



# Guide des recommandations ODATIS : données & métadonnées

Guide des recommandations du pôle Océan - ODATIS pour les métadonnées et les données















#### **Titre court**

Guide des recommandations ODATIS: données & métadonnées

#### **Titre long**

Guide des recommandations du pôle Océan - ODATIS pour les métadonnées et les données

#### Auteur

Cécile NYS, OCEANSCOPE Joël SUDRE, CNRS UMS CPST Valérie HARSCOAT, IFREMER IRSI/SISMER

Dissémination	Copyright
Publique	© ODATIS, 2021

#### DOI

10.13155/81375

#### Citation

Nys Cecile, Sudre Joel, Harscoat Valerie (2021). Guide des recommandations du pôle Océan - ODATIS pour les métadonnées et les données. https://doi.org/10.13155/81375

#### Référence projet

Cette publication a été réalisée avec le soutien de l'ANR dans le cadre du projet ANR-19-DATA-0020





anr.copilote@ifremer.fr

## Historique

Version	Auteurs	Date	Commentaires
0.9	Cécile NYS, Joël SUDRE & Valérie HARSCOAT	22/06/2021	Reprise du chapitre §3 du guide Copilote en version draft (version 0.9)
0.10	Valérie HARSCOAT	24/06/2021	Ajout des chapitres §1 et §2
0.11	Cécile NYS	25/06/2021	Modification titre ; Révision des ajouts et corrections ; Ajout d'une section Références
1.0	Cécile NYS, Joël SUDRE & Valérie HARSCOAT	25/06/2021	Version diffusable en ligne
1.1	Marine VERNET	28/03/2022	Ajout DOI, citation et référence projet
1.2	Cécile NYS	14/03/2023	Correction signification EML





#### anr.copilote@ifremer.fr

### Table des matières

1. OBJET	6
2. VOCABULAIRES ET METADONNEES	6
2.1. VOCABULAIRES DES PARAMÈTRES MESURÉS	
2.1.1. Description	<i>6</i>
2.1.2. Outils	7
2.2. MÉTADONNÉES	7
3. FORMATS, ATTRIBUTS, CONVENTIONS	8
3.1. FORMATS [OBLIGATOIRE]	
3.2. Attributs de paramètres [Fortement recommandée]	8
3.2.1. Quelle convention de format pour quelle discipline ? [Obligatoire]	
3.2.2. Convention paramètres et données de SeaDataNet (SDN) [Obligatoire]	<u>9</u>
3.2.3. Convention paramètres et données du NetCDF-CF [Fortement recommandée]	10
3.3. PARAMÈTRES DE POSITIONNEMENT [OBLIGATOIRE]	10
4. REFERENCES	11







## **Abréviations**

Abréviations	Signification
CDS	Centre(s) de Données et Services
CF	Convention « Climate and Forecast »
EML	Ecological Metadata Language
EV	Essential Variable
EBV	Essential Biological Variable
ECV	Essential Climate Variable
EOV	Essential Ocean Variable
NetCDF	Network Common Data Format
ODV	Ocean Data View
SDN	SeaDataNet







### 1. Objet

Le présent guide du pôle de données Océan – ODATIS (IR Data Terra) est produit dans le cadre du projet ANR COPILOtE, reprend et affine:

- Les recommandations générales spécifiées dans le cahier des charges des CDS<sup>1</sup>;
- Les recommandations détaillées établies lors des Ateliers Techniques du pôle ODATIS<sup>2</sup> et tracées dans les compte-rendu.

#### Ces recommandations concernent

- D'une part les vocabulaires et métadonnées, qu'elles soient celles de découverte des données ou dans les données;
- Les formats, attributs et conventions pour les fichiers de données.

### 2. Vocabulaires et métadonnées

### 2.1. Vocabulaires des paramètres mesurés

#### 2.1.1. Description

Pour la description des paramètres mesurés, plusieurs référentiels SeaDataNet sont utilisés avec leurs vocabulaires associés. Ainsi, pour un paramètre donné, il est nécessaire de trouver sa **discipline** (P08) puis sa **catégorie** (P03), son **paramètre de découverte** (P02) et enfin son **vocabulaire propre** (*P01*) :

- le P08 pour la discipline du paramètre (12 Sélections Possibles),
- le PO3 pour la catégorie du paramètre (76 Sélections Possibles),
- le P02 pour faciliter la découverte de paramètres (462 Sélections Possibles),
- le P01 rassemble le vocabulaire pour chaque paramètre (42911 Sélections Possibles).

Un paramètre est défini par un vocabulaire contrôlé et par un modèle sémantique (définis dans SeaDataNet depuis 2004). Ce modèle et ce vocabulaire ont été adoptés par plusieurs projets (nationaux, européens et internationaux) pour faciliter l'échange de données et créer de l'interopérabilité entre les différents projets. Il est nécessaire de bien comprendre le modèle sémantique adopté par SeaDataNet (provenant du British Oceanographic Data Centre -BODC) ainsi que ses composants, sa structure et sa logique pour arriver à trouver aisément le vocabulaire associé aux paramètres recherchés.

<sup>&</sup>lt;sup>1</sup> https://doi.org/10.13155/81220

<sup>&</sup>lt;sup>2</sup> https://www.odatis-ocean.fr/activites/ateliers-techniques





anr.copilote@ifremer.fr

Le modèle sémantique pour le P01 est constitué de trois éléments principaux :

- Une propriété quantitative ou nominative d'une entité (concentration, abondance, ...)
  (« property »);
- Une entité physique, biologique ou chimique (« object of interest »);
- Une entité environnementale à laquelle l'entité physique, biologique ou chimique se rapporte ou dans laquelle elle est intégrée (colonne d'eau, sédiment, ...) (« *Matrix* ») ;
- Des champs optionnels (« statistical qualifier, sample preparation, analytical method, processing method »).

Le nom du paramètre est structuré de la façon suivante :

## <Property> {statistical qualifier} <object of interest> <Measurement to Matrix relationship> <Matrix> {optional fields}

ex : <Concentration>{standard deviation} of <ammonium{NH4 + CAS 14798-03-9}><per unit of mass of the> <water body> [dissolved plus reactive particulate].

Pour chaque élément constitutif du modèle sémantique, un vocabulaire est associé et il est disponible sur le site de SeaDataNet<sup>3</sup>.

#### 2.1.2. Outils

- Il est possible de visualiser le contenu des vocabulaires au travers du site web vocab.nerc.ac.uk pour le PO2 par exemple, pour changer de référentiels il suffit de changer le PO2 de ce lien vocab.nerc.ac.uk/collection/PO2/current/.
- Librairie (API) développée par IFREMER en Java 1.8 qui permet de stocker les listes de vocabulaires sur un disque local et de récupérer les codes de vocabulaires et les libellés associés pour les utiliser dans des programmes.

#### 2.2. Métadonnées

Au sein des métadonnées de découverte du catalogue ODATIS :

- Variables observées: les « Essential Variable » (EV) du thésaurus Odatis sont obligatoires. Il existe plusieurs types d'EV: EOV (Essential Ocean Variable), EBV (Essential Biological Variable), ECV (Essential Climate Variable). Ces EV peuvent être affinés par les paramètres PO2 SeaDataNet si besoin
- L'utilisation des codes **EDMO (European Directory of Marine Organizations**) est **recommandée** pour identifier les parties prenantes, les producteurs de données en particulier (lien vers EDMO sur le site de SeaDataNet).

<sup>&</sup>lt;sup>3</sup> https://www.seadatanet.org/





anr.copilote@ifremer.fr

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
- Taxonomie : l'identification des espèces (et d'une manière générale des taxons) utilise la taxonomie WoRMS (World Ocean Registery of Marine Species, voir le site de WoRMS). La création de taxons temporaires (nouvelle espèce, groupes morphologiques ...) est autorisée sous réserve d'un accord des parties prenantes.

### 3. Formats, attributs, conventions

### 3.1. Formats [Obligatoire]

Le cahier des charges des CDS spécifie l'un des deux formats ci-dessous pour les données 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

- ASCII TSV (ODV speadsheet normalisé SDN): les données *in situ* sont à minima à mettre dans un fichier dans ce format et avec l'ensemble des attributs globaux en en-tête de fichier
- NetCDF: le NetCDF version 4 (sans les groupes) est le format NetCDF à privilégier (Les anciens fichiers en NetCDF 3 sont à migrer en NetCDF 4 dès que cela est possible). Pour les métadonnées, il est recommandé d'utiliser la version de convention CF 1.6 à minima ou une version supérieure.

Ainsi, il est recommandé **d'utiliser** à minima le format ASCII TSV (ODV speadsheet normalisé SDN) et si possible du NetCDF convention CF (v1.6 ou supérieure).

### 3.2. Attributs de paramètres [Fortement recommandée]

Afin d'avoir la liste des attributs globaux et des attributs de paramètres qui est nécessaire a minima de mettre dans un fichier NetCDF ou ASCII TSV, le lecteur peut se référer aux recommandations :

- de la conventions CF<sup>4</sup> et en particulier le chapitre 2 du document « NetCDF Climate and Forecast (CF) Metadata Conventions »<sup>5</sup> ;
- du site SeaDataNet qui donne accès au document décrivant les formats (NetCDF et ODV)
  « SeaDataNet. Datafile formats. ODV, MEDATLAS, NETCDF <sup>6</sup>». Sur le site SeaDataNet, il y a aussi des exemples qu'on peut télécharger à la page « Data Transport Formats »<sup>7</sup>

<sup>4</sup> http://cfconventions.org/

<sup>&</sup>lt;sup>5</sup> http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html

<sup>&</sup>lt;sup>6</sup> https://doi.org/10.13155/56547

<sup>&</sup>lt;sup>7</sup> https://www.seadatanet.org/Standards/Data-Transport-Formats





anr.copilote@ifremer.fr

#### 3.2.1. Quelle convention de format pour quelle discipline ? [Obligatoire]

La convention « Climate and Forecast » (CF) ainsi que la convention SeaDataNet (SDN) sont les deux conventions à utiliser pour les données océanographiques. Les attributs des deux conventions sont à insérer dans les fichiers NetCDF et ASCII TSV :

- Pour les paramètres physiques où la convention CF est largement répandue, l'utilisation de la convention SDN est facultative mais il est fortement recommandé de les insérer en attribut supplémentaire décrivant le paramètre, le « long name » correspondant au paramètre décrit en convention SDN (liste P01 de SDN);
- Pour les paramètres de biogéochimie, de chimie, de plastique et micro-plastique, etc. (tous les paramètres non physiques), il est obligatoire d'utiliser la convention SDN (liste P01 de SDN) et d'insérer en attribut supplémentaire le « long name » en convention CF correspondant au nom du paramètre en convention SDN (liste P07);
- **Pour les paramètres de biologie**, qui sont des cas particuliers, où actuellement deux voire trois standards co-existent (BioODV, Darwin Core, EML). Il est possible d'utiliser les conventions de format:
  - BioODV: promu par SDN, format tabulaire avec fichier unique. Ce format ne permet pas actuellement de décrire avec une finesse identique tous les paramètres d'acquisition ou tous les traits de vie liés aux spécimens observés. De plus, il n'est pas utilisé dans les projets majeurs de publication de données de biodiversité (EMODNet Biology alimente OBIS et donc GBIF);
  - Darwin Core: promu par EMODNet Biology, format tabulaire comprenant 3 fichiers: un pour les caractéristiques de base des événements d'échantillonnage (lieu, date), un pour les caractéristiques de base des spécimens observés (identification du spécimen, nombre d'occurrences), un pour les descripteurs additionnels qu'ils soient liés à l'échantillonnage (caractéristiques des engins de prélèvement) ou aux occurrences (caractéristiques des spécimens mesurés: taille, poids, sexe, stade de développement etc...);
  - o EML (Ecological Metadata Language): promu par le PNDB.

#### A noter qu'à ce jour :

- Il n'existe pas de passerelle automatique permettant l'ingestion des données de biodiversité publiées dans SDN dans EMODNet Biology;
- Il est plus facile de convertir le Darwin Core en SDN que l'inverse.

#### 3.2.2. Convention paramètres et données de SeaDataNet (SDN) [Obligatoire]

Les unités des paramètres (physique ou autres) sont à prendre dans la liste P06 de SDN et la discipline dans la liste P08 de SDN.

Pour la nomenclature d'un paramètre il est **fortement conseillé** de mettre la nomenclature la plus précise possible (**liste P01 de SDN**) plutôt que la nomenclature générique de la liste de découverte de paramètre (liste P02 de SDN). Dans le cas d'un paramètre précis manquant dans la liste P01, il est conseillé de se rapprocher de SDN pour le faire insérer dans la liste.







#### 3.2.3. Convention paramètres et données du NetCDF-CF [Fortement recommandée]

Bien que le NetCDF version 4 apporte 5 nouveaux types de donnée utilisateur (« UserDefinedType » : « Enum », « Opaque », « Compound », « VariableLength »), il est fortement déconseillé de les utiliser.

Le NetCDF 4 apporte un niveau de hiérarchisation supplémentaire avec la notion de groupe dans son modèle (afin d'être conforme avec le HDF-5), il est **fortement déconseillé** de l'utiliser. En effet, l'utilisation de plusieurs groupes dans un fichier NetCDF complexifie grandement l'utilisation de ce fichier.

En ce qui concerne les paramètres date et heure dans les fichiers NetCDF, il est fortement recommandé de les insérer sous forme d'entier (type « long ») avec un offset (optionnel) et un scale-factor (obligatoire). L'échelle de temps à adopter est obligatoirement l'UTC (Universel Temps Coordonné).

Lorsque des axes sont nécessaires dans un fichier NetCDF, il est **fortement conseillé** de bien définir les axes et l'orientation. La profondeur est souvent insérée en positif dans les fichiers NetCDF (ex 2000m), ce qui pose des problèmes lorsque l'on va utiliser ces fichiers avec des fichiers ayant des paramètres atmosphériques qui eux vont être aussi en altitude positive (2000m)

### 3.3. Paramètres de positionnement [Obligatoire]

Il est nécessaire de préciser dans les attributs globaux, le système de coordonnées géodésique qui est utilisé comme référentiel pour les paramètres de positionnement. Le code EPSG (European Petroleum Survey Group) est à insérer dans les attributs globaux. A noter que ce standard est aussi utilisé par OGC (Open Geospatial Consortium).





anr.copilote@ifremer.fr

#### 4. Références

- CF Conventions and Metadata (2021). NetCDF CF Metadata Conventions. Available at: http://cfconventions.org/
- Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Blower, J., et al. (2017). NetCDF Climate and Forecast (CF) Metadata Conventions. Available at: http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html.
- Lowry Roy Fichaut Michele, S. R. M. G. B. S. (2019). SeaDataNet. Datafile formats. ODV, MEDATLAS, NETCDF. UK, FRANCE, GERMANY doi:10.13155/56547.
- Maudire, G., Schmidt, S., and Sudre, J. (2020). Cahier des charges des Centres de Données du Pôle de Données Océan ODATIS de l'Infrastructure de Recherche Data Terra. 1–23. doi:10.13155/81220.
- ODATIS (2021). Ateliers Techniques. www.odatis-ocean.fr. Available at: https://www.odatis-ocean.fr/activites/ateliers-techniques [Accessed June 25, 2021].
- SeaDataNet AISBL (2021). SeaDataNet Pan-European infrastructure for ocean & marine data management. Available at: https://www.seadatanet.org/ [Accessed June 25, 2021].
- SeaDataNet AISBL (2020). Data Transport Formats. www.seadatanet.org. Available at: https://www.seadatanet.org/Standards/Data-Transport-Formats [Accessed June 25, 2021].
- SeaDataNet AISBL, and British Oceanographic Data Centre (2021). SeaDataNet Parameter Discovery Vocabulary (P02). *vocab.nerc.ac.uk*. Available at: http://vocab.nerc.ac.uk/collection/P02/current/[Accessed June 25, 2021].