



## Table ronde

Cycle de Vie de la Donnée

Sabine Schmidt (EPOC) et Joël Sudre (LEGOS)



Ce document présente

- le support utilisé pour animer la table ronde
- La restitution des échanges, **en orange**



# Objectifs de la table ronde

Echanger sur:

- les pratiques,
- les besoins
- les attentes

en matière de gestion des données de recherche et d'aider à construire une offre de services adossée à l'IR Data Terra et les pôles de données et services, comme ODATIS pour l'océan. Les discussions porteront notamment sur le besoin de méta-données, la place et liens avec les producteurs de données, le contrôle de l'usage des données sous forme de DOI ou de licences, et le vocabulaire.

Le but est de sensibiliser la communauté à la nécessité de rendre les données de recherche accessibles (directives INSPIRE, DCE ....).



# Contexte

## La directive INSPIRE du 14 mars 2007

- vise à établir une infrastructure d'information géographique pour favoriser la protection de l'environnement.

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.



# Contexte

## Le principe FAIR

- le partage et l'ouverture des données doivent de surcroît respecter la notion de:

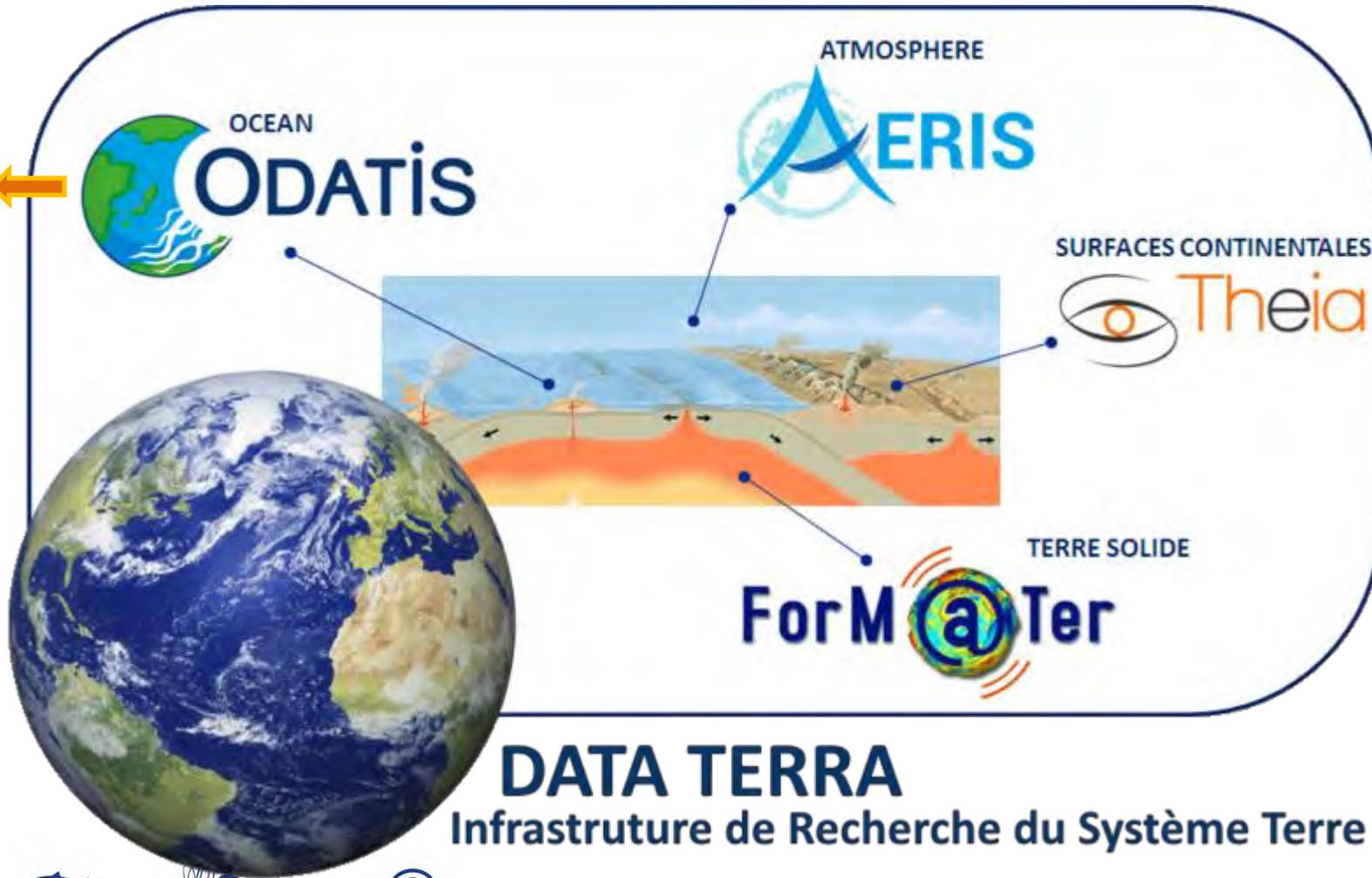
FAIR data

À savoir les manières de construire, stocker, présenter ou publier des données de manière à permettre que la donnée soit: trouvable, accessible, interopérable et réutilisable.

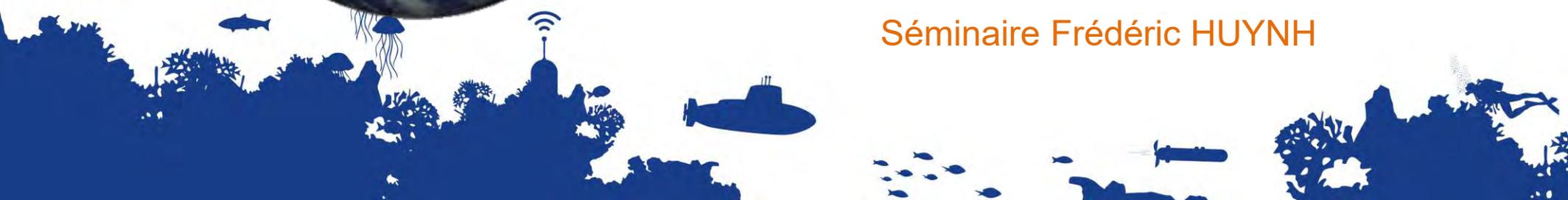


# La réponse française

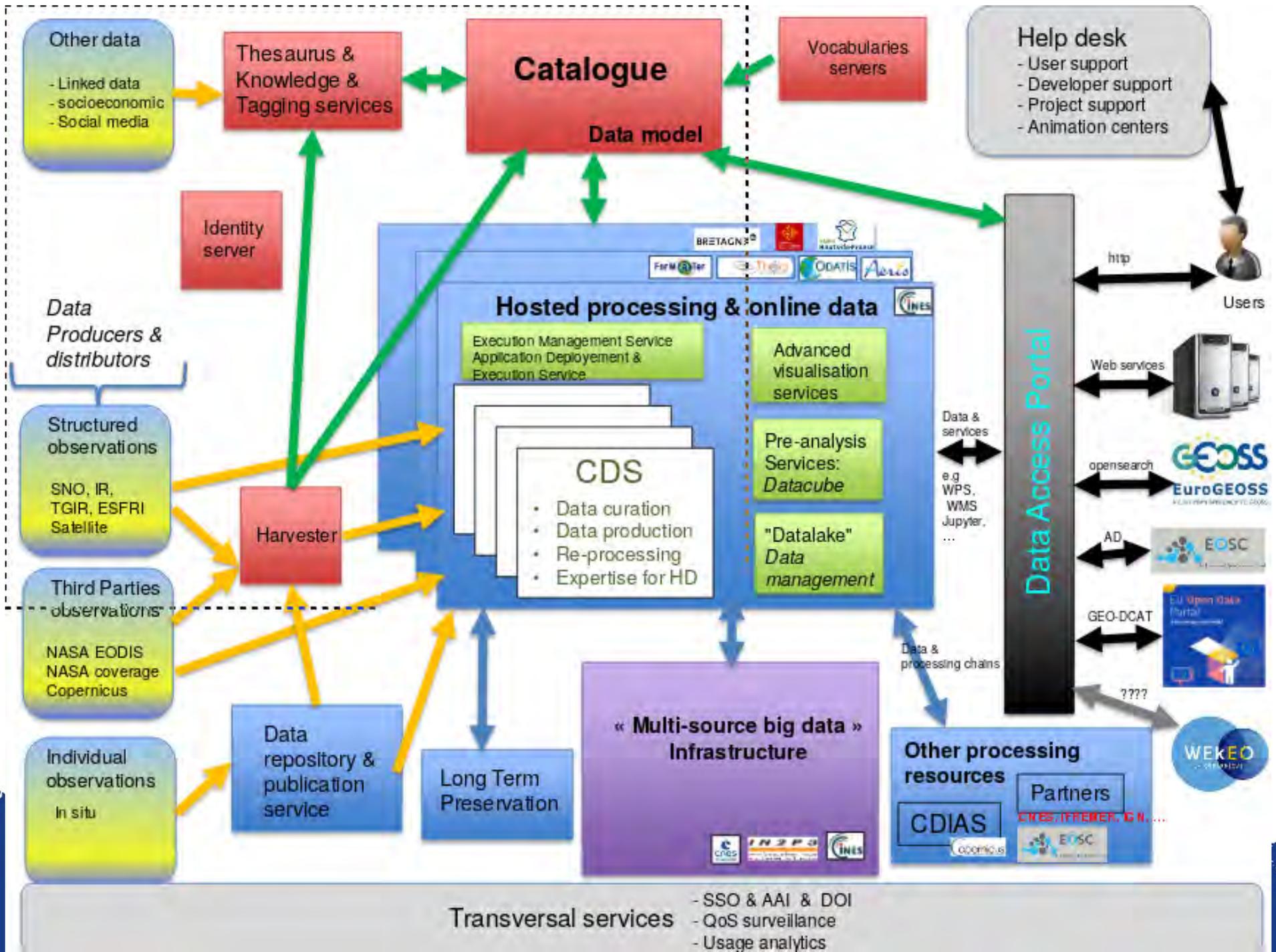
Poster ←



Séminaire Frédéric HUYNH



# A terme !





# Vous causez DPM ?

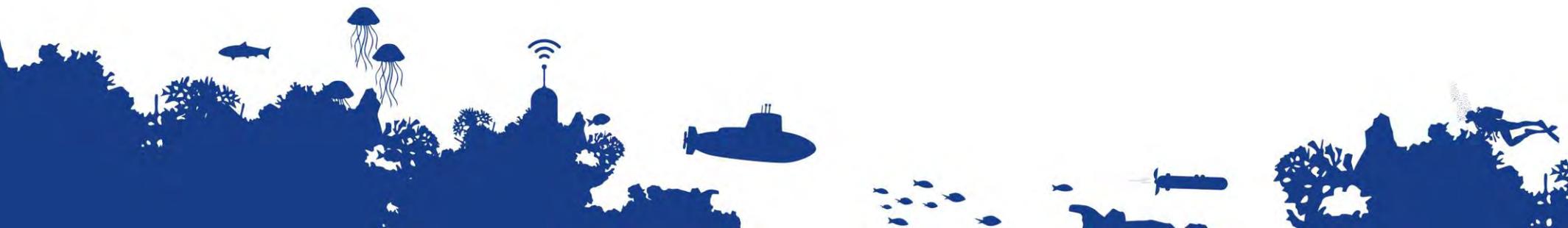
Multiplication de l'instrumentation in situ (capteurs innovants, drones, gliders, ...)

→ production de données très diverses (stations fixes ou lagrangiennes, fréquence d'acquisition, paramètres mesurés).

→ crucial de prévoir en amont

le plan de gestion des données ou DMP data management plan

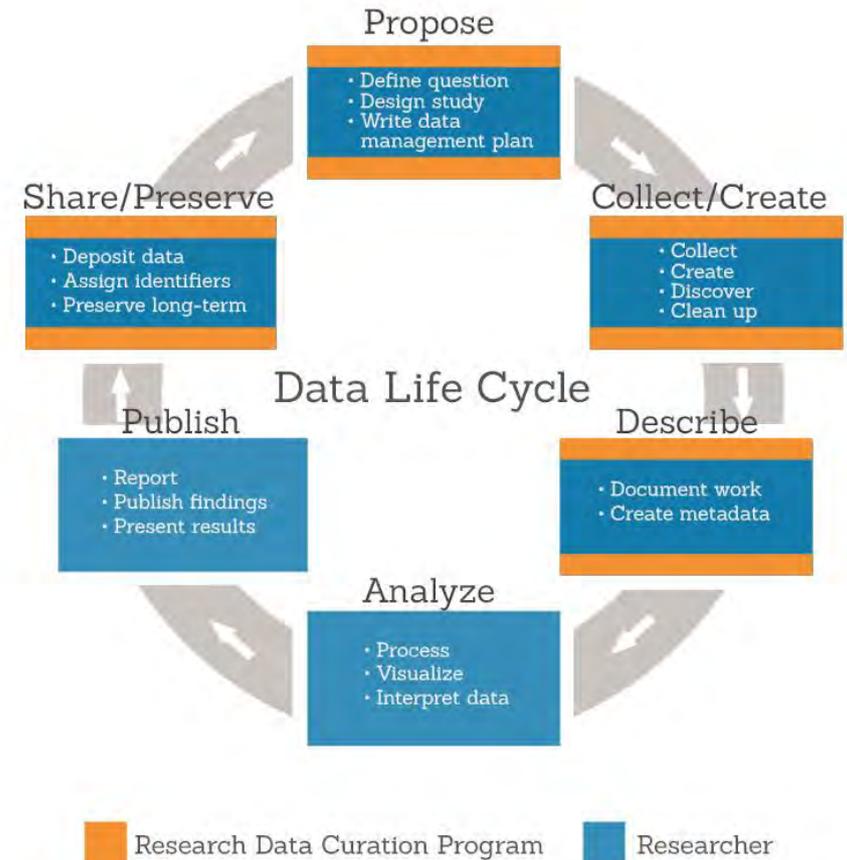
pour garantir l'acquisition de bases de données, bien documentées en vue d'exploitation dans le futur, par des politiques d'archivage, d'interopérabilité et de licences adaptées.



# Vous causez DPM ?

Avant toute manipe, voire dépose de projet de recherche:

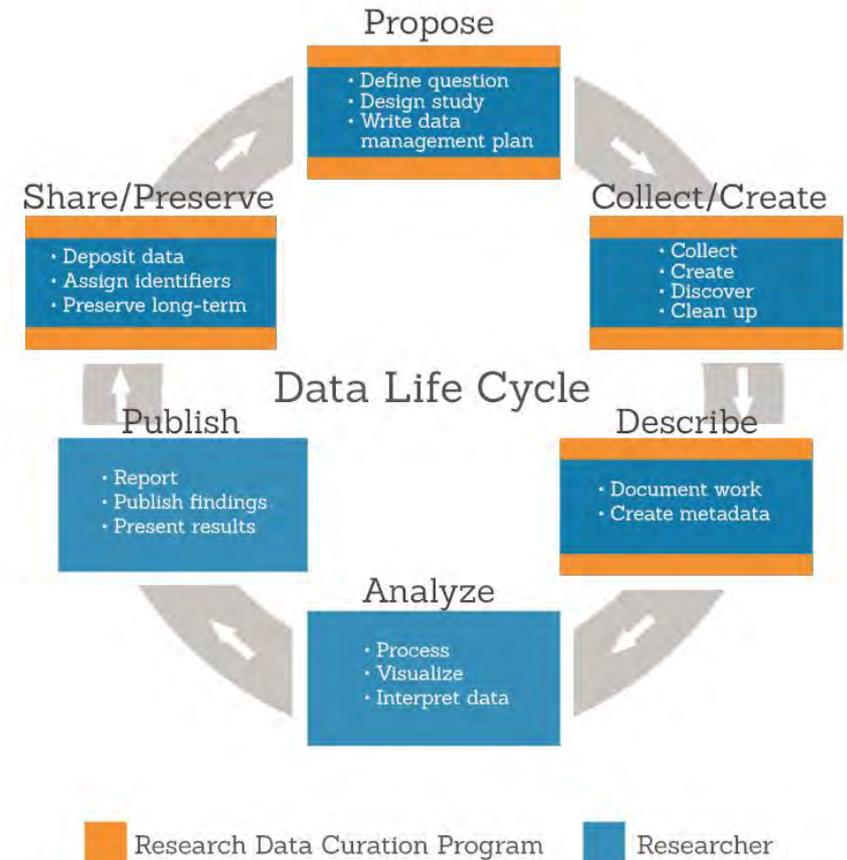
- Nécessité d'établir un data management plan pour établir comment les données seront gérées pendant le projet mais surtout à la fin (metadata, bancarisation, préservation pour le futur).



# Vous causez DPM ?

Retour d'expérience ?  
Contrainte ou Bénéfice

Suggestions sur les services que  
pourraient offrir les pôles pour aider ?



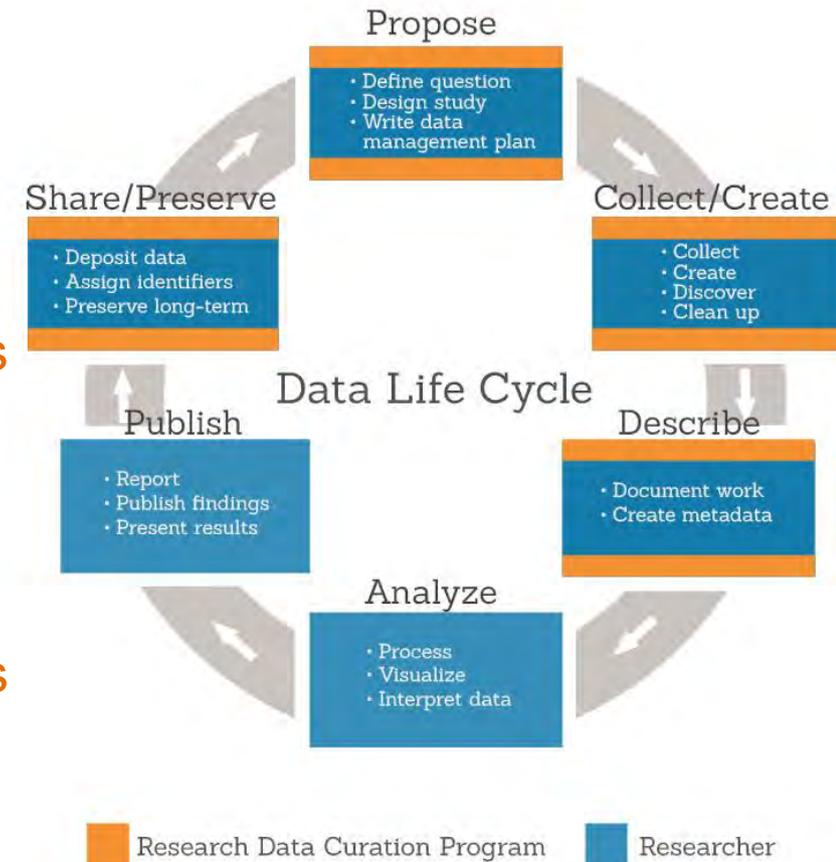
# Data Management Plan (DPM)

Peu d'expérience dans la salle

OPIDor service du CNRS en ligne qui propose des templates

→ Besoin d'accompagnant

→ A quel niveau ? l'échelle semble au niveau des pôles



# La donnée

## Comment définir une donnée ?

Vocabulaire: différentes disciplines  
différents sens,

Les unités

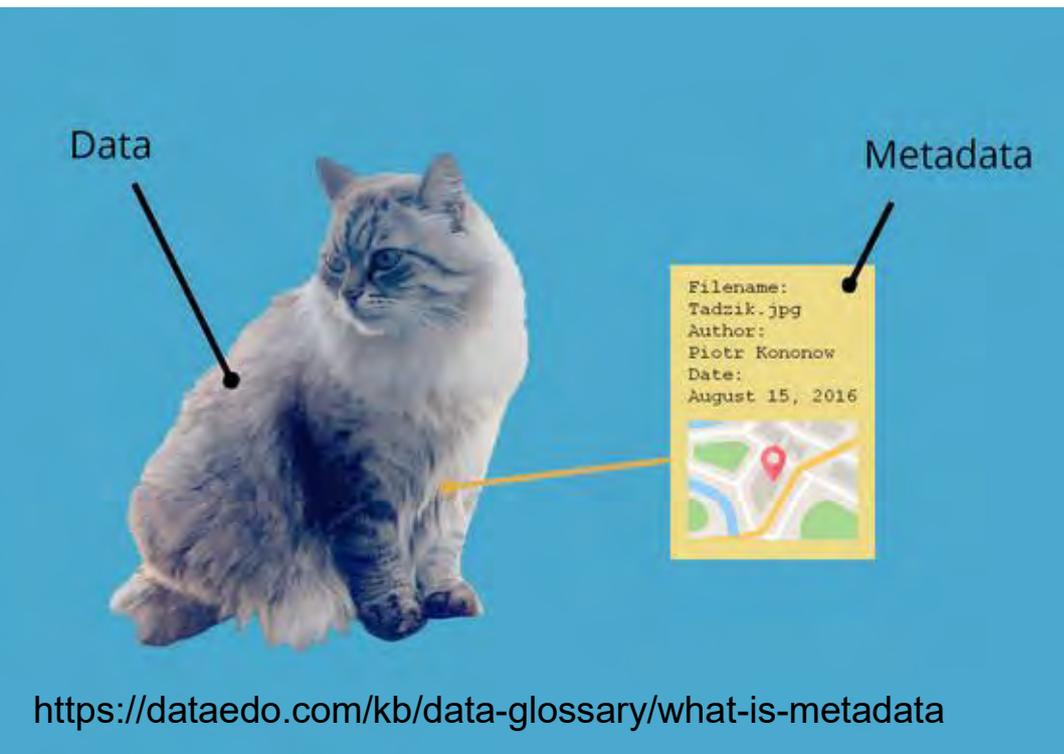
Les capteurs

## Les méta-données

- Données sur les données



# La donnée



Comment définir une donnée ?  
Vocabulaire, capteurs

Les méta-données  
→ Données sur les données



# La donnée

Les données à stocker:

la donnée et ses méta-données dans un seul fichier (recommandation du groupe inter-pôles)

Expérience du spatial: hiérarchisation des données traitées

Discussion sur le « bouchage » des trous de données dans des séries , cela est possible mais doit être considéré comme un produit transformé

Quid des données de pailleasse (hors campagnes, SNO) et modèles → pas d'actualité dans les pôles, questionnement sur quelles données archiver ?

Problèmes des formats de fichiers: il y a des conventions mais multiples selon les communautés



# La donnée

Indispensable de conserver les données bruts  
le traitement de la donnée peut évoluer, avec des re-calibrations

Bien documenter les méta-données: coordonnées, heure UTC, capteurs ...

Aller vers une norme avec une convention des meta-données au niveau de  
l'IR data terra



# La donnée

Les données à stocker:

la donnée et ses méta-données dans un seul fichier (recommandation du groupe inter-pôles)

Expérience du spatial: hiérarchisation des données traitées

Discussion sur le « bouchage » des trous de données dans des séries , cela est possible mais doit être considéré comme un produit transformé

Quid des données de pailleasse (hors campagnes, SNO) et modèles → pas d'actualité dans les pôles, questionnement sur quelles données archiver ?

Problèmes des formats de fichiers: il y a des conventions mais multiples selon les communautés



# La valorisation

L'obligation pour soumettre dans de nombreuses revues (ex AGU) de déposer les données sur un accès libre.

Quelles offres de data repository connaissez-vous ?



# La valorisation

Exemple de Seanoe (données marines)

[www.seanoe.org](http://www.seanoe.org)

**SEANOE** Sea scientific open data public

Search everywhere Search documents

483 Result(s)

Reset filters

**PUBLICATION YEAR**

- 2019 (58)
- 2018 (157)
- 2017 (121)
- 2016 (72)
- 2015 (52)
- 2014 (16)
- 2013 (2)
- 2012 (1)

**DISCIPLINE**

- Administration and dimensions (41)
- Atmosphere (13)
- Biological oceanography (83)
- Chemical oceanography (55)
- Cross-discipline (16)

**Growth parameters data (size and weight) and sexual maturity considering fish, cephalopods and crustaceans of the Gulf of Lions**  
Bensebaini Cyria Meriem, Billet Norbert, Certain Gregoire, Jadaud Angelique  
These data refers to biological data of some of the main commercial marine species of the Gulf of Lions. The biological informations collected refer to growth curves in size and weight and sexual maturity. The species selected are commercial species such as fish (*Dicentrarchus labrax*, *Engraulis encrasicolus*, *Liza ramada*, *Mugil cephalus*, *Lophius budegassa*, *Merluccius merluccius*, *Mullus barbatus barbatus*, *Mullus surmuletus*, *Pagellus erythrinus*, *Raja clavata*, *Sardina pilchardus*, *Scylliorhinus* [...])  
2019-07. Dataset [Open access](#)

**Bathymetric survey on the tropical South Atlantic continental shelf**  
Álvares-Da-Silva Orione   
The bathymetric dataset was accumulated during twenty-two oceanographic research expeditions, with a 51 feet vessel between December 2016 and April 2017. The research expeditions covered a diverse array of transects with a total length of approximately 1,800 km.  
2019. Dataset [Open access](#)

**Genetic, morphometric, and additional data for albacore tuna in the**

# La valorisation

L'obligation de déposer les données sur un entrepôt certifié pour soumettre dans de nombreuses revues (ex AGU) :

→ Des solutions existent pour les jeux de données hors éditeur:  
pour les données marines: SEANOE (doi + licences avec des moratoires)

À généraliser à tous les pôles ?

Quid des codes → changer de revue ?

Question des licences: sujet complexe, non commerciales ou pas, a-t-on le droit de la mettre, en contradiction avec la directive Inspire)

Le législateur n'a encore rien dit sur le type de licences qui doit être utilisées → cela doit être traité au niveau du Ministère (MESRI)



# L'archivage

dark data → chaque collègue a des données ponctuelles dans son micro qui pourraient être mises en base

La perte de données estimée à 60%

→ Trop de données ne sont pas publiées et sont sous-utilisées



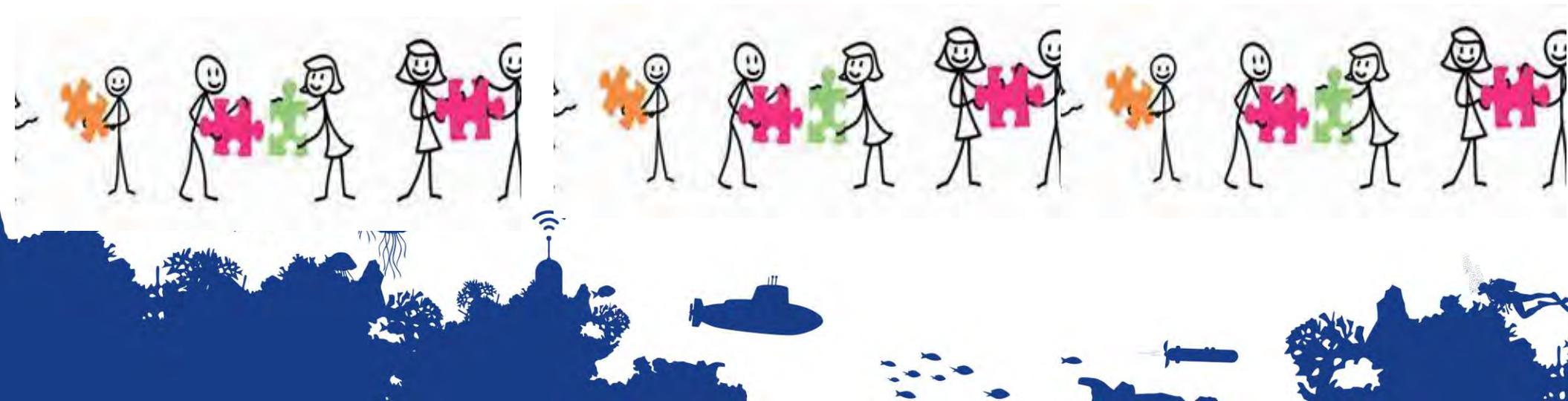
# L'archivage

**Ce qui pourrait inciter les producteurs à mettre en base ?**

lien avec le producteur de données

assurer la protection (propriété intellectuelle)

retour sur l'usage des données



# L'archivage

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)



# L'archivage

Les collègues ont besoin d'aide pour la création de Data Management Plan et la bancarisation

C'est le but de l'IR DATA TERRA qui doit prévoir des moyens et coordonner les compétences pour accompagner les communautés.

Besoin d'assistance locale

→ échelle labo , OSU ? car il y a un besoin de proximité avec les producteurs

mais aussi avoir une masse critique dimensionnée pour créer et maintenir des bases à long terme

