

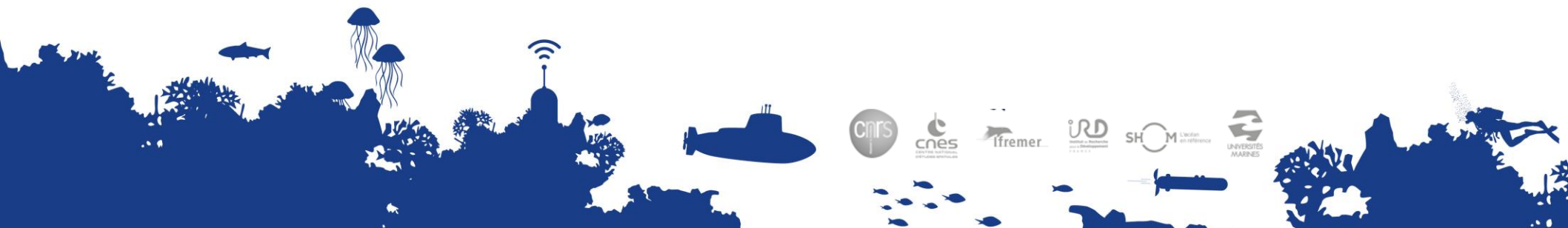
# Cahier des charges des Centres de Données et de Services



# Un cahier des charges : pourquoi?



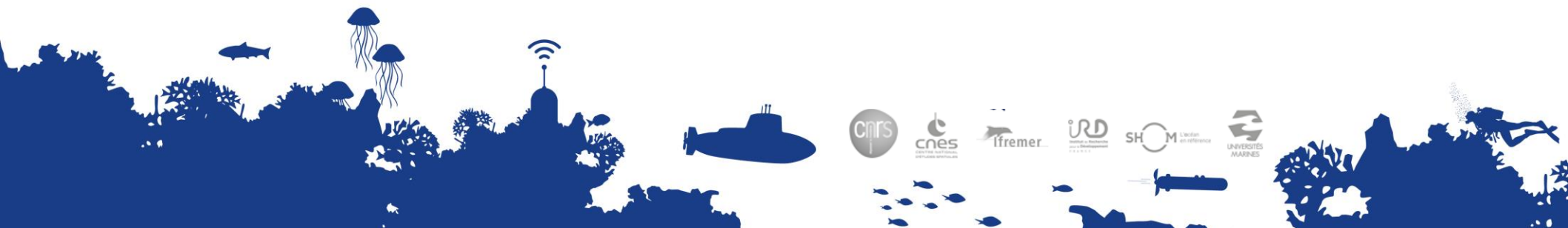
- Définir au sein du pôle Odatis, pour un CDS
  - Les tâches qu'il doit accomplir
  - Les interfaces avec le Pôle
  - ➔ Le CDS est libre de la manière de les mettre en oeuvre
- Un document pour les responsables de CDS
  - Permettant de se projeter sur les fonctions à mettre en place
  - Permettant de justifier les besoins matériels et humains
- Un document « applicable »
  - Rapports d'activité vers le pôle et l'IR
  - ➔ Elaboration des coûts complets
  - ➔ Disponibilité des services
  - ➔ Jeux de données mis à disposition : mise à jour du catalogue du Pôle



# Fonction d'un CDS



- Se base sur le modèle OAIS - Modèle de référence pour un Système ouvert d'archivage d'information  
[http://pin.association-aristote.fr/lib/exe/fetch.php/public/documents/norme\\_oais\\_version\\_francaise.pdf](http://pin.association-aristote.fr/lib/exe/fetch.php/public/documents/norme_oais_version_francaise.pdf)
- Avec des adaptations (un CDS ne fait pas que de l'archivage...)

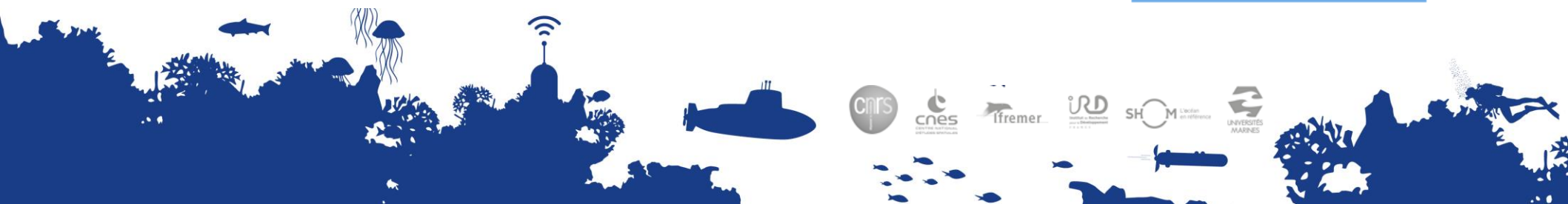


# Fonctions d'un CDS (OAIS)



Les 6 fonctions du modèle OAIS

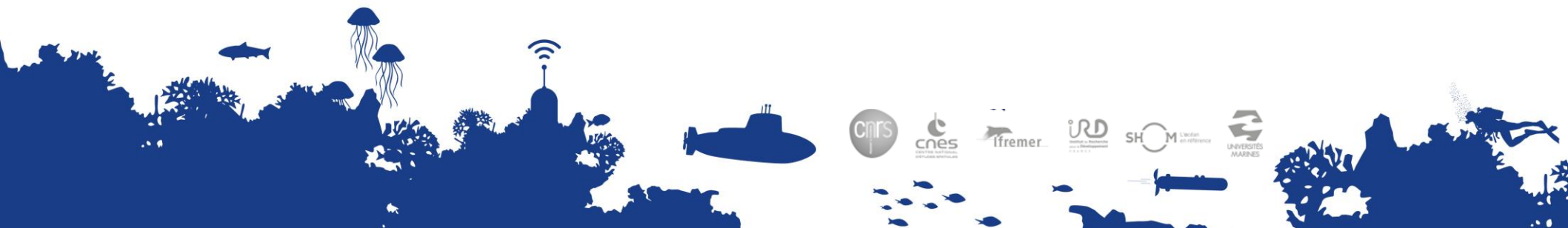
*(Source CINES)*



# Fonction d'un CDS (1/5)



- **La fonction « entrées »** reçoit, contrôle et valide les données à gérer, soit :
- Une harmonisation des données du même type provenant de différentes équipes et/ou de différents systèmes d'observation : utilisation de référentiels communs, d'une même codification de variables et paramètres, d'une structuration homogène.
- Un contrôle qualité des données reçues, quand cela est préconisé par les programmes d'observation et de gestion de données pour ce type ... A minima, un contrôle de complétude

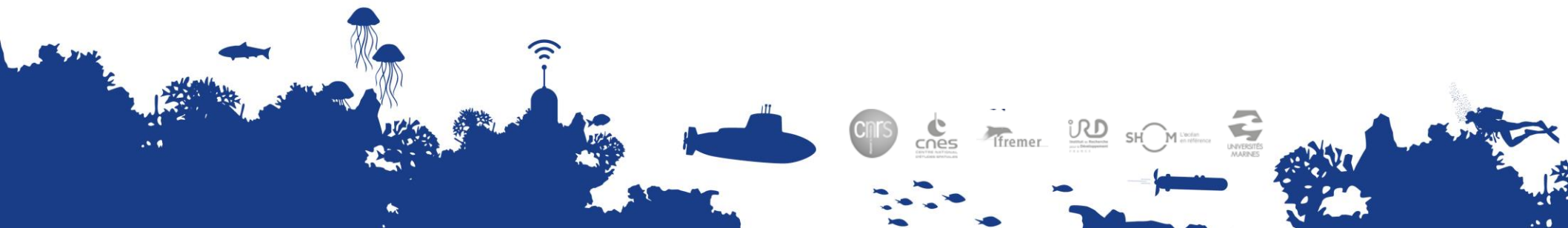




# Fonction d'un CDS (2/5)



- **La fonction « stockage »** assure la conservation physique des données gérées. Il s'agit au sens du Pôle de la copie primaire, dite de référence, des données. Elle s'appuie sur la fonction de pérennisation pour en assurer la préservation et la sécurité.
- **La fonction « planification de la pérennisation »**, assure la pérennisation physique des informations (réalisation de copies multiples, renouvellement des supports anciens, transition vers de nouvelles génération de supports/lecteurs informatiques ...) soit en mettant en œuvre ses moyens propres, soit en déléguant, conventionnant et supervisant cette tâche auprès d' un service externe



# Fonction d'un CDS (3/5)



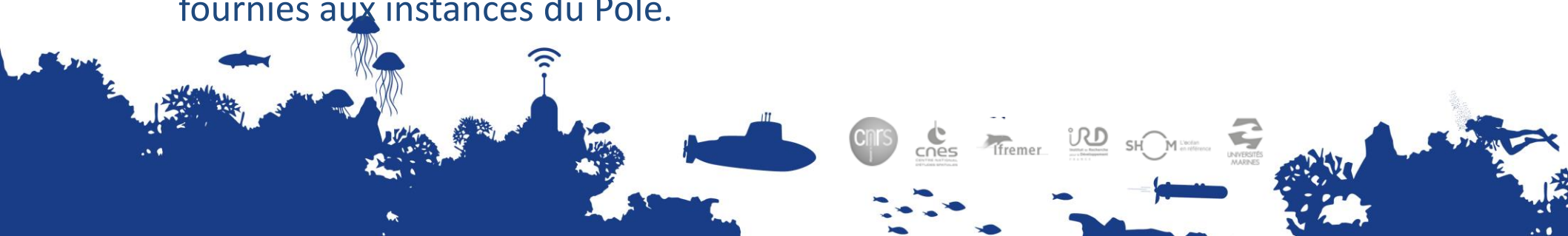
- La fonction « gestion des données » : tenue à jour des informations nécessaires à la gestion et à l'utilisation ultérieure des données:
  - La description des données gérées, en support aux fournisseurs de données, conformément à la définition du catalogue du pôle (métadonnées à la norme ISO 19115 par exemple, utilisation des vocabulaires communs définis). Ces descriptions doivent notamment comprendre : l'origine de la donnée (fournisseur individuel ou système d'observation), les conditions techniques et organisationnelles d'observation, la liste des variables observées, les conditions et limites d'utilisation des données (licences d'utilisation, domaines d'applicabilité, ...).
  - La description des résultats des chaînes de traitement conformément à la définition du catalogue du pôle.
  - La gestion des droits d'accès et des conditions d'accès aux données, conformément à la réglementation en vigueur et en accord avec les producteurs de données : période d'exclusivité à l'équipe d'observation, ...



# Fonction d'un CDS (4/5)



- **La fonction « administration »** assure la coordination générale du système. Elle veille à la qualité globale du service rendu et à son amélioration. Elle rend compte au management et au Pôle : évolution des données gérées, interruptions de services éventuelles, ...
  - Mise en œuvre des évolutions demandées par le Pôle.
  - Les moyens techniques informatiques, soit en propre, soit par délégation, conventionnement ou contractualisation, conformément aux exigences de service,
  - Le maintien en conditions opérationnelles, le maintien en configuration (tests, gestion des versions, ...), les passages en exploitation des logiciels utilisés par le CDS : outil de formatage et de contrôle des données, chaînes de traitement, ...
  - Production des statistiques d'activité (production, utilisation) qui seront fournies aux instances du Pôle.

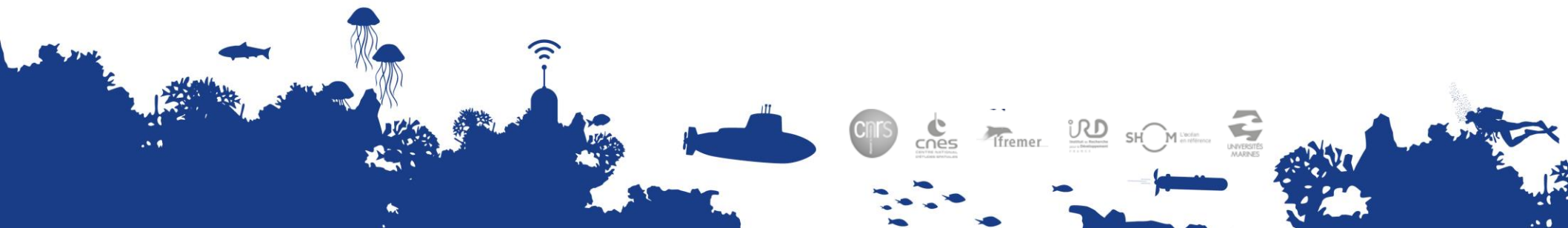




# Fonction d'un CDS (5/5)



- **La fonction « accès »** regroupe tous les services qui sont en interface avec le Pôle soit directement avec les utilisateurs, soit par délégation auprès d'une infrastructure externe partagée. Outre les fonctions de contrôle d'accès, il s'agit d'opérer en continu, avec un taux de disponibilité opérationnel, les services d'accès et/ou de téléchargement des données gérées.



# Les conditions d'exercice

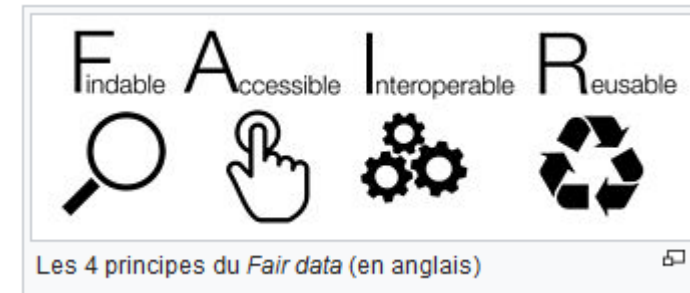


- Un contexte général à respecter

- FAIR Data principes

[https://fr.wikipedia.org/wiki/Fair\\_data](https://fr.wikipedia.org/wiki/Fair_data)

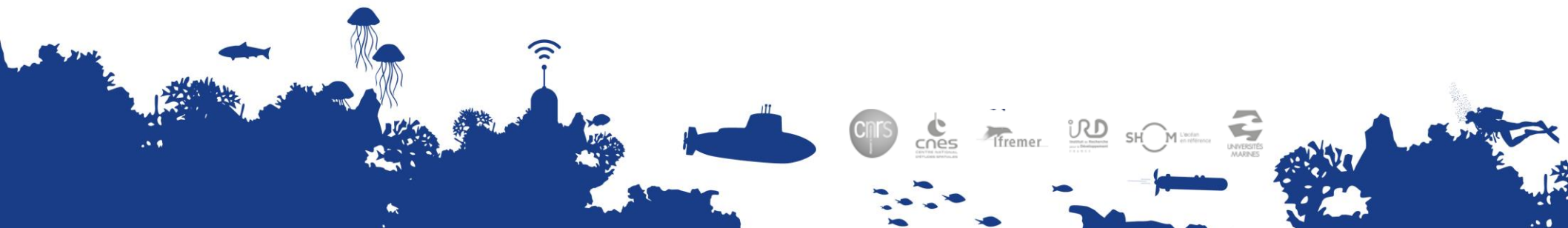
- Aspects réglementaires : droits des données  
Open Data, EtatLab, Creative Commons, ...



- Tâches à mener en continu, en routine  
selon un périodicité appropriée aux données gérées

➔ Contrat d'« Engagement de service » (Service Level Agreement) avec des clauses dérivées de celles édictées par la certification « Core Trust Seal for Trustworthy Data Repositories » de la Research Data Alliance (cf. <https://www.coretrustseal.org/why-certification/requirements> )

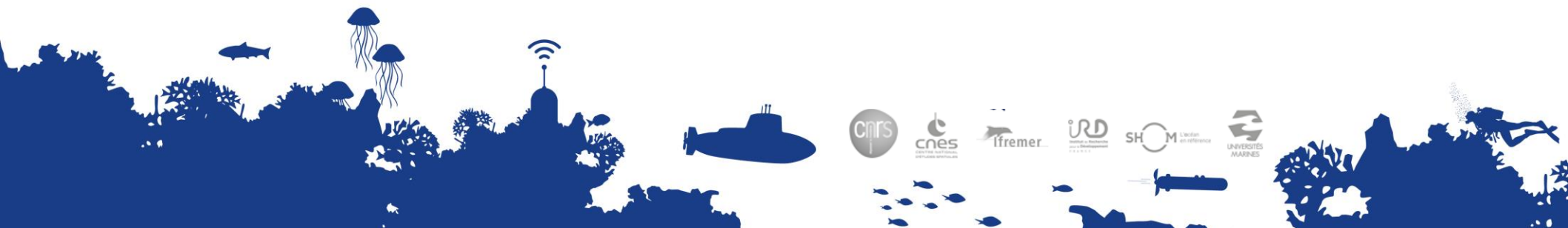
- Mandat, Sécurité, Pérennisation, Disponibilité, ...



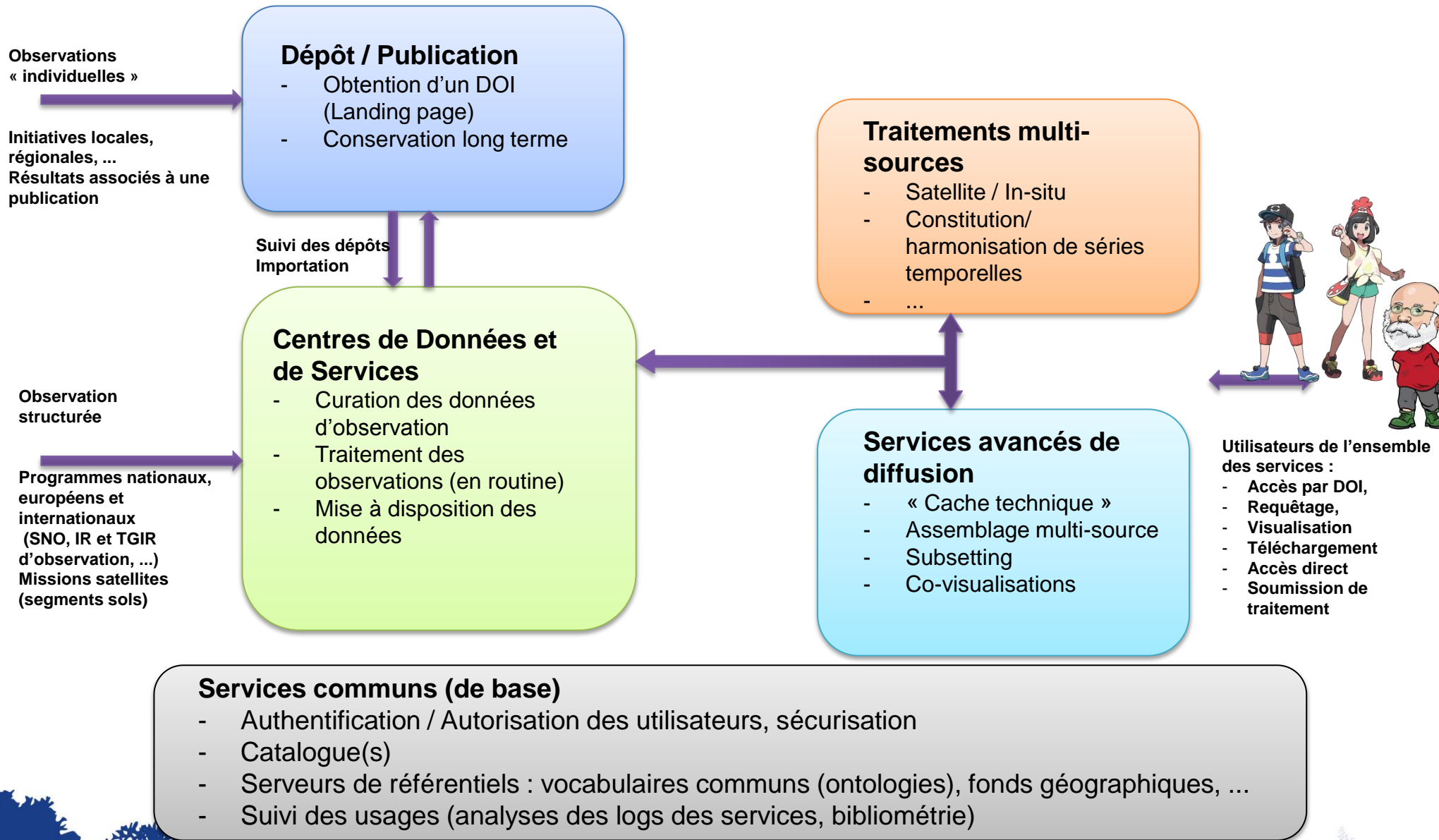
# Des contraintes fortes



- **La disponibilité des services en ligne est un des points critiques (objectif 7j/7, 24/24)**
  - ➔ Idée : soulager les CDS des tâches opérationnelles (donc extrêmement coûteuses)
  - ➔ Idée : Fédérer les données (« cache technique ») en avant des services de diffusion, mise en commun des services de diffusion, ...



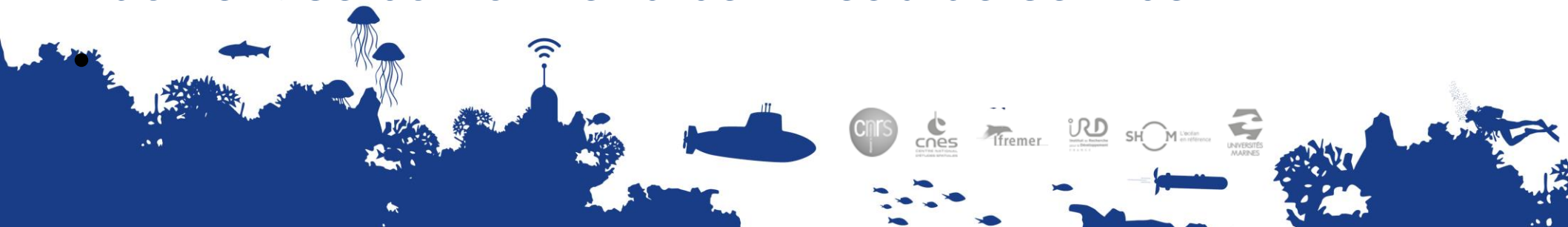
# Evolution vers des services hiérarchisés



# Vers deux types de CDS?



- Les **Centres d'Assemblage et de Traitement (CDS-CAT)**, consacrés aux tâches de gestion de données et de génération de produits , mais déléguant la diffusion des données,
- Les **Centres d'Assemblage, de Traitement et de Diffusion (CDS-CATD)**, qui outre les tâches de gestion de données, propose également des services opérationnels de diffusion. Les Centres d'Assemblage gérant des données de volumes importants, pouvant difficilement être répliquées par le réseau internet, doivent se conformer à ce niveau de service.





# Interfaces CDS / Pôles

## « Accès aux données »



- Pour les **Centres d'Assemblage et de Traitement** (CDS-CAT), réplication des données vers le cache technique, avec un niveau de disponibilité de type « Best Effort » :  
ex :  
Moissonnage des métadonnées + service de type ErrDap, interrogé une fois par jour ouvrable
- Pour les **Centres d'Assemblage, de Traitement et de Diffusion** (CDS-CATD), mise en œuvre de services en ligne avec taux de disponibilité exigé (98%, 7j/7, 24/24)  
Téléchargement, subsetting, visualisation, ...

