



DATA
TERRA



ODATIS

Data Terra et le pôle ODATIS Une Infrastructure de Recherche pour les données marines

Sabine Schmidt¹, Gilbert Maudire², Cécile Nys², Joël Sudre³, Valérie Harscoat²,
Caroline Mercier⁴, Gérald Dibarboure⁵, Frédéric Hyunh³

¹ CNRS, OASU, UMR5805 EPOC, Université de Bordeaux, Pessac, France

² Ifremer, Centre de Bretagne, Plouzané, France

³ UMS CPST, Coordination Pôles de données et de services pour le Système Terre, Montpellier, France

⁴ Akka, Toulouse, France.

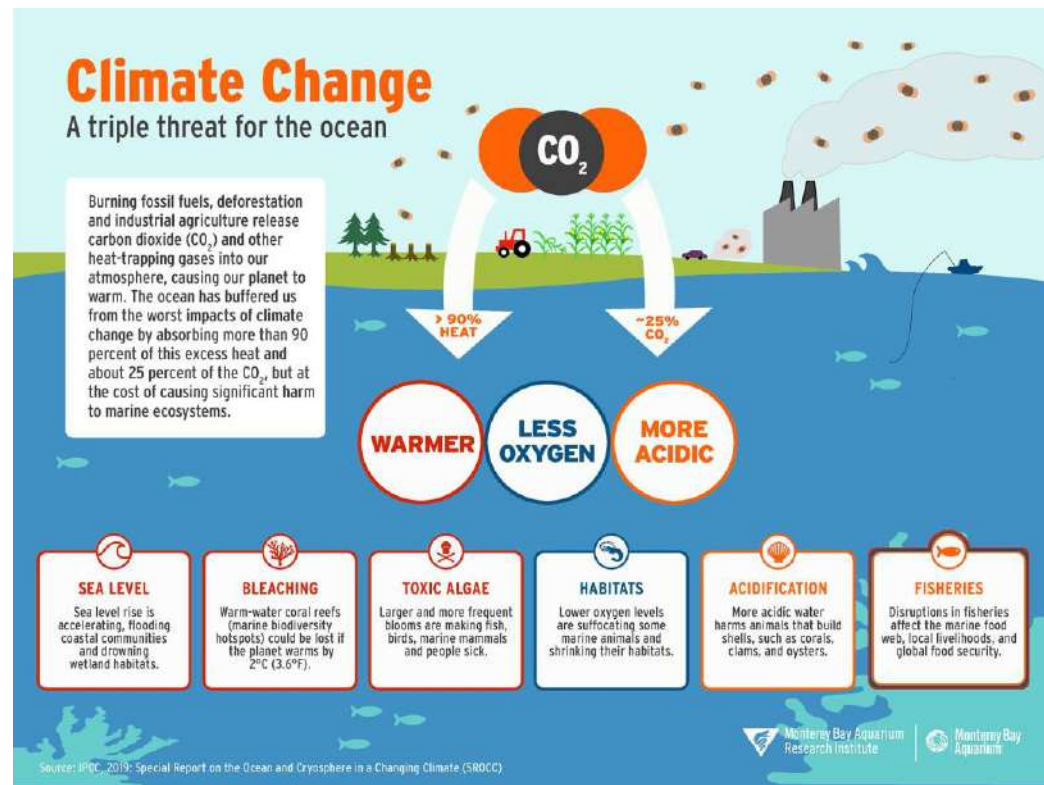
⁵ CNES, Toulouse, France.



L'anthropocène

Depuis la révolution industrielle, l'empreinte des activités humaines sur l'environnement mondial s'est accrue.

Les conséquences actuelles et attendues du changement global sur l'océan sont multiples.



Un besoin crucial de mieux comprendre pour prévoir les impacts du changement global

Des observations sont nécessaires à tous les stades du processus scientifique :
description, compréhension, modélisation et prévision

Considérant que:

- l'acquisition des données marines est difficile et onéreuse :

Nécessité d'accéder à des sites distants et d'utiliser de nombreux moyens techniques (flotte océanographique, sites instrumentés, gliders, bouées Argo, ...)

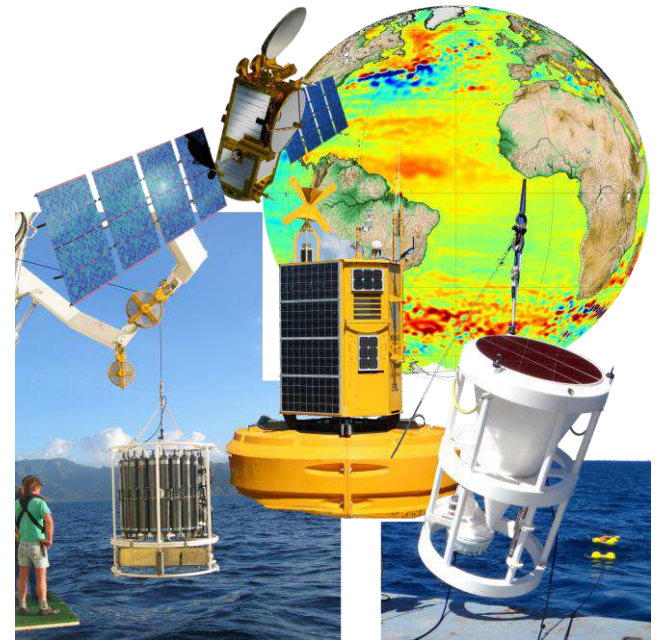
- sans archivage, > 30% des données sont perdues ou inutilisables 10 ans après leur acquisition (source: Ifremer).

➡ La préservation des données marines est un enjeu majeur,

Les défis de l'accès aux données marines

Augmentation du nombre des observations marines (*in situ* et par télédétection) au cours des dernières décennies.

Pour tirer le meilleur parti de ce flux de données au profit de la connaissance et de la société, les centres de données doivent respecter des principes communs.

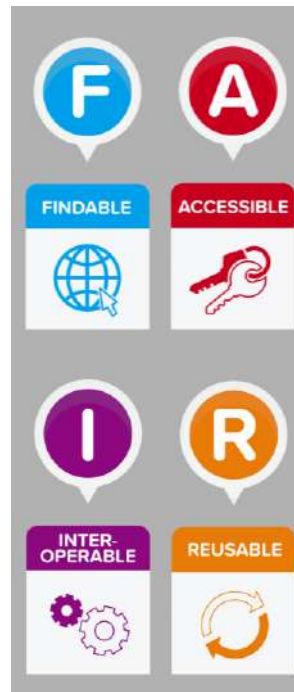


Les défis de l'accès aux données marines

Défi 1 : la qualité des données (principes FAIR)

Les métadonnées et les données doivent être faciles à trouver et à (ré)utiliser (décrire vos données, appliquer des identificateurs persistants)

Les données doivent être intégrables à d'autres données (format ouvert, vocabulaire cohérent, normes de métadonnées)



Envisager ce qui sera partagé et comment on peut y accéder

Les données doivent être réutilisables, avec des métadonnées bien décrites et une licence appropriée

Les défis de l'accès aux données marines



Défi 2: certification des dépôts de données

La Research Data Alliance (RDA) fournit un cadre commun pour la mise en œuvre et la maintenance des dépôts numériques.

the CoreTrustSeal requirements

La certification est importante pour garantir: :

- ✓ la fiabilité et la durabilité des dépôt de données,
- ✓ l'archivage et le partage à long terme des données pour les utilisateurs et les financeurs.



www.coretrustseal.org

Les défis de l'accès aux données marines

Défi 3: Amener les producteurs à partager leur données

L'ICSU (International Council for Science) promouvait déjà en 2011

« un accès complet et ouvert aux données scientifiques, en particulier lorsque la recherche est financée par des fonds publics ».

Les chercheurs peuvent être réticents à partager leurs données **en raison des coûts réels et/ou perçus**

Sentiment de - perte de contrôle sur les données,
- contraintes sans bénéfice,

Manque de - ressources informatiques/humaines
- formation.

Les dépôts de données doivent rendre le partage de données

**plus faciles pour le producteur
pas seulement pour l'utilisateur.**



Le besoin d'infrastructures interoperables

Afin d'accélérer la collecte et l'utilisation des données, il est nécessaire de disposer d'infrastructures interoperables pour aider :

- les producteurs à archiver et à partager leurs données,
- les utilisateurs à obtenir un accès relativement facile aux données



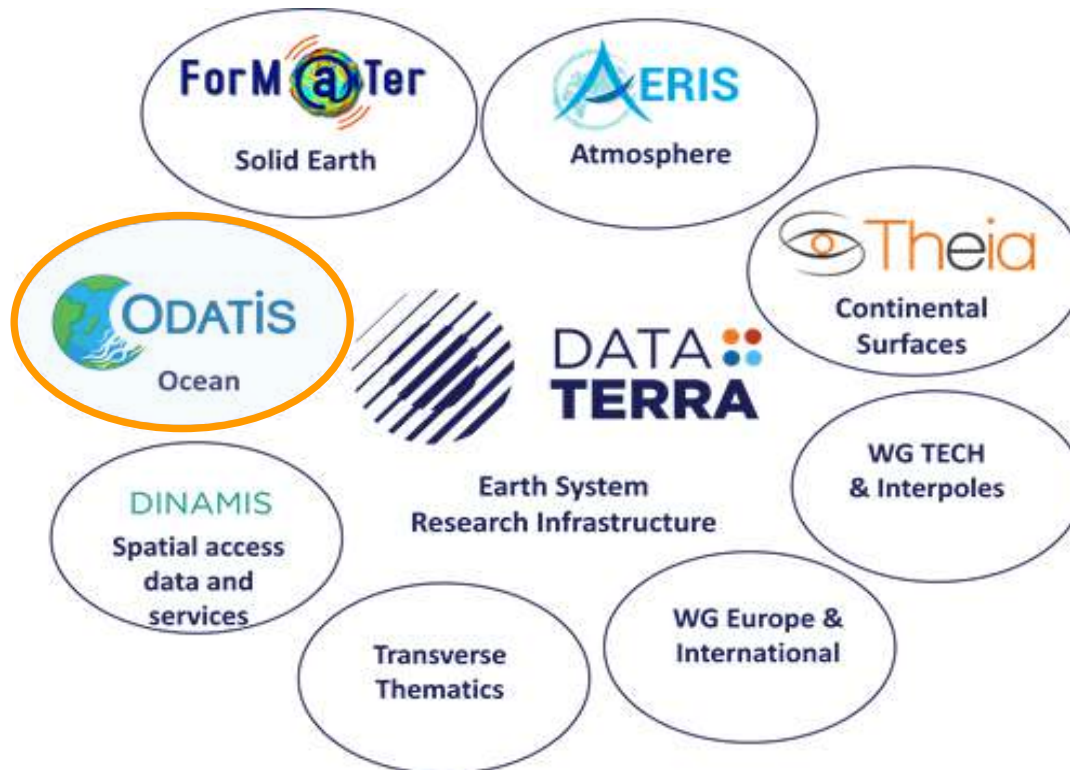
coordonnées au moins au niveau national, voire international.

Ici, nous illustrons une initiative nationale d'un portail dédié aux données marines françaises



L'IR Data Terra

Infrastructure de recherche « *pôle de données et services pour le Système Terre* » :
de la donnée à la connaissance du Système Terre



Composée de:

- 4 pôles thématiques
- groupes de travail
- services transverses

ODATIS: Ocean DATA Information and Services

www.odatis-ocean.fr

Mission Préserver, gérer et diffuser les données et produits d'observations des océans

Partenaires



Demande en cours
d'instruction



Source des Données

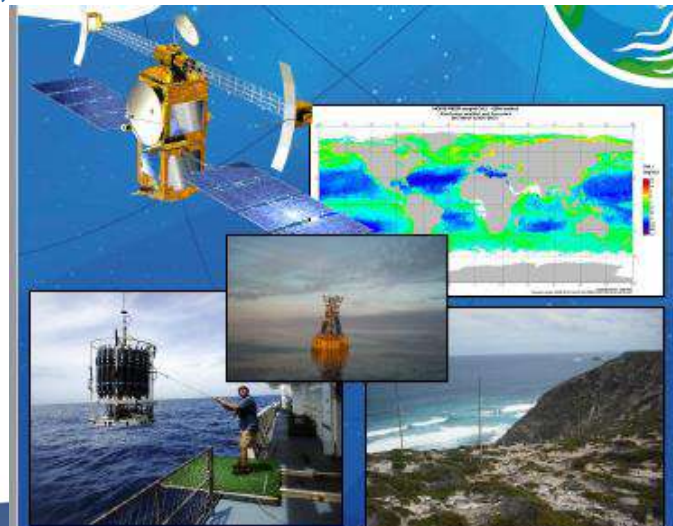
missions satellites «marines»

altimètre, radiomètre, diffusiomètre, optique, ...

observations *in situ*

SNO/SO côtiers (IR ILICO, régionaux) et hauturiers (Ohis, Argo, SSS ...)

TGIR Flotte Océanographique, LEFE-CYBER, ...



Merigéo – webconférence – Novembre 2020

Feuille de route d'ODATIS

- Offrir une vision globale des observations (*in situ*; satellite) et de leurs produits;
- Faciliter l'accès, via un portail unique, aux données, entièrement décrites et qualifiées en accord avec les normes en vigueur (FAIR, TRUST) ;
 - ce qui implique d'assurer
 - la conservation à long terme des jeux de données;
 - l'interopérabilité des jeux de données (espace; temps; disciplines);
- Promouvoir des utilisations combinées de données de nature (*in situ*; satellite) ou d'origine différentes (réseaux opérationnels/expériences scientifiques);
- Offrir la possibilité d'explorer, d'extraire et d'analyser des données en proposant des outils et des ressources informatiques.

Organisation d'ODATIS

STRATEGIC LEVEL

Inter-institution Steering Committee



Data Terra Executive Board

Defines a common strategy for all the thematic clusters to provide an unified access to data (interoperability, vocabulary), products, softwares, tools and services.

Scientific Council

Experts on marine and coastal sciences.
Represent the user scientific community, express the needs and recommend scientific consortia.

EXECUTIVE BOARD

Management team



Representative of each DSC

Director
Technical director
Scientific director
Assistants

Data and Service Centres (DSCs)

Catalog: web & client: CNES
server & catalog: Ifremer

Data:

Satellite (CDS-SAT)
CERSAT, AVISO

In situ (CDS-in situ)
Coriolis, SHOM, SISMER
OMP, OASU, IMEV. IMEV, SBR

Technical Workshop

Proposes technical guidelines (interoperability, services, ...), Organizes practical sessions to test technics and exchange experiences among CDSs.

Scientific Consortia

Propose and instruct new thematic data compilation (ex: EOVI) or new product on the request of the scientific community.

OPERATIONNAL LEVEL



Les Centres de Données et Services

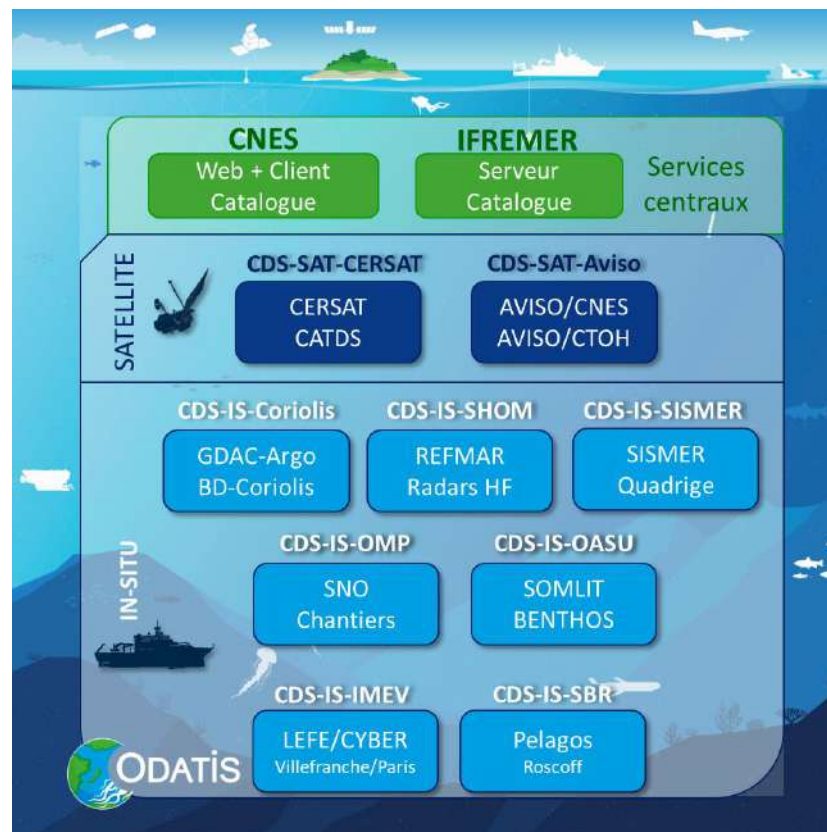
En charge de:

- gestion des données:
bancairisation, pérennisation, diffusion, ...
- élaboration de produits au quotidien au plus proche des utilisateurs;

Soutenus par les tutelles du pôle

Dans le cadre d'un mandat (type de données)
et d'un cahier des charges

Mis en relation par des modules communs :
Portail web et Catalogue des données



Le portail ODATIS

www.odatis-ocean.fr



LANGUAGES | CONTACT |

QUI SOMMES-NOUS? ▾ DONNÉES ET SERVICES ▾ ACTIVITÉS ▾ ACTUALITÉS ▾ RESSOURCES ▾

↓ Accès au catalogue

The screenshot displays the ODATIS web application interface, which is divided into several functional areas:

- Rechercher (Search):** Located at the top left, it features a search bar, a map of the world with a highlighted region, and filters for variables, type of data, and data source.
- Décrire (Describe):** A detailed view of a dataset, including a title, creation date (2010-01-01T00:00:00), source (VORTEX/Galvaneur / SHOM), and a description of the system (Tide Database (TDB)). It also includes a metadata table and a 'Télécharger' (Download) form with fields for email, phone, name, organization, and usage.
- Visualiser (Visualize):** A central area showing a map with a data overlay, a data table, and a 'Fond de carte' (Map background) selection tool.
- Accès aux données (Data Access):** A section on the right side of the 'Décrire' page, providing a direct link to the data.

The interface is designed with a blue and white color scheme, featuring satellite and ocean-themed illustrations. The bottom of the page includes a decorative border with coral and fish, and the ODATIS logo is repeated in the bottom right corner.

DATA
TERRA

ODATIS

Le portail ODATIS

www.odatis-ocean.fr



LANGUAGES | CONTACT |

QUI SOMMES-NOUS? ▾ DONNÉES ET SERVICES ▾ ACTIVITÉS ▾ ACTUALITÉS ▾ RESSOURCES ▾

Accueil



Accès aux données >

Services aux producteurs de données

Services aux producteurs



Le portail ODATIS

www.odatis-ocean.fr



LANGUAGES | CONTACT |

QUI SOMMES-NOUS? ▾ DONNÉES ET SERVICES ▾ ACTIVITÉS ▾ ACTUALITÉS ▾ RESSOURCES ▾

Accueil



Accès aux données >
Services aux producteurs de données

Services aux producteurs

spécificité ODATIS → SEANOE

<https://www.seanoe.org/>

dépôt de données sous licences Creative Commons

Merigéo – webconférence – Novembre 2020



Publish your marine data

SEANOE offers you the possibility of publishing free of charge your scientific data in the field of the marine sciences.

Each data set published by SEANOE has a DOI. Your data may well be cited in scientific articles in a reliable and sustainable way.

SEANOE offers a fast responding service : if your dataset is well described¹, **you should get a DOI within 24 hours.**

Possibilité d'embargo (2 ans max.) pour restreindre l'accès aux données, par exemple d'une publication en cours d'évaluation, par exemple.

Les ateliers techniques

réunissent les partenaires autour de présentations et tables rondes pour:

- définir les orientations techniques du pôle;
- réaliser des mises en pratique sur la base d'exercices d'application de logiciels de visualisation de la donnée.

Tous les ateliers sont mis en ligne

Merigéo – webconf



Atelier Technique #7 : Gestion de données biologiques marines

ORGANISATION : Joël SUDRE (LEGOS/UMR5566), Cécile NYS (IFREMER IRSI/SISMER).

DATE : mardi 22 septembre (9h-18h) et mercredi 23 septembre (9h-16h30) 2020.

LIEU : Visioconférence GoToMeeting / Salle de visioconférence du bâtiment Jules Barrois à l'IMEV (ex-OOV). Suite à votre inscription, les informations de connexion ont été transmises par mail.



Agenda

Compte-rendu

Liste des participants

- Présentation du **Pôle de données Océan** – ODATIS – *Gilbert MAUDIRE*
- Présentations des **initiatives dans la gestion des données biologiques marines** (20min/présentation + 10min de questions) :
 - **PNDB, de la gestion de données à son analyse, présentation des outils et services**– *Yvan LE BRAS*
 - **EMBRC** (France) Données biologiques marines dans le Centre National de Ressources Biologiques Marines - *Gemma GIMENEZ-PAPIOL* - *présentation accessible exclusivement à destination des participants pendant l'atelier.*
 - **SAR/SIMM** : **référentiels utilisés dans le cadre du Système d'Information sur le Milieu Marin** – *Clémence RABEVOLO*
 - **ECOTAXA & Zooplancton** - *Jean-Olivier IRISSON* - *présentation accessible sur authentification; veuillez renseigner vos identifiants dans la marge en haut à droite de cette page pour voir apparaître le titre et le lien de cette présentation à la suite de cette liste à puces.*
 - **SeaDataCloud/EMODNet Bio/PhytOBS** : **Mise en place des chaînes de traitement à Roscoff pour la publication des données des suivis à long terme**– *Mark HOEBEKE*
 - **SNO BENTHOBS** - *Vincent BOUCHET / Nicolas DESROY*
 - **Quadrige** - *Arnaud ROUILLY*
 - **APA & SI MORSE** - **Le projet MORSE, vers un nouveau portail de suivi des échantillons biologiques** - *Sylvie VAN ISEGHEM*
 - **Imagerie** - *Catherine BORREMANS*
 - **Cytométrie** - *Gérald GREGORI, Felipe ARTIGAS, Maurice LIBES, Marc Sourisseau, Melilotus Thyssen*
 - Qu'est-ce que la cytométrie en flux (diapos 1-10).
 - Cytométrie en flux automatisée et accessibilité des données pico-nano-microplancton (diapos 11-30).
 - Avancement chaîne de traitement de cytométrie en flux (diapos 31-45).
- Ateliers de discussions :
 - Atelier de discussion « **Référentiels** »
 - Atelier de discussion « **Format de données** »
 - Atelier de discussion « **Outils** »
 - Atelier de discussion « **Imagerie** »
- Autres points & Préparation du prochain Atelier Technique

Les Consortiums d'Expertises Scientifiques (CES)

Objectif: promouvoir ou/et valoriser des méthodes de traitement et des produits innovants d'observation spatiale, aéroportée ou in-situ de l'océan et de ses interfaces (atmosphère, littoral et sous-sol sous-marin) avec les autres pôles de données,

Regroupent les experts nationaux du domaine (inter-organismes, inter-laboratoires)

Dans le cadre d'une démarche projet : calendrier, mode de financement, livrables (durée 2 ans)

Plusieurs catégories:

CES de prospection (financement des missions)

CES techniques:

- définition de produits, d'algorithmes
- élaboration de prototypes de traitement (financement de la réalisation du prototype)
- implémentation (ex d'un collègue qui aurait développé un algorithme)

CES Cytométrie en flux

CES Cytométrie en flux

Le Consortium d'Expertise Scientifique (**CES**) **Cytométrie en flux** a été mis en place début 2020 afin de fédérer les acteurs scientifiques au niveau national, voire international, autour de la cytométrie en flux (**CMF**), technologie très largement utilisée de nos jours pour caractériser la structure des assemblages de **micro-organismes planctoniques**, aussi bien en milieu marin qu'en eau douce. La **CMF** permet de **classer ces micro-organismes** en différents groupes fonctionnels, définis chacun par des cellules aux propriétés optiques similaires, et d'en **déterminer les abondances**. En règle générale, ces groupes cytométriques appartiennent au pico-, nano- ou microphytoplancton, ainsi qu'aux virus, procaryotes et nanoflagellés hétérotrophes. Certains instruments sont également capables d'associer une prise de photo à l'analyse individuelle des cellules, offrant la possibilité d'une identification taxonomique supplémentaire.

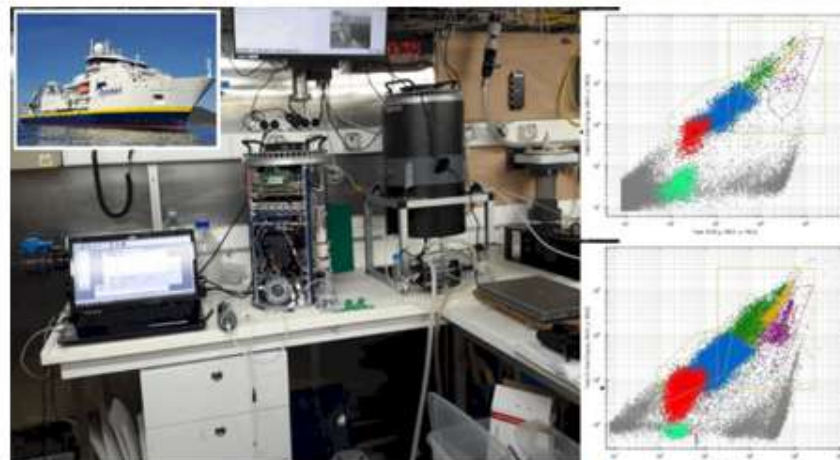
Contexte : Le **phytoplancton** inféodé à la couche euphotique de l'Océan est responsable par la photosynthèse de la moitié de la production primaire annuelle globale. Le **bactérioplancton hétérotrophe** présent dans toute la colonne d'eau et la partie superficielle des sédiments est le principal responsable de la minéralisation de la matière organique et de la régénération des sels nutritifs indispensables au développement du phytoplancton en surface. L'ensemble de ces micro-organismes affecte les cycles du carbone, de l'azote, du phosphore et d'autres éléments clés. Indispensables au maintien de l'écosystème, ils sont toutefois très sensibles au changement de leur environnement, aussi bien dans l'océan côtier qu'en océan ouvert. Leur étude est donc fondamentale dans le contexte du changement global et de l'impact des activités anthropiques.

Objectif : L'objectif de ce **CES**, piloté par Gérald Grégori (**MIO, CNRS**), est de générer une initiative communautaire pour bancariser, partager et diffuser à l'échelle nationale et internationale les données de cytométrie en flux environnementales. Afin que ces données jouent leur rôle de levier pour la science et l'innovation, il est indispensable qu'elles soient facilement accessibles et réutilisables selon les principes **FAIR**. Il est également indispensable que ces données soient qualifiées et pour cela répondent aux protocoles d'analyse standardisés reconnus par la communauté.

Ce **CES** comporte trois groupes de travail (WP) :

- **WP Vocabulaire** : animé par Melilotus Thyssen (**MIO, CNRS**) et David Pecqueur (**OOB, CNRS**),
- **WP Chaîne de traitement des données de cytométrie** : animé par Maurice Libes (O.S.U Pytheas, **CNRS**) et Marc Sourisseau (**IFREMER**),
- **WP Bonnes pratiques** : animé par Felipe Artigas (**LOG, ULCO**), Christophe Lambert (**LEMAR, CNRS**)









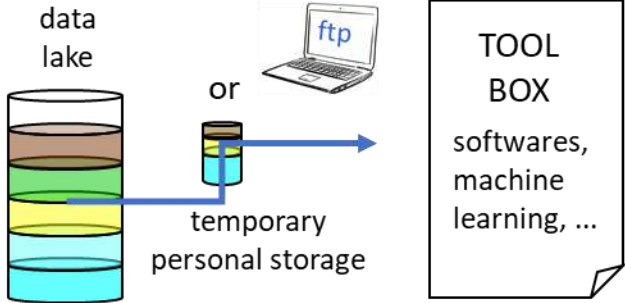
Le **CES** aura aussi en charge d'identifier et promouvoir ces protocoles en réalisant une veille technologique grâce à des échanges d'informations avec la communauté. Il pourra également identifier les verrous ainsi que les outils susceptibles d'améliorer la procédure de génération ou de visualisation de ces données, et leur archivage sur des bases de données interopérables.



Cytomètres en flux en cours d'installation à bord du R/V « Pourquoi Pas ? » lors de la campagne océanographique Peacetime, en 2017 en mer Méditerranée et cytogrammes (à droite). Crédits M. Thyssen (MIO).

Le besoin de développer une hiérarchisation des centres de données

La mise en œuvre d'une telle e-infrastructure nécessite d'optimiser les ressources humaines et informatiques, sur la base de trois types de centres de données:

SERVICE	data repository	production	on demand
Involved structures			 
USER	data repositories doi, licences reporting on data use	combination of different marine dataset (<i>in situ</i> / satellite) from the same thematic or area.	data analyses and interpretation cross analyses of different data from all Earth compartments
BACK OFFICE	<p>Data Assembling Centres:</p> <ul style="list-style-type: none">  Close to the producer  Common catalogue and vocabulary servers  Long-term archive 	<p>Data & Service Centres:</p> <ul style="list-style-type: none">  National data hub Aggregates large collections at the national minimum level 	<p>Virtual Research Environments (VRE)</p> 

Dépôt du projet Equipex PIA3 Gaia Data porté par l'IR Data Terra



DATA
TERRA



Merci pour votre attention

