil entier flagué à 4 |
| Inversion de densité > seuil fixé | Flag 4 | Pour la température et/ou la salinité |

3.1.4 Diffusion

Une fois bancarisées sur la banque de physique-chimie, si les données ne sont sous aucun moratoire, elles sont alors consultables et téléchargeables à partir du portail d'accès Nautilus¹, et sont également accessibles via l'infrastructure européenne SeaDataNet². Les fichiers de sortie se présentent sous trois formats possibles : medatlas ou odv (ascii), ou NetCDF.

¹ Portail Nautilus : <http://donnees-campagnes.flotteoceanographique.fr/>

² Portail SeaDataNet : http://seadatanet.maris2.nl/v_cdi_v3/search.asp

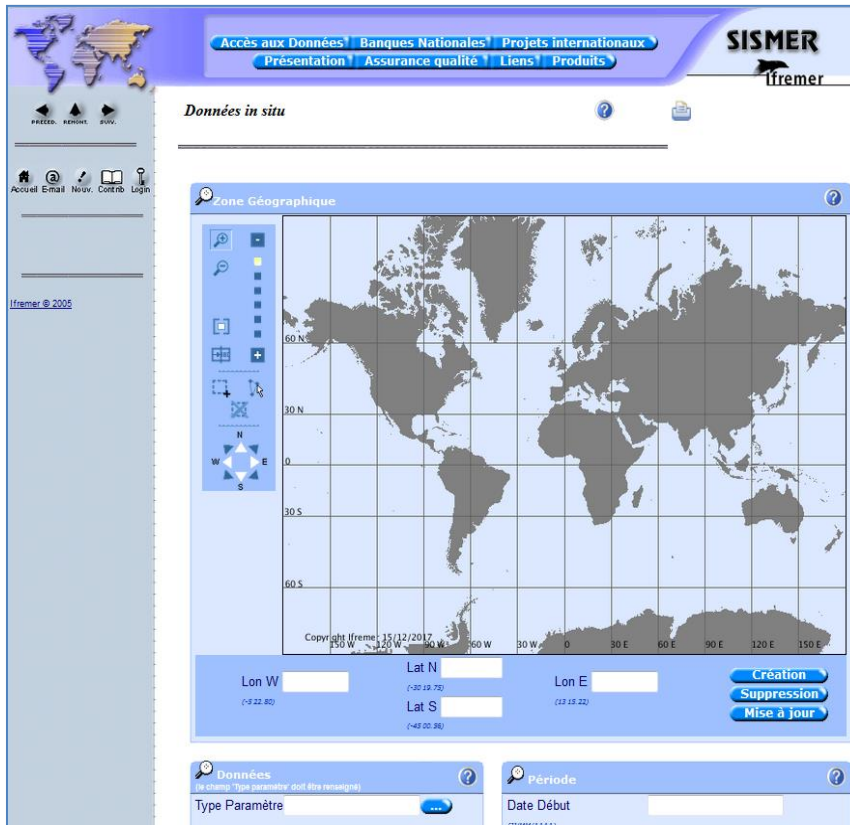


Figure 2 : Portail d'accès aux données Nautilus (SISMER)

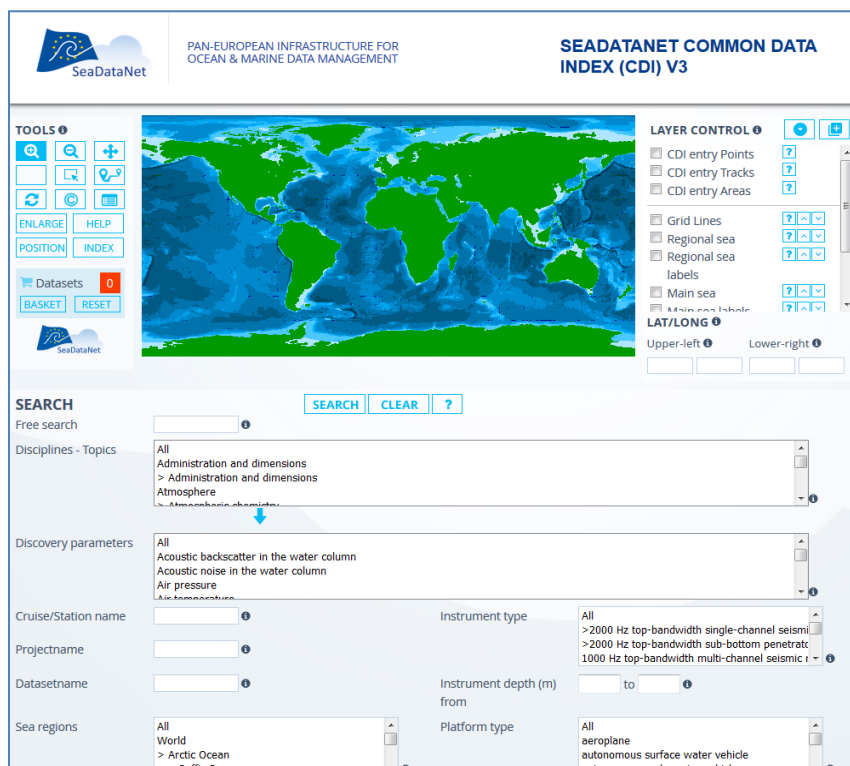


Figure 3 : Portail d'accès aux données SeaDataNet

3.2 Quadrige² (VIGIES)

3.2.1 Présentation

Quadrige² (Q²) est un système d'information développé et géré au sein du service VIGIES de l'Ifremer. Cette plateforme à la fois de bancarisation et de valorisation (grâce aux outils d'interprétation qu'elle met à disposition), gère les données de la surveillance du littoral dont les données issues des réseaux de surveillance mis en œuvre par l'Ifremer (REPHY, REBENT, REMI, ROCCH). La base de données mise à jour constamment intègre ainsi des données physico-chimiques, de phytoplancton, mais également de zooplancton (e.g. données IGA).

3.2.2 Organisation des données

L'administration et la bancarisation des données sont organisées autour de référentiels permettant la cohérence avec le référentiel national du SANDRE (Service d'Administration Nationale des Données et Référentiels sur l'Eau). Ces données de référence concernent en particulier la liste des taxons, les lieux de surveillance, les agents, les engins de prélèvement, les paramètres, supports, fractions, méthodes d'analyse, etc.

Les données intégrées à Quadrige sont réparties par programme et stratégie, et identifiées par lieux de surveillance, passages, prélèvements, échantillons. Ces informations sont importantes pour l'identification des résultats et cela nécessite donc de renseigner les métadonnées telles que le lieu (code, libellé, coordonnées), la date, et l'engin utilisé. Les valeurs quantitatives et qualitatives des résultats sont ensuite exprimées par PSFM (Paramètre - Support - Fraction - Méthode) et son unité de mesure.

Au niveau de la bancarisation, les données peuvent être intégrées au SI par deux méthodes :

- **la saisie manuelle** (23) : guidée par des grilles de saisie développées spécifiquement pour chaque programme (liste des PSFM prédéfinis, lieux de surveillance, etc.)
- **l'intégration automatisée** : résultant de l'importation des données formatées, par exemple au format Quadrilabo (24).

Pour la première méthode, un temps-agent doit être alloué régulièrement pour la saisie manuelle de nouvelles données. Les temps de saisie estimés pour les données physico-chimiques sont de 1 min par passage (un lieu à une date, généralement appelé station) pour la saisie, le contrôle et la validation, et de 500 résultats par jour.

La seconde méthode dispense de formation des agents à l'utilisation du logiciel Quadrige², mais nécessite un travail plus important en amont pour formater les données selon le modèle nommé Quadrilabo, basé sur une codification SANDRE¹ des données de référence. Il peut être nécessaire de développer un programme adapté spécifiquement aux données pour convertir les fichiers source des producteurs de données en ce format Quadrilabo, notamment pour faciliter la conversion des libellés des données de référence en code SANDRE correspondant. Si la cellule Quadrige gère le script d'import des fichiers Quadrilabo dans Quadrige, elle ne maintient pas les programmes générant les formats Quadrilabo. Cette maintenance est à la charge des fournisseurs des données.

Q2 peut en outre stocker des fichiers de mesures régulières ou ponctuelles, pour des lieux de surveillance géoréférencés. En revanche, en ce qui concerne les données haute fréquence du type FerryBox et les données de profils du type CTD, Q2 n'est pas particulièrement adapté à leur bancarisation qui nécessiterait alors des ajustements importants. De même, le SI n'est pas adapté

¹ https://wwz.ifremer.fr/quadrige2_support/Mes-donnees/J-integre-mes-donnees-a-l-aide-de-Quadrilabo

non plus à la bancarisation de gros volumes d'imagerie. D'autres systèmes performants et adaptés existent déjà par ailleurs.

3.2.3 Processus de qualification

Quadrige assure un contrôle sur les données saisies et importées. Ce contrôle consiste en la vérification de la cohérence des données vis-à-vis du cahier de paillasse ou des fiches terrain. A l'issue de ce contrôle, les données sont validées et verrouillées (elles ne sont plus modifiables). Leur extraction est alors possible pour tout utilisateur ayant accès à la base de données et la diffusion via le portail de Surval¹ est effective pour certaines données de programmes bien identifiés (REPHY, réseaux hydrologiques régionaux, REMI, ROCCH), sauf si les données sont protégées par un moratoire.

La qualification intervient après ce premier processus de vérification de la donnée. Elle correspond à la recherche des données douteuses voire aberrantes du point de vue scientifique ; la correction des données lorsque cela est possible ; l'attribution d'un niveau de qualification aux données (bon, douteux, faux). Seules les données qualifiées « Bonnes » et « Douteuses » sont alors diffusées dans Surval (les données qualifiées de « faux » sont disponibles par extraction dans l'application Quadrige²).

La qualification est composée de deux étapes :

- 1) **qualification « automatique »** qui consiste à rechercher des erreurs de cohérence au sein d'une donnée prise individuellement. Cela correspond à des erreurs « grossières » et facilement identifiables. Exemples : erreur de paramètre, de support analytique, erreur de saisie de nombre (température de 100°C au lieu de 10°C par exemple) ou incohérences (données saisies sur le niveau « surface » avec une immersion de 20 m par exemple). Ces erreurs peuvent être décelées informatiquement en définissant des règles de contrôle simples (ex : heure = 00:00:00).
- 2) **qualification « experte »** qui consiste à vérifier la cohérence d'une donnée au sein de l'ensemble des autres données via des méthodes adaptées (séries temporelles, tests statistiques...). Par exemple, pour déceler qu'une salinité de l'eau est anormale pour la saison à laquelle elle a été mesurée, il faut replacer la donnée dans une série temporelle et tenir compte de la saisonnalité (alors que le même résultat pris de façon isolée pourrait sembler correct).

3.2.4 Diffusion

Les données bancarisées sur Quadrige peuvent ensuite être consultées à partir de n'importe quel poste ayant l'application Q2 installée. Un compte utilisateur valide est cependant requis pour utiliser l'outil. Plusieurs moyens de diffusion sont par ailleurs utilisés pour mettre à disposition les données. Le schéma ci-dessous résume les différentes plateformes de diffusion nationales et internationales alimentées par les données de Q2.

¹ Surval : http://envlit.ifremer.fr/resultats/acces_aux_donnees/presentation

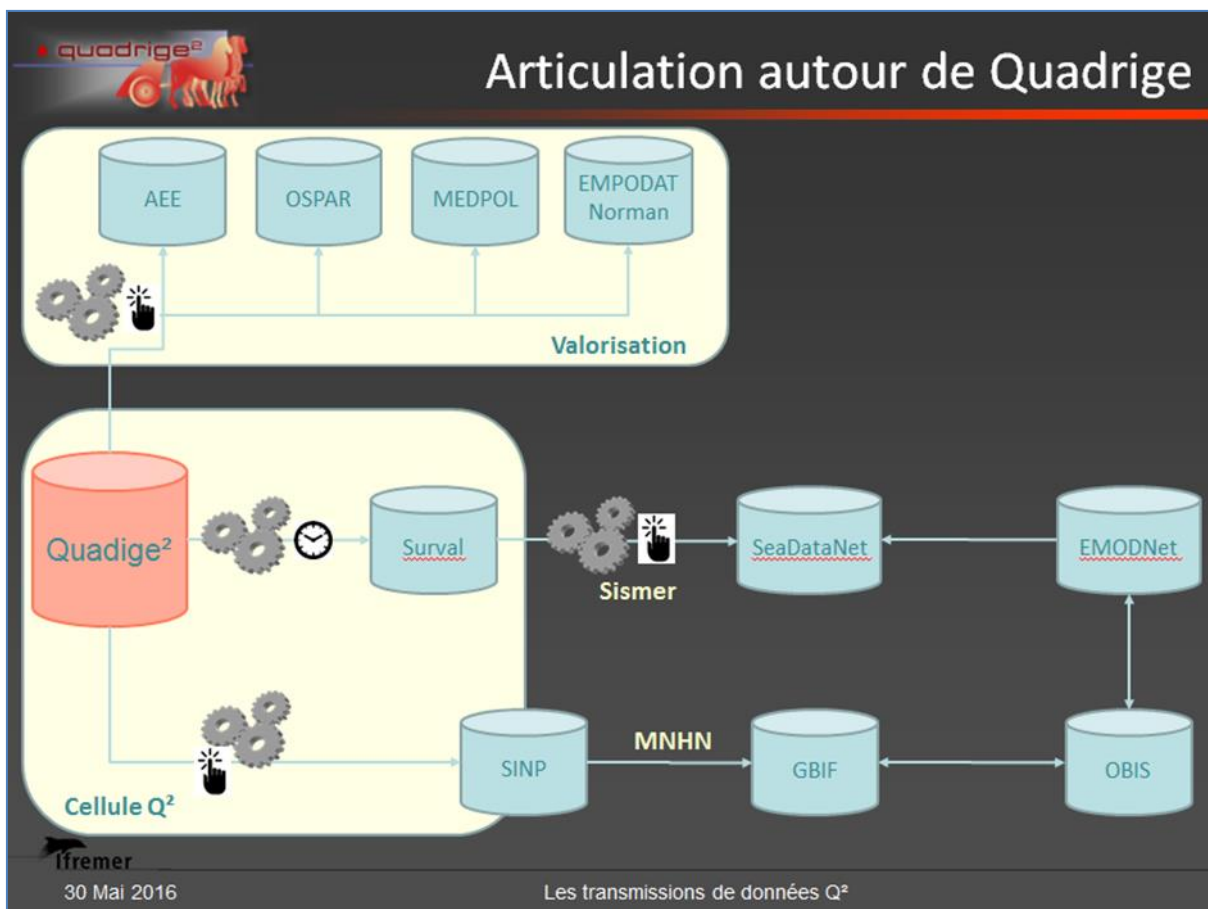


Figure 4 : Diffusion des données bancarisées sur Quadrige²

Entre autre, le site Surval hébergé par Ifremer¹ est un portail d'accès aux données en ligne et public qui permet de visualiser et télécharger les données de la surveillance du littoral issues des programmes bancarisés dans Q2. Les données validées des réseaux de surveillance (REMI, REPHY, ROCCH, etc.), non-soumises à un moratoire, sont ainsi consultables par lieu de surveillance et peuvent être filtrées par paramètre et par zone géographique grâce à l'outil cartographique.

¹ Portail Surval: <http://www.ifremer.fr/surval2/>



Figure 5 : Interface actuelle Surval (visualisation)

Une refonte totale du site Surval est en développement et intégrera bientôt l'outil Sextant pour la visualisation et le téléchargement des données en ligne. Cette nouvelle interface sera effective d'ici l'été 2018. Néanmoins, le catalogage dans Sextant est d'ores et déjà assuré et connecté directement aux couches cartographiques disponibles sur Surval.

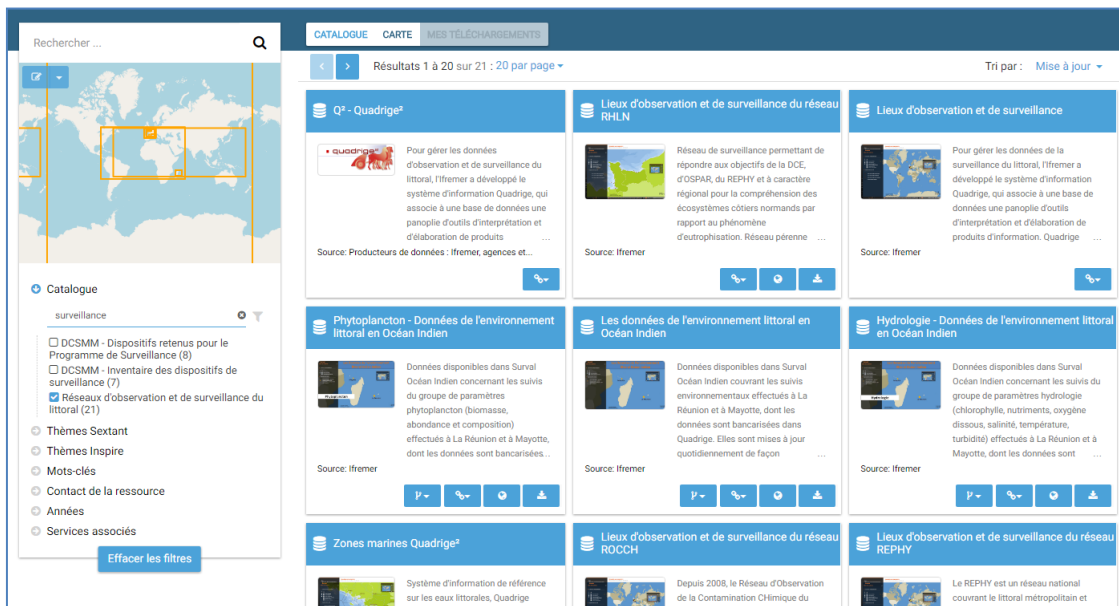


Figure 6 : Interface type Sextant (catalogue de données de Surval)

3.3 Portail Coriolis

Le portail Coriolis est destiné à centraliser la mise à disposition de données de mesures hydrologiques et biogéochimiques *in situ* (navires, flotteurs, mouillage, etc.) pour l'océanographie opérationnelle. Il compile ainsi des données nationales, européennes et internationales. Le portail¹ (data center) est géré par l'Ifremer (SISMER).

Cette plateforme de bancarisation et de mise à disposition en ligne est envisagée pour la bancarisation des données hautes fréquences issues des (Pocket) FerryBox collectées sur les six campagnes halieutiques. Le protocole d'archivage/bancarisation est en cours de développement et sera commun aux six campagnes. Il devrait être finalisé au cours du premier semestre 2018.

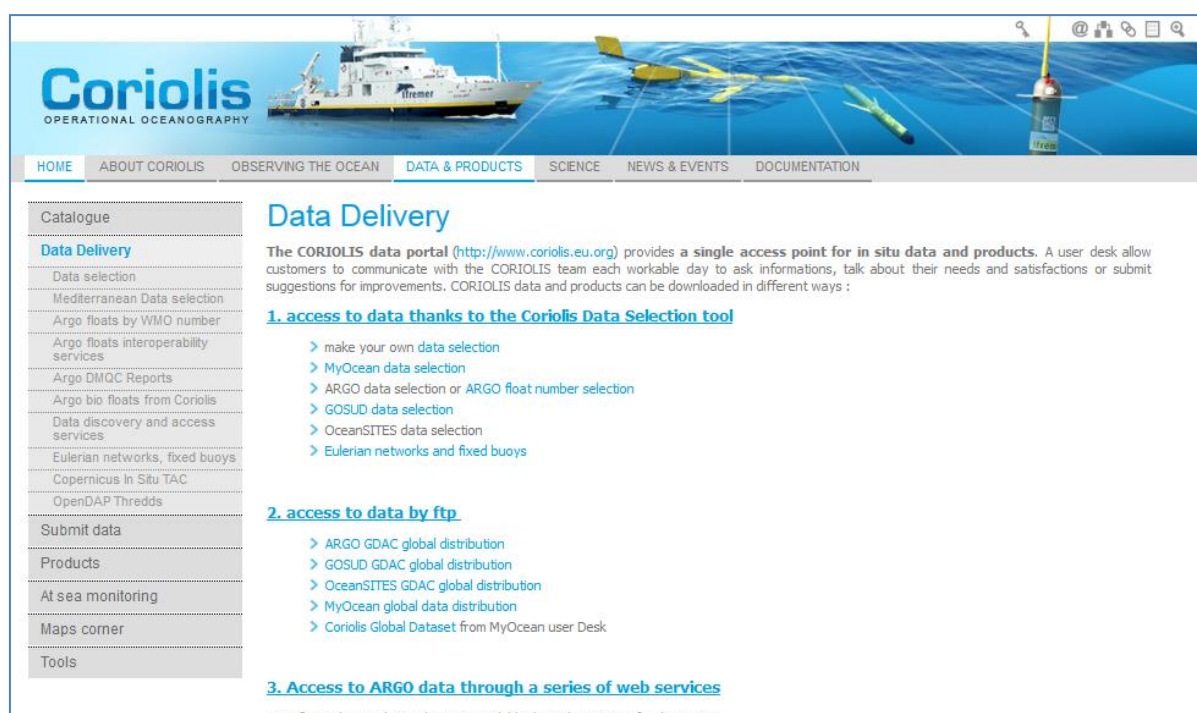


Figure 7 : Data center de Coriolis

En termes de qualification, le SISMER s'assure du contrôle en temps différé et/ou réel (section 3.1.3 pour plus d'information).

Le contrôle en temps réel est assuré pour les données à flux direct depuis le bord vers CORIOLIS. Il concerne les données servant à l'océanographie opérationnelle (observation et modélisation de l'océan en temps réel).

3.4 Autres plateformes de travail

3.4.1 Ecotaxa (OOV)

La plateforme Ecotaxa (25) est une application web thématique gérée par l'Observatoire Océanologique de Villefranche-sur-Mer (OOV - CNRS). C'est un **outil de travail** qui permet la bancarisation, la classification automatique, le tri manuel et le partage de données d'imagerie taxonomiques de plancton. Ecotaxa est utilisée pour dépouiller, formater et analyser les données

¹ <http://www.coriolis.eu.org/Data-Products>

d'imagerie du zooplancton des campagnes PELGAS et EVHOE (depuis 2016), pour les échantillons de zooplancton collectées par filet WP2 et CUFES. L'intégration des données de zooplancton sur Ecotaxa nécessite que les données aient été traitées par des outils d'imagerie tels que le ZooCam, le ZooScan, le FlowCam, etc.

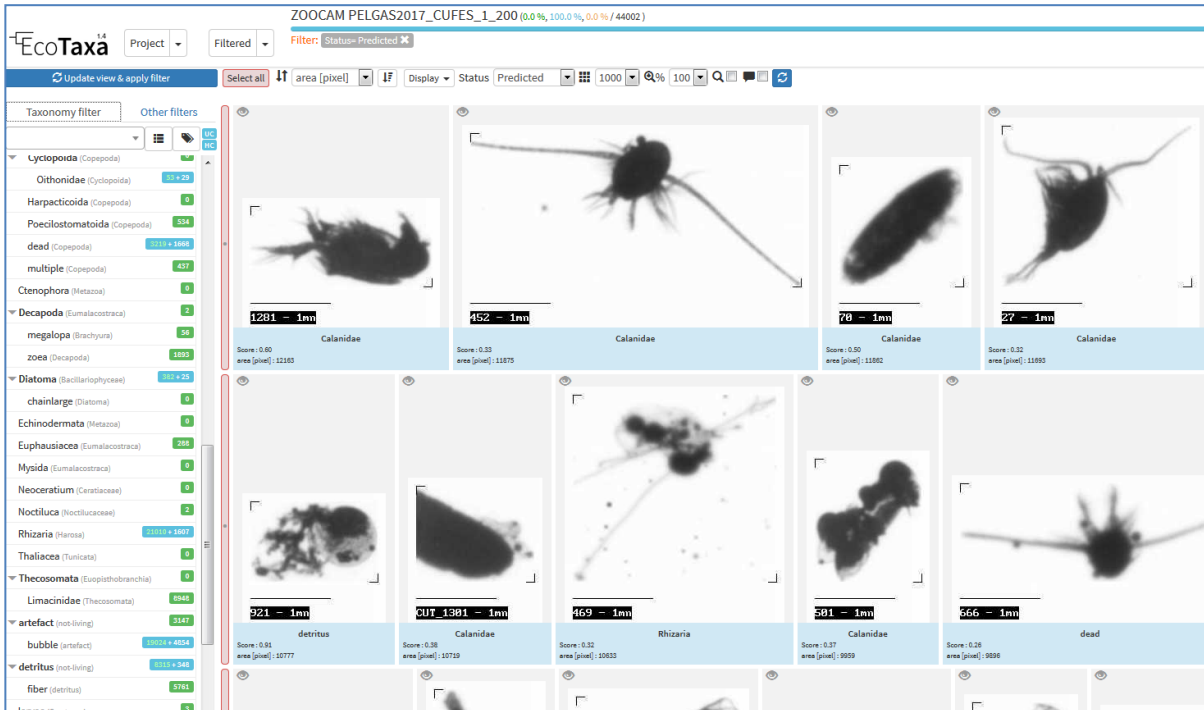


Figure 8 : Ecotaxa - Interface de consultation de la base de données d'imagerie en ligne

Outre le stockage d'images individuelles d'organismes planctoniques et des métadonnées associées, Ecotaxa offre un module d'**identification taxonomique automatique** faisant intervenir des méthodes d'apprentissage machine (machine learning) avancées (Random Forest et Deep Learning par réseau de neurone convolutionnel) et efficaces. Les images importées dans la base de la plateforme peuvent ainsi être identifiées automatiquement (prédiction) et classées par groupes taxonomiques dans une taxonomie de référence (uniEuk), puis l'utilisateur peut valider et/ou corriger les prédictions manuellement si nécessaire.

Concernant les campagnes PELGAS et EVHOE, les données d'imagerie par ZooCam sont dans un premier temps envoyées et stockées sur un disque dédié pour les campagnes DCF (SISMER). Un premier tri des images issues du CUFES est réalisé à bord avec le logiciel du ZooCam (prédiction et validation des œufs de poissons essentiellement). Ces images (triées et non triées) et leurs données associées sont ensuite importées sur Ecotaxa pour permettre un second tri détaillé et la validation manuelle. Les données issues des prélèvements par filet WP2 sont elles aussi importées pour être triées et identifiées.

L'application permet une organisation des données par projet (ici les campagnes), et permet la bancarisation des vignettes avec leurs annotations, la consultation, le partage, et l'export automatisé des résultats sous forme de **tableaux de comptage**. Le détail des données caractéristiques des individus imagés (taille, couleur, forme, identification taxonomique) est également une sortie standard d'Ecotaxa. On retrouve ainsi toutes les métadonnées nécessaires à l'identification de l'observation (campagne, station, coordonnées, date, heure et profondeur, engin de prélèvement), associées aux données de chaque objet imagé du projet. La base de données d'images peut ainsi être filtrée selon ces critères.

L'application permet également l'export des données vers des fichiers au format tsv (ascii avec séparateur en tabulation) accompagnés ou non des vignettes. Ce format structuré par colonne et

standard simplifie l'**importation vers des systèmes d'information** tels que Quadrige² ou les banques du SISMER, grâce au développement de programmes d'intégration spécifiques.

A noter qu'EcoTaxa n'est pas un outil de bancarisation en tant que tel. En revanche, cette application thématique est très **utile en amont de la bancarisation** comme outil de travail pour l'analyse, le formatage et le partage des données d'imagerie planctonique.

En 2017, la plateforme hébergée à l'OOV comptait plus de 30 millions de vignettes, dont 30 % étaient annotées et validées. L'Ifremer est propriétaire d'environ 10 % de ces données (données de campagnes PELGAS et EVHOE depuis 2016). Etant donné le volume de données généré par imagerie sur les campagnes un projet d'implantation d'une **instance d'Ecotaxa** au sein du SISMER est en cours (avec l'infrastructure de l'IDM) pour rapatrier les données de campagnes à l'Ifremer.

3.4.2 Datarmor (SISMER)

L'outil de stockage/calcul/traitement de l'Ifremer anciennement appelé Caparmor a évolué en **Datarmor**. Cette infrastructure possède désormais une puissance de calcul plus élevée et se divise en deux espaces de stockage : Datawork, un espace de projet permettant de stocker et traiter des données pour des travaux d'équipes en cours ; et Dataref, un serveur destiné à stocker des « données de référence » dans le but de les rendre accessible à une plus ou moins grande communauté (accès contrôlé). Il s'agit d'un espace **non sauvegardé**. Les espaces Datawork et Dataref peuvent communiquer de manière à intégrer des scripts de calculs pour la création d'indicateurs par exemple.

Les procédures d'archivage des données à bord des navires sont actuellement en cours d'évolution pour intégrer les données ZooCam aux données de référence « données navires » stockées sur Dataref. Ces images (ou « frames ») et vignettes seront transmises par Genavir au SISMER dans un délai de deux mois après la fin de la campagne via le flux de données navires déjà existant. Chaque lot de données reçu est **archivé de manière pérenne** sur des infrastructures IDM (sur **bandes magnétiques**) avant d'être stocké sur Dataref.

4 Discussion

L'état de l'art sur la collecte des données hydrologiques au cours des six campagnes DCF a permis d'identifier une importante hétérogénéité en termes de méthodes de collecte et d'outils de bancarisation. Cette hétérogénéité est particulièrement notable entre les différents types de données collectées :

- **Physico-chimie de la CTD**
- **Physico-chimie de la (Pocket) FerryBox (PFB)**
- **Physico-chimie des bouteilles Niskin**
- **Phytoplancton des prélèvements par bouteilles Niskin**
- **Zooplancton des prélèvements par filets / CUFES**

4.1 Vers une bancarisation multiplateforme

L'objectif premier étant d'harmoniser les données mises à disposition, il est prioritaire que les différents types de données soient uniformisés au moment de la restitution auprès de la DCSMM. Pour cela, deux possibilités se présentent en vue du portail SIMM :

- 1) une uniformisation et une **restitution** de chaque type de données **au niveau de chaque SI** où elles sont bancarisées : les données sont regroupées par type sur les SI les plus pertinents, et la restitution auprès des pilotes de la DCSMM se fait directement à ce niveau. Ils devront ensuite fusionner les différentes sources.
- 2) une uniformisation pour chaque type de données et une **restitution** au niveau du **portail SIMM** : les données sont regroupées par type sur les SI les plus pertinents, et la restitution auprès des pilotes de la DCSMM se fait via le portail SIMM qui irait rechercher les données et les fusionnerait selon la requête d'extraction.

La première solution n'est *a priori* pas l'alternative la plus pertinente selon les besoins des pilotes de la DCSMM. En effet, envisager une uniformisation et une restitution de chaque type de données en amont du portail SIMM implique que les données soient rassemblées sur un minimum de banques, voire une seule, afin d'éviter la multiplication des extractions avant de pouvoir exploiter les données dans leur ensemble. Or, aucun SI n'est aujourd'hui capable de bancariser de manière optimale l'ensemble de ces données dont les thématiques sont bien différentes (Tableau 15). Par conséquent, les extractions devraient être accompagnées d'un traitement pour fusionner les différentes bases ainsi récupérées. La multiplication des outils et des tâches que les utilisateurs auraient donc à faire avant de pouvoir exploiter les données est un obstacle important du point de vue de l'optimisation et de l'harmonisation des données.

Tableau 15 : Compétences actuelles des banques en fonction du type de données

| | Coriolis | Banque(s) SISMER | Quadrigé ² |
|--|----------|------------------|-----------------------|
| Données physico-chimique haute-fréquence (type PFB) | V | X | X |
| Données physico-chimique de profils verticaux (type sonde CTD) | X | V | X |
| Données physico-chimique (type bouteilles Niskin) | X | V | V |
| Données taxonomiques de phytoplancton | X | X / V | V |
| Données taxonomiques de zooplancton | X | X | V |

X : n'est pas en mesure d'accueillir ce type de données à l'heure actuelle ; V : est en mesure et/ou bancarise déjà ce type de données de manière structurée

La seconde solution préconisant la restitution des données au niveau du portail SIMM semble donc la plus pertinente. Or, chacun des systèmes de bancarisation envisageables présente des avantages et des inconvénients selon le type de données. Plusieurs alternatives peuvent donc être envisagées pour la bancarisation des données en amont du portail SIMM (alternatives développées plus bas).

4.2 Assurer la fusion des bases de données

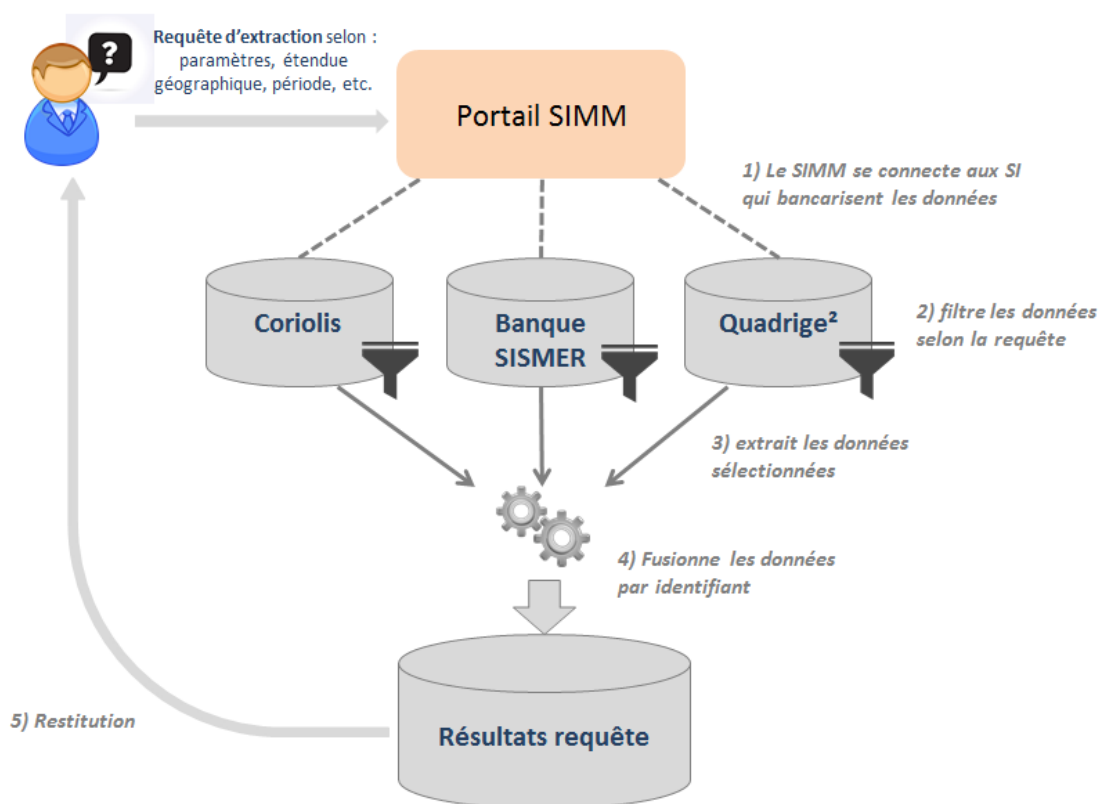


Figure 9 : Fusion et restitution des données par simple requête sur le portail SIMM

Quelle que soit l'alternative retenue, la bancarisation des données sur différentes bases implique de **conserver des passerelles** (ou des liens) permettant simplement de fusionner les

données au niveau du SIMM, pour que la restitution se fasse en **une seule extraction** selon la requête formulée. Cette **clé commune** devra être créée à partir des éléments qui caractérisent les mesures et permettent de les identifier de manière unique. Ces éléments concernent en particulier la station, associée à ses coordonnées (longitude / latitude), la date, et l'heure de l'opération. Chaque responsable de campagne utilise déjà plus ou moins ses propres identifiants de station en interne pour pouvoir relier les tables et/ou les fichiers entre eux, mais ceux-ci sont très hétérogènes et leur utilisation n'est pas systématique.

Actuellement, le SISMER utilise un algorithme pour générer des identifiants uniques de campagnes composés de 13 caractères (ex : FI35201600240 pour la campagne PELGAS 2016). Cet identifiant peut être prolongé d'une série de 5 caractères identifiant les stations (ex : FI35201600240U0182 pour la station U0182 de la campagne PELGAS 2016). Cet identifiant de station est repris de manière unique au sein du SISMER afin de pouvoir fusionner les données physico-chimiques des bouteilles Niskin (ex : sels nutritifs) et les données hydrologiques des CTD (ex : température).

De la même manière, la présente étude préconise de reprendre ces identifiants de stations pour la bancarisation des données sur le SI Quadrigé² dans le champ « mnémonique du passage », spécialement créé pour stocker ce genre d'information (section 3.2.2). De plus, l'outil de travail Ecotaxa permet lui aussi de créer des champs supplémentaires pour l'annotation des imagettes. Le champ d'identification issu du SISMER pourra donc là aussi être repris aisément.

4.3 Bancarisation des données physico-chimiques

Dans un premier temps et en conclusion des travaux menés, l'étude préconise la bancarisation de toutes les **données physico-chimiques** actuellement collectées **au SISMER** par l'utilisation de 1) la banque physique-chimie et 2) le portail Coriolis.

4.3.1 FerryBox et Pocket FerryBox

Concernant les nouvelles données **haute fréquence** issues des **(Pocket) FerryBox**, la collecte et l'archivage/bancarisation sont sur le point d'être mis en place de manière harmonisées. Le SISMER aura pour tâche de réceptionner les données depuis le bord, pour une qualification et une bancarisation vers le portail **Coriolis**, qui est adapté pour ce type de données actuellement collectées. Les **protocoles** de collecte et de bancarisation, **communs aux six campagnes**, sont en cours de développement et seront finalisés dans le premier semestre 2018. Les données seront alors bancarisées sous format netcdf.

4.3.2 Sonde CTD

S'agissant de la sonde CTD, la plateforme de bancarisation qui recevra l'intégralité des données est d'ores et déjà identifiée pour les six campagnes. En effet, les trois façades envoient dès à présent leurs données CTD au SISMER, afin qu'il les stocke et les archive sur la **banque physique-chimie**, même si ce travail n'est pas encore finalisé pour toutes les campagnes (*i.e.* MEDITS). En revanche, bien que la banque soit définie, les **traitements et la mise en forme** des données envoyées au SISMER ne sont pas encore communs à l'ensemble des campagnes. Il est donc important de définir un **format unique** en fonction des besoins pour une bancarisation homogène.

Les traitements réalisés en amont de l'envoi au SISMER doivent être identiques pour toutes les campagnes. Actuellement, ces traitements sont en grande partie les mêmes sur les façades Manche-Mer du Nord et Atlantique, mais absents sur la façade Méditerranée. Outre la suppression des lignes aberrantes, l'alignement et le moyennage des données sont requis, et ces traitements peuvent être facilement appliqués à l'aide du programme SBE avant l'envoi au SISMER. Afin de garder une résolution la plus fine possible, la présente étude préconise un moyennage au demi-

décibar pour les besoins de la DCSMM (comme réalisé sur IBTS et CGFS), bien que les standards adoptés par le SISMER préconisent le décibar. De plus, ces traitements en amont doivent être accompagnés d'une mise en forme spécifique afin d'assurer une bancarisation effective par le SISMER. Ces conditions concernent en particulier la conversion du fichier hex au **format csv**, et l'intégration des **métadonnées indispensables en en-tête du fichier** (Figure 10) incluant :

- le nom de la campagne et de la station
- la date et heure (TU)
- latitude/longitude
- nom des paramètres et leur unité
- les paramètres de référence verticale du type pression (dbar)
- les types de capteurs utilisés (avec les numéros de série)
- les traitements opérés
- les coefficients de calibration des capteurs (à partir du fichier xmlcon)

Concernant le nombre de décimales à conserver, les capteurs de la sonde n'ayant pas systématiquement la même précision, il est préconisé de garder **le maximum de décimales** pour chaque capteur, afin de pouvoir accéder à l'information la plus précise possible et éviter toute perte d'information. Cette précision des capteurs est à renseigner lors de chaque envoi de données au SISMER.

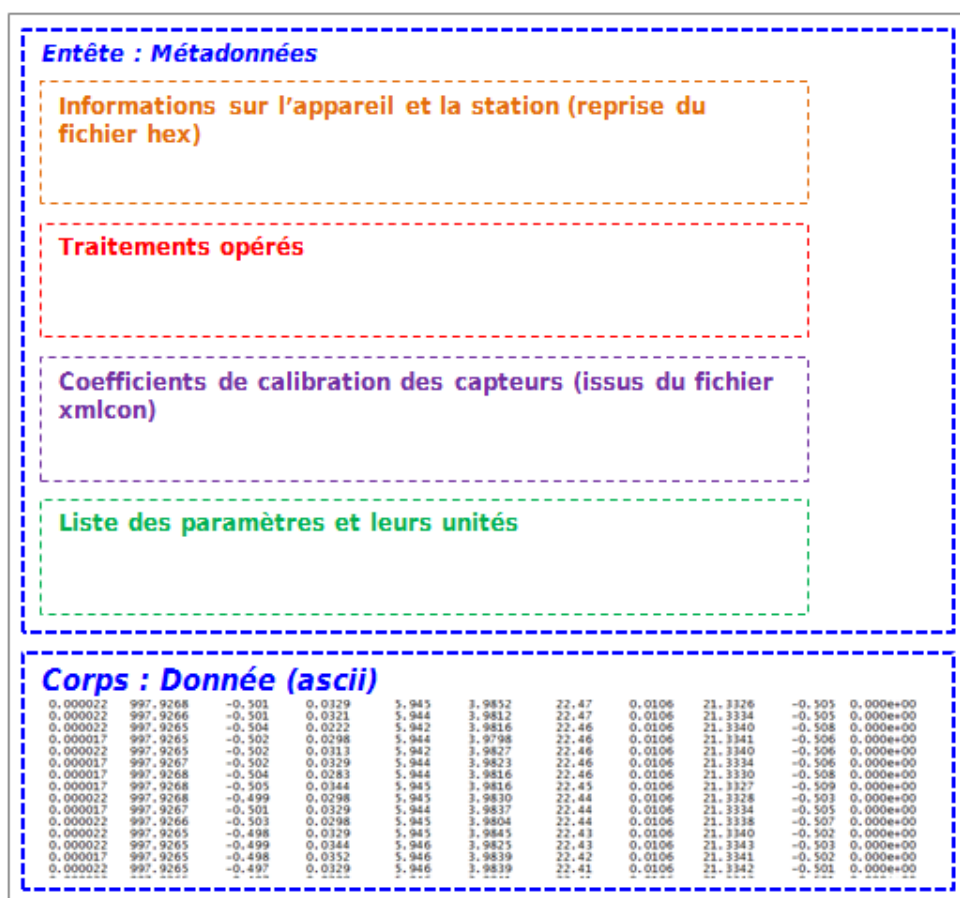


Figure 10 : Structure du fichier csv de la CTD envoyé au SISMER

4.3.3 Bouteilles Niskin

Cette solution envisage également la bancarisation des données physico-chimiques des bouteilles Niskin dans la **banque physique-chimie** du SISMER, aux côtés des données du même type de la CTD. A l'instar de ce qui est fait pour la campagne PELGAS (section 1.2.3), le **fichier csv** (ou xlsx) envoyé intègre les données de concentration de matière en suspension, chlorophylle-a, ainsi que les concentrations en phéopigments et sels nutritifs lorsque ces paramètres sont collectés. Ce fichier doit également être **accompagné des métadonnées** correspondantes :

- le nom de la campagne et de la station
- la date et heure (TU)
- latitude/longitude
- les paramètres de référence verticale du type profondeur (m)
- le nom des paramètres et leur unité

L'exemple ci-dessus (Tableau 16) illustre le format dans lequel pourront être envoyées les données physico-chimiques issues des bouteilles Niskin. On retrouve ainsi les colonnes de métadonnées permettant d'identifier les données et leur origine (colonnes 1 à 6), suivi des paramètres mesurés avec leur unité précisée dans l'en-tête (colonnes 7 et plus).

Tableau 16 : Données physico-chimiques des bouteilles - Format du tableur à fournir au SISMER

| 1 | 2 | 3 | 4 | 5 | | 6 | 7 | 8 | 9 |
|-----------|-------------|------------|------------|-------------|-----|------------|-------------------|-----------------|-----------------|
| Campagne | Station | Date | Heure (TU) | Coordonnées | | Profondeur | Param 1 (unité) | Param 2 (unité) | Param 3 (unité) |
| NOM_ANNEE | IDENTIFIANT | jj/mm/aaaa | hh:mm:ss | LON | LAT | M | Concentrations... | | |

Il est primordial que les formats qui seront envoyés au SISMER suivent la même structure pour toutes les campagnes et stations afin que le travail d'uniformisation et de formatage des données à l'aide de l'outil NEMO soit efficace et optimal (voir section 3.1.2).

Concernant, les **données biologiques** (phyto- et zooplancton), le choix est plus large puisqu'aucun système ne bancarise encore de données issues de ces campagnes en routine. De plus, **l'aspect taxonomique** et les méthodes d'analyse par microscope et/ou à l'aide d'un appareil à imagerie complexifient le processus de bancarisation. Il est donc important de considérer l'ensemble des **alternatives** possibles afin d'identifier les **avantages et les inconvénients** pour chacune d'elles. Par conséquent, la suite de ce rapport se concentre uniquement sur la bancarisation des données biologiques, les données physico-chimiques demeurant telles que présentées dans la section 4.3.

4.4 Bancarisation des données biologiques : Alternative Quadrige²

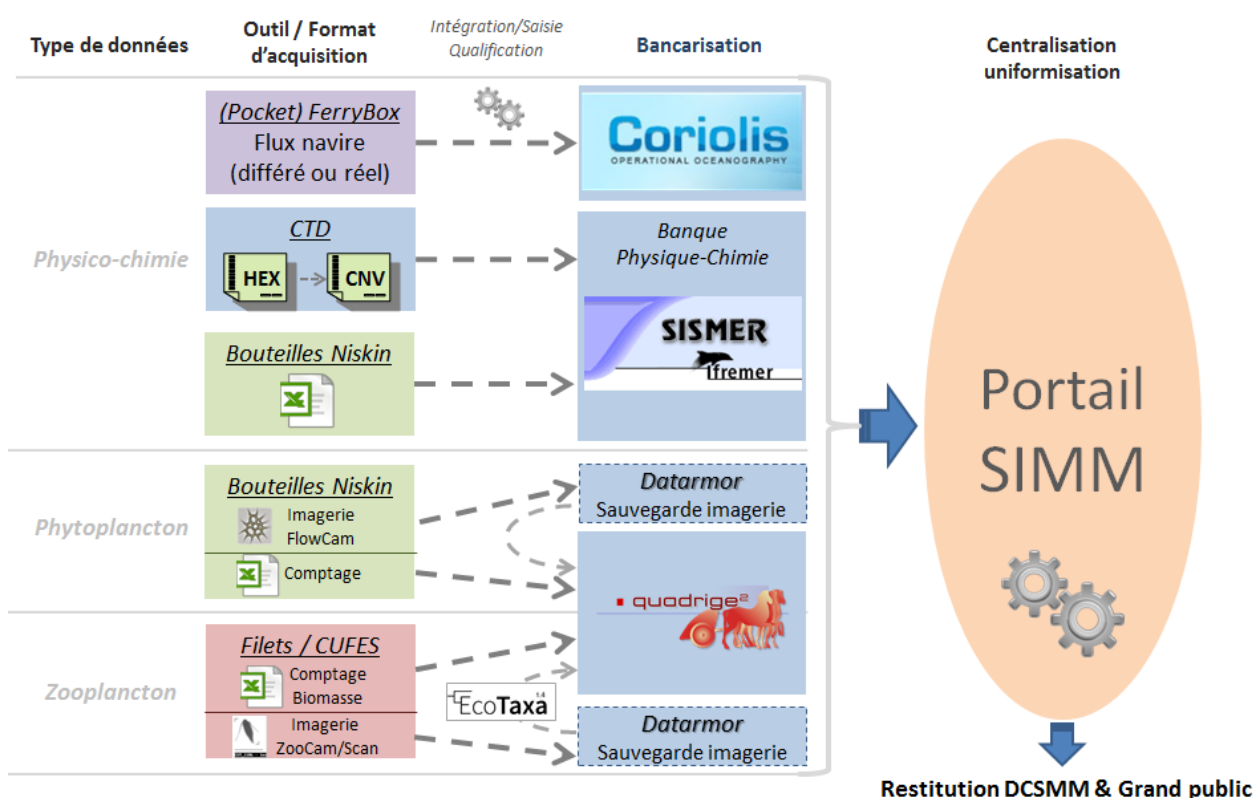


Figure 11 : Alternative Quadrige² - processus de la collecte à la restitution par type de données

4.4.1 Données de phytoplancton

Les données de phytoplancton sont collectées à partir des prélèvements de bouteilles Niskin, au même titre que les données physico-chimiques (sels nutritifs, chlorophylle-a, etc.). En revanche, s'agissant de données taxonomiques, le format et les contraintes liées à la bancarisation de ce type de données imposent une structure et un format différent.

En outre, **Quadrige²** a entre autre été développé autour de la gestion de référentiels taxonomiques, et certains programmes de surveillance du littoral issus du REPHY intègrent déjà des données taxonomiques de phytoplancton issues des prélèvements de bouteilles Niskin. Le SI paraît donc tout à fait approprié pour ce type de données de comptage.

4.4.1.1 Comptages

Les données de phytoplancton issues des analyses au microscope concernent en particulier les abondances taxonomiques pour les stations Niskin réalisées actuellement sur les campagnes CGFS, IBTS, et PELMED. Les données sont actuellement stockées sur différents formats mais les

informations à bancariser sont toutes disponibles. Sur le même principe que pour les autres programmes de Q2, les résultats seraient bancarisés par passage (= station) et prélèvement (= niveau du prélèvement). Il est donc important que les métadonnées nécessaires à leur identification soient fournies :

- nom de la station (sera attribué au mnémonique du passage)
- niveau de prélèvement (sera attribué au mnémonique du prélèvement)
- latitude / longitude
- date et heure (TU)
- profondeur (m)

Concernant les résultats, le paramètre d'abondance de chaque espèce sera exprimé par un PSFM qui sera à créer sous Q2, et dont l'unité sera en « nombre de cellules/L ». En outre, il sera nécessaire de **renseigner toutes les informations liées à la méthode** ayant servi à analyser les prélèvements.

Ce paramètre d'abondance étant du type taxonomique, la gestion des référentiels est essentielle dans la bancarisation de données de plancton car elle permet de maintenir les listes de taxons à jour dans la base et d'assurer l'utilisation correcte des noms d'espèces et de leur classification (arbre) selon les référentiels internationaux tels que le WoRMS et ERMS. L'administration du référentiel sous Quadrige² nécessite d'établir en amont la liste des taxons concernés par la bancarisation, afin de prévoir l'intégration éventuelle de nouveaux noms absents du référentiel.

4.4.1.2 Intégration / saisie

La bancarisation sera dépendante du volume des données, et du temps-agents allouable aux différentes tâches associées. Bien que l'intégration par formatage des données au format Quadrilabo semble a priori une solution simple et pérenne, le développement des programmes en amont sera important et devra être opéré par un agent maîtrisant des outils comme R. Or, une fois développés et opérationnels, ces programmes devront être maintenus et régulièrement mis à jour en fonction de l'évolutivité des données en entrée et des éventuels cas particuliers. En outre, la mise en place de nouvelles règles de contrôles telles que le contrôle des maximum/minimum ne seront pas applicables dans un Quadrilabo. De plus, il faut noter qu'il sera nécessaire de développer un programme d'intégration spécifique pour chaque format de données en entrée ; c'est-à-dire que les fichiers actuellement stockés pour PELMED devront avoir leurs propres programmes d'intégration différents de ceux pour la façade Manche-Mer du Nord, dû au fait que les données ne se présentent pas de la même manière (sections 1.1.3 et 1.3.3).

La saisie manuelle, en revanche, est beaucoup plus flexible. Les grilles de saisie sous Q2 sont facilement évolutives, et la gestion des listes de référentiels et des PSFM permet des saisies semi-automatisées évitant par exemple au saisisseur de taper manuellement des noms de taxons erronés. Comme pour l'intégration, cette méthode impliquera l'allocation à ce titre d'un temps-agent pour chaque façade concernée.

Quel que soit la méthode retenue, il sera indispensable de désigner un responsable de programme pérenne qui puisse faire le lien entre les équipes des campagnes et la cellule Quadrige, afin que la bancarisation soit assurée.

La présente étude recommande d'adopter des structures de fichiers d'entrée communes à toutes les campagnes et stations, afin de limiter le nombre de méthodes différentes pour intégrer les données de phytoplancton.

4.4.1.3 Problématique de l'imagerie

Bien que les analyses soient actuellement réalisées au microscope (sections 1.1.3 et 1.3.3), de futurs outils d'analyse par imagerie, comme le FlowCam, sont à prévoir. A l'instar des données d'imagerie par ZooCam et ZooScan, ces imagerie de phytoplancton pourront être directement envoyées sur un disque dédié du SISMER (sections 3.4.1 et 3.4.2). L'imagerie sera ainsi archivée sur bande magnétique et pourra être récupérée à tout moment pour valider les premières identifications et/ou continuer le tri sur des outils de travail comme Ecotaxa (section 3.4.1). Un identifiant ou code présent à la fois dans le fichier d'imagerie et dans les données du prélèvement dont elle est issue permettra de relier chaque imagerie à ses métadonnées sources et aux résultats de comptage de l'échantillon auquel elle appartient.

4.4.2 Données de zooplancton

Pour les données d'abondance et de comptage de zooplancton, il existe d'ores et déjà deux cas de figure au niveau des données stockées : les données d'analyse par ZooCam ou ZooScan sous forme d'**imagerie** annotée, et les données issues de l'analyse au microscope sous forme de **tableaux de comptage**. Comme pour les données de phytoplancton (section 4.4.1), **Quadrigé²** est un système adapté et éprouvé pour la bancarisation de données de comptages zooplanctoniques, grâce à la gestion des référentiels taxonomiques et des paramètres sous forme de PSFM. En revanche, l'imagerie ne peut pas être intégrée au SI.

4.4.2.1 Comptages et biomasses

Les données de zooplancton concernent en particulier les abondances taxonomiques analysées à partir des prélèvements par filets et CUFES, et les biomasses par classe de taille. Bien que plusieurs méthodes d'identification et de comptage soient actuellement pratiquées (identification par ZooCam ou ZooScan, identification visuelle, identification par EcoTaxa, microscope), les résultats peuvent être mis en commun sous un même format puisqu'il s'agit finalement de tables d'abondance et de biomasse par taxon. Quadrigé² peut ainsi stocker toutes données de ce type en utilisant des paramètres déjà pris en charge pour d'autres programmes recueillant le même type de données (IGA). Ainsi, à l'instar des données phytoplanctoniques, il sera nécessaire de renseigner les métadonnées utiles à l'identification des résultats :

- nom de la station (sera attribué au mnémonique du passage)
- niveau de prélèvement (sera attribué au mnémonique du prélèvement)
- latitude / longitude
- date et heure (TU)
- profondeur (m)

Concernant les résultats, le paramètre d'abondance de chaque espèce sera exprimé par un PSFM créé sous Q2, et dont l'unité pourra être en « nombre par 10 m³ ». En outre, il sera nécessaire de **renseigner toutes les données liées à la méthode** ayant servi à analyser les prélèvements.

Ce paramètre d'abondance étant du type taxonomique, la gestion des référentiels est essentielle dans la bancarisation de données de plancton car elle permet de maintenir les listes de taxons à jour dans la base et d'assurer l'utilisation correcte des noms d'espèces et de leur classification (arbre) selon les référentiels internationaux tels que le WoRMS et ERMS. L'administration du référentiel nécessite d'établir en amont la liste des taxons concernés par la bancarisation, afin de prévoir l'intégration éventuelle de nouveaux noms absents du référentiel.

4.4.2.2 Intégration / saisie

Concernant la bancarisation, les données étant actuellement analysées et stockées de manière différente selon les campagnes et les paramètres, il sera nécessaire de recourir à plusieurs méthodes pour intégrer les données. En effet, les données issues des analyses de l'imagerie (ZooCam et ZooScan) sont stockées sur table Access pour la façade Manche-Mer du Nord, et sur l'application EcoTaxa pour la façade Atlantique. Or, les données issues des analyses au microscope pour la façade Méditerranée sont quant à elles stockées sur fichiers csv individuels.

L'intégration automatisée par formatage des données nécessite de développer un programme spécifique pour chaque source de données différente. Pour les données stockées sur ExoTaxa et base Access, cette méthode paraît pertinente car ces outils permettent des exports de données sur fichier de type ascii, suivant des structures régulières. En revanche, pour les autres données entrées manuellement sur des tableurs csv individuels, le développement de programmes d'intégration spécifiques risque d'être plus aléatoire du fait que les informations sont plus disparates. Dans ce cas, il faudra probablement privilégier la saisie manuelle sur le SI, à l'aide de grilles spécialement préparées. Il en est de même pour certaines données historiques ou de biomasses restant encore sur fichier csv (*i.e.* façade Atlantique - section 1.2.4).

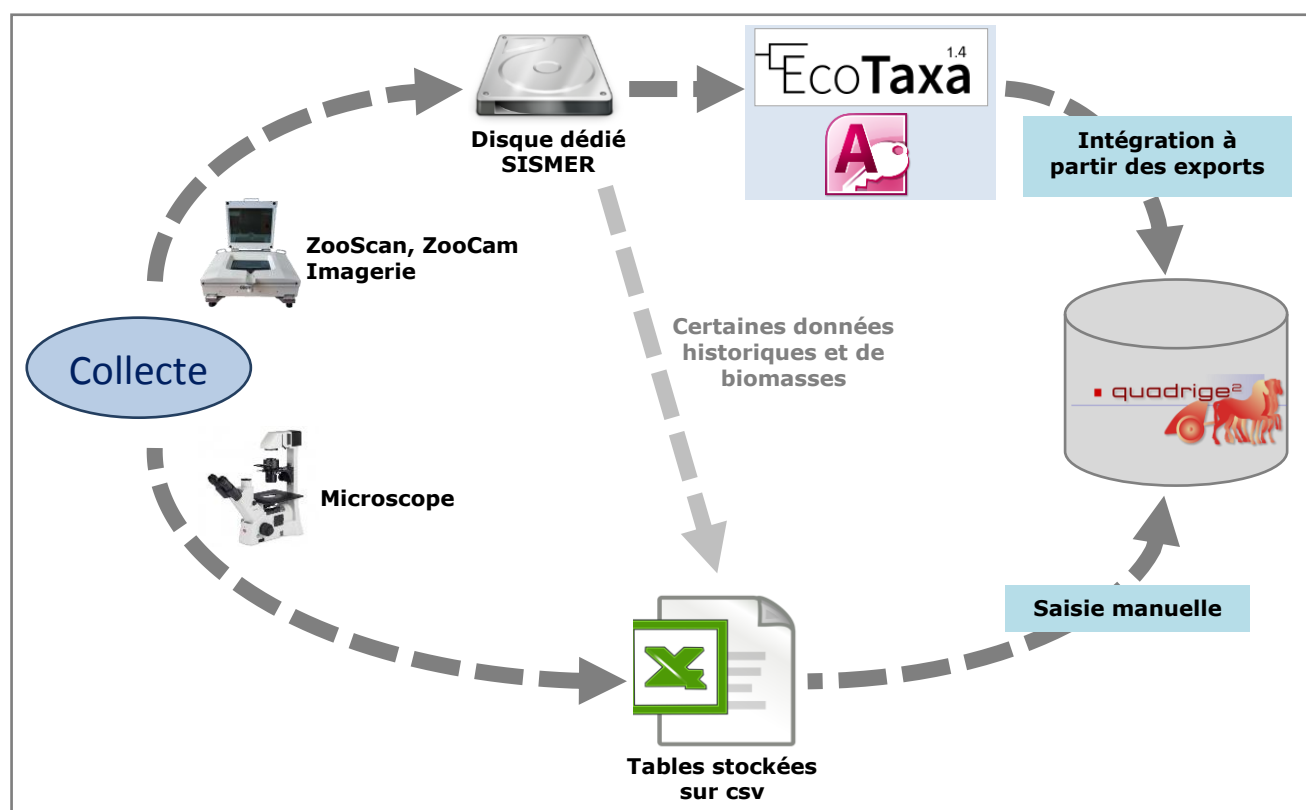


Figure 12 : Processus de bancarisation des données de zooplancton

La présente étude recommande d'adopter des structures de fichiers d'entrée communes à toutes les campagnes et les stations, afin de limiter le nombre de méthodes différentes pour intégrer les données de zooplancton.

4.4.2.3 Problématique de l'imagerie

Comme décrit plus haut les analyses par ZooCam et ZooScan sont faites à partir d'images qui sont directement envoyées depuis le bord vers disque de stockage au SISMER dédié aux campagnes DCF. Quadrige² n'étant pas adapté pour stocker ce type de données, seules les résultats de comptages et de biomasses taxonomiques seront bancarisés. Les images à

disposition des scientifiques des campagnes pourront continuer d'être stockées au niveau du SISMER, avec une sauvegarde sur bande magnétique, mais leur restitution au niveau du portail SIMM n'est pas nécessaire.

4.4.3 Validation des critères de choix

Tableau 17 : Alternative Quadrigé² - Liste des avantages et des inconvénients

| Critères de choix | Alternative Quadrigé ² |
|---|--|
| <i>Facilité d'accès aux données</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Diffusion publique des données physico-chimiques des campagnes d'ores et déjà bancarisées au SISMER via les portails Nautilus, SeaDataNet et Coriolis. - Forte visibilité de Quadrigé² pour les données biologiques marines - Diffusion des données via l'application Q2 et le futur portail Surval - Le portail SIMM prévoit de faire le lien pour aller chercher des données sur Q2 et au SISMER dans le cadre d'autres programmes de surveillance |
| <i>Uniformité des données</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Quadrigé² bancarise également les données de phyto- et zooplancton issues de la surveillance du littoral - Toutes les données biologiques sont rassemblées sur une même base de données <p>MOINS</p> <ul style="list-style-type: none"> - Les données physico-chimiques des bouteilles Niskin ne sont pas bancarisées au même endroit que les données biologiques issues de ces mêmes bouteilles - Nécessité de développer un outil permettant de rendre disponibles les données biologiques issues de Quadrigé² sur le catalogue national des campagnes géré par le SISMER - Les données de phytoplancton issues des autres campagnes et bancarisées au SISMER seront potentiellement séparées de celles des campagnes DCF bancarisées sur Quadrigé² |
| <i>Limiter le nombre d'interlocuteurs</i> | <p>PLUS</p> <ul style="list-style-type: none"> - La cellule Quadrigé² est l'interlocuteur unique pour les données biologiques (phytoplancton et zooplancton) <p>MOINS</p> <ul style="list-style-type: none"> - L'ensemble des données hydrologiques des campagnes DCF sont bancarisées auprès de deux équipes différentes de l'Ifremer (SISMER et VIGIES) |
| <i>Qualification des données</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Le SISMER et Quadrigé² intègrent tous deux des processus de qualification éprouvés et adaptables |
| <i>Coût de mise en place</i> | <p>PLUS</p> <ul style="list-style-type: none"> -Le SISMER bancarise déjà les données physico-chimiques issues de la CTD et des bouteilles - Quadrigé² est déjà bien adapté à la bancarisation des données taxonomiques, et la gestion des référentiels est opérationnelle - Pas de temps-agent supplémentaire à allouer à la création du programme sous Quadrigé² et des PSFM correspondants. Cette charge revient à la cellule Quadrigé² déjà en place <p>MOINS</p> <ul style="list-style-type: none"> - Temps-agent pour le développement des programmes d'intégration des données (si appliqué) |
| <i>Coût de</i> | MOINS |

| | |
|--------------------|--|
| <i>maintenance</i> | - Temps-agent à allouer pour la saisie manuelle et/ou le maintien des programmes d'intégration |
|--------------------|--|

L'utilisation de Quadrigé² pour les données biologiques paraît la solution de bancarisation la plus pertinente et **la plus optimale** du point de vue des possibilités et de l'expertise offertes par le SI et sa cellule pour ce type de données taxonomiques. La création de nouveaux programmes pour accueillir ces données des campagnes sera facilitée par le fait qu'il existe déjà des données similaires pour la surveillance du littoral (REPHY, IGA). Les **délais de développement** seraient donc éventuellement réduits par rapport à une bancarisation sur une nouvelle banque du SISMER, l'outil Quadrigé² étant aujourd'hui le plus adapté à ce type de données.

Concernant les coûts, la cellule Quadrigé² du service VIGIES a entre autre pour mission de créer les nouveaux programmes et PSFM nécessaires aux nouvelles bancarisations. Il n'y aura donc pas de temps-agent supplémentaire à allouer pour cette tâche concernant les données de phyto- et zooplancton. Cependant, il sera nécessaire de prévoir du **temps-agent** pour la création des **programmes d'intégration** et/ou la **saisie manuelle** des données dans le SI. De plus, bien que ce soit le cas pour n'importe quelle alternative, un **responsable de programme** devra être désigné afin de permettre le lien entre les équipes scientifiques de campagnes et la cellule Q2, notamment pour la mise à jour des référentiels et pour s'assurer de la bonne bancarisation des données sur le long terme.

En revanche, **d'autres contraintes**, propres à l'alternative Quadrigé², devraient en partie atténuer les avantages de cette solution en termes de délais et de coût de développement :

La solution Quadrigé² ne pourra être adoptée que si plusieurs aspects techniques sont résolus. En effet, le SISMER étant en charge d'archiver l'ensemble des données de campagnes, cette bancarisation au sein de Quadrigé² pose la question de la **mise à disposition des données sur le catalogue national des campagnes océanographiques**¹ géré par le SISMER. Ainsi, il faudra discuter des aspects techniques permettant aux données de phyto- et zooplancton d'être directement accessible à partir de ce catalogue en ligne, ce qui pourrait a priori être le cas grâce au développement des services web.

De plus, se pose également la question des données phytoplanctoniques issues des autres campagnes océanographiques. En effet, le SISMER aura toujours pour tâche de bancariser ces données qui seront alors séparées de celles des campagnes DCF collectées au titre de la DCSMM. La possibilité de **généraliser à l'ensemble des campagnes océanographiques** la bancarisation des données de phytoplancton sur Quadrigé² devra donc être discutée par la suite.

Bien que la présente étude privilégie l'alternative Quadrigé² comme étant la plus pertinente sous plusieurs aspects énoncés plus haut, ces deux principales réserves ne permettent pas à l'heure actuelle d'être catégorique sur le caractère « coût-efficacité » de cette alternative.

¹ Catalogue des campagnes océanographiques : <http://campagnes.flotteoceanographique.fr/>

4.5 Bancarisation des données biologiques : Autres alternatives

Bien que la solution de bancarisation des données biologiques vers le SI Quadrige² soit privilégiée par le groupe de travail, plusieurs points restent encore à résoudre et l'alternative de bancarisation au sein du SISMER n'est donc pas à exclure définitivement.

4.5.1 Alternative 2

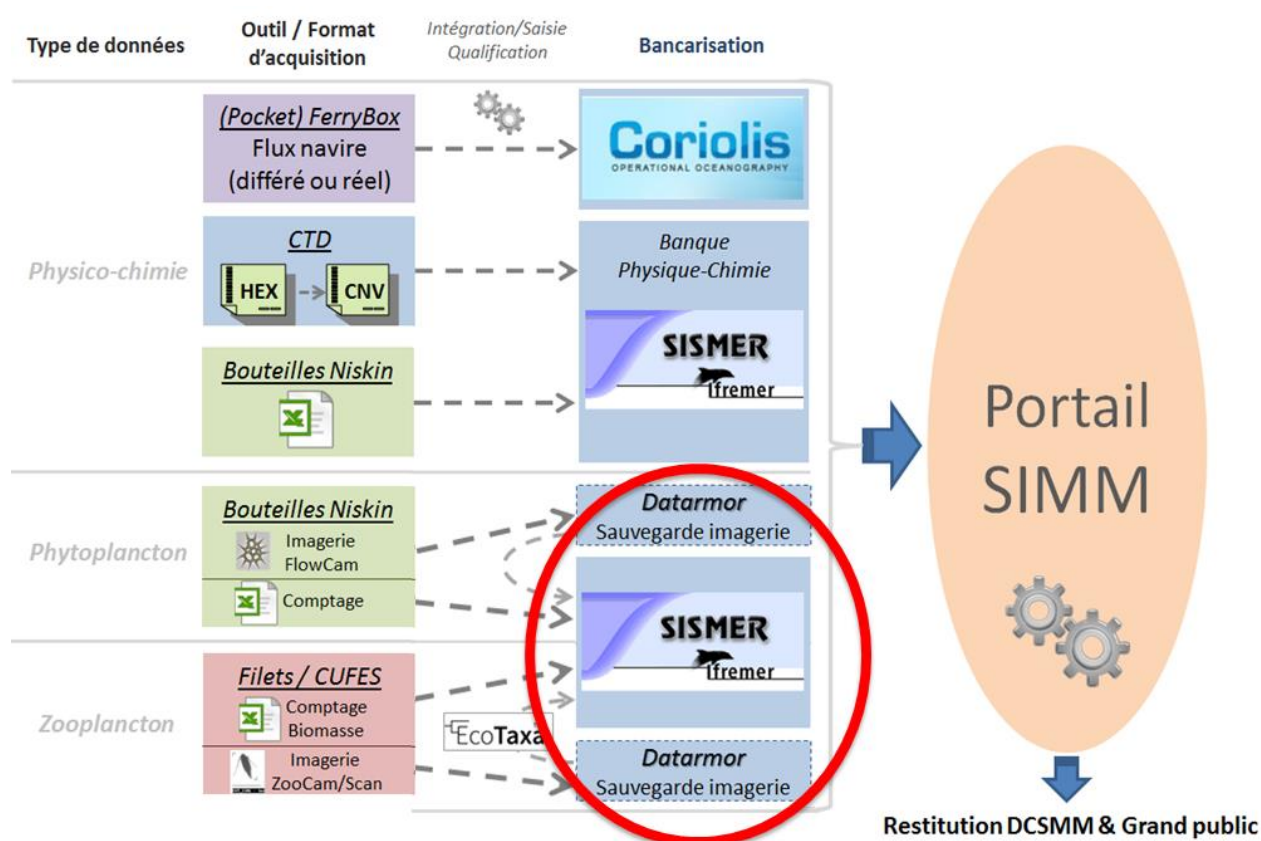


Figure 13 : Alternative 2 - processus de la collecte à la restitution par type de données

Cette alternative serait de centraliser la bancarisation de l'ensemble des données collectées à bord des campagnes DCF au sein d'une seule et même entité : le SISMER. A l'heure actuelle, le SISMER reçoit et bancarise une grande partie des données hydrologiques issues des sondes CTD et des bouteilles Niskin, et va être également en charge des données haute fréquence des FerryBox. Bien que la bancarisation des données de phytoplancton soit encore marginale, la gestion de ce type de données taxonomique peut être développée au sein du SISMER afin notamment d'accueillir les données d'autres campagnes océanographiques. En outre, **les aspects techniques et les contraintes présentées pour l'alternative Quadrige² (section 4.4) seront en tous points les mêmes si l'alternative du SISMER était retenue** (métadonnées nécessaires, intégration, gestion de l'imagerie, etc.).

Le principal avantage de cette alternative serait d'avoir l'ensemble des données hydrologiques des campagnes DCF rassemblées et gérées au niveau d'une seule entité de l'Ifremer, permettant potentiellement une réduction du nombre d'interlocuteurs. De plus, les réserves énoncées plus haut au sujet de la mise à disposition des données sur le catalogue des campagnes (section 4.4.3) seraient automatiquement levées.

Tableau 18 : Alternative 2 - Liste des avantages et des inconvénients

| Critères de choix | Alternative 2 |
|---|---|
| <i>Facilité d'accès aux données</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Diffusion publique des données physico-chimiques des campagnes d'ores et déjà bancarisées au SISMER via les portails Nautilus, SeaDataNet et Coriolis - Le portail SIMM prévoit de faire le lien pour aller chercher des données au SISMER dans le cadre d'autres programmes de surveillance |
| <i>Uniformité des données</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Le SISMER bancarise l'intégralité des données hydrologiques des campagnes halieutiques <p>MOINS</p> <ul style="list-style-type: none"> - Les données de phyto- et zooplancton de la surveillance au large sont séparées des données de la surveillance du littoral (sur Q2) |
| <i>Limiter le nombre d'interlocuteurs</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Le SISMER est l'interlocuteur unique pour toutes les données hydrologiques des campagnes halieutiques |
| <i>Qualification des données</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Le SISMER intègre des processus de qualification éprouvés et adaptables <p>MOINS</p> <ul style="list-style-type: none"> - Le SISMER ne bancarise pas autant de métadonnées que dans Quadrigé² pour le phyto et zooplancton (ex : les protocoles de mesure sont archivés en tant que pièces jointes mais ne sont pas bancarisés) |
| <i>Coût de mise en place</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Le SISMER bancarise déjà des données physico-chimiques issues de la CTD et des bouteilles <p>MOINS</p> <ul style="list-style-type: none"> - Les banques pour accueillir des données taxonomiques pour le phyto- et le zooplancton au SISMER, et la gestion de référentiels en interne sont encore à étendre - Temps-agent pour le développement des programmes d'intégration pour le phyto- et le zooplancton au SISMER |
| <i>Coût de maintenance</i> | <p>MOINS</p> <ul style="list-style-type: none"> - Temps-agent à allouer pour le maintien des programmes d'intégration |

Cette centralisation de toutes les données hydrologiques des campagnes DCF au niveau du SISMER impliquera un **travail plus important en amont** pour permettre à la base d'accueillir davantage de données biologiques de phyto- et zooplancton. Se pose également la question du niveau de métadonnées nécessaire à une valorisation de ces données (rapportage DCSMM). Actuellement le SISMER ne bancarise pas les mêmes métadonnées que dans Quadrigé, et ceci pourrait rendre difficile et non automatisée l'interprétation de certaines "ruptures" dans les données. En comparaison avec l'alternative privilégiant Quadrigé², cela nécessiterait des **délais et des coûts de développement supplémentaires** avant que la bancarisation soit effective, bien que l'alternative Quadrigé² ait également des contraintes de développement qui lui sont propre (section 4.4.3).

De plus, du point de vue optimisation, le développement de la bancarisation d'abondance planctonique au sein du SISMER, similaire à ce qui est déjà existant à VIGIES, ne paraît pas entièrement pertinent, même si cela permettrait de centraliser toutes les données au sein d'une même entité de l'Ifremer.

4.5.2 Alternative 3

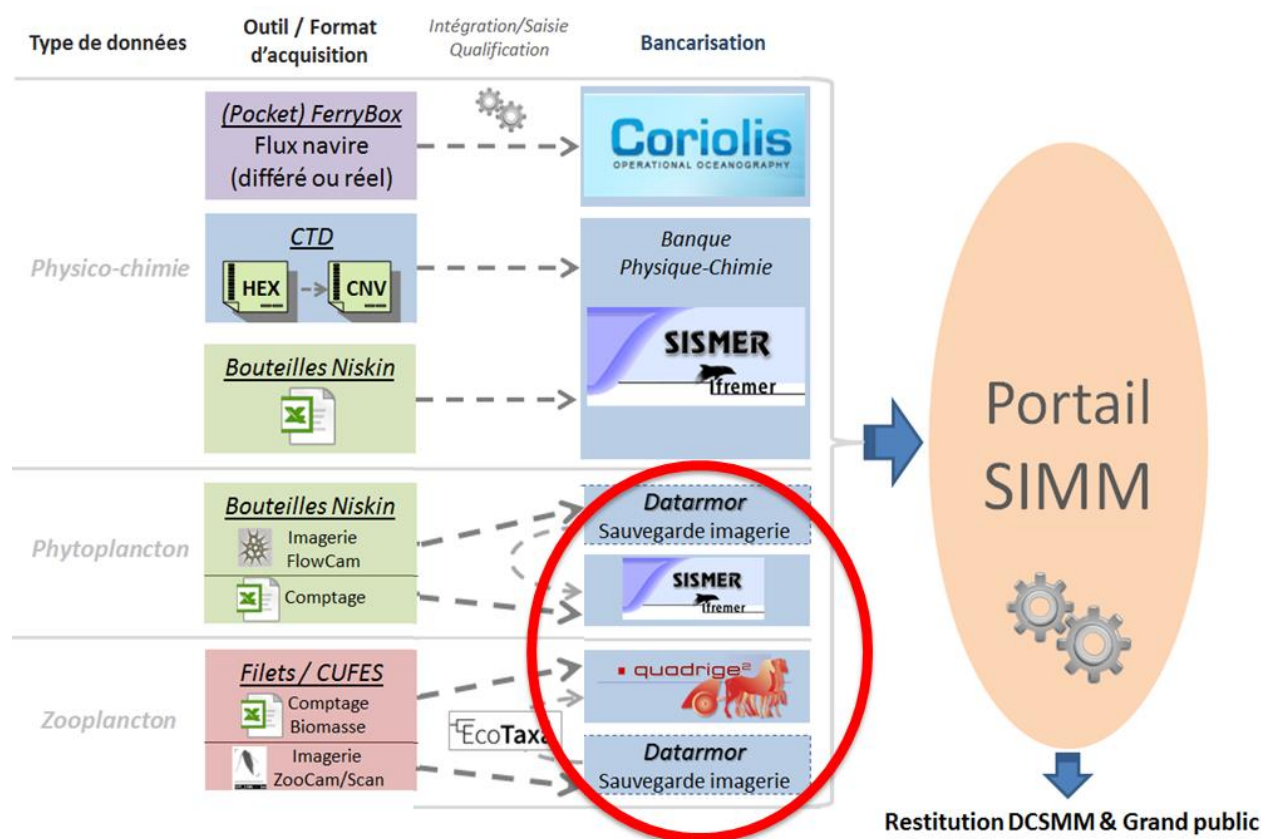


Figure 14 : Alternative 3 - processus de la collecte à la restitution par type de données

Une troisième alternative consisterait à partager les données biologiques au sein des deux entités de l’Ifremer présentées plus haut. La partie phytoplancton des bouteilles Niskin rejoindrait les données physico-chimiques collectées par la sonde CTD au SISMER, comme décrit en section 4.5.1 ; tandis que le zooplancton serait lui bancarisé sur la base Quadrige², comme décrit en section 4.4.2.

Le principal avantage de cette alternative serait de ne pas séparer le jeu de données bouteilles NISKIN entre deux SI. Cette solution serait aussi la plus coût-efficace à la vue de l’expérience de chacun des systèmes d’information suivant les types de données, et de la diffusion la plus optimale. En revanche, cela isolerait la bancarisation des données de zooplancton du reste des données hydrologiques collectées à bord des campagnes DCF, en dispersant par la même occasion les données biologiques sur deux SI.

Tableau 19 : Alternative 3 - Liste des avantages et des inconvénients

| Critères de choix | Alternative 3 |
|------------------------------|--|
| Facilité d'accès aux données | <p>PLUS</p> <ul style="list-style-type: none"> - Diffusion publique des données physico-chimiques des campagnes d’ores et déjà bancarisées au SISMER via les portails Nautilus, SeaDataNet et Coriolis - Forte visibilité de Quadrige² pour les données biologiques marines - Diffusion des données via l’application Q2 et le futur portail Surval - Le portail SIMM prévoit de faire le lien pour aller chercher des données sur Q2 et au SISMER dans le cadre d’autres programmes de surveillance |
| Uniformité des données | <p>PLUS</p> <ul style="list-style-type: none"> - Quadrige² bancarise également les données de zooplancton issues de la |

| | |
|--|--|
| | <p>surveillance du littoral</p> <ul style="list-style-type: none"> - Le SISMER bancarise l'intégralité des données des bouteilles Niskin <p>MOINS</p> <ul style="list-style-type: none"> - Les données biologiques de phyto- et de zooplancton sont dispersées sur deux SI différents - Les données de phytoplancton de la surveillance au large sont séparées des données de la surveillance du littoral - Nécessité de développer un outil permettant de rendre disponibles les données biologiques issues de Quadrigé² sur le catalogue des campagnes géré par le SISMER |
| <i>Limitier le nombre d'interlocuteurs</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Le SISMER est l'interlocuteur unique pour les données de la CTD et des bouteilles <p>MOINS</p> <ul style="list-style-type: none"> - Les données biologiques sont bancarisées auprès de deux entités de l'Ifremer (SISMER et VIGIES) |
| <i>Qualification des données</i> | <p>PLUS</p> <ul style="list-style-type: none"> - Le SISMER et Quadrigé² intègrent tous deux des processus de qualification éprouvés et adaptables <p>MOINS</p> <ul style="list-style-type: none"> - le SISMER ne bancarise pas autant de métadonnées que dans Quadrigé pour le phytoplancton (ex : les protocoles de mesure sont archivés en tant que pièces jointes mais ne sont pas bancarisés) |
| <i>Coût de mise en place</i> | <p>PLUS</p> <ul style="list-style-type: none"> -Le SISMER bancarise déjà des données physico-chimiques issues de la CTD et des bouteilles - Pour les données de zooplancton, Quadrigé² est déjà bien adapté à la bancarisation des données taxonomiques, et la gestion des référentiels est déjà opérationnelle - Pour le zooplancton, pas de temps-agent supplémentaire à allouer à la création du programme sous Quadrigé² et des PSFM correspondants. Cette charge revient à la cellule Quadrigé² déjà en place <p>MOINS</p> <ul style="list-style-type: none"> - La banque pour accueillir des données taxonomiques au SISMER, et la gestion de référentiels en interne est encore à étendre - Temps-agent pour le développement des programmes d'intégration pour le phytoplancton au SISMER et le zooplancton sur Quadrigé² (si appliqué) |
| <i>Coût de maintenance</i> | <p>MOINS</p> <ul style="list-style-type: none"> - Temps-agent à allouer pour la saisie manuelle et/ou le maintien des programmes d'intégration |

En définitive, si un outil devant accueillir des données d'abondance taxonomique devait être créé au SISMER, la bancarisation de toutes les données biologiques des campagnes DCF devrait alors être envisagée, afin d'éviter une dispersion supplémentaire des données de zooplancton sur Quadrigé² d'un côté, et du phytoplancton au SISMER de l'autre (= Alternative 3). Les problématiques liées à la gestion des référentiels en particulier seront les mêmes pour ces deux types de données.

5 Conclusion

La présente étude préconise une **bancarisation multiplateforme** des données hydrologiques collectées à bord des campagnes DCF. Cette bancarisation en amont étant suivie d'une **restitution unique** par le futur **portail SIMM** développé dans le cadre de la DCSMM, qui centralisera les requêtes et la fusion des données issues des différentes bases.

Suite à l'examen des différents scénarios, les plateformes de bancarisation se partageraient ainsi les données physico-chimiques d'un côté, et les données biologiques de l'autre, selon les spécialités et les contraintes de chacun des systèmes d'information :

- Les **données physico-chimiques** de la sonde CTD, des bouteilles Niskin et de la (Pocket) FerryBox seraient centralisées au niveau du **SISMER** qui réceptionnerait les fichiers dans des **formats uniformes et complets**. Des recommandations ont été émises pour harmoniser l'effort dans ce sens au cours de la présente étude. Les données de profils rejoindraient la banque physique-chimie, tandis que les données haute fréquence seraient bancarisées sur le portail Coriolis.
- Les **données biologiques** de phytoplancton et de zooplancton rejoindraient la base **Quadrige²**, qui accueille et gère déjà des données taxonomiques de ce type. Les fichiers issus de l'imagerie continueraient d'être stockés au SISMER sur un disque dédié aux campagnes DCF.

Cette solution privilégiant l'**alternative Quadrige²** pour les données biologiques semble la plus **pertinente** à la vue de ce que propose déjà le système d'information et de l'expérience de sa cellule pour ce type de données. C'est par ailleurs la solution qui a été privilégiée par les différents acteurs utilisant ces données (des responsables de données hydrologiques des campagnes aux pilotes scientifiques de la DCSMM). Cependant, cette solution de bancarisation est sous réserve de garantir un accès direct aux données, à partir du catalogue national des campagnes dont le SISMER a la charge (<http://campagnes.flotteoceanographique.fr>). De plus, il sera nécessaire d'étudier la possibilité d'étendre cette bancarisation à l'ensemble des campagnes océanographiques. Cette faisabilité devra être déterminée quelle que soit la solution de bancarisation retenue.

Enfin, quelles que soient les solutions de bancarisation retenues, l'étude a montré l'intérêt d'adopter des **formats de données uniformisés et réguliers d'une campagne et d'une station à l'autre**, car cela représente un élément primordial dans l'harmonisation de l'ensemble des données collectées à bord des six campagnes halieutiques. De plus, systématiquement, un(des) responsable(s) de programme devra(ont) être désigné(s) afin de permettre un lien pérenne entre les équipes scientifiques de campagnes et les équipes responsables des systèmes d'information, notamment pour la mise à jour des référentiels et pour s'assurer de la bonne bancarisation des données sur le long terme.

Les perspectives à l'issue de cette première étude et du choix des outils de bancarisation sont de définir précisément les aspects techniques entourant l'harmonisation, ainsi que les coûts et les délais de réalisation, afin d'établir les ressources nécessaires pour une mise en chantier à partir de 2019.

6 Références

1. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive) (Text with EEA relevance). 164, 32008L0056 juin 25, 2008.
2. Baudrier J., Wessel N., Lecomte J.-P., Bertrand J., 2015. Optimisation des campagnes halieutiques pour la construction d'indicateurs DCSMM. Colloque de l'Association Française d'Halieutique, 1-3 juillet 2015, Aquarium Mare Nostrum Montpellier. Poster.
3. Baudrier J., 2015. Mise en œuvre du programme de surveillance de la DCSMM : synthèse des essais réalisés à bord des campagnes halieutiques. Rapport Ifremer, Dyneco-Vigies, 22 p.
4. MEDDE (Ministère de l'écologie, du développement durable et de l'énergie), 2015a. Plan d'action pour le milieu marin, Programme de surveillance, sous-région marine Manche - mer du Nord. Coordination technique: AAMP, Ifremer. 438 p.
5. MEDDE (Ministère de l'écologie, du développement durable et de l'énergie), 2015b. Plan d'action pour le milieu marin, Programme de surveillance, sous-région marine mers celtiques. Coordination technique: AAMP, Ifremer. 390 p.
6. MEDDE (Ministère de l'écologie, du développement durable et de l'énergie), 2015c. Plan d'action pour le milieu marin, Programme de surveillance, sous-région marine golfe de Gascogne. Coordination technique: AAMP, Ifremer. 394 p.
7. MEDDE (Ministère de l'écologie, du développement durable et de l'énergie), 2015d. Plan d'action pour le milieu marin, Programme de surveillance, sous-région marine Méditerranée occidentale. Coordination technique: AAMP, Ifremer. 384 p.
8. Baudrier J., 2017. Bilan de l'optimisation des campagnes halieutiques réalisée dans le cadre du programme de surveillance de la directive cadre stratégie pour le milieu marin - année 2016. Rapport scientifique Ifremer ODE/VIGIES/DCSMM/17-009, p. 26.
9. Commission Decision (EU) 2010/477 of 1 September 2010 on criteria and methodological standards on good environmental status of marine waters (notified under document C(2010) 5956) Text with EEA relevance. OJ L, 32010D0477(01) sept 2, 2010.
10. Commission Implementing Decision (EU) 2016/1251 of 12 July 2016 adopting a multiannual Union programme for the collection, management and use of data in the fisheries and aquaculture sectors for the period 2017-2019 (notified under document C(2016) 4329). OJ L, 32016D1251 août 1, 2016.
11. Verin Y., 1992. IBTS INTERNATIONAL BOTTOM TRAWL SURVEY (IBTS),, <http://dx.doi.org/10.18142/17>.
12. Coppin F., Travers-Trolet M., 1989. CGFS: CHANNEL GROUND FISH SURVEY, <http://dx.doi.org/10.18142/11>.
13. Doray M., Duhamel E., Huret M., Petitgas P., Masse J., 2012. PELGAS, <http://dx.doi.org/10.18142/18>.
14. Mahe J.-C., 1987. EVHOE EVALUATION HALIEUTIQUE DE L'OUEST DE L'EUROPE, <http://dx.doi.org/10.18142/8>.
15. Bourdeix J.-H., Saraux C., 1985. PELMED - PELAGIQUES MEDITERRANÉE, <http://dx.doi.org/10.18142/19>.
16. Jadaud A., Souplet A., Bertrand J., 1994. MEDITS, <http://dx.doi.org/10.18142/7>.
17. Schroeder F., Mizerkowski B., Petersen W., 2008. The pocketFerryBox - A new portable device for water quality monitoring in oceans and rivers. J Oper Oceanogr. 1 janv 2008;1(2):51-7.

18. Aiken C.M., Petersen W., Schroeder F., Gehrung M., Ramírez von Holle P.A., 2011. Ship-of-Opportunity Monitoring of the Chilean Fjords Using the Pocket FerryBox. *J Atmospheric Ocean Technol.* oct 2011;28(10):1338-50.
19. Beutler M., Wiltshire K.H., Meyer B., Moldaenke C., Lüring C., Meyerhöfer M., et al., 2012. A fluorometric method for the differentiation of algal populations in vivo and in situ. *Photosynth Res.* 72(1):39-53.
20. Ruser A, Popp P, Kolbowski J, Reckermann M, Feuerpfeil P, Egge B, et al., 1998. Comparison of chlorophyll-fluorescence-based measuring systems for the detection of algal groups and the determination of chlorophyll-a concentrations. In: Workshop at the FTZ Büsum from October 1998. In.
21. Senn C., Romagnan J.-B., 2017. Cartographie à haute résolution du zooplancton dans le golfe de Gascogne, par imagerie. IFREMER. Poster.
22. Colas F., Tardivel M., Perchoc J., Lunven M., Forest B., Guyader G., 2017. The ZooCAM, a new in-flow imaging system for fast onboard counting, sizing and classification of fish eggs and metazooplankton. *Prog Oceanogr.* 24 oct 2017.
23. Cellule d'Administration Q², 2014. Guide de l'utilisateur Quadrige²: Configuration de l'application et saisie des données. Quadrige² - Référentiel National de gestion des données de la surveillance littorale. Déc 2014.
24. Lamoureux A., 2018. Intégration des données Quadrilabo à Quadrige - Version 1.7. Quadrige² - Référentiel National de gestion des données de la surveillance littorale, ODE/VIGIES.18-01/AL, Fév 2018.
25. Picheral M., Colin S., Irisson J.O., Elineau A., de Vargas C., Stemmann L., 2017. ECOTAXA, a web based application for collaborating on large plankton image datasets. *Marine Imaging Workshop*, 20-24 February 2017, GEOMAR Helmholtz Centre for Ocean Research Kiel. Poster.