



DATA
TERRA



ODATIS

Campus de la Mer – 14 mai 2024 !

Cycle de vie des données

Sabine Schmidt

Directrice scientifique d'ODATIS



Objectif et plan de l'exposé

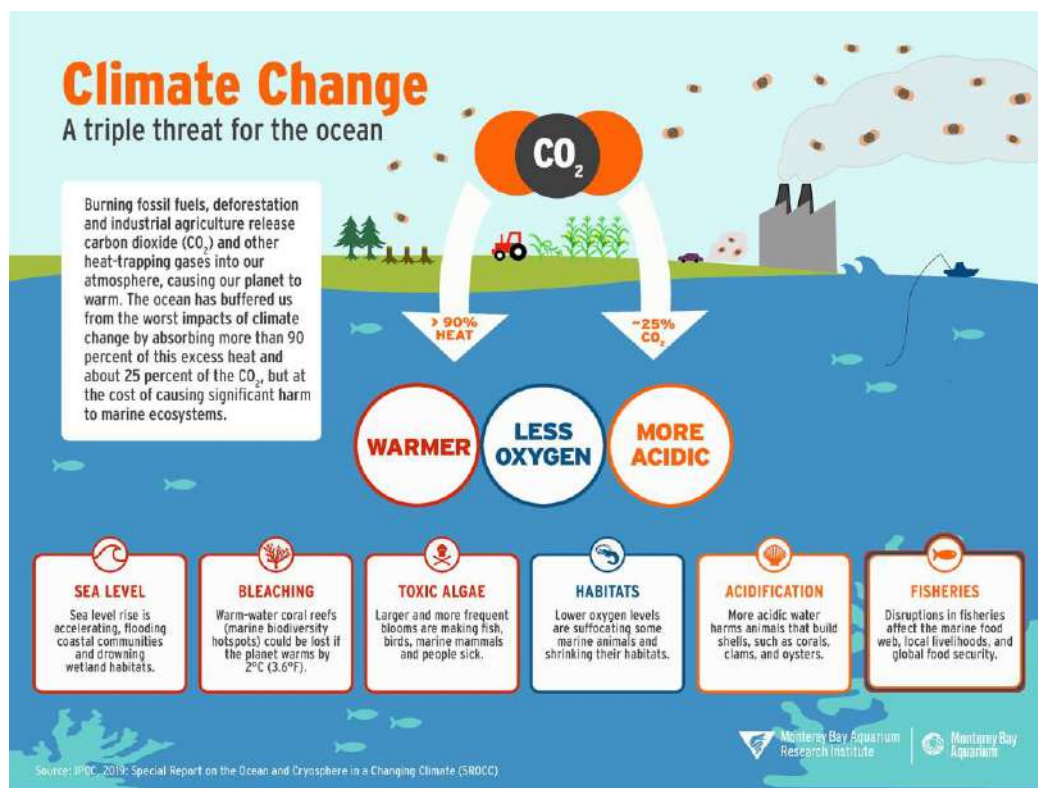
Sensibiliser sur le cycle de vie et la gestion des données d'observation,

L'importance des données
Le cadre réglementaire
Les principes FAIR
Le besoin d'infrastructures nationales
Le DMP

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

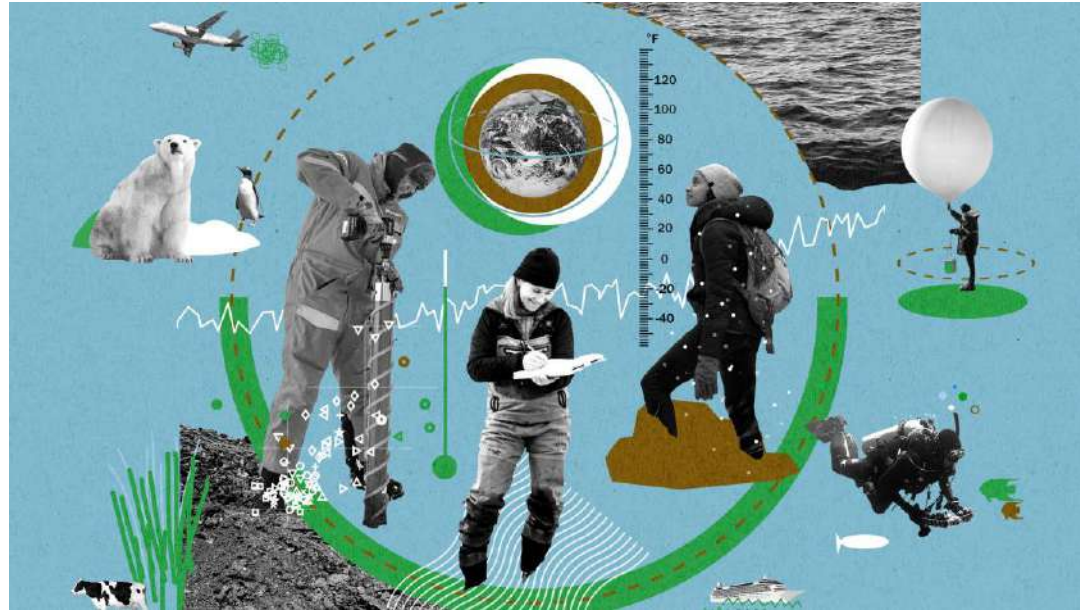
Considérant que:

- 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.

Le cadre réglementaire



La directive 2007/2/CE INSPIRE du 14 mars 2007:

Directive du Parlement européen visant:

- à rendre disponible une information géographique, appropriée, harmonisée et de qualité, pour aider à l'élaboration, l'exécution, la surveillance et l'évaluation des décisions politiques environnementales européennes.
- Ce qui impose d'établir une infrastructure d'information géographique = ensemble de services d'information disponibles sur Internet, répartis sur les sites des différents acteurs concernés, et permettant la diffusion et le partage de données géographiques.
- **concerne les autorités publiques (État, collectivités territoriales et leurs groupements, établissements publics, ainsi que toute personne physique ou morale fournissant des services publics en rapport avec l'environnement).**
- **impose aux autorités publiques, de rendre ces données accessibles au public en les publiant sur Internet, d'autre part de les partager entre elles.**

Le cadre réglementaire

La science ouverte



Le Plan national pour la science ouverte
2021-2024 : vers une généralisation de la
science ouverte en France



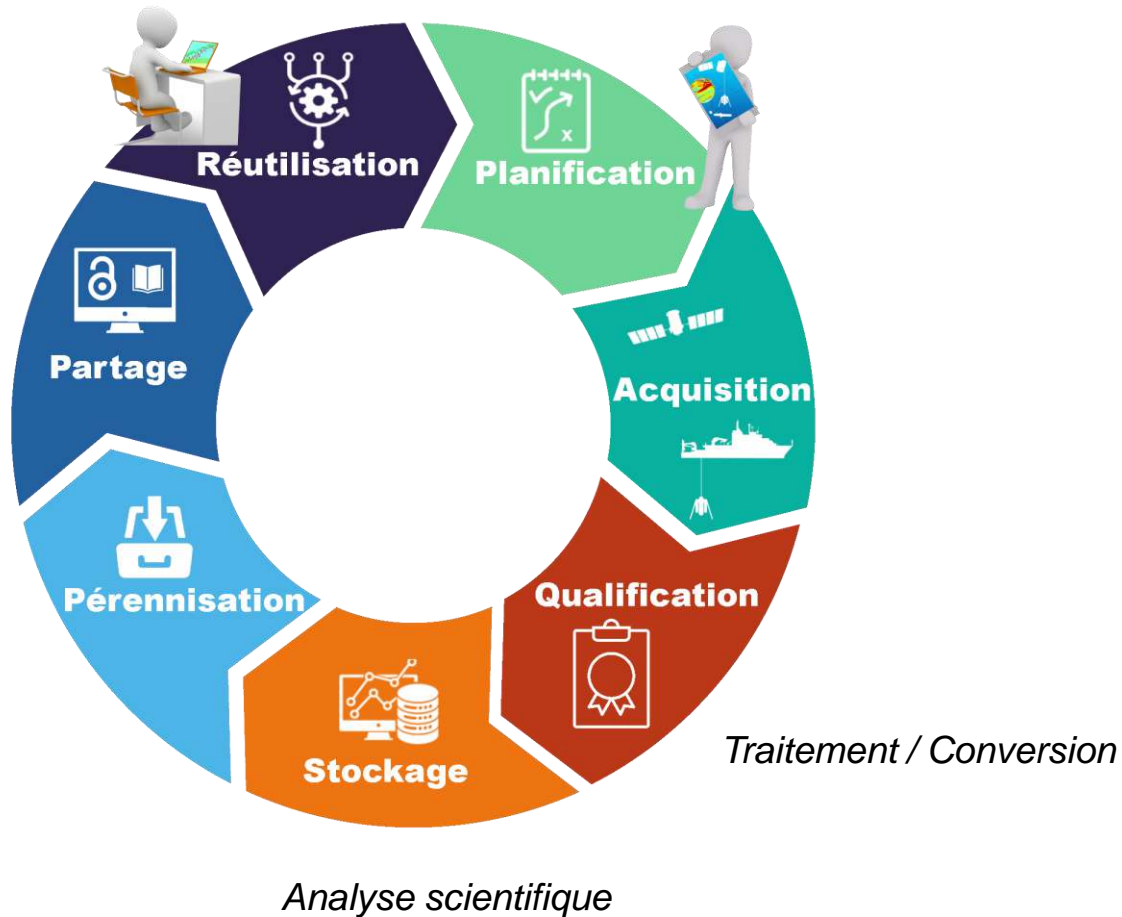
visent à généraliser les pratiques de science ouverte, à partager et ouvrir les données de la recherche, et à promouvoir les codes sources produits par la recherche.

→ création de la plateforme nationale des données:

www.data.gouv.fr

données environnementales : IR Data Terra

Le cycle de vie de la donnée

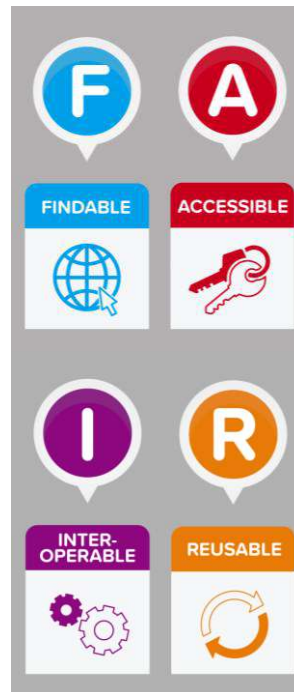


Les défis de l'accès aux données marines

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.

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.

the CoreTrustSeal requirements



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

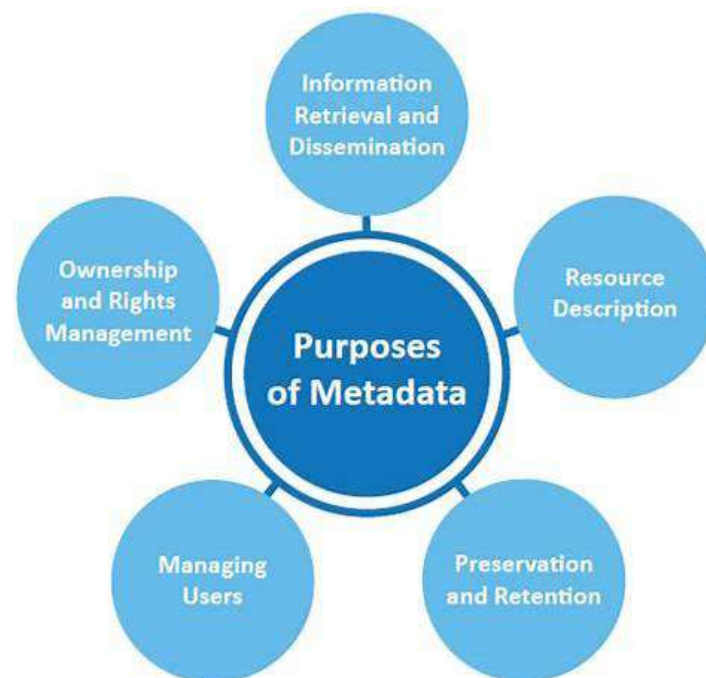
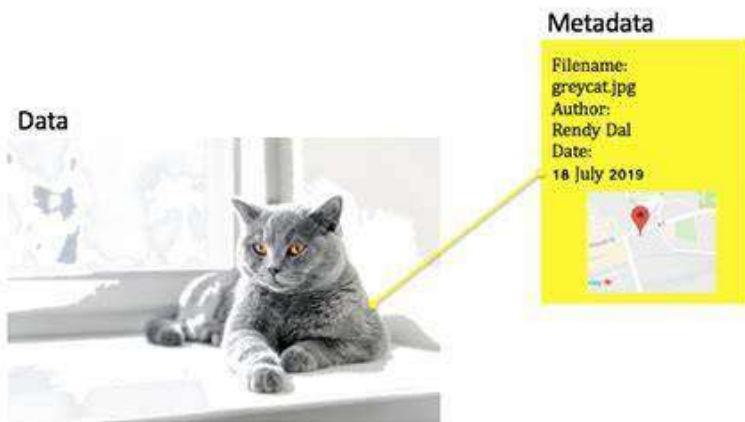
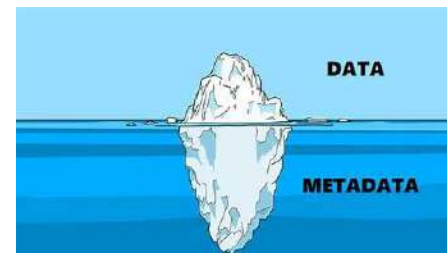
**plus faciles pour le producteur
pas seulement pour l'utilisateur.**

Campus de la Mer, le mardi 14 mai 2024

| 11

La donnée

Les méta-données
Données sur les données



Le vocabulaire

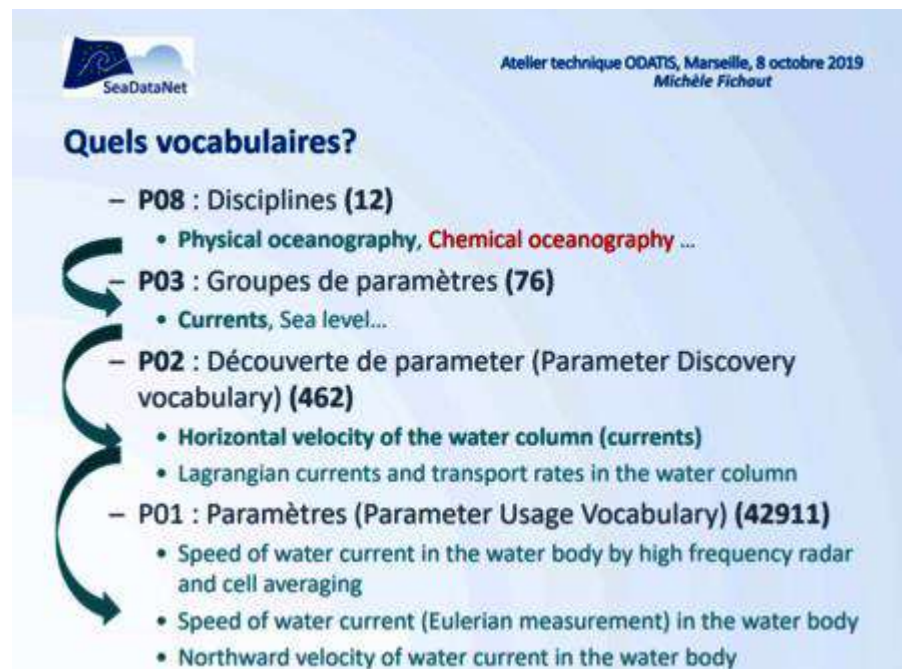
Un paramètre est défini par:

un vocabulaire contrôlé et un modèle sémantique pour faciliter l'échange de donnée et l'interopérabilité

Exemple: serveur de vocabulaire British Oceanographic Data Centre-BODC).
adoptés par plusieurs projets (nationaux et européens)

Pour plus de détails:

<https://www.odatis-ocean.fr/donnees-et-services/principes-de-gestion-des-donnees/referentiels-et-voculaires>



SeaDataNet

Atelier technique ODATIS, Marseille, 8 octobre 2019
Michèle Fichaut

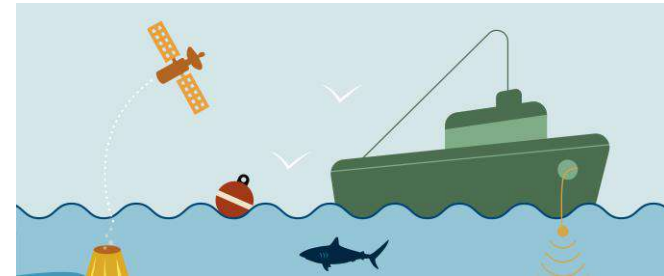
Quels vocabulaires?

- P08 : Disciplines (12)
 - Physical oceanography, Chemical oceanography ...
- P03 : Groupes de paramètres (76)
 - Currents, Sea level...
- P02 : Découverte de parameter (Parameter Discovery vocabulary) (462)
 - Horizontal velocity of the water column (currents)
 - Lagrangian currents and transport rates in the water column
- P01 : Paramètres (Parameter Usage Vocabulary) (42911)
 - Speed of water current in the water body by high frequency radar and cell averaging
 - Speed of water current (Eulerian measurement) in the water body
 - Northward velocity of water current in the water body

Favoriser le partage et l'archivage des données

Quels sont les freins à la mise en base de données ?

complexité
méconnaissance des bases, format
le temps
enjeux de propriétés intellectuelle
pb de la transformation de la donnée (licences)



Favoriser le partage des données

Ce qui pourrait inciter les producteurs à mettre en base ?



simplifier le format des fichiers de données
lien avec le producteur de données

assurer la protection (propriété
intellectuelle)

retour sur l'usage des données



Le cycle de vie de la donnée implique un plan de gestion des données (DMP)





Merci pour votre attention

