

Atelier Technique#17

Données bioinformatiques de diversité

11-13 mars 2024

Station Biologique de Roscoff

Bonnes pratiques de gestion de données

Metadonnées, ressources terminologiques et vocabulaire contrôlé (Bioportal, ...), formats standards, définition claire des workflows, entrepôts de référence

Outils

Recommandations, exemples d'utilisations, gestion de méta-données, publications dans des entrepôts, visualisation, moteurs de workflows, génération de rapports d'analyse



DATA
TERRA



Pôle National
de Données de Biodiversité



Atelier#17 : Données bioinformatiques de diversité

Station biologique de Roscoff

11-12-13 mars 2024



DATATERRA



contact@data-terra.org

+33 (0)4 67 54 87 08

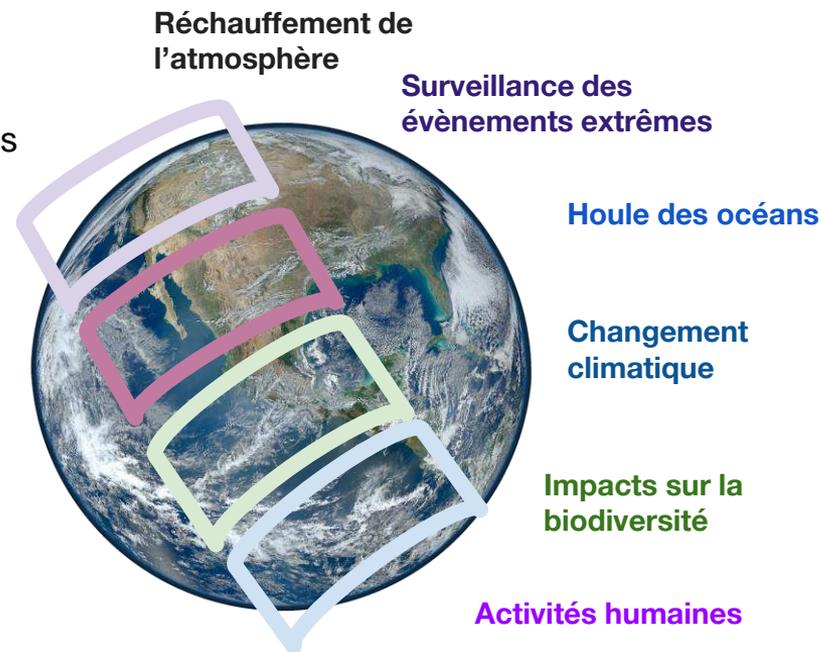
www.data-terra.org

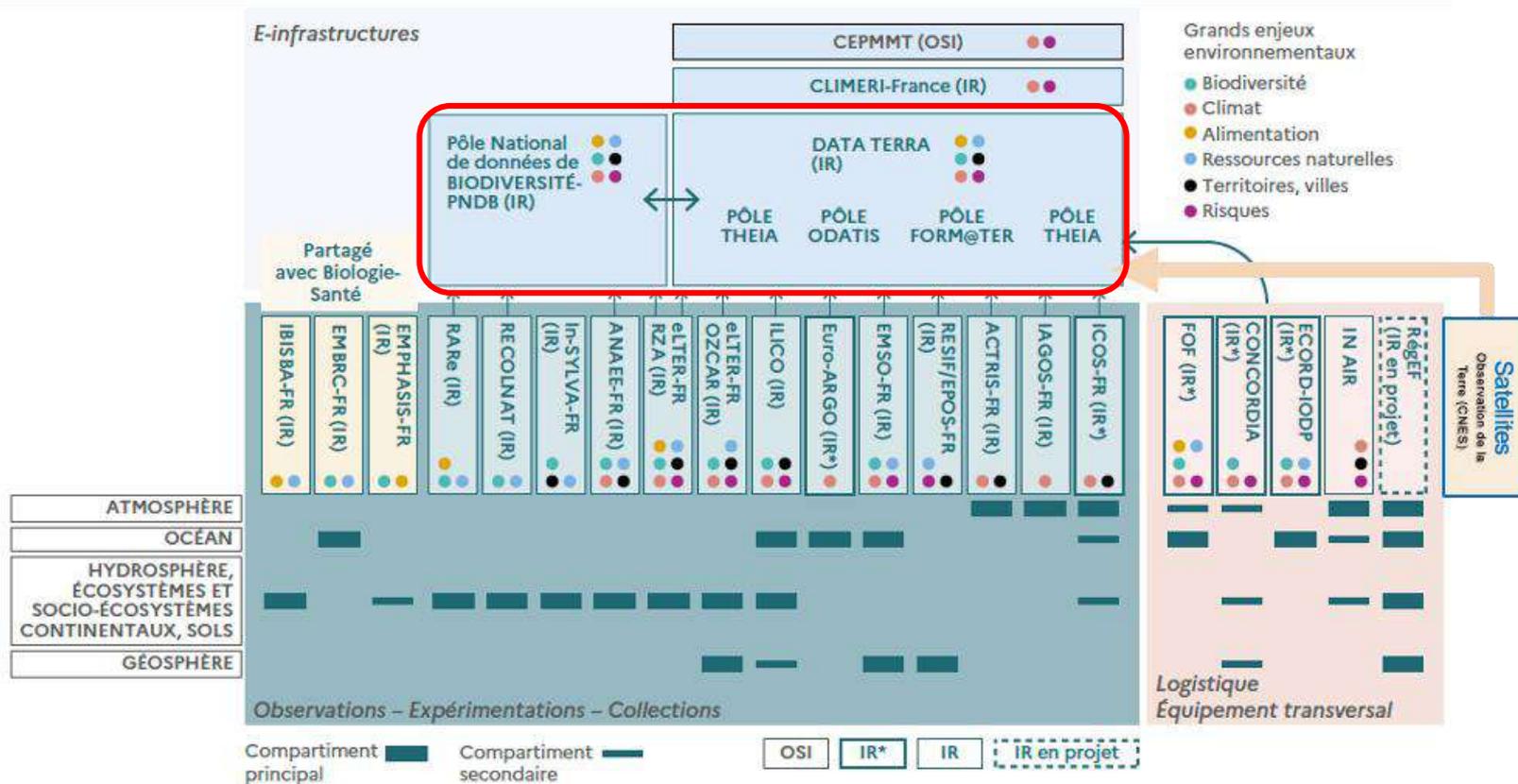
LA TERRE, UN SYSTÈME DYNAMIQUE COMPLEXE

- De nombreux processus géophysiques et environnementaux,
- A différentes échelles spatiales et temporelles,
- Interactions permanentes entre la Terre solide, les surfaces continentales, l'océan, les compartiments atmosphériques et anthroposphère.

ÉVOLUTION DES BESOINS

- **Approches plus intégrées de la complexité**
- **Données multi-sources, multi-capteurs permettant une multi-échelle (in situ, spatiale, spectrale et temporelle), séries long terme.**
- **Gestion de la diversité et volume : services cloud, Big Data et IA.**
- **Interdisciplinarité et transversalité des communautés scientifiques**





DATA TERRA : 5 PÔLES compartiments du système Terre-Environnement et dispositifs transversaux



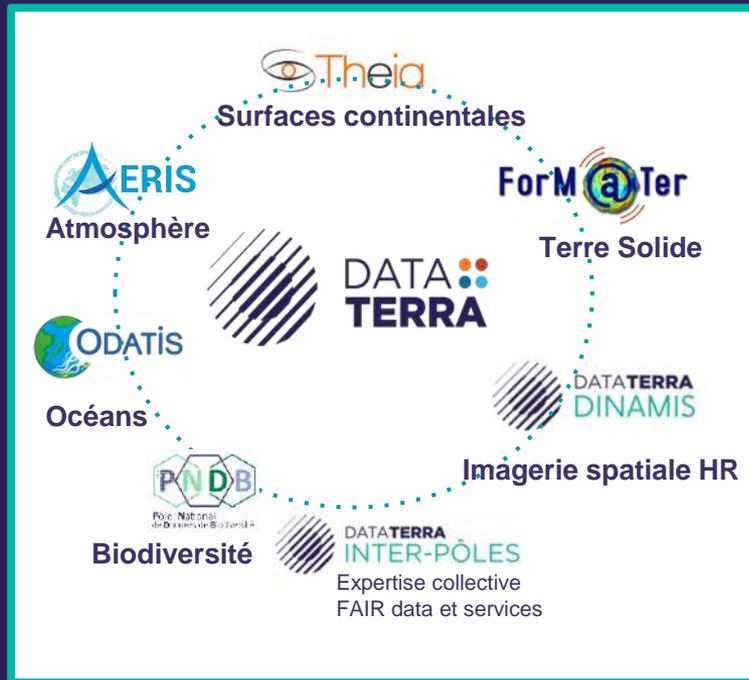
Faciliter l'accès et l'utilisation des données et produits de qualité sur l'ensemble des compartiments du système Terre (Données spatiales, aéroportées, sols, in-situ)

Développer des services de visualisation et de traitements adaptés aux besoins, à l'accroissement de la volumétrie et aux avancées technologiques

Favoriser la mutualisation, l'interopérabilité, émergence d'approches multi- et inter-disciplinaires

Servir les communautés scientifiques, les acteurs de l'action publique et de l'innovation

Mettre en œuvre une stratégie nationale, européenne et internationale



PROJETS SCIENTIFIQUES ET TECHNIQUES

Nationaux, Européens et Internationaux

- EOSC :
- FAIR-EASE, ...
- GAIA DATA
- PEPR ...



GRUPE DE TRAVAIL
SCIENCE, TECH, FORMATION,
EUROPE, COMMUNICATION,
COOPÉRATION Pays du Sud, ...



DATA
TERRA



ForM@Ter



AERIS



ODATIS

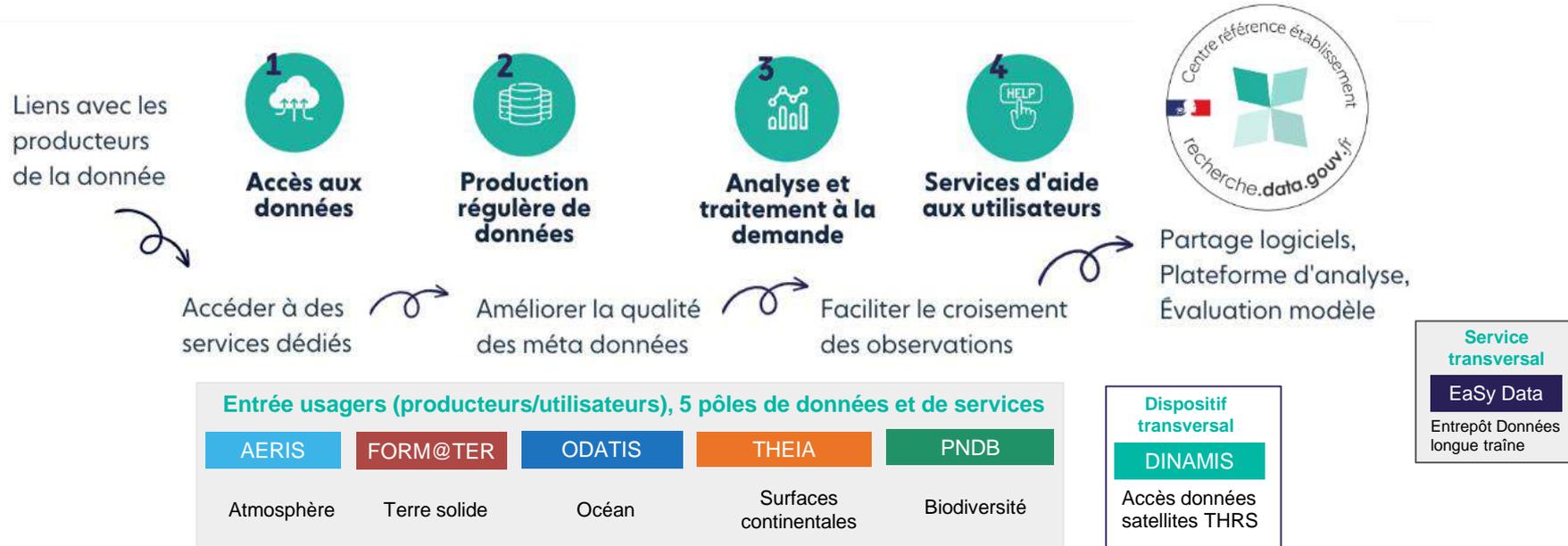


Theia

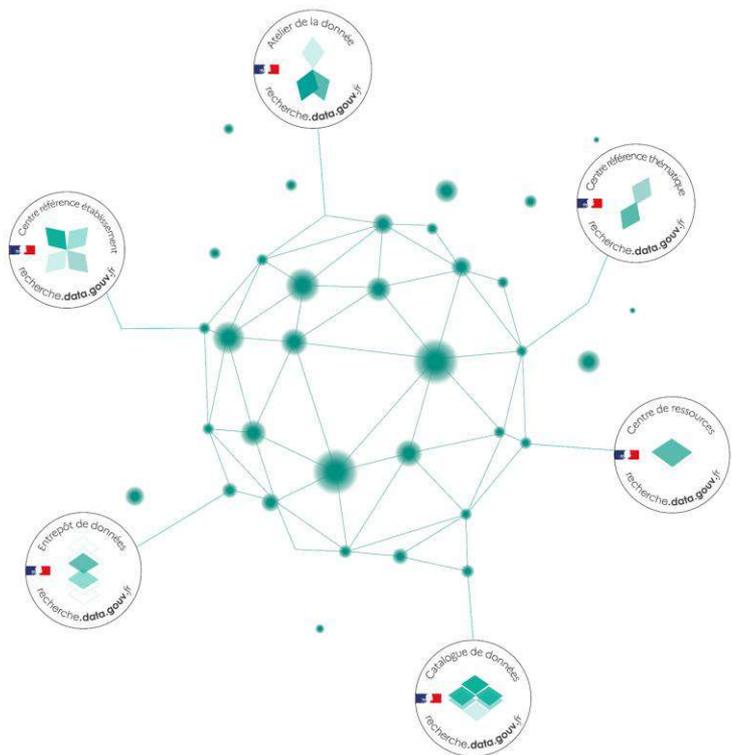


DATA TERRA, une e-Infrastructure de Recherche dédiée aux systèmes Terre-Environnement

L'IR DATA TERRA propose des **services** autour des données d'observation du **système Terre interopérables** et **interdisciplinaires** à tous les niveaux



Recherche Data Gouv : un écosystème pour et par les communautés de recherche



Missions des centres de références thématiques

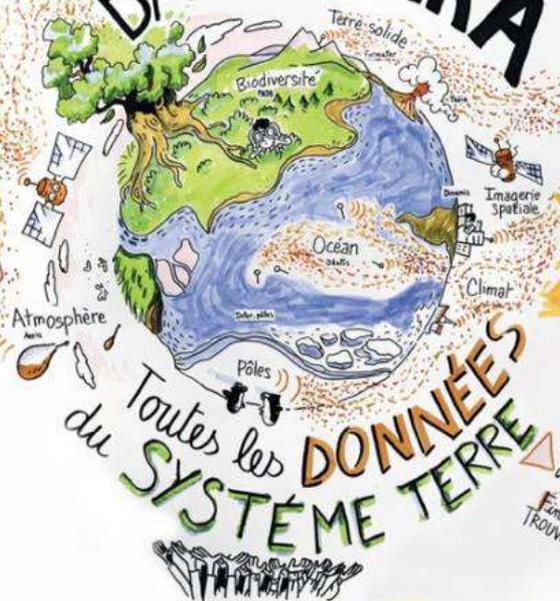
Définir

- les **normes internationales** de description et de diffusion des données (ouverture, période d'embargo, accès restreint).
- et **diffuser les bonnes pratiques** de collecte, documentation, traitement, et diffusion des données.
- la **liste des entrepôts de données de référence** (nationaux et/ou internationaux) vers lesquels orienter les chercheurs pour le dépôt des données et que *Recherche Data Gouv* moissonnera

Contribuer à la définition de l'**arborescence** des données de l'entrepôt *Recherche Data Gouv* thématiques **Systeme Terre et Environnement**

Soutenir l'**articulation entre les dispositifs thématiques** spécialisés et *Recherche Data Gouv* et l'**inscription dans l'écosystème international**

DATA TERRA



Toutes les **DONNÉES** du **SYSTÈME TERRE**

26 Tutelles

Institutions, Instituts, Universités
interactions européennes

Gaïa Data

PLATEFORME INTÉGRÉE numérique



Pour **ANALYSER** des problèmes **COMPLEXES**

pour **FACILITER** le **TRAVAIL** de recherche

multi sources
Grandes ENSEUX de SOCIÉTÉ
multi impacts

des **DONNÉES FAIR***
FONDABLE TROUVABLES
ACCESSIBLE ACCESSIBLES
INTEROPÉRABLE INTEROPÉRABLES
RENTABLES RENTABLES

TRANSMETTRE les BONNES PRATIQUES pour GARANTIR la QUALITÉ des DONNÉES

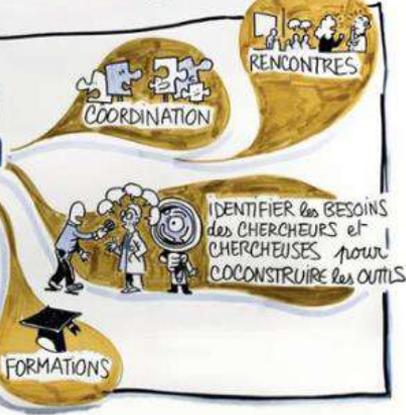
Elles sont FAIR mais données
Elles sont FAIR mais données

Infrastructures TECHNIQUES



Coordination HUMAINE

Créer une CULTURE COMMUNE





DATA
TERRA

Pôle de Données et Services pour l'Océan



Le pôle ODATIS dans l'IR DATA TERRA

Dans le cadre de l'IR Data Terra, ODATIS s'occupe de la gestion des données marines:

- Du littoral au hauturier, de la surface au plancher océanique,
- Physique, Chimie, Biologie dans les différents compartiments : Eau, Sédiment, Biota
- Aux interfaces terre/mer (**THEIA**), océan/atmosphère (**AERIS**), sous-sol sous-marin (**Form@ter**) et le vivant (**PNDB**)

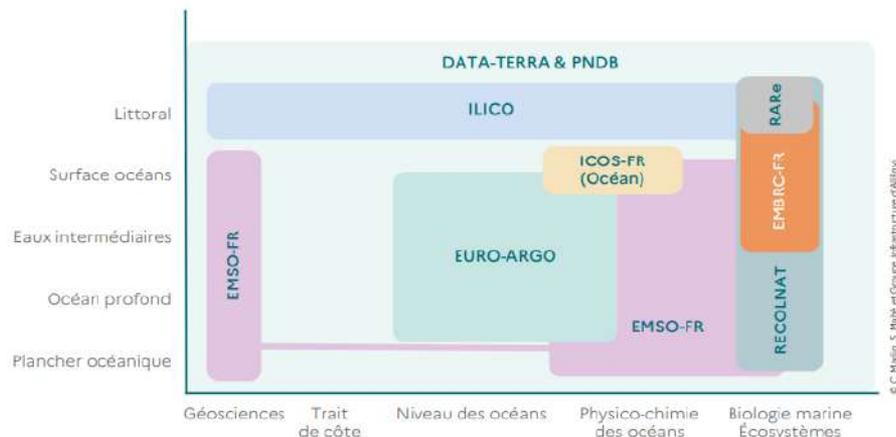


Figure 4 : infrastructures de recherche françaises dans le domaine « Océan et littoral ». L'encadré orange (EMBRC-FR) indique une infrastructure commune avec le domaine Biologie-Santé. Les infrastructures de logistique pour l'accès au milieu océanique ne figurent pas sur cette figure.

Tutelles partenaires



Les Centres de Données et de Services (CDS)

Chargés de la gestion des données au quotidien

- Bancarisation
- Métadonnées
- Contrôle qualité
- Service d'accès en ligne
- Préservation

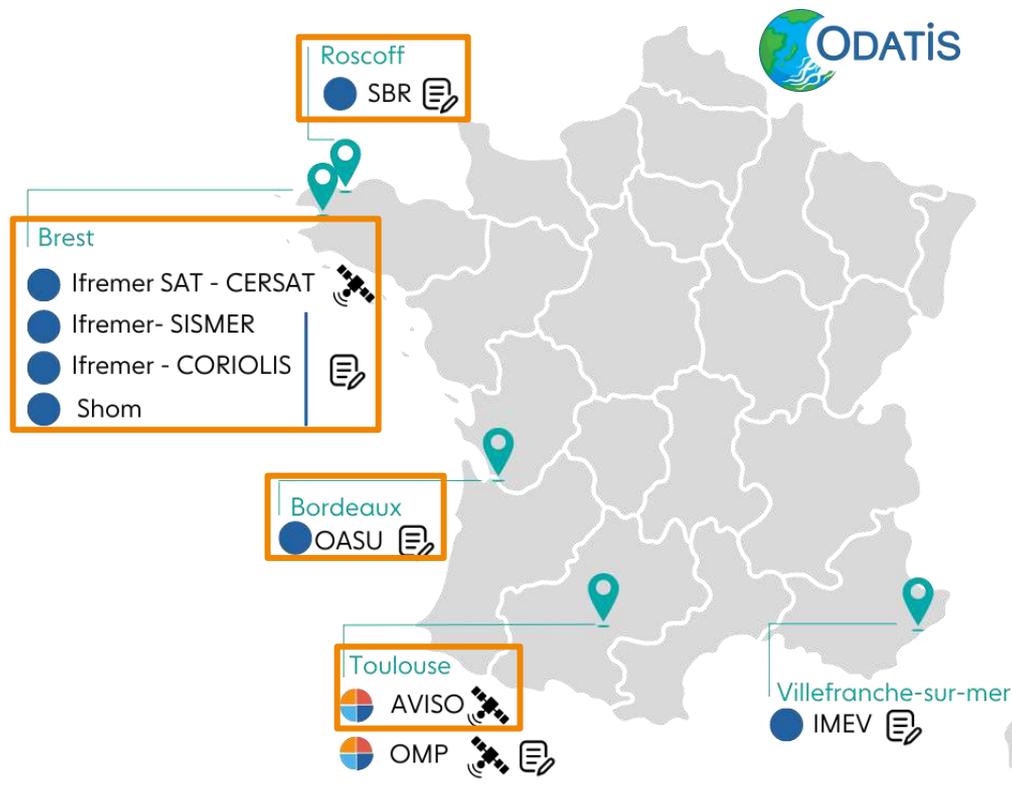
Soutenus par les tutelles du pôle

Dans le cadre d'un mandat et d'un cahier des charges

<https://doi.org/10.13155/81220>

ANR Flash Copilote (2020-2022)

- Amener à la certification Core Trust Seal
- Evaluation FAIRitude des données



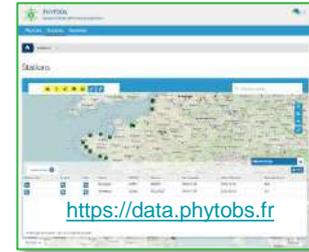


SNO PHYTOBS : réseau d'observation du microphytoplancton

- Bancarisation des données pour CNRS/Universités dans la base PELAGOS
- Mise en accès des données du SNO (agrégation des extractions de PELAGOS & Quadrige)
- Variables mesurées :
 - Flore totale « initiale » (détermination & occurrences)
 - Flore « compilée » reposant sur une **table de référence de groupements taxonomiques**
- En complément : Paramètres hydrologiques si disponibles



phytobs



<https://data.phytobs.fr>

SNO BenthOBS : réseau d'observation de la macrofaune benthique

- Bancarisation des données pour les producteurs CNRS/Universités dans la base BenthOBS-B
- Mise en accès des données du SNO (une version par an agrégeant des extractions de BenthOBS-B & Quadrige)
- Variables mesurées :
 - Macrofaune benthique (détermination & occurrences)
 - Granulométrie
 - *Matière organique sèche*
 - *Paramètres hydrologiques*



<https://www.benthobs.fr>



<https://data.benthobs.fr>

SNO COAST-HF

- Bancarisation des données de la bouée instrumentée au point ASTAN
- Mise à disposition des données pour leur intégration dans CORIOLIS
- Variables mesurées : T° et Pression atmo, Vitesses Vents, T° CTD, Conductivité CTD, Pression CTD, Salinité
- *Mise à disposition de représentations graphiques de l'évolution au cours du temps des param. mesurés*



CDS IS SISMER

Systèmes d'Informations Scientifiques
pour la MER



Gestion (collecte, contrôle, archive et diffusion) **des données in situ** et services associés, organisée en processus/filières (démarche qualité ISO 9001)

- « Recueillir et mettre à disposition des données sur le milieu marin » pour

- **Catalogue** (filière F1) et **Données** (filière F2) **des campagnes à la mer**

- **Entrepôt et catalogue des données scientifiques – Sextant et Seanoe** (filière F6)

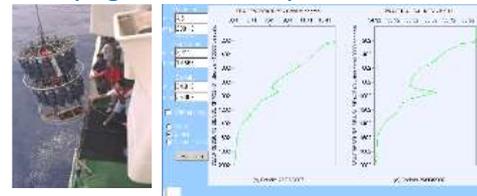
activités assurées par trois services dont SISMER pour l'exploitation dans le département IRSI (Infrastructures de Recherche et Systèmes d'Information)

- « Observer l'environnement littoral et les ressources biologiques » pour les **données d'environnement littoral**

activités assurées par le service VIGIES (Valorisation de l'Information pour la Gestion Intégrée Et la Surveillance) pour l'exploitation et IRSI pour le maintien en conditions opérationnelles du système d'information Quadrigé.

Banque de Physique/Chimie marines

Données des campagnes scientifiques



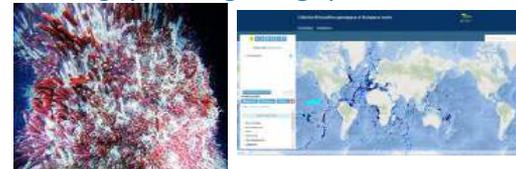
Banque de Géophysique/Géologie marines

Levés du plateau continental, talus, dorsales, etc.



Bigood : Biology and Geology Ocean Database

Echantillons biologiques et géologiques



Quadrigé/produit Surval : Environnement littoral

Hydrologie littorale, plancton, contaminants, etc.



Les Consortium d'Expertises Scientifiques (CES)

ODATIS s'appuie sur des CES pour **promouvoir et développer des méthodes de traitements et des produits innovants** pour les observations satellitaires, aéroportées et *in situ* de l'océan et à ses interfaces.

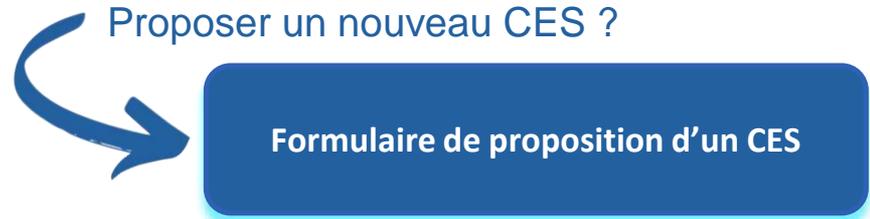
Ces CES mettent **en réseau et fédèrent des experts scientifiques et techniques** autour d'une thématique.

Le CES s'engage



Charte des Consortiums d'expertise scientifique

Proposer un nouveau CES ?



Formulaire de proposition d'un CES

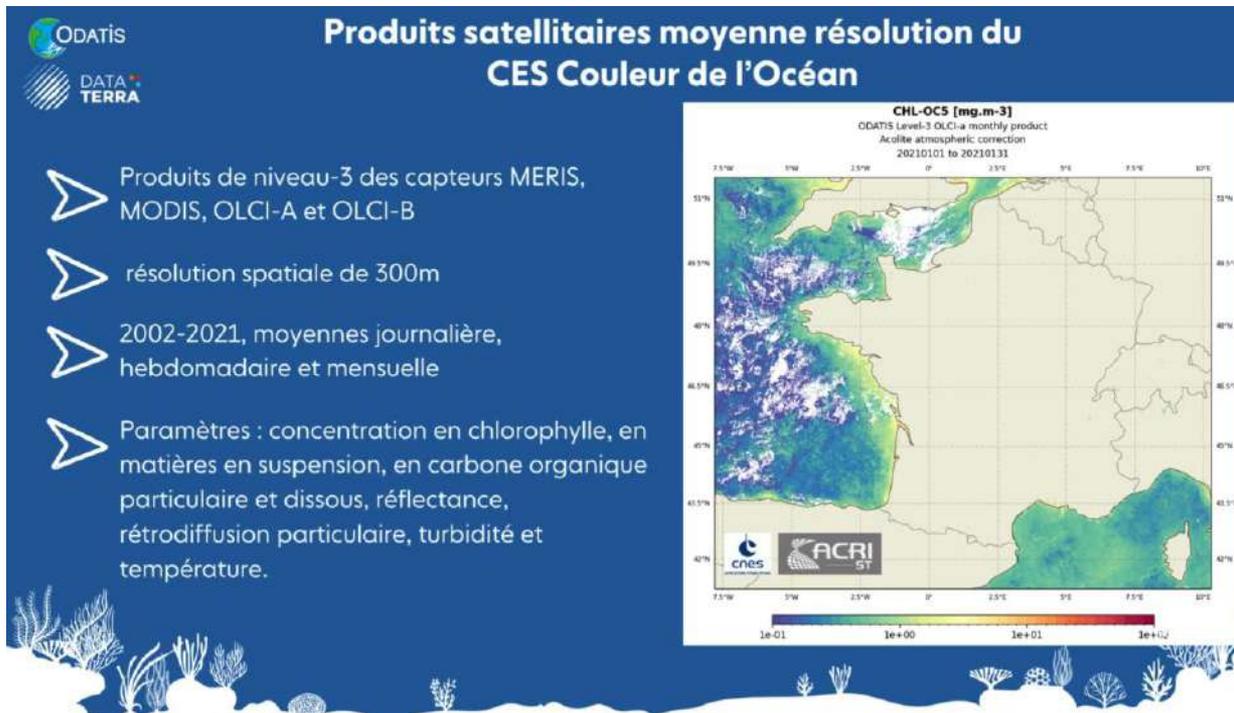
5

CES Couleur de l'Océan
CES Cytométrie en flux
CES Oxygène dissous
CES CO2 & pH marin
CES Imagerie optique benthique

CES Couleur de l'océan

David Doxaran (LOV) et Vincent Vantrepotte (LOG)

- Une réunion annuelle du CES
- Deux Groupes de Travail :
 - Groupe de travail **d'imagerie hyperspectrale** en milieu aquatique
 - Groupe de travail d'imagerie satellitaire de couleur de l'eau et **infrarouge thermique** à haute résolution spatiale
- 2024 : Distribution des produits satellitaires « **ODATIS-MR** »



bit.ly/42Xallm

Offre de services du pôle ODATIS

Stockage | Entrepôt | Catalogue | Visualisation | VRE | Accompagnement | Support

L'infrastructure informatique du pôle ODATIS repose sur 2 centres de données et calcul de type HPC alliant ressources de calcul et stockage dédié à l'hébergement et l'exploitation massive de données.

Supercalculateur Trex du CNES



Crédits CNES/Grimault

Supercalculateur DATARMOR à l'Ifremer



Crédits Ifremer/Olivier Dugonay

Offre de services du pôle ODATIS

Stockage | Entrepôt | Catalogue | Visualisation | VRE | Accompagnement | Support

SEANOE Sea scientific open data edition

SEANOE SEA SCIENTIFIC OPEN DATA PUBLICATION

PHYTOBS dataset - French National Service of Observation for Phytoplankton in coastal waters

Date: 2021
Temporal extent: 1987-2019
Author(s): PHYTOBS
Contributor(s): Lemoine, Maucé, Cloquin Pascal, Abadie Eric, Amaud Christophe, Arrigas Felipe, Blondel Camille, Breton Elsa, Carpentier Liliane, Chaumerat Nicolas, Christaki Urania, Conan Pascal, Cornet Véronique, Coste Laurence, Courty Gaelle, Dagault Françoise, Del Arno Yolanda, Delebecq Gaspard, Diener Anne, Dupuy Christine, Fauchot Juliette, Francoise Sylvaine, Gabellac Rabou,

Affiliation(s): Ifremer, France
Ifremer, VIGIES, Nantes
Ifremer, DYNECO-PELAGOS, Brest
Ifremer LER : Boulogne sur mer, Port en Bessin, Dinard, Concarnau, Lorient, La Tremblade, Sète, La Seyne sur Mer
CNRS
Université de Caen Normandie
Université de Lille - CNRS - UMR 8187 LOG ULCO
Université de Caen Normandie - CREC

Click to download the data DATA



Download metadata TXT, RIS, XLS, RTF, BIBTEX

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



SEDOO
Service de données
Observatoire Midi-Pyrénées



Entrepôt de données des CDS



EaSy Data
Earth System Data Repository

Publier, partager et découvrir les données du Système Terre et de l'Environnement

Déposer des données

Rechercher des données

<https://www.easydata.earth>

Offre de services du pôle ODATIS

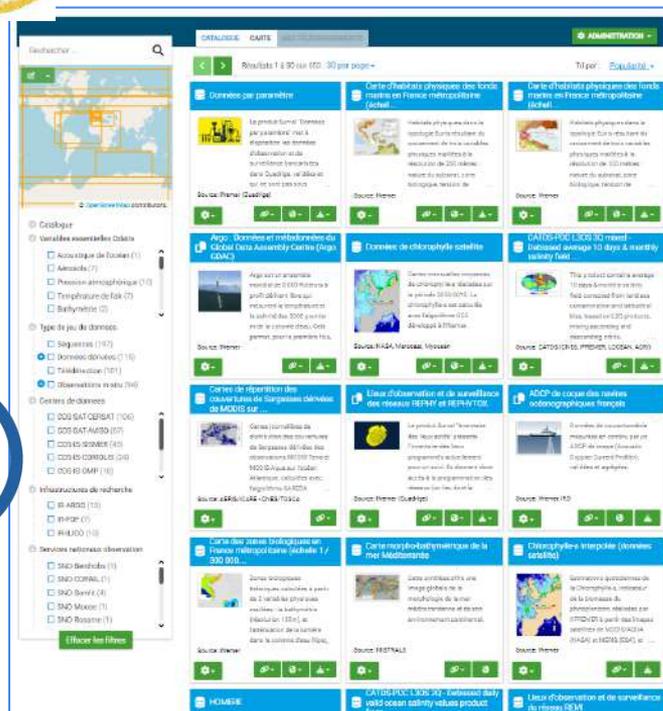
Stockage | Entrepôt | Catalogue | Visualisation | VRE | Accompagnement | Support



- Moissonnage des catalogues existants dans les CDS, OSU et autres sources,
- Service « API catalogue » proposé aux projets, laboratoires
- Elargissement du périmètre du catalogue
 - Intégration des données de SEANOE (2024), données bioinformatiques (fin 2023)
 - Création d'un catalogue de services | outils | scripts



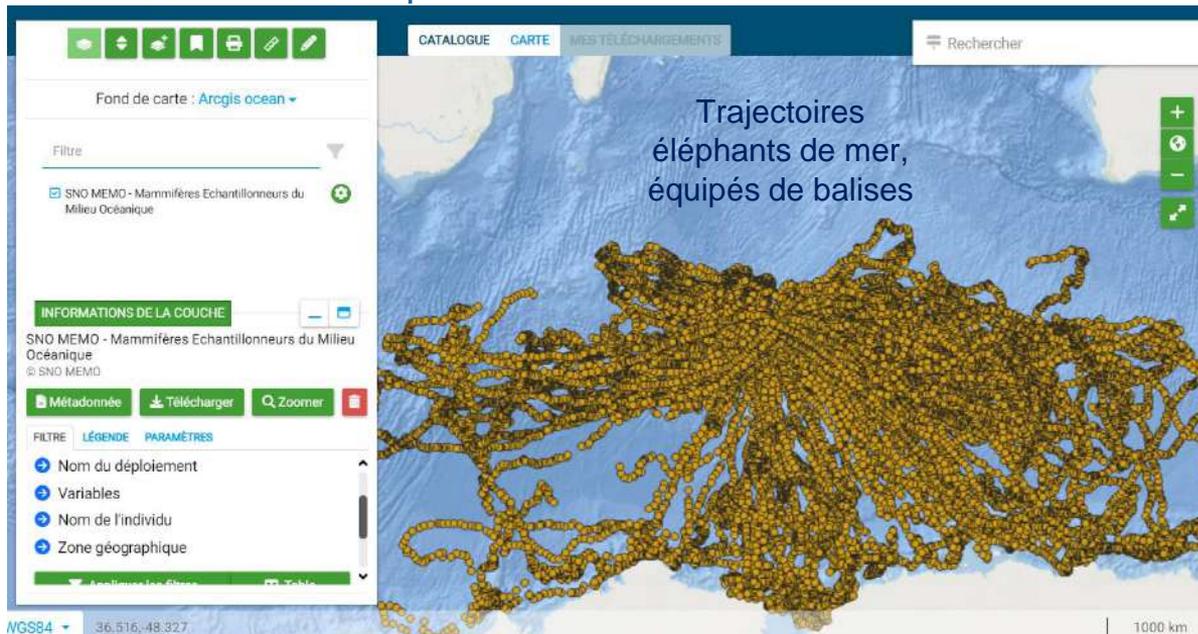
Infrastructure Sextant pour le catalogage des données



Offre de services du pôle ODATIS

Stockage | Entrepôt | Catalogue | **Visualisation** | VRE | Accompagnement | Support

- Création de services de visualisation interopérables
- Création de cartes thématiques interactives – Uses Cases



Offre de services du pôle ODATIS

Stockage | Entrepôt | Catalogue | **Visualisation** | VRE | Accompagnement | Support



Atelier NEO : niche écologique optimale

- Approche multi-sites et inter-SNOs pour **étudier de longues séries temporelles**,

- **Mutualisation** et couplage des bases de données,

- Atelier statistique pour **tester de nouvelles méthodes** de traitement de données.



Support ODATIS



Pressions multiples

- salinisation et ensablement des estuaires,
- baisse des régimes hydriques des fleuves,
- élévation du niveau marin,
- hausse de la stratification thermique,
- augmentation de l'intensité des événements météorologiques extrêmes.

Trajectoires écologiques

Nombreux **cas d'usages** développés dans le cadre de Gaia Data

J.F.Piollé, G.Caer, C.Germineaud

- Manipulation et exploration des données SWOT, croisement entre observations satellites + *in situ*
- Support aux campagnes en mer
(visualiser les traces satellites passées et futures pour planifier une campagne en mer)
- Traitement des données de fibre optique
- Traitement de données d'imagerie benthique
- Correction de mesures TSG de salinité...

Offre de services du pôle ODATIS

Stockage | Entrepôt | Catalogue | Visualisation | VRE | **Accompagnement** | Support

Le pôle ODATIS organise de nombreux ateliers techniques



Atelier Technique #12

Avancement COPILOTE
Présentation des guides FAIR et PID
Evaluation croisée Core Trust Seal
Evaluation FAIR

08 juin 2022 (matin) - Plouzané



Atelier Technique #13

Services d'accès aux données
Données d'imagerie optique sous-marine benthique

08 juin (pm) et 09 juin 2022 - Plouzané + visio



15 juin 2023 - Brest

Atelier Technique #15

Données de la Flotte Océanographique Française

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

ODATIS DATA TERRA FLEET Océanographique Française



Atelier Technique #17

Données bioinformatiques de diversité

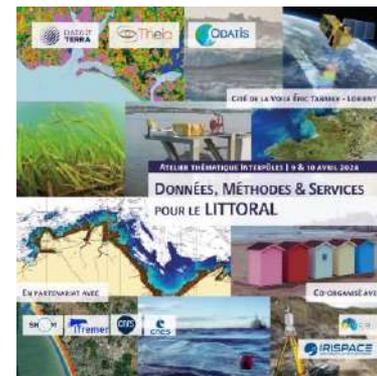
11-13 mars 2024
Station Biologique de Roscoff

Bonnes pratiques de gestion de données
Métadonnées, ressources terminologiques et vocabulaire communs (bioportals, ...), formats standards, définition claire des workflows, entrepôts de référence

Outils
Recommandations, exemples d'utilisations, gestion de métadonnées, publications dans des entrepôts, visualisation, montage de workflows, génération de rapports d'analyse

DATA TERRA ODATIS PDI DSI

Adms mio SeBIMER



DATA TERRA TheP ODATIS

CITE DE LA VIE ET DE L'ÉCOLOGIE - LORENT

ATELIER THÉMATIQUE INTERVÈNES | 9 & 10 AVRIL 2024

DONNÉES, MÉTHODES & SERVICES POUR LE LITTORAL

EN PARTENARIAT AVEC

CO-ORGANISÉ AVEC

IRISPACE

Atelier
thématique

Offre de services du pôle ODATIS

Stockage | Entrepôt | Catalogue | Visualisation | VRE | Accompagnement | Support

ANR COPILOtE

Certification CoreTrustSeal + Auto-évaluation FAIR

- 4 CDS demande de certification
- 3 demandes de renouvellement

ANR PPR

Un Océan de Solution

RIOMAR

FUTUROBS

PIA3 / EQUIPEX+

DeepSea Innovation

GAIA DATA



ODATIS



A VENIR

PEPR BRIDGES/Grands fonds marins
EMR

eosc | FAIR-IMPACT



Activités éditoriales

Portail web



www.odatis-ocean.fr/

Newsletter (250 abonnés)



bit.ly/3URbsyM

Rapport annuel 2022



[DOI 10.13155/95127](https://doi.org/10.13155/95127)



contact@odatis-ocean.fr



Pôle National de Données de Biodiversité

COMPRENDRE

?!

PARTAGER



UTILISER



www.pndb.fr



Le Pôle National de Données de Biodiversité : préambule



2018, Inscription sur la **feuille de route des Infrastructures** nationales de recherche du Ministère de l'Enseignement Supérieur, de la Recherche (MESR), pilotée par l'Allenvi.



9 organismes
10 universités

Portage par **18 établissements partenaires**

(+ Fondation pour la recherche sur la biodiversité, **FRB**).

=> Mise en œuvre confiée au **Muséum National d'Histoire Naturelle** via PatriNat (OFB-MNHN-CNRS-IRD) et un co-portage de l'animation avec la FRB



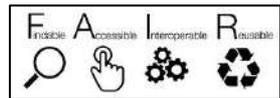
2022 : **Centre de référence thématique** "système Terre-Environnement" (MESR & Recherche Data Gov.) => Structuration des communautés autour des données

2024 : 5ème pôle de données de l'IR Data Terra, aux côté des pôles :

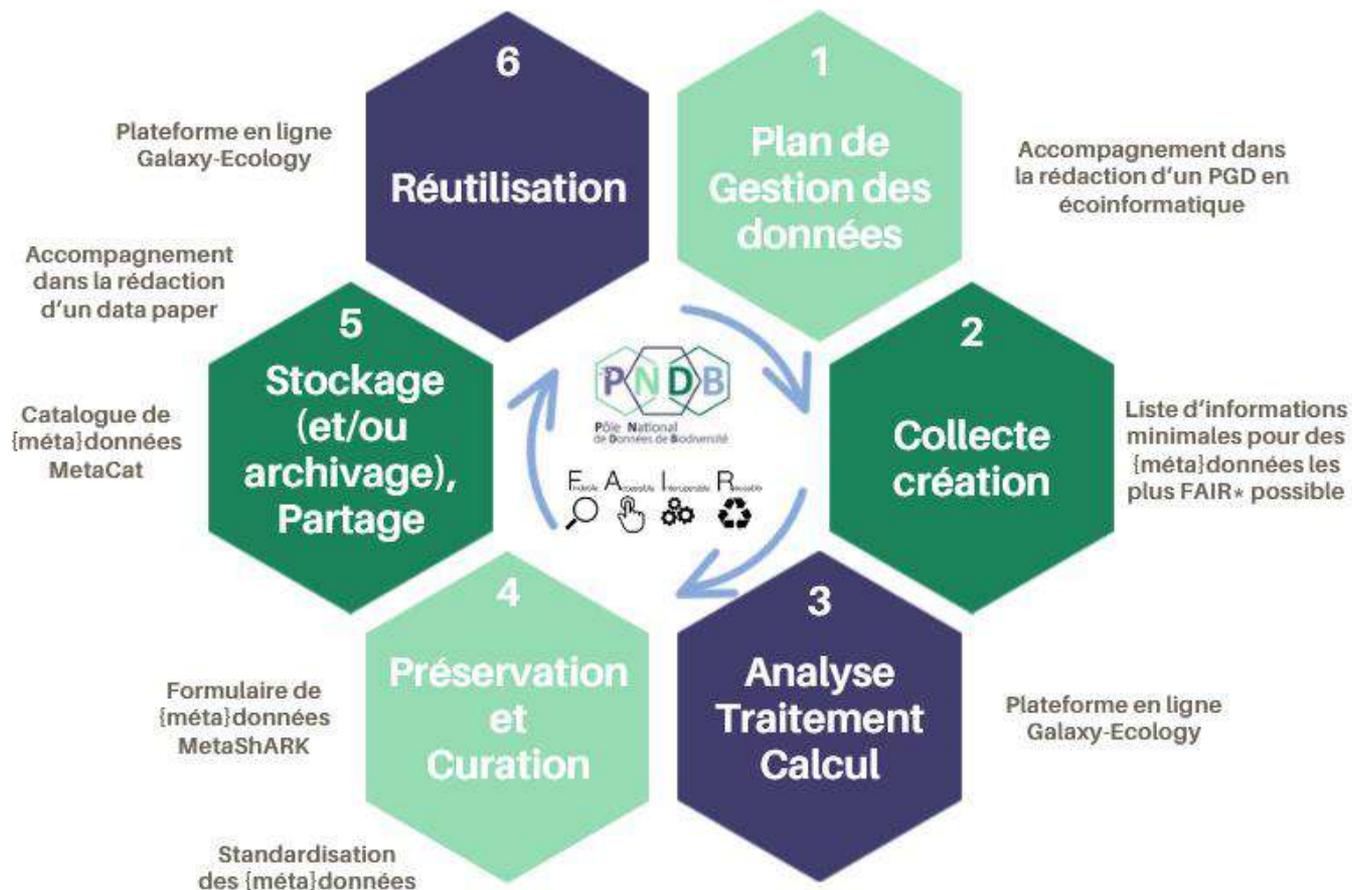
Aeris (atmosphère), Formater (terre solide), Odatis (océan) et Theia (surfaces continentales)



=> Un hub pour et par les communautés de Recherche pour accompagner et faciliter la compréhension, le partage et l'utilisation des {méta}données de biodiversité



Le Pôle National de Données de Biodiversité : *outils & services FAIR*



*Facile à trouver, Accessible, Interopérable, Réutilisable

COMPRENDRE

?!

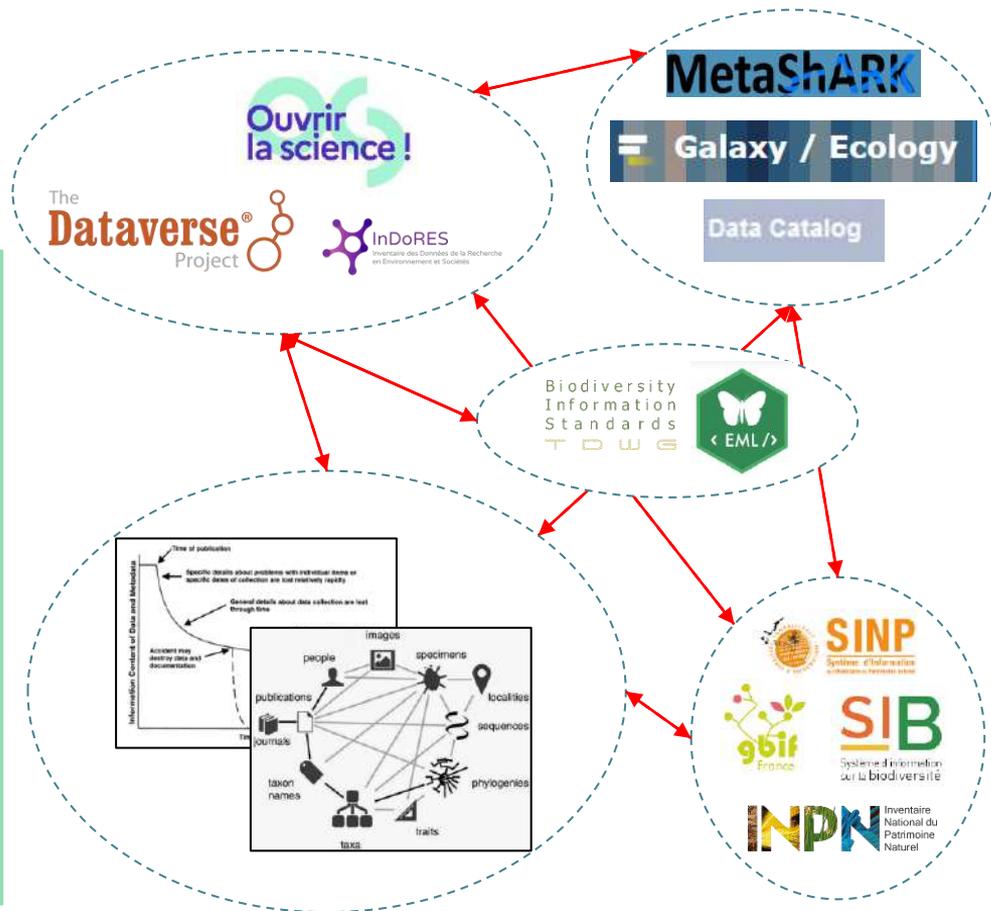


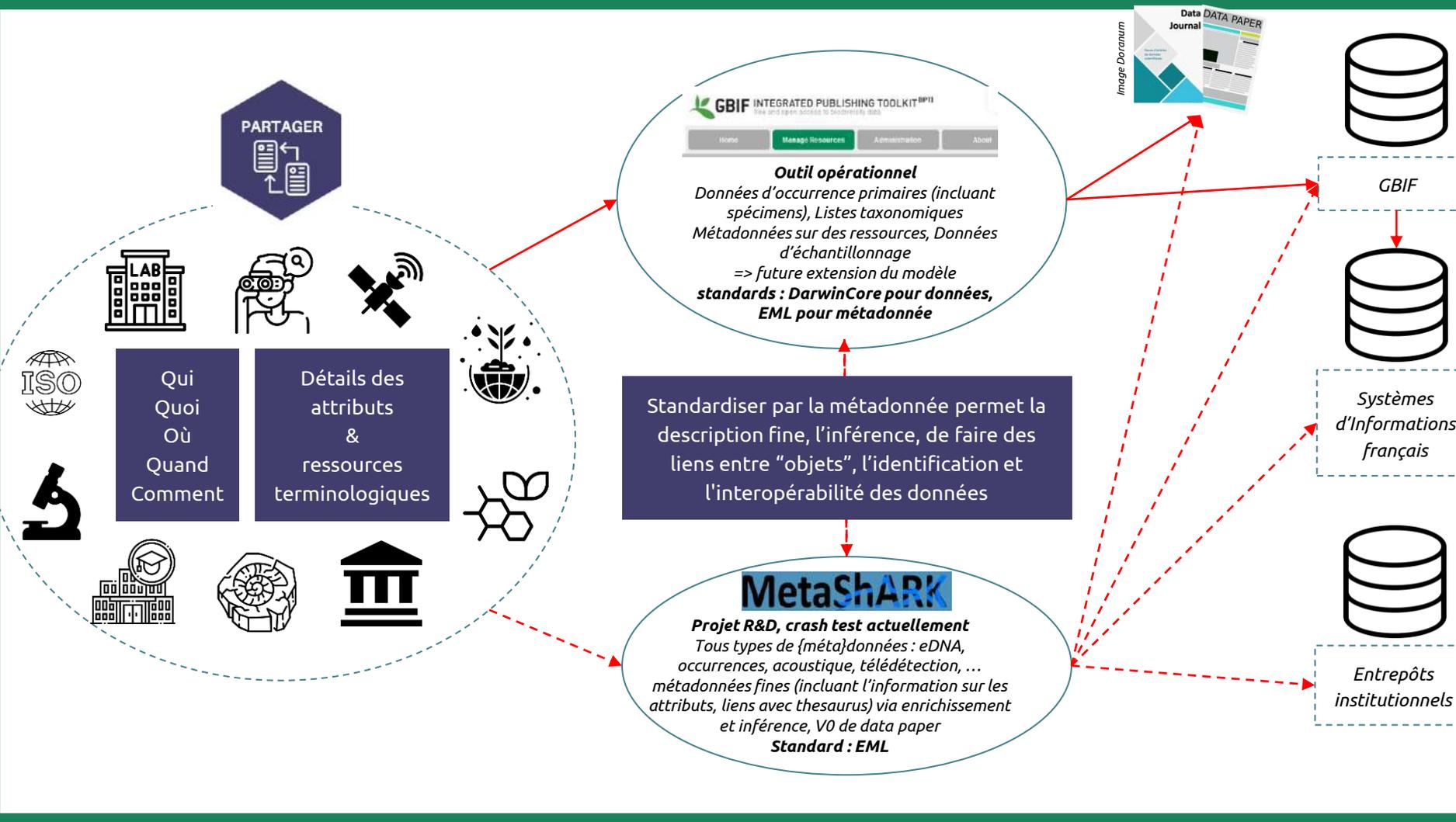
Le PNDB en tant que centre de références thématique, informe, forme, et oriente les communautés de recherche sur :

- Le **contexte et les enjeux autour des {méta}données de biodiversité**,
- les paysages de l'**écoinformatique (SI, entrepôts, IR, ...)**
- les bonnes pratiques (principes **FAIR, cycle de la donnée, ...)**
- les **standards inter-nationaux**
- les **outils** (e.g. Galaxy-Ecology) & **services** (DMP, hackathons, ...)
- etc...



recherche.data.gouv.fr



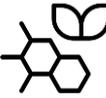


PARTAGER



Qui
Quoi
Où
Quand
Comment

Détails des attributs & ressources terminologiques



Outil opérationnel
Données d'occurrence primaires (incluant spécimens), Listes taxonomiques
Métadonnées sur des ressources, Données d'échantillonnage
=> future extension du modèle
standards : DarwinCore pour données, EML pour métadonnée

Standardiser par la métadonnée permet la description fine, l'inférence, de faire des liens entre "objets", l'identification et l'interopérabilité des données

MetaShARK
Projet R&D, crash test actuellement
Tous types de {méta}données : eDNA, occurrences, acoustique, télédétection, ...
métadonnées fines (incluant l'information sur les attributs, liens avec thesaurus) via enrichissement et inférence, VO de data paper
Standard : EML



GBIF



Systèmes d'Informations français



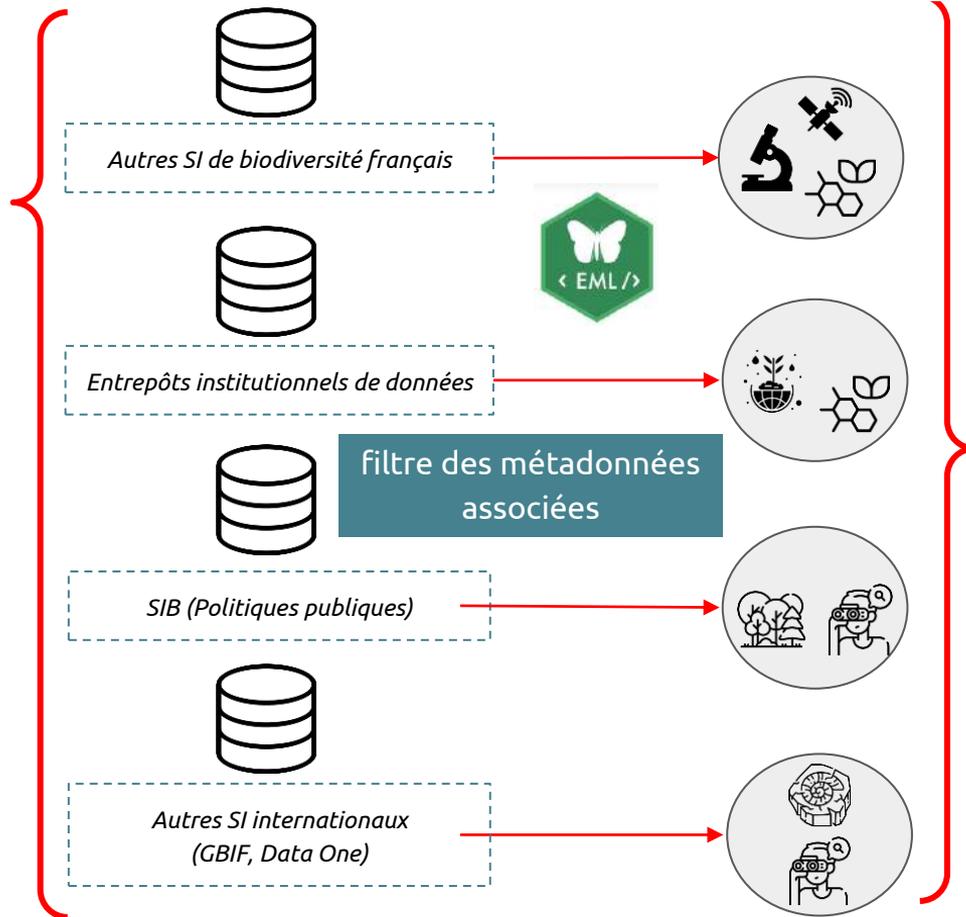
Entrepôts institutionnels

Des outils & services adaptés : utiliser les {méta}données : focus sur le catalogue

requête par filtres
sur le catalogue -
haut degré de
FAIRitude (attribute)

Filter by:

- ▶ Data attribute
- ▶ Data files
- ▶ Creator
- ▶ Year
- ▶ Identifier
- ▶ Taxon
- ▶ Location
- ▶ Access



Clear all filters

Search

My Search: IA Biodiv

Filter by:

- ▶ Data attribute
- ▶ Data files
- ▶ Creator
- ▶ Year
- ▶ Identifier

Datasets 1 to 3 of 3

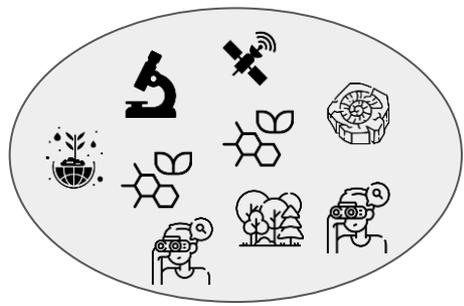
Sort by: Most recent

Francesco Aori, Mauro Bastianini, Fabrizio Bernardi Aubry, Elisa Camati, Caterina Bergami, et al. 2022. LTER Northern Adriatic Sea marine data from 1965 to 2015. PNDB Data Repository. urn:uuid:281b9c25-f13-fa92-6b9e-80e526b0701.

Graham Edger and Rick Stuart-Smith. 2022. Reef Life Survey Fish Mediterranean sample. PNDB Data Repository. urn:uuid:81b39ef5-016-4ec5-8fb5-e40775a8d1a.

Observatoire Pelagos University La Rochelle - CNRS Ministère de la Transition Ecologique (MTE) Office français de la Biodiversité (OFB), 2022. SAMM observatoire Pelagos oiseaux. PNDB Data Repository. urn:uuid:903885c2-4524-45a0-8125-f81754909e11.

sélection de jeux de données
pertinents et accessibles
répondant à la requête



Plateforme d'analyses en écologie



Galaxy
PROJECT

PNDDB Pôle National
de Données de Biodiversité

Une plateforme en ligne pour
obtenir, traiter, analyser et visualiser
des données écologiques

Accès via :

- ecology.usegalaxy.eu et/ou
ecology.usegalaxy.fr

Tutoriels

- training.galaxyproject.org/training-material/topics/ecology/

