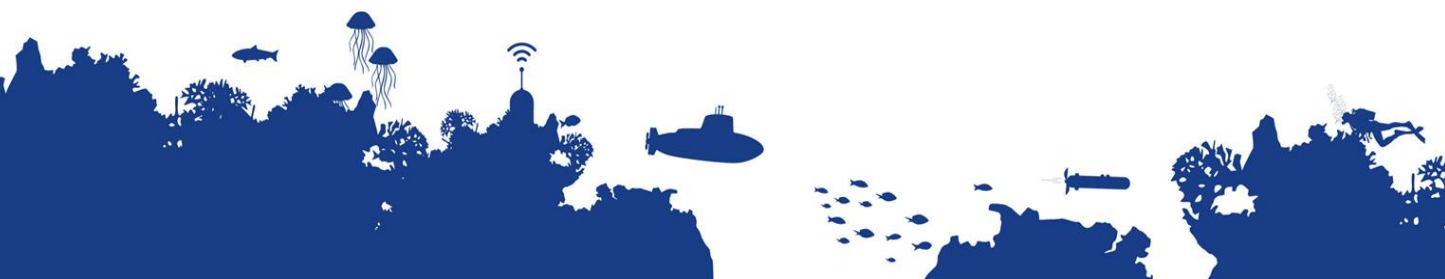


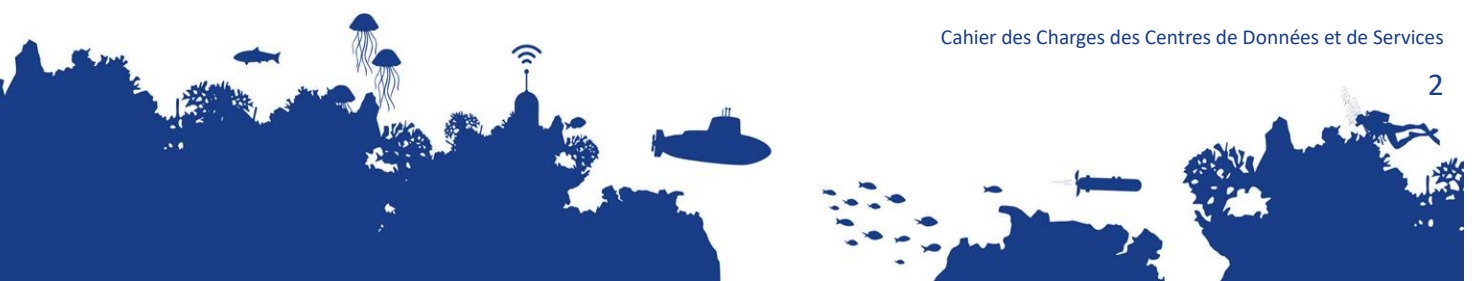


**Cahier des charges
des Centres de Données
du Pôle de Données Océan – Odatis
de l'Infrastructure de Recherche Data Terra**

Version 3.0 du 6 avril 2020



Numéro du livrable	Titre court
	Cahier des charges CDS
Titre long	
	Cahier des charges des Centres de Données du Pôle Océan-Odatis de l'IR Data Terra
Description courte	
	Spécifications des fonctions, des critères de bon fonctionnement, des principes de gestion de données et des interfaces d'un Centre de Données du Pôle de Données Océan – Odatis.
Auteur	Groupe de travail
Gilbert Maudire, Sabine Schmidt, Joël Sudre	
Dissémination	Copyright
Tous Centres de Données du Pôle	Pôle ODATIS



Historique

Version	Auteurs	Date	Commentaires
1.0	Gilbert Maudire	12/11/2019	Version initiale
1.1	Cécile Nys	12/11/2019	Relecture et corrections & Ajout table des abréviations
2.0	Gilbert Maudire	12/03/2020	Ajout des types de centres de données (hiérarchisation)
2.1	Gilbert Maudire	25/03/2020	Corrections après relecture par CNRS/INSU (Bruno Blanke)
3.0	Gilbert Maudire, Sabine Schmidt, Joël Sudre	05/04/2020	Version applicable pour le lancement du projet Copilote

Table des matières

Table des abréviations.....	5
1) Introduction	6
2) Spécifications des fonctions d'un Centre de Données.....	8
3) Fonctions à assumer.....	11
4) Critères de bon fonctionnement.....	12
5) Principes de gestion de données : FAIR Data.....	13
6) Spécifications des interfaces d'un CDS avec le Pôle Odatis.....	15
7) Spécifications des interfaces d'un Centre de Collecte.....	17
8) Annexe 1 : Préconisations techniques	18
9) Annexe 2 : Conditions et Licences d'accès aux Données	21
10) Annexe 3 : Contexte général des CDS au sein de l'IR Data Terra	22

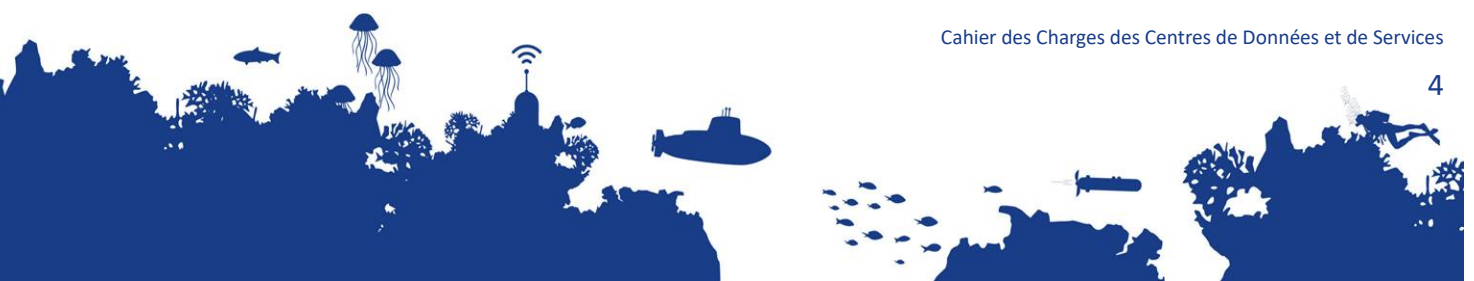
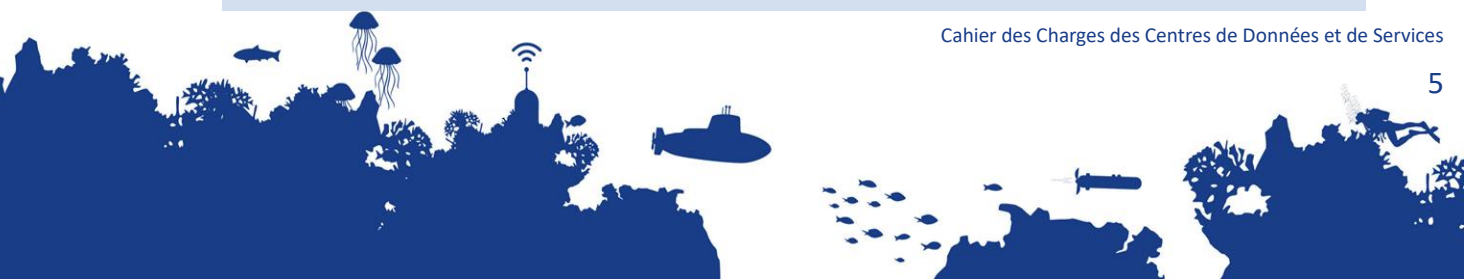


Table des abréviations

Abréviations	Signification
CD	Centre de Données – Data Centre, infrastructure de gestion de données, composante du Pôle Odatis
CDS	Centre(s) de Données et Services – Data and Service Centre, d'échelle nationale ou internationale proposant des services de production régulière, de diffusion et, éventuellement, de traitement de données à la demande
Centre de Collecte	Centre de Données assurant principalement l'interface avec les producteurs de données – Assembling Centre
COI	Commission Océanographique Intergouvernementale
CTS(A)	Certification « Core Trust Seal (of Approval) de RDA
EDMO	European Directory of Marine Organizations
EOV	Essential Ocean Variables
FAIR	Findable Accessible Interoperable Reusable
IODE	International Oceanographic Data Exchange (Programme de la COI)
IR	Infrastructure de Recherche
OAIS	Open Archival Information System
OSU	Observatoire des Sciences de l'Univers (CNRS)
RDA	Research Data Alliance
SNO	Service National d'Observation
WORMS	World Ocean Registry of Marine Species



1) Introduction

Ce document spécifie les fonctions qu'un Centre de Données (CD) doit assumer afin d'être éligible à une intégration au sein du Pôle Odatis, le pôle de données et de services Océan de l'Infrastructure de Recherche Data Terra (Pôle de données et services pour le système Terre).

Ce document doit donc être considéré comme les clauses d'un « Engagement de Service » (Service Level Agreement), tel que décrit par la norme ISO20000.

Cet engagement définit en particulier le périmètre thématique des données à gérer telles que :

- Données provenant d'un ou de plusieurs systèmes d'observation labellisés par les partenaires (par exemple, données observées par un ou plusieurs SNO ou un sous-ensemble de celles-ci) ;
- Données observées par une organisation nationale ou supranationale (Infrastructure de Recherche nationale ou européenne, composante nationale d'un programme d'observation international).

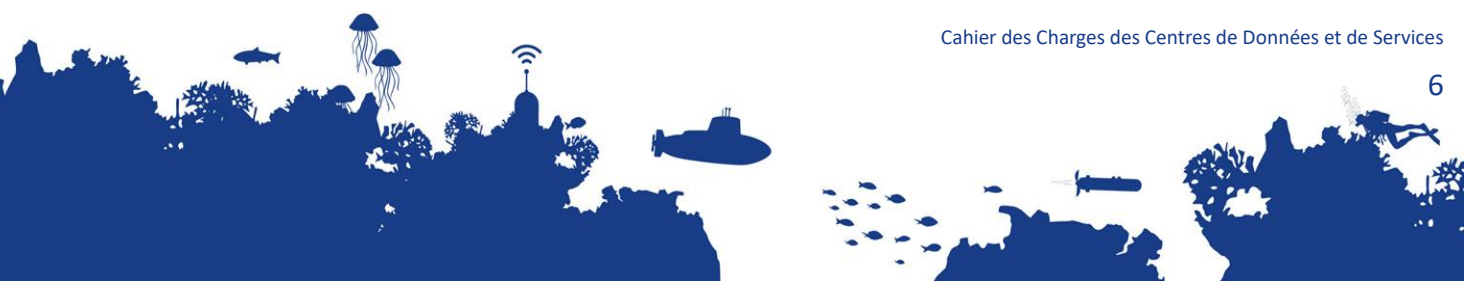
Un CD assemble, harmonise, maintient et rend accessible le jeu de données pour le périmètre dont il a la charge.

Les fonctions à assurer sont des fonctions opérationnelles, devant être conduites dans la durée et dans le cadre d'un mandat du ou des organisme(s) partenaire(s) qui s'engage(nt) à en garantir les moyens humains, financiers et techniques adaptés à ces fonctions.

Tous les Centres de Données n'assumeront pas la totalité des fonctions. Aussi, seront distingués **deux types de centres** :

- Les **Centres de Collecte (« Assembling Centres »)**, qui géreront principalement l'interface de proximité avec les équipes d'observation productrices de données, au sein des OSUs par exemple,
- Les **Centres de Données et de Services (« Scientific Data Centres »)**, assumant l'intégralité des fonctions, et en particulier, l'organisation de la pérennisation du jeu dont ils ont la charge et qui est considérée comme la copie de référence, les services de diffusion et de traitement vers l'ensemble des utilisateurs dans le cadre d'une exploitation certifiée.

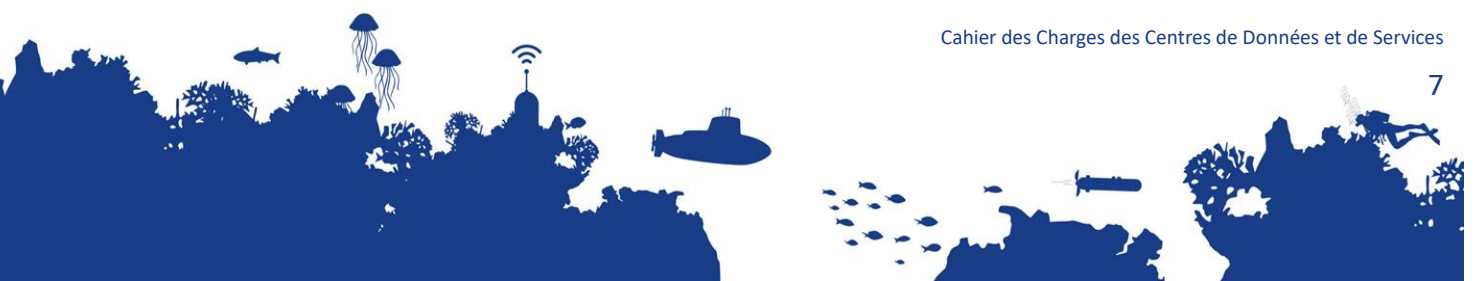
Ces fonctions peuvent être assurées suivant différentes approches organisationnelles et techniques, selon les types de données et les modes d'exercice des organismes partenaires hôtes du CD. Ainsi, sur le plan organisationnel, les moyens mis en œuvre peuvent être internes aux organismes ou externalisés, mais, dans ce cas, conventionnés ou contractualisés selon un mode compatible avec les objectifs spécifiés. Un CD peut également être opéré conjointement par plusieurs organismes partenaires du Pôle, dans le cadre d'une convention.



Sur le plan technique, les données peuvent être gérées en bases de données relationnelles, en fichiers ou selon tout autre dispositif adapté.

En conséquence, ne sont spécifiées dans ce document que :

- La typologie des fonctions (mais pas leur implémentation),
- Les critères de bon fonctionnement,
- Les principes de gestion de données à respecter (principes FAIR),
- La normalisation des interfaces vis-à-vis du Pôle.



2) Spécifications des fonctions d'un Centre de Données

Les fonctions demandées à un Centre de Données s'inspirent du modèle « Open Archival Information System » (OAIS) qui cartographie un système d'information selon 6 grands domaines fonctionnels. Il s'agit d'un modèle conceptuel destiné à la gestion, à l'archivage et à la préservation à long terme de documents numériques. Ce modèle OAIS est enregistré comme norme ISO sous la référence 14721:2012.

Le modèle OAIS constitue une référence décrivant dans les grandes lignes les fonctions, les responsabilités et l'organisation d'un système qui préserve de l'information, en particulier des données numériques, sur le long terme, pour en garantir l'accès à une communauté d'utilisateurs. Le long terme est défini comme suffisamment long pour être soumis à l'impact des évolutions technologiques (évolution des supports informatiques notamment).

Il s'agit d'un modèle conceptuel, laissant libre le choix de l'implémentation des fonctions (schéma source CINES) :



Les 6 fonctions du modèle OAIS

La fonction « **ENTREES** » reçoit, contrôle et valide les données à gérer. Il s'agit principalement d'assurer :

- Une **harmonisation des données du même type** provenant de différentes équipes 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**, selon les préconisations des programmes d'observation et de gestion de données pour ce type : programmes internationaux de type

ARGO, COI/IODE... A minima, un contrôle de complétude des métadonnées et un contrôle de la localisation spatiale et temporelle des données sont nécessaires.

Cette fonction organise la collaboration avec les équipes scientifiques qui fournissent les données gérées : définition et conventionnement avec les activités/systemes d'observation, retours éventuels sur les anomalies détectées...

Selon les types de données, la fonction « **ENTREES** » peut être accolée à une fonction de « **PRODUCTION** » :

- La mise en œuvre, sur une base opérationnelle et en accord avec la périodicité exigée, des **chaînes de traitement** nécessaires à la génération du niveau de données attendu ;
- Le **contrôle de cohérence des résultats** obtenus, selon les règles de production ;
- La **description des résultats** (production des métadonnées) conformément à la fonction « **GESTION DES DONNÉES** » ci-après.

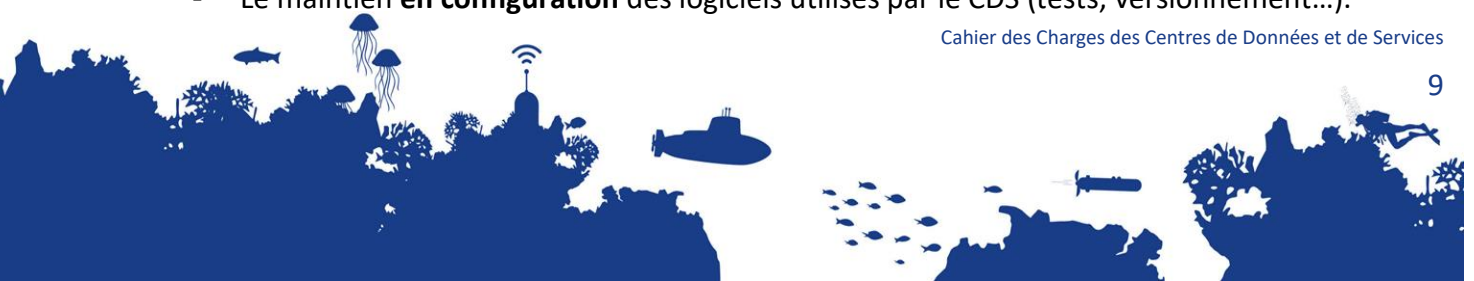
La fonction « **STOCKAGE** » assure la **conservation physique des données gérées**. Il s'agit, au sens du Pôle Odatis, de la copie primaire, dite de référence, des données. Elle s'appuie sur la fonction de pérennisation pour en assurer la préservation à long terme et la sécurité.

La fonction « **GESTION DES DONNÉES** » prend en charge la **tenue à jour de toutes les informations** nécessaires à la gestion et à l'utilisation ultérieure des données, selon les principes FAIR (cf. §3). À ce titre, elle assure notamment :

- La **description des données gérées**, en support aux fournisseurs de données, conformément à la définition du catalogue du pôle ;
- 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.

La fonction « **ADMINISTRATION** » assure la **coordination générale** du Centre de Données. Elle planifie et met en œuvre les règles d'exploitation. 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... Elle instruit et, après accord des partenaires tutelles dans le cadre du Comité Directeur, s'assure de la mise en œuvre des évolutions demandées par le Pôle. Elle assure la **gestion des outils** nécessaires au bon fonctionnement du CDS et notamment :

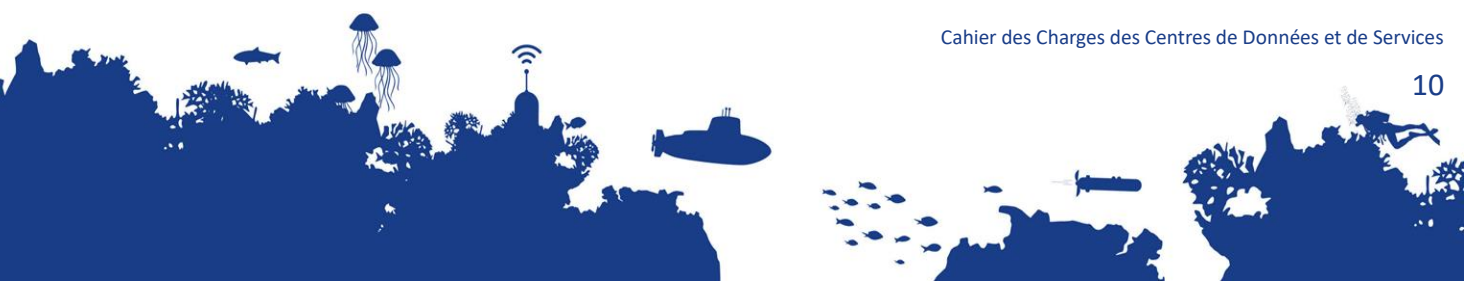
- Les **moyens techniques informatiques**, soit en propre, soit par délégation, conventionnement ou contractualisation, conformément aux exigences de service ;
- Le maintien en condition opérationnelle ;
- Les **passages en exploitation** des nouvelles versions des logiciels : outil de formatage et de contrôle des données, chaînes de traitement... ;
- Le maintien en **configuration** des logiciels utilisés par le CDS (tests, versionnement...).



La fonction « **PLANIFICATION de la PÉRENNISATION** », conformément aux règles établies par la Research Data Alliance (RDA), et en particulier par les dispositions techniques prévues par la certification « Core Trust Seal of Approval », s'assure de la **pérennisation physique des informations** (réalisation de copies multiples, renouvellement des supports anciens, transition vers de nouvelles générations 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. La **pérennisation** concerne :

- Les **données et produits non reproductibles**, dont le Centre de Données et de Service est gestionnaire de la copie primaire ;
- Les **métadonnées** décrivant les données et produits ;
- Les **descriptions** des formats, des vocabulaires... utilisés pour encoder les données quand ceux-ci ne sont pas standardisés par une entité externe pérenne ;
- Les **logiciels** permettant la lecture des données.

La fonction « **ACCÈS** » regroupe tous les services qui sont en interface avec les utilisateurs : autres composantes du Pôle, utilisateurs internes ou externes... Outre les fonctions de **contrôle d'accès et d'établissement des statistiques d'utilisation**, il s'agit de mettre en place les **services de téléchargement ou d'accès direct aux données gérées**.

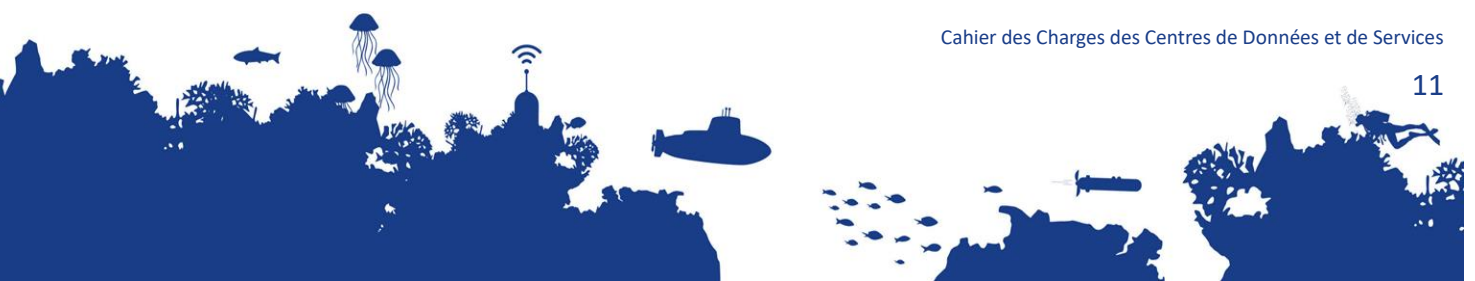


3) Fonctions à assumer

Parmi les tâches décrites au chapitre 2 :

- Les **Centres de Collecte** (« Assembling Centre ») doivent assumer a minima les fonctions « ENTRÉE » et « GESTION DES DONNÉES ». Les fonctions de « PRODUCTION », « STOCKAGE », « PLANIFICATION de la PÉRENNISATION » et «ACCÈS» seront déléguées. Un centre de collecte aura donc un rôle privilégiant les relations avec les producteurs de données et les systèmes d'observations locaux et régionaux, afin d'assurer une prise en compte régulière de ces observations par le Pôle Odatis, en termes de description (métadonnées), de complétude et d'exhaustivité pour le périmètre dont il a la charge. Leurs obligations en matière de diffusion se limiteront aux retours vers leurs propres utilisateurs et aux transferts des données vers le Centre de Données et de Services adapté qui intégrera ces contributions à l'échelle nationale et en assurera une large diffusion.
- Les **Centres de Données et de Services** (« Scientific Data Centre ») doivent assumer l'ensemble des fonctions décrites au chapitre 2, dans le respect des critères de bon fonctionnement décrits au chapitre 4 et les principes de gestion des données décrits au chapitre 5. Les Centres de Données et de Services ont la charge d'agréger, avec le support éventuel de Centres de Collecte, et d'organiser la pérennité de **la copie de référence de niveau national pour les données dont ils ont la charge.**

De plus, ils devront s'inscrire dans une **démarche de certification des services** proposés. Aussi, ils doivent pouvoir s'appuyer sur une infrastructure informatique adaptée (sécurité de fonctionnement, disponibilité...) opérée par une équipe dimensionnée pour assumer ces contraintes.



4) Critères de bon fonctionnement

Un Centre de Données et de Services doit pouvoir s'appuyer sur un ensemble d'acteurs, parties prenantes de son bon fonctionnement :

- Administrateurs du CDS (représentants des directions des organismes porteurs du CDS),
- Producteurs des données gérées (équipes scientifiques, services d'observation...),
- Gestionnaires des données (opération des systèmes d'information),
- Infrastructures informatiques,
- Utilisateurs de la donnée.

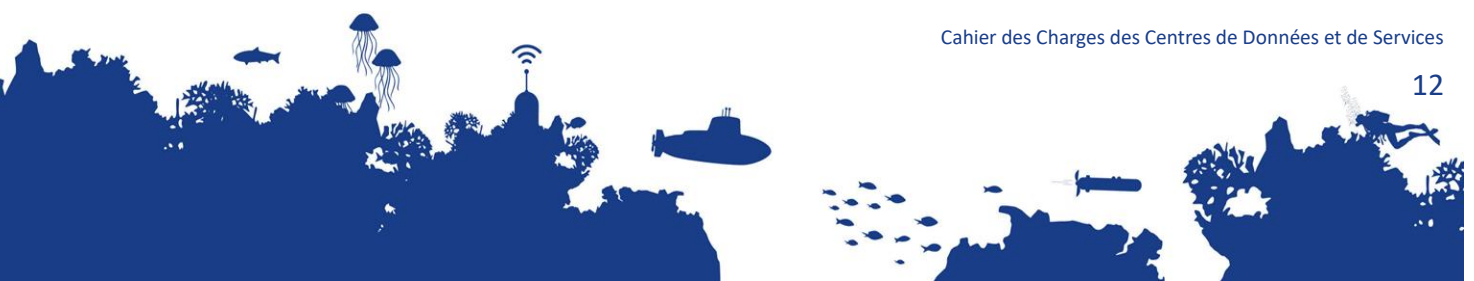
Dans ce contexte, afin de s'assurer de la coordination des acteurs et du bon fonctionnement d'ensemble, doivent être suivis les points suivants :

- Mission et périmètre des données gérées : mandat initial et évolutions éventuelles ;
- Méthode de gestion et de préservation des données : intégrité et authenticité des données, procédures de stockage et plan de préservation ... ;
- Qualité de la donnée ;
- Conditions d'accès : licences, identification des jeux de données (DOI), découverte (méta-données)... en application des principes FAIR (cf. §4) ;
- Qualité des services d'accès : continuité de service, disponibilité de service... ;
- Infrastructures informatiques nécessaires.

Le niveau de prise en compte par le Centre de Données et de Services de ces points est évalué par la certification « Core Trust Seal », portée en particulier par la Research Data Alliance (cf. <https://www.coretrustseal.org/why-certification/requirements/>). Aussi, est-il demandé à chaque CDS de :

- 1) S'évaluer par rapport aux critères de la certification Core Trust Seal ;
- 2) Soumettre un dossier de certification quand est atteint le niveau 3 « Critère en phase d'implémentation » pour l'ensemble des critères.

Ces critères de bon fonctionnement s'appliquent également aux Centres de Collecte avec, toutefois, un mode d'application allégé : en particulier, n'offrant pas de services à un large public, l'évaluation par rapport aux critères de la certification « Core Trust Seal » n'est pas obligatoire.



5) Principes de gestion de données : FAIR Data

Dans le respect des dispositions légales régissant l'accessibilité aux données de la recherche (Open Data) et des sciences ouvertes (Open Science), les principes de FAIR data recouvrent les manières de construire, stocker, présenter ou publier des données de manière à permettre que la donnée soit «trouvable, accessible, interopérable et réutilisable» (en anglais : Findable, Accessible, Interoperable, Reusable, d'où l'acronyme «FAIR»). L'ensemble des fonctions détaillées au §2 ci-devant, en particulier les fonctions « ENTREES », « GESTION DES DONNÉES » et « ACCES » doivent en conséquence implémenter les principes FAIR.

Dans le contexte du Pôle Odatis, ces principes seront déclinés suivant les règles suivantes :

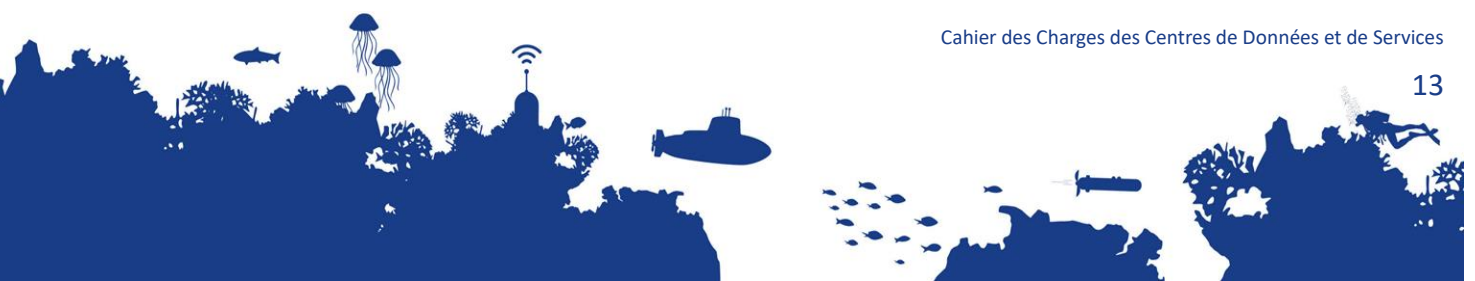
- **Findable** ("Trouvable") : Tout jeu de données géré doit être décrit dans le catalogue Odatis, en utilisant la famille de normes ISO 19115. Ces métadonnées sont des métadonnées de découverte, c'est-à-dire décrivant des jeux de données homogènes (des "familles" de données) et non les observations individuelles. Si les jeux de données sont décrits dans un catalogue local propre au Centre de Données, ce catalogue local doit pouvoir être moissonné pour mise à jour régulière du catalogue Odatis, en utilisant le protocole OGC-CSW (Catalogue Service for the Web).

Les descriptions (métadonnées) doivent notamment comprendre :

- L'origine de la donnée (fournisseur individuel ou système d'observation) ;
- Les conditions techniques et organisationnelles d'observation ;
- Les emprises géographiques et temporelles du jeu de données ;
- La liste des variables observées. Cette liste sera établie en utilisant le vocabulaire « Variables Océaniques Essentielles » (EOV – cf. Annexe 1) ;
- Les conditions et limites d'utilisation des données (licences d'utilisation, domaines d'applicabilité...) (cf. « Accessible » ci-dessous).

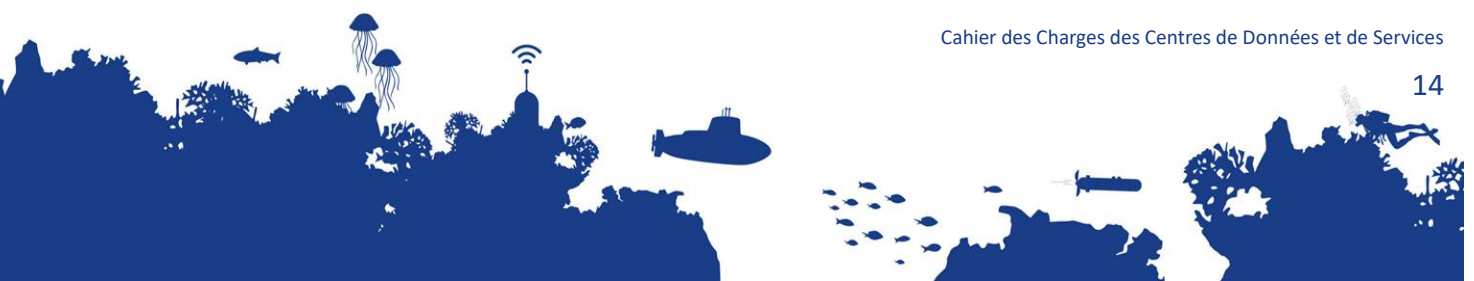
- **Accessible** (« Accessible ») : Tout jeu de données géré doit être rendu accessible en ligne via un ou des service(s) de téléchargement adapté(s) à la nature et au volume des données (cf. Annexe 1 – Préconisations techniques). Ce service doit être décrit dans les métadonnées. Le format des données fournies doit être compatible avec un des formats préconisés par Odatis (cf. Annexe 1 – Préconisations techniques).

Afin d'harmoniser les conditions d'accès aux jeux de données gérés, l'adoption d'une licence commune est préconisée (cf. Annexe 2 – Conditions et Licences d'accès aux Données).



- **Interopérable** (« Interopérable ») : Afin de proposer aux utilisateurs du Pôle Odatis des interfaces d'accès homogènes aux différents jeux de données, les services de téléchargement ou d'accès aux données doivent respecter les préconisations techniques précisées en annexe 2. Sont en particulier normalisés :
 - Les vocabulaires utilisés pour décrire les jeux de données (métadonnées de découverte) et les données elles-mêmes (fichiers de données), afin de garantir l'interopérabilité sémantique entre sources différentes ;
 - Les formats de fichiers transmis aux utilisateurs ;
 - Les protocoles de visualisations des plans de position et des données ;
 - Les protocoles des services de téléchargement ou d'accès aux données.

- **Reusable** ("Réutilisables") : Des identificateurs permanents (DOI – Digital Object Identifiers) sont assignés aux données. Pour ce faire, une politique de versionnement adaptée à la nature et au volume des données gérées sera mise en œuvre, telle que :
 - Agrégation du jeu de données complet pour des données n'évoluant pas ou plus ;
 - Génération d'images stables (« snapshots ») selon des intervalles temporels à définir pour les données évolutives (séries temporelles actives par exemple).



6) Spécifications des interfaces d'un CDS avec le Pôle Odatis

Chaque Centre de Données et de Services doit **mettre en place les interfaces ci-dessous** avec le Pôle de Données Odatis. Ces interfaces sont de nature organisationnelle ou technique :

Interfaces organisationnelles :

Chaque CDS doit être en mesure d'élaborer les documents suivants :

- **Description du contexte d'activité** : mandat, organisme(s) de tutelle, périmètre des données gérées, points de contact (responsables), résumé des dispositions de mise en œuvre. Ce document sera écrit à la création du CDS et maintenu à jour.
- **Descriptif des coûts complets** engendrés par son activité, au format exigé par Ministère de l'Enseignement supérieur de la Recherche et de l'Innovation (MESRI). Ce descriptif est à établir annuellement, sauf instruction contraire, et porte sur l'exercice budgétaire précédent (l'année calendaire précédente).
- **Rapport d'auto-évaluation de conformité** par rapport aux critères de la certification Core Trust Seal, évaluant la qualité et la robustesse des procédures mises en œuvre et des services proposés. Ce rapport sera mis à jour tous les deux ans.
- **Rapport d'activité** résumant l'évolution des jeux de données gérés (volumétrie, procédures de gestion...) et l'évolution de l'utilisation des données et services (nombre d'accès par exemple). Ce rapport sera établi annuellement.

Interfaces techniques :

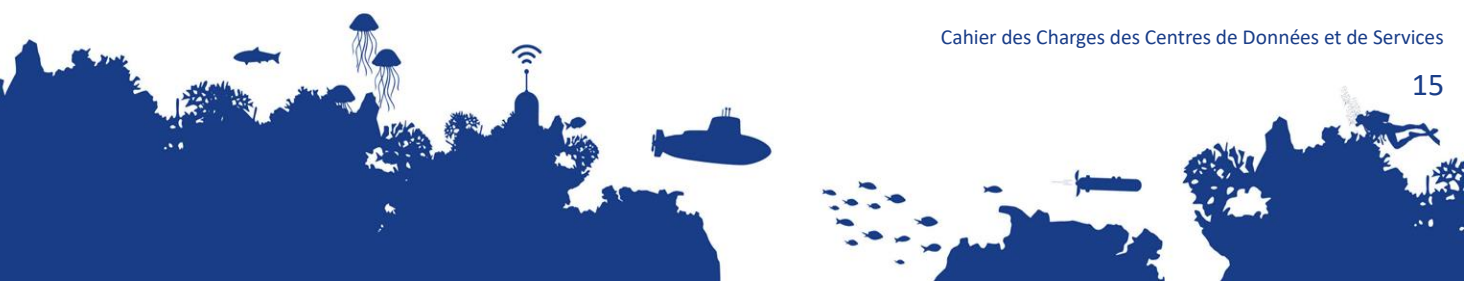
Chaque CDS doit se conformer aux **recommandations techniques spécifiées à l'annexe 1** : Préconisations techniques.

Ces recommandations ont pour objectif d'harmoniser l'interface avec les utilisateurs et donc de faciliter l'usage intégré de l'ensemble des données et services du Pôle. Elles concernent essentiellement :

- Les **descriptifs de données** : métadonnées et vocabulaires communs associés ;
- Les **formats de données proposés** aux utilisateurs et les vocabulaires associés ;
- Les **services de données** et les protocoles associés ;
- **L'assignation de DOI.**

Ces recommandations sont évolutives, car sujettes à l'évolution de la réglementation, des recommandations des programmes internationaux et européens d'observation et de gestion des données, de l'état de l'art informatique, des données gérées et de leur volumétrie et des besoins des utilisateurs.

Deux instances permettent la concertation et l'établissement de préconisations :

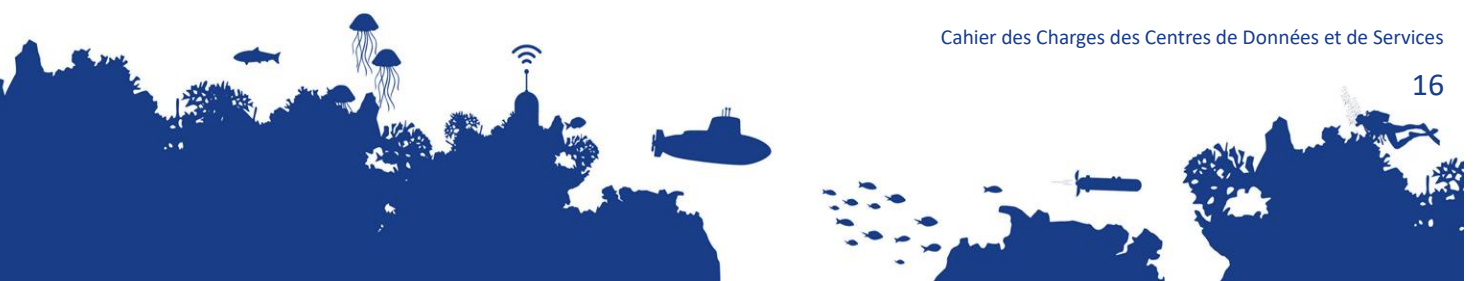


- Le Groupe Technique InterPôle de l'Infrastructure de Recherche Data Terra qui étudie les recommandations s'appliquant à l'ensemble des Pôles de données (cf. <https://www7.obs-mip.fr/interpole/>)
- L'Atelier Technique du Pôle Odatis (cf. <https://www.odatis-ocean.fr/activites/ateliers-techniques>).

Il est nécessaire que chaque Centre de Données et de Services délègue des représentants à l'Atelier Technique Odatis.

La participation au Groupe Technique InterPôle, et à ses sous-groupes de travail, est proposée par le Bureau Exécutif du Pôle et basée sur l'expertise. Les organismes de tutelle du Pôle peuvent également désigner un représentant.

À l'issue des études menées par ces deux instances, le Bureau Exécutif du Pôle est l'instance de décision pour leur mise en application.



7) Spécifications des interfaces d'un Centre de Collecte

Chaque Centre de Collecte doit **mettre en place les interfaces ci-dessous** avec le Pôle de Données Odatis (en pratique, avec le Centre de Données et de Services de sa thématique) :

Interfaces organisationnelles :

Chaque CC doit être en mesure d'élaborer les documents suivants :

- **Description du contexte d'activité** : mandat, organisme(s) de tutelle, périmètre des données gérées, points de contact (responsables), résumé des dispositions de mise en œuvre. Ce document sera écrit à la création du Centre de Collecte et maintenu à jour.
- **Descriptif des coûts complets** engendrés par son activité, au format exigé par Ministère de l'Enseignement supérieur de la Recherche et de l'Innovation (MESRI). Ce descriptif est, sauf instruction contraire, à établir annuellement et porte sur l'exercice budgétaire précédent (l'année calendaire précédente). Si l'activité exercée au titre du « Centre de Collecte » est très fortement liée (mêmes personnes...) aux activités d'observation (SNO...), un coefficient forfaitaire (par exemple 20% du temps de travail) pourra être appliqué.
- **Résumé d'activité** résumant l'évolution des jeux de données collectées (volumétrie, procédures de gestion...). Ce résumé sera établi annuellement.

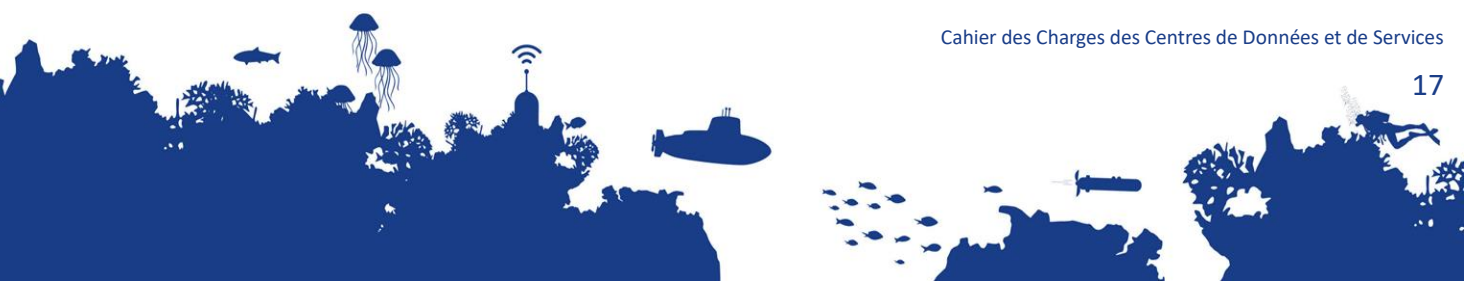
Interfaces techniques :

Chaque Centre de Collecte doit se conformer aux **recommandations techniques spécifiées à l'annexe 1** : Préconisations techniques.

Ces recommandations ont pour objectif d'harmoniser les interfaces et concernent uniquement :

- Les **descriptifs de données** : métadonnées et vocabulaires communs associés ;
- Les **formats de données proposés** aux utilisateurs et les vocabulaires associés.

Il est souhaitable qu'un représentant de chaque Centre de Collecte participe à l'Atelier Technique du Pôle Odatis (cf. <https://www.odatis-ocean.fr/activites/ateliers-techniques>). Il y sera invité en particulier quand les thématiques étudiées lors de la session se rapporteront à son contexte d'activité.



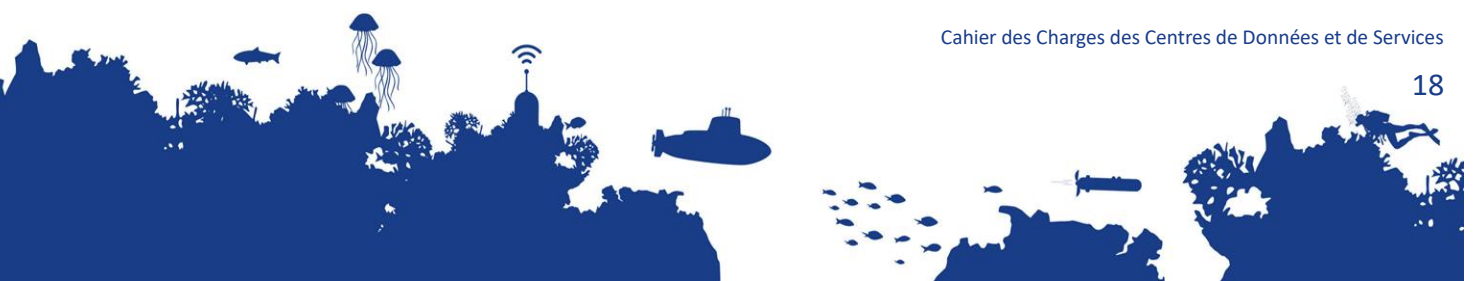
8) Annexe 1 : Préconisations techniques

Cette annexe résume les préconisations techniques. Pour une information détaillée et actualisée, se référer aux conclusions et recommandations de l'atelier technique à l'adresse : <https://www.odatis-ocean.fr/activites/ateliers-techniques>.

- **Identificateurs permanents** : Le Pôle de Données Océan – Odatis utilise les **DOI – Digital Object Identifiers** comme identificateurs permanents des jeux de données. La nomenclature des DOI n'est pas imposée. Toutefois, il est recommandé de mettre en œuvre les recommandations d'harmonisation édictées par le Groupe Technique InterPôles et le groupe ad hoc de la Research Data Alliance (RDA), telles que par exemple, l'utilisation d'un DOI identique et de suffixes pour identifier les différentes occurrences de données au sein d'une même famille (les différents snapshots d'une série temporelle par exemple). Une page d'atterrissage sera maintenue par DOI. Cette page doit proposer une méthode d'accès à la donnée.
- **Métadonnées de découverte (fiches de description & catalogue)**
Au sein du catalogue Odatis, les métadonnées de découverte décrivent des filières ou des jeux de données, mais pas les occurrences de données afin de limiter le nombre d'entrées. La sélection fine des données par les utilisateurs se fait à un deuxième niveau via des web services proposés par les métadonnées de découverte et adaptés à chaque filière.
 - **ISO 19115 / 19139** (avec extensions V3 pour les données satellites)
 - Moissonnage par le **protocole Catalogue Service for the Web (CSW)**, le cas échéant : en effet il n'est pas obligatoire dans le cadre du Pôle, pour chaque CDS, de mettre en place un serveur de métadonnées, les métadonnées pouvant être saisies directement dans le catalogue Odatis.
- **Formats pour les données**

Les données physico-chimiques rendues accessibles par les CDS doivent l'être à un des deux formats suivants :

- **NetCDF**, transition vers la version 4 recommandée, **convention CF** (Climate & Forecast). Le respect du guide de bonnes pratiques SeaDataNet est recommandé (cf. Guide « Data Transport Formats »,



<https://www.seadatanet.org/Standards/Data-Transport-Formats>,

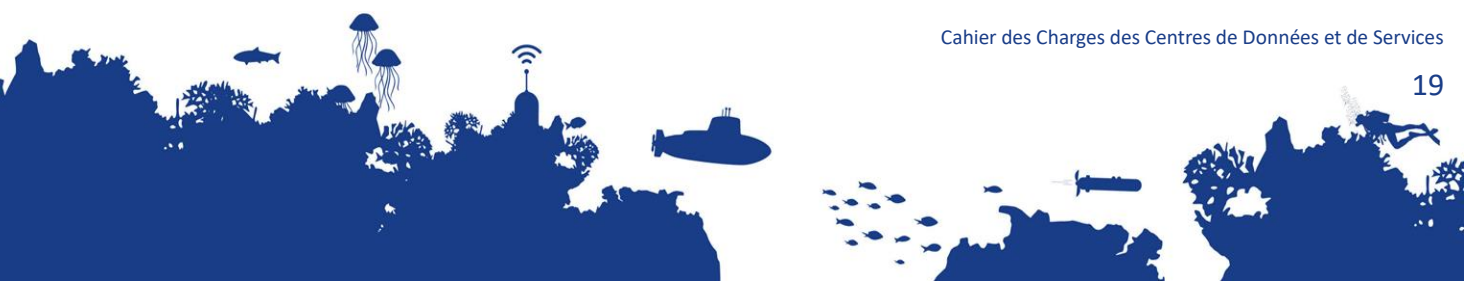
DOI : [10.13155/56547](https://doi.org/10.13155/56547)), en particulier pour affiner, par rapport à la convention CF, la définition des variables observées en utilisant le vocabulaire SeaDataNet quand cela est nécessaire (variables chimiques en particulier).

- **Ocean Data View Spreadsheet** (compatible CSV), (cf. Guide « Data Transport Formats », <https://www.seadatanet.org/Standards/Data-Transport-Formats>, DOI : [10.13155/56547](https://doi.org/10.13155/56547)).
- D'autres formats normalisés adaptés à la nature des données sont également admis. Par exemple, formats vidéo et images en particulier images géoréférencées de type GeoTIFF, formats géographiques (Shape File). Se référer dans ce cas aux résultats des travaux du Groupe Technique InterPôles.

Remarque importante : Ces recommandations de formats s'appliquent essentiellement aux observations physico-chimiques. Les recommandations concernant les formats et les services pour les données à caractère biologique (espèce dépendante) ne sont, à ce jour, pas formellement établies.

- **Vocabulaires**

- Au sein des métadonnées de découverte :
 - Variables observées : **Essential Ocean Variables (EOV)** (cf. http://www.goosocean.org/index.php?option=com_content&view=article&id=14&Itemid=114), éventuellement affinées par les paramètres P02 SeaDataNet.
 - L'utilisation des codes **EDMO (European Directory of Marine Organizations)** est recommandée pour identifier les parties prenantes, les producteurs de données en particulier (cf. <https://www.seadatanet.org/Metadata/EDMO-Organisations>).
- Au sein des fichiers (métadonnées d'usage) :
 - **Convention CF** pour les fichiers NetCDF, complétée le cas échéant (données chimiques en particulier) par le **vocabulaire SeaDataNet P01**.
 - Variables observées / Unités / Méthodes : **Vocabulaire SeaDataNet P01**. (<https://www.seadatanet.org/Standards/Common-Vocabularies>)
 - **Taxonomie** : l'identification des espèces (et d'une manière générale des taxons) utilise la taxonomie **WORMS** (World Ocean Regis-



tery of Marine Species, [cf. http://www.marinespecies.org/](http://www.marinespecies.org/)). La création de taxons temporaires (nouvelle espèce, groupes morphologiques...) est autorisée sous réserve d'un accord des parties prenantes.

- **Services (préconisations à finaliser)**

Les CDS doivent proposer les trois types de services spécifiés ci-dessous. Chaque service sera décrit dans la fiche de métadonnées et, en conséquence, rendu accessible depuis le catalogue.

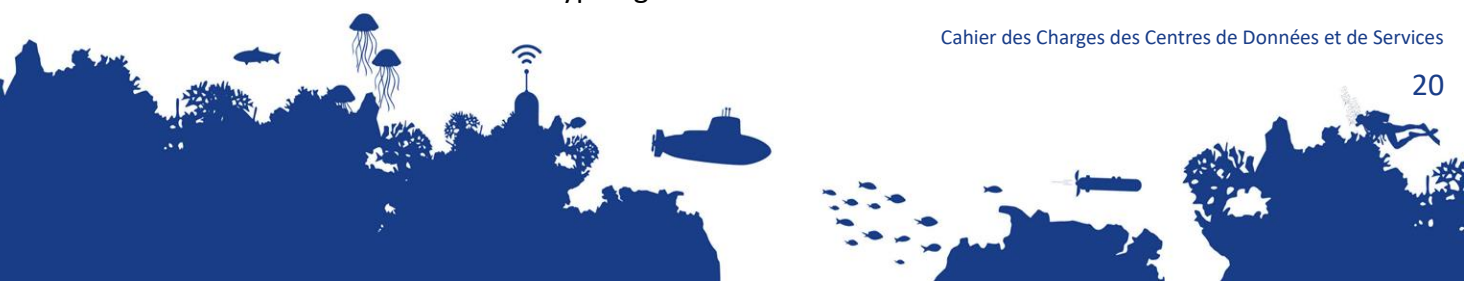
- **Visualisation** : Il s'agit de visualiser soit, a minima, la position des observations ou leur emprise géographique globale, soit d'offrir une prévisualisation des données elles-mêmes. Les visualisations de nature géographique (plan de position, données interpolées sur une grille régulière, images...) doivent respecter les protocoles OGC : WMS, WFS, WCS, WMTS.

Les autres services de visualisation, dépendant de la nature des données (séries temporelles par exemple), ne sont pas normalisés.

- **Accès aux données** : Deux types d'accès aux données devront être proposés suivant la nature et la volumétrie des données :

- Pour les jeux de données dont la volumétrie n'excède pas la capacité des réseaux, un **service de téléchargement direct** du jeu de donnée (protocole HTTP, HTTPS ou FTP). Ce service est notamment mis en œuvre en cas d'accès via la page d'atterrissage du DOI.
- Pour tous les jeux de données, un **service de sélection et d'accès distant conforme au protocole OpenDAP** doit être mis en œuvre.

Le choix du serveur est laissé libre : Thredds Data Server (<https://www.unidata.ucar.edu/software/tds/>), Hyrax (<https://www.opendap.org/software/hyrax-data-server>), Errdap (<https://coastwatch.pfeg.noaa.gov/erddap/index.html>). Toutefois, pour des jeux de données dont les distributions géographique et temporelle sont aléatoirement distribuées, dont le nombre de paramètres est important, et dont la volumétrie est réduite, le serveur ErrDAP est recommandé, car plus spécifiquement adapté à cette typologie.



9) Annexe 2 : Conditions et Licences d'accès aux Données

Les conditions d'accès à la donnée sont définies par chaque organisme producteur.

Toutefois, afin de faciliter l'utilisation des données, et de placer le Pôle Odatis dans le respect des règles « Données Ouvertes » pour une « Science Ouverte », il est souhaitable que l'ensemble des CDS du Pôle de Données Odatis propose des conditions d'accès harmonisées aux jeux de données. Pour cela, l'adoption d'une licence commune par défaut est préconisée et doit être relayée auprès des organismes producteurs.

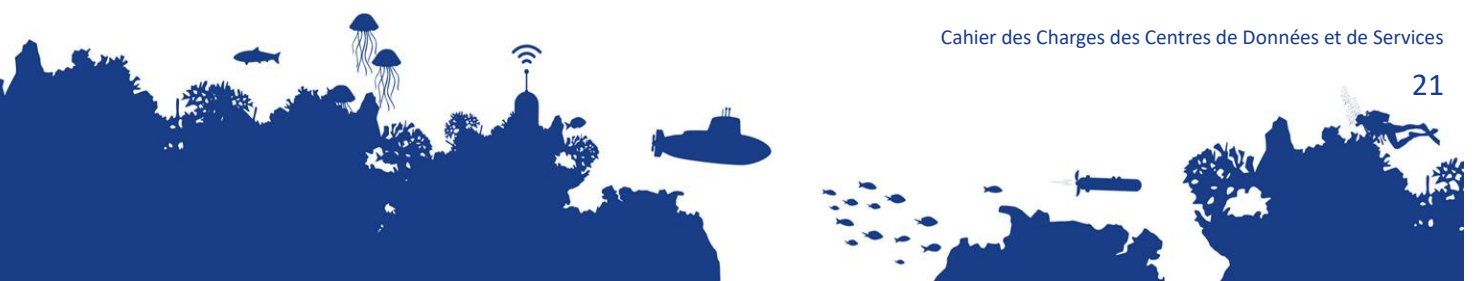
Compte tenu du contexte international de l'observation et de la gestion des données marines, la recommandation est de s'orienter vers l'adoption de la licence « Creative Commons », avec les restrictions suivantes :

- **Attribution [BY]** : le jeu de données peut être utilisé à la condition de l'attribuer à l'auteur en citant son nom. Cela ne signifie pas que l'auteur est en accord avec l'utilisation qui est faite de ses œuvres.
- **Pas d'utilisation commerciale [NC]** (Non Commercial) : le titulaire de droits restreint aux utilisations non commerciales (les utilisations commerciales restent soumises à son autorisation). Elle autorise à reproduire, diffuser, et à modifier une œuvre, tant que l'utilisation n'est pas commerciale (usage de type recherche ou appui aux politiques publiques par exemple). Il ne s'agit pas d'interdire l'utilisation commerciale, mais de la subordonner à la mise en place d'un accord spécifique entre producteur et utilisateur.



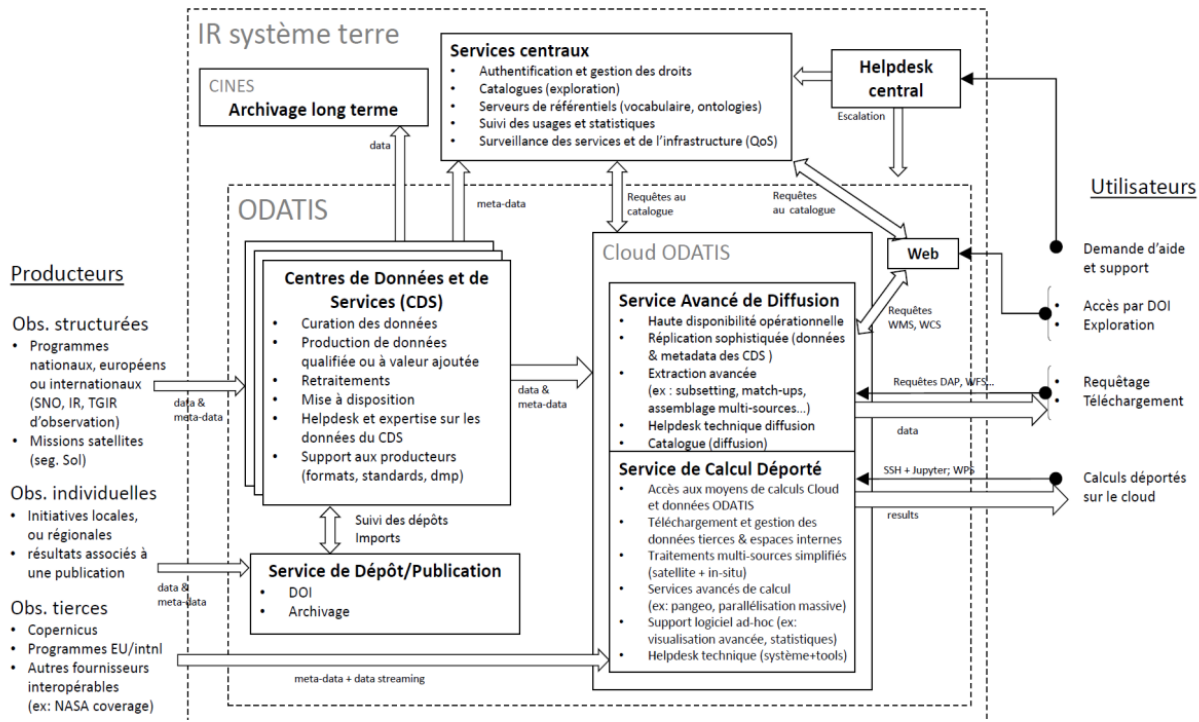
Certaines conditions supplémentaires sont également prévues en droit français, et pourront le cas échéant être appliquées, comme :

- Une **période d'exclusivité d'utilisation** pour les équipes de recherche ayant produit les données.
- Un **accès restreint** pour des données ayant un caractère sensible et notamment : données individuelles, données environnementales portant sur des habitats sensibles ou des espèces protégées, protection des intérêts nationaux stratégiques et économiques...



10) Annexe 3 : Contexte général des CDS au sein de l'IR

Data Terra



Nota : L'interface entre CDS et Producteurs de Données pourra être mise en œuvre via un ou plusieurs Centres de Collecte : par exemple, un ou plusieurs centres régionaux pour les données in situ.