

Modèle de données pivot Theia/OZCAR

Fichier de métadonnées au format JSON

Version 1.2

Changelog :

« *producer.contacts.** » : Suppression de la contrainte d'avoir un seul Project leader.

« *datasets.*.observations.*.featureOfInterest.samplingFeature.name* » : Modification de recommandé à obligatoire. Le nom choisi doit être persistant dans le temps.

« *datasets.*.metadata.portalSearchCriteria* » : Modification d'obligatoire à optionnel

« *datasets.*.metadata.portalSearchCriteria.geologies* » : Modification d'obligatoire à optionnel/déprécié compte tenu que l'enrichissement des données selon la géologie est réalisé automatiquement au niveau de chaque observation.

« *datasets.*.metadata.portalSearchCriteria.climates* » : Modification d'obligatoire à optionnel/déprécié compte tenu que l'enrichissement des données selon le climat est réalisé automatiquement au niveau de chaque observation.

« *datasets.*.observations.*.procedure.dataProduction.sensors.*.model* » : Modification d'obligatoire à recommandé.

« *datasets.*.observations.*.procedure.dataProduction.sensors.*.manufacturer* » : Modification d'obligatoire à recommandé.

« *datasets.*.observations.*.procedure.dataProduction.sensors.*.sensorType* » : Modification d'optionnel à obligatoire.

« *datasets.*.observations.*.observedProperty.keywords* » : Ajout – champ optionnel.

« *datasets.*.observations.*.observedProperty.gcmdKeywords* » : déprécié car remplacé par le champ plus générique *keywords* .

« *datasets.*.observations.*.observedProperty.theiaCategories* » : Il n'est plus obligatoire de renseigner les URI de catégories de variable de dernier niveau hiérarchique. Les URI des catégories renseignées doivent permettre d'identifier le compartiment de la zone critique dans laquelle a été effectuée l'observation. Alternativement, il est désormais possible de référencer une URI de variable plutôt qu'une URI de catégorie si ce terme de type "Variable" existe déjà dans le thésaurus et décrit assez précisément la variable producteur. Dans ce cas, le tableau JSON contiendra une et une seule URI représentant la variable Theia/OZCAR avec laquelle associer la variable producteur.

Avant-propos :

Le flux d'information à mettre en place entre les producteurs de données et le SI Theia/OZCAR s'appuie sur un modèle de données "pivot" rassemblant un set de métadonnées à échanger afin de favoriser la découverte, le partage et la réutilisation des données. Il permettra d'alimenter les différentes fonctionnalités du portail d'accès aux données Theia/OZCAR et d'implémenter différents services comme le catalogue de métadonnées ou la publication de DOI pour les jeux de données.

L'analyse des métadonnées à fournir s'est focalisée dans un premier temps sur des données de séries temporelles pour les observations basées sur des points/trajectoires. Le format pourra être adapté dans un deuxième temps à d'autres types de données.

L'objectif de ce document est de décrire le contenu du modèle de données pivot à implémenter au format JSON. Ce fichier texte au format JSON sera produit par chacun des producteurs de données. Le portail Theia/OZCAR analysera ensuite les informations contenues dans ce fichier pour alimenter son SI. Ces informations comportent les informations sur le producteur de données et les jeux de données dont les variables qui les composent. Un jeu de données regroupe les données géographiquement référencées d'une ou plusieurs variables observées sur une période temporelle. Le jeu de données est décrit par des métadonnées décrivant son contenu et les variables mesurées.

JSON (JavaScript Object Notation – Notation Objet issue de JavaScript) est un format léger d'échange de données. Il est facile à lire ou à écrire pour des humains. Il est aisément analysable ou générable par des machines. Il est basé sur un sous-ensemble du langage de programmation JavaScript ([JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999](#)). JSON est un format texte complètement indépendant de tout langage. Le JSON se base sur les structures suivantes :

Un **objet**, qui est un ensemble de couples nom/valeur non ordonnés. Un objet commence par { (accolade gauche) et se termine par } (accolade droite). Chaque nom est suivi de : (deux-points) et les couples nom/valeur sont séparés par , (virgule). L'ensemble de couples nom/valeur sont les membres de l'objet.

Un **tableau** est une collection de valeurs ordonnées. Un tableau commence par [(crochet gauche) et se termine par] (crochet droit). Les valeurs sont séparées par , (virgule). La collection de valeurs ordonnées sont les éléments du tableau.

Une **valeur** peut être soit une chaîne de caractères entre guillemets, soit un nombre, soit `true` ou `false` ou `null`, soit un **objet** soit un **tableau**. Ces structures peuvent être imbriquées.

Les objets géographiques sont décrits selon le standard [GeoJSON](#) qui permet de décrire une grande variété de structure géométrique.

Choix de la langue pour les métadonnées :

Le choix de la langue (anglais ou français) décrivant les métadonnées est laissé à l'initiative du fournisseur de données en fonction de l'existant dans son système d'information. S'il n'y a pas d'existant, l'anglais est préféré.

En revanche, **les valeurs possibles des énumérations (cf : annexe)** doivent quant à elles être **fournies en anglais**.

Si les 2 langues existent, il y a possibilité de nous fournir les 2; les noms de fichiers indiquant la langue utilisée avec un code sur 2 digit (ex: CATC_fr.json)

Nomenclature du nom de fichier json :

Le nom de fichier doit respecter la convention suivante :

XXXX_yy.json

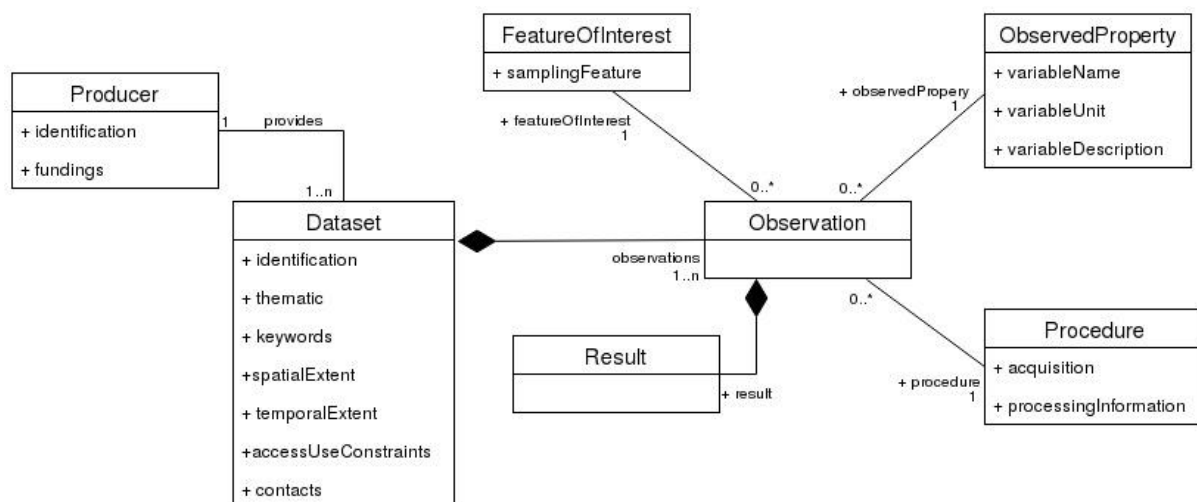
XXXX : l'identifiant Theia/OZCAR du fournisseur de données. La liste des identifiants pour chacun des observatoires est disponible en annexe.

yy : le code en deux lettres de la langue utilisée dans le fichier. « **en** » pour l'anglais ou « **fr** » pour le français. Dans l'optique de mettre en œuvre l'interopérabilité avec les infrastructures Européennes, un fichier en langue anglaise est demandé.

Encoding des caractères à utiliser dans les fichiers :

L'encoding à utiliser est UTF-8.

Modèle de classe simplifié du format pivot



Les informations décrivant un jeu de données (classe Dataset) sont issues de la norme ISO 19115/Inspire.

Les concepts décrivant une observation sont dérivés du standard OGC O&M : classes Observation, FeatureOfInterest, ObservedProperty, Procedure, Result

Métadonnées obligatoires / recommandées / optionnelles:

Le caractère obligatoire ou non des métadonnées est indiqué dans le document avec le code suivant :

- **M** (mandatory) = **information obligatoire**
- **R** (recommended) = **information recommandée**
- **O** (optional) = **information optionnelle**

Un objet JSON peut être constitué d'un ensemble d'objets JSON. Un objet JSON **O** optionnel peut être constitué d'objets JSON obligatoires **M**. Dans ce cas de figure, si l'objet JSON optionnel est implémenté il sera obligatoirement constitué de ses membres obligatoires.

Gestion des non valeurs pour les champs json optionnels

Pour les champs optionnels que le producteur ne souhaite pas renseigner, et afin d'être conforme aux schémas json définis, il est attendu que ces **champs soient absents du fichier json**. Des valeurs égales à une chaîne vide (pour le type string) ou null ne sont pas autorisées.

Ex : **O** email

A faire : Absence du champ

A ne pas faire : "email": " " ou "email": null

Code de non-valeur pour les unités de mesure des variables

Pour les **variables sans unité** (ex : pH), la valeur du champ "**unit**" doit être renseigné avec le **code "N/A"**.

Ex : "observedProperty": {

 "name": "pH",

 "unit": "N/A",

 }

Description détaillée des objets du fichier json :

Dans ce document, chaque nom d'un couple nom/valeur est indiqué sur fond gris (**nom**).

Au plus haut niveau du fichier sont définis 3 objets JSON ayant les noms suivants :

- **M producer** : est la fiche de métadonnées relatives au fournisseur de données publiant le fichier
- **M datasets** : est la collection de jeux de données de l'observatoire. Cette collection est représentée par un tableau JSON dont les éléments représentent chaque jeu de données du fournisseur.
- **M version** : est la chaîne de caractère correspondant au numéro de version du format pivot. Cette première version est la « 1.0 »

1. L'objet JSON **producer**

L'objet **producer** est composé des membres suivants :

- **M producerId** : Identifiant du fournisseur de données. Les identifiants des observatoires sont fournis en annexe.
- **M name** : couple nom/valeur dont la valeur est le nom de l'observatoire. *Exemple* : « AMMA-CATCH ». Champ texte libre.
- **M title** : couple nom/valeur dont la valeur est le titre de l'observatoire. *Exemple* : « AMMA-CATCH: a hydrological, meteorological and ecological observatory on West Africa ». Champ texte libre.
- **M description** : couple nom/valeur dont la valeur est un résumé des activités scientifiques du fournisseur de données. Champ texte libre obligatoire. *Exemple* : « L'observatoire Draix-Bléone a été créé en 1983 pour étudier les processus hydrologiques et érosifs en montagne. Il est situé dans les Alpes du Sud, sur un substrat très érodable constitué de marnes noires, et caractérisé par un climat montagnard et méditerranéen. Une dizaine

de bassins versants ont été équipés pour la mesure des flux solides et liquides. Ces bassins couvrent des surfaces allant de 0.1 ha à plusieurs centaines de km², avec des couvertures végétales, des lithologies et des topographies variées. Le Groupement d'Intérêt Scientifique Draix-Bléone créé en 2002 rassemble actuellement 18 laboratoires avec des compétences d'hydrologie, hydraulique, géomorphologie, écologie, géochimie, ingénierie, géographie. »

- **R objectives** : couple nom/valeur dont la valeur est un résumé des objectifs scientifiques du fournisseur de données. Exemple : « *L'observatoire Draix-Bléone a pour objectif d'étudier les processus hydrologiques et érosifs en montagne.* »
 - **R measuredVariables** : couple nom/valeur dont la valeur est un résumé des variables mesurées par le fournisseur de données. Exemple : *Les données disponibles sont les chroniques de pluie et les données climatiques (température, humidité, rayonnement), les chroniques de débit à l'exutoire des bassins suivis, les flux de matières en suspension et les dépôts de matériaux grossiers pour chaque crue, le suivi de l'état hydrique des sols, le modèle numérique de terrain et la couverture végétale.*
 - **O email** : couple nom/valeur dont la valeur est la chaîne de caractères représentant l'email du fournisseur de données.
 - **M contacts** : collection de contacts pour le producteur de données représenté sous la forme d'un tableau JSON. **Les contacts doivent être des personnes physiques** : deux rôles sont possibles. Le **Data manager** est la personne qui assure le rôle de gestionnaire de données. Le **Project leader** est le responsable scientifique de l'observatoire. Ces rôles sont utiles à la création de DOI ainsi qu'à la publication de fiche de métadonnées INSPIRE. Au moins un contact avec le rôle **Project leader** est requis. Les éléments de la collection **contacts** représentant des personnes physiques sont des objets possédant les membres suivants :
 - o **M firstName** : le prénom du contact. Chaîne de caractère libre. Cet élément est obligatoire.
 - o **M lastName** : le nom du contact. Chaîne de caractère libre. Cet élément est obligatoire.
 - o **M email** : adresse email du contact. Chaîne de caractères. Cet élément est obligatoire.
 - o **M role** : Le rôle du contact pour le jeu de données. Ce champ est obligatoire. Le champ peut prendre une des valeurs suivantes : "Project leader", "Data manager"
 - o **O orcid** : l'identifiant orcid de la personne s'il existe.
 - o **O organisation** : l'organisation d'appartenance du contact physique. Un objet JSON composé des membres suivants :
 - **M role** : Le rôle de l'organisation, pour le moment un seul rôle est possible « Research group ».
 - **M iso3166** : le code à deux lettres du pays de l'organisation. Voir http://kirste.userpage.fu-berlin.de/diverse/doc/ISO_3166.html
 - **R name** : le nom de l'organisation d'appartenance.
 - **R acronym** : l'acronyme du nom de l'organisation d'appartenance.
 - **R idScanR** : id ScanR de l'organisation (SIREN ou id RNSR)
- Un membre de l'objet Organisation doit être composé d'un acronym ou d'un name ou d'un idScanR**
- **M fundings** : un tableau JSON représentant la liste des financeurs du fournisseur de données (fonctionnement et masse salariale). Ce champ est obligatoire. Il est utile pour la déclaration de DOI ainsi que pour produire des statistiques sur les tutelles. Chaque élément du tableau JSON **fundings** est un objet JSON composé des membres suivants :
 - o **M type** : le type de financeur respectant l'énumération en annexe. Le champ peut prendre une des valeurs de l'énumération « EnumFundingType ». Un financeur peut être un institut de recherche (ex : CNRS, IRD, BRGM, IGN...), une université, un OSU est

de type "French Federative structure". Les UMR sont de type "Research unit".

- **M iso3166**: le code à deux lettres du pays du financeur. Voir http://kirste.userpage.fu-berlin.de/diverse/doc/ISO_3166.html
- **R acronym**: l'acronyme du financeur
- **R name**: le nom de financeur
- **R idScanR**: id ScanR de l'organisation (SIREN ou id RNSR). Visible sur <https://scanr.enseignementsup-recherche.gouv.fr/recherche/structures>

Un membre de l'objet Funding doit être composé d'un acronym ou d'un name ou d'un idScanR

- **O onlineResource**: Les différentes ressources en ligne liées au producteur de données. Cet objet JSON est composé des membres suivants:
 - **R urlDownload**: Cet élément de métadonnée fournit un lien vers le portail d'accès aux données du producteur de données. Champ de texte sous forme d'une url valide.
Ex: <http://sevnol.ohmcv.fr/Sevnol2>
 - **R urlInfo**: Cet élément de métadonnée fournit un lien vers le site Web du producteur de données (observatoire). Champ de texte sous forme d'une url valide.
Ex: <http://www.ohmcv.fr>
 - **R doi**: L'identifiant doi référençant de manière unique le producteur (doi de l'observatoire tout entier). Champ de texte sous forme d'un identifiant doi valide.
Ex: 10.17178/AMMA-CATCH.all
 - **R webservices**: tableau JSON répertoriant les URLs des différents services web permettant l'interopérabilité du producteur de données. Chaque élément du tableau JSON **webservices** est un Objet JSON composé de deux couples nom/valeur. Le premier couple nom/valeur avec nom=**description** est un champ texte libre décrivant le webservice. Le deuxième couple nom/valeur avec nom=**url** est un champ de texte libre fournissant l'url valide du webservice. Ce champ permettra de diffuser les webservices mis en place sur les informations du producteur de données. Une fiche ISO19115/INSPIRE publié dans un catalogue CSW pour le producteur de données sera décrite dans le champ **description** par l'url du *getRecordById* dont l'id correspond à la fiche du fournisseur de données. Ex: <https://sig.oreme.org:8443/geonetwork/srv/eng/csw?SERVICE=CSW&VERSION=2.0.2&REQUEST=GetRecordById&ID=8dbdb4d7-161a-45e3-b3c3-51e6058f6e8a>. Un catalogue CSW des jeux de données du producteur de données sera décrit dans le champ **description** par l'url du *getCapabilities*. Ex: <https://sig.oreme.org:8443/geonetwork/srv/eng/csw?SERVICE=CSW&REQUEST=GetCapabilities&VERSION=2.0.2>

2. L'objet JSON datasets:

L'objet **datasets** est la collection des jeux de données du fournisseur de données. L'objet JSON **datasets** est représenté sous forme d'un tableau JSON. Chaque valeur du tableau JSON est un objet JSON décrivant un jeu de données. Chaque objet JSON représentant un jeu de donnée possède lui-même les 3 membres suivants:

- **M datasetId**: une chaîne de caractère composée d'un code formé par les 4 premières lettres du nom du fournisseur de données conforme à la liste fournie en annexe 1, suivi par un la chaîne de caractère « _DAT_ » et d'un **identifiant pérenne du jeu de données. Il est obligatoire que l'identifiant choisi pour le jeu de données n'évolue pas dans le temps.** Ce peut être un identifiant d'une table de la base de données ou un code identifiant le jeu de données
Ex: AGRH_DAT_1
- **M metadata**: un objet JSON regroupant les métadonnées décrivant le jeu de données

- **M observations**: un objet JSON rassemblant les observations associées au jeu de données.

a) L'objet **metadata**:

Cet objet JSON est un membre de chacun des objet JSON élément de la collection **datasets**. Il contient lui-même les membres suivants :

- **M title** : le titre du jeu de données. Il est recommandé de faire figurer une indication de la zone géographique couverte, notamment si le même type d'enregistrement de données scientifiques est proposé sur plusieurs territoires.
Il est recommandé de ne pas faire figurer les parties responsables c'est-à-dire les organisations exerçant une quelconque responsabilité (gestionnaire, propriétaire, ...).
Il est recommandé de ne pas indiquer uniquement l'acronyme de la ressource (à moins de l'expliciter dans le champ description).
Champ texte libre obligatoire qui sera utile à la création de fiche de métadonnées conforme INSPIRE.
Exemple: Niveau de la nappe dans les puits ruraux, méso-site d'Oueme, Benin.
- **M description**: un résumé du jeu de données. Le résumé doit décrire la ressource de façon compréhensible par l'utilisateur. Pour un producteur, il s'agit en particulier de définir au mieux l'information ou le phénomène représenté dans la donnée. On va donc y trouver des éléments de définition, mais aussi éventuellement une indication sommaire de la zone couverte ou le cas échéant, des informations sur les particularités de la version du jeu de données.
Champ texte libre obligatoire qui sera utile à la création de fiche de métadonnées conforme INSPIRE.
Exemple: Mesures du niveau de la nappe et mesures de la conductivité manuelle pour certains puits. Mesures horaires pour les puits équipés. Le niveau de la nappe phréatique est exprimé depuis le haut du puits.
- **O objective** : Résumé des objectifs scientifiques du jeu de donnée. Cet élément est optionnel. Champ de texte libre.
- **O onlineResource** : Les différentes ressources en ligne du jeu de données. Cet objet JSON est composé des membres suivants :
 - o **R urlDownload** : Cet élément de métadonnée fournit un lien vers une page Web de téléchargement du jeu de donnée. Cet élément est recommandé si une telle possibilité existe. Champ de texte sous forme d'une url valide. Il permettra de rediriger l'utilisateur lors d'une action de téléchargement.
Exemple: <ftp://ftp.ifremer.fr/ifremer/argo/>
 - o **R urlInfo** : Cet élément de métadonnée fournit un lien vers une page Web décrivant le jeu de données. Champ de texte sous forme d'une url valide.
Exemple: <http://www.seanoe.org/data/00311/42182/>
 - o **R doi** : L'identifiant doi référençant de manière unique le jeu de données. Champ de texte sous forme d'une identifiant doi valide. Il permettra la citation du jeu de données.
Exemple: 10.13155/29825
 - o **R webservices** : tableau JSON répertoriant les URLs des différents services web permettant l'interopérabilité du jeu de données. Chaque élément du tableau JSON **webservices** est un Objet JSON composé de deux couples nom/valeur. Le premier couple nom/valeur avec nom=**description** est un champ texte libre décrivant le webservice. Le deuxième couple nom/valeur avec nom=**url** est un champ de texte libre fournissant l'url valide du webservice. Ce champ permettra de diffuser les webservices des jeux de données. Ceux-ci pourrait être utilisés par l'application lors de prochaines versions. Une fiche ISO19115/INSPIRE publié dans un catalogue CSW pour le jeu de données sera décrite dans le champ **description** par l'url du *getRecordById*

dont l'id correspond à la fiche du jeu de donnée. Ex :
<https://sig.oreme.org:8443/geonetwork/srv/eng/csw?SERVICE=CSW&VERSION=2.0.2&REQUEST=GetRecordById&ID=131142c1-cb88-4510-9f52-e0fe5f3c18be>

- **M datasetLineage**: décrit la généalogie d'un jeu de données, i.e. l'historique du jeu de données et, s'il est connu, le cycle de vie de celui-ci, depuis l'acquisition et la saisie de l'information jusqu'à sa compilation avec d'autres jeux et les variantes de sa forme actuelle. Il s'agit d'apporter une description littérale et concise soit de l'histoire du jeu de données, soit des moyens, procédures ou traitements informatiques mis en œuvre au moment de l'acquisition du jeu de données. La généalogie fait état de l'historique du traitement et/ou de la qualité générale de la série de données géographiques. Le cas échéant, elle peut inclure une information indiquant si la série de données a été validée ou soumise à un contrôle de qualité, s'il s'agit de la version officielle (dans le cas où il existe plusieurs versions) et si elle a une valeur légale. Cet élément est une chaîne de caractère libre et est obligatoire servant à création de fiche de métadonnées conforme INSPIRE.
- **M contacts** : collection de contacts pour le jeu de données représenté sous la forme d'un tableau JSON. Un membre de la collection contact peut être une personne physique, soit une organisation.

Dans le cas d'une personne physique : quatre rôles sont possibles. Deux rôles « techniques », Data Manager ou Data Collector. Le Data Manager est la personne qui assure le rôle de gestionnaire de données. Le Data Collector est la personne qui récolte les données. Si un contact avec le rôle Data Manager est précisé, celui-ci sera le point de contact pour le jeu de données. Deux rôles « scientifiques » sont possibles : Principal Investigator (référant scientifique pour le jeu de données), ou Project Member (scientifique associé à la création du jeu de données). Ces rôles sont utiles à la création de DOI ainsi qu'à la publication de fiche de métadonnées INSPIRE. Au moins un contact avec le rôle Principal Investigator est requis, tous les autres rôles sont optionnels. Les éléments de la collection **contacts** représentant des personnes physiques sont des objets possédant les membres suivants :

- o **M firstName**: le prénom du contact. Chaîne de caractère libre. Cet élément est obligatoire.
- o **M lastName**: le nom du contact. Chaîne de caractère libre. Cet élément est obligatoire.
- o **M email**: adresse email du contact. Chaîne de caractères. Cet élément est obligatoire.
- o **M role** : Le rôle du contact pour le jeu de données. Ce champ est obligatoire. Le champ peut prendre une des valeurs suivantes : "Principal investigator", "Project member", "Data manager", "Data collector". Le rôle "Project leader" est défini pour l'objet Producer seulement.
- o **O orcid** : l'identifiant orcid de la personne s'il existe.
- o **O organisation** : l'organisation d'appartenance du contact physique. Un objet JSON composé des membres suivants :
 - **R name** : le nom de l'organisation d'appartenance.
 - **R acronym** : l'acronyme du nom de l'organisation d'appartenance.
 - **M role** : rôle de l'organisation ayant pour valeur "Research group".
 - **M iso3166** : le code à deux lettres du pays de l'organisation. Voir http://kirste.userpage.fu-berlin.de/diverse/doc/ISO_3166.html
 - **R idScanR** : id ScanR de l'organisation (SIREN ou id RNSR). Visible sur <https://scanr.enseignementsup-recherche.gouv.fr/>

Dans le cas d'une organisation : un seul rôle est possible, "Research group". Les éléments de la collection contacts représentant des organisations sont des éléments représentés par des objets JSON composés des membres suivants :

- o **R name** : le nom de l'organisation d'appartenance.

- **R** **acronym** : l'acronyme du nom de l'organisation d'appartenance.
- **M** **role** : role de l'organisation ayant pour valeur "Research group".
- **M** **iso3166** : le code à deux lettres du pays de l'organisation. Voir http://kirste.userpage.fu-berlin.de/diverse/doc/ISO_3166.html
- **R** **idScanR** : id ScanR de l'organisation (SIREN ou id RNSR). Visible sur <https://scanr.enseignementsup-recherche.gouv.fr/>

Un membre de l'objet **Organisation** doit être composé d'un **acronym** ou d'un **name** ou d'un **idScanR**

- **M** **dataConstraint** : Cet élément de métadonnées définit les contraintes d'accès et d'utilisation, soit, les conditions applicables aux limitations d'accès public et à l'accès et à l'utilisation des séries de données géographiques. L'objet **dataConstraint** est composé des membres suivants :
 - **M** **AccessUseConstraint** : un champ texte libre décrivant les conditions d'utilisation. Si le jeu de données est soumis à des conditions d'utilisation (ex : utilisation sous condition de citation via phrase de remerciement), elles doivent être indiquées dans ce champ. Si aucune condition ne s'applique, le champ prendra la valeur "Aucune condition ne s'applique", si les conditions sont inconnues, le champ prendra la valeur "Conditions inconnues". Si les éléments de métadonnées sont décrit en anglais le champ pourra prendre "[no conditions to access and use](#)" ou la valeur "[conditions to access and use unknown](#)". Ce champ est obligatoire.
 - **O** **urlDataPolicy** : fournit un lien vers l'url de la data policy. Champ de texte sous forme d'une url valide.
 - **O** **licence** : objet JSON représentant la licence à laquelle le jeu de données est associé. L'objet JSON est composé des membres suivants :
 - **M** **title** : le nom de la licence. Ce champ est obligatoire si une licence a été souscrite pour le jeu de données.
Ex : CC BY 4.0
 - **M** **url** : URL de la licence du jeu de données. Ce champ est obligatoire si une licence a été souscrite pour le jeu de données
Ex : <https://creativecommons.org/licenses/by/4.0/>
 - **O** **embargo** : une période d'embargo peut être définie pour limiter l'accès aux données d'un jeu. Cette période d'embargo est la durée en jours, depuis la date actuelle, pour laquelle l'accès aux données est restreint à un nombre d'utilisateur limité. Ainsi pour une durée d'embargo de 365 jours, si une requête est effectuée sur le jeu de données en date du 01/03/2018, les données mesurées le 01/05/2017 ne seront pas disponibles, tandis que les données mesurées le 28/02/2017 le seront. L'objet JSON **embargo** est constitué de deux membres :
 - **M** **duration** : la durée en jours de l'embargo sur le jeu de données.
 - **M** **privilegedUsers** : tableau JSON de champ textuelle. Chacun des champs étant l'email d'un utilisateur privilégié autorisé à outrepasser l'embargo sur le jeu de données.
- **O** **portalSearchCriteria** : Ce champ permet de filtrer les jeux de données sur le portail. Les critères de recherche suivants sont proposés :
 - **O** **climates** : **déprécié**. Une collection des climats pouvant être associés au jeu de données décrite sous la forme d'un tableau JSON. La liste des climats possibles est disponible en annexe.
 - **O** **geologies** : **déprécié**. Une collection des géologies pouvant être associées au jeu de données décrite sous la forme d'un tableau JSON. La liste des géologies possibles est disponible en annexe.

Les champs **climates** et **geologies** sont **dépréciés** compte tenu que l'enrichissement des données selon le climat et la géologie est réalisé automatiquement au niveau de chaque observation.

- **M topicCategories**: Catégorie thématique du jeu de données issue du standard ISO 19115/Inspire. Valeur : tableau JSON de champs texte conforme à la liste des différentes catégories thématiques INSPIRE fournie en annexe. Au moins une valeur de la catégorie thématique doit être fournie. Ce champ permettra la création de fiche de métadonnées INSPIRE.
- **M inspireTheme**: Thématique du jeu de données requis par INSPIRE. Valeur : texte conforme à la liste des différents thématiques INSPIRE fournie en annexe. Ce champ permettra la création de fiche de métadonnées INSPIRE. Le mapping des champs **topicCategories** et **inspireTheme** a été suggéré dans [le guide de saisie des éléments de métadonnées-INSPIRE v2.0](#) mais n'est pas rendu obligatoire.
- **O keywords** : Tableau JSON d'Objet JSON. Chacun des objets JSON représentant un mot clé. L'Objet JSON représentant un mot clé est composé de 2 couples nom/valeur. Le premier couple avec nom = « keyword » est un champ de texte libre décrivant le mot-clé, et le second couple avec nom= « uri » est un champ de texte libre décrivant l'URI du mot clé dans un thésaurus publié.
- **O documents** : collection des documents associés au jeu de données (publications ou manuels) représentée par un tableau JSON. Les documents peuvent être de deux types : « Publication » ou « Manual ». L'objet **documents** est un tableau JSON. Chacun des éléments de ce tableau est un objet JSON comprenant deux membres. Un membre « type » pouvant prendre la valeur « Publication » ou « Manual », un membre « url » représentant l'url à laquelle est accessible le document. Cet élément optionnel permet de documenter les publications de références associées au jeu de données, les protocoles d'acquisition du jeu de données...
- **M spatialExtent** : Étendue du jeu dans l'espace géographique, exprimée sous la forme d'un ou plusieurs rectangles de délimitation (BBOX). Ce rectangle de délimitation est défini par les longitudes est et ouest et les latitudes sud et nord en degrés décimaux, avec une précision d'au moins deux chiffres après la virgule. Ces coordonnées sont exprimées selon le référentiel de coordonnée WGS84 (ESPG:4326). Si plusieurs emprises sont indiquées, elles ne doivent pas se superposer. Ce champ permettra la création de fiche de métadonnées conforme INSPIRE et de documenter l'emprise des jeux de données. **spatialExtent** est un objet GeoJSON et les membres suivants servent à le décrire.
 - o **M** La constante **"type" : "Feature"** est attribué afin de faire correspondre l'objet au schéma GeoJSON.
 - o **M** La constante **"properties" : {}** est attribué afin de faire correspondre l'objet au schéma GeoJSON.
 - o **M geometry** décrit l'emprise spatiale du jeu de données. L'objet **geometry** est composé des membres suivant.
 - **M type** : le type de la géométrie de l'emprise. Les deux types d'emprise possible sont **"Polygon"** et **"MultiPolygon"** dans le cas où plusieurs emprises sont nécessaires pour le jeu de données.
 - **M coordinates** : pour un type polygone, le membre **coordinates** doit être une collection de collection de positions. Chaque position étant décrite par le doublet de coordonnées [longitude,latitude]. Les coordonnées sont exprimées en degré décimaux. L'emprise rectangulaire comprendra 5 positions, la cinquième identique à la première. Les coordonnées du polygone sont définies dans la collection dans le sens anti-horaire. Pour un type MultiPolygon, le membre **coordinate** est une collection de Polygon.

b) L'objet **observations**

Dans le format pivot, une observation contient la série temporelle d'une variable donnée à une localisation donnée. L'objet **observations** est la collection des observations associées au jeu de données. Il est décrit par un tableau JSON. D'après le standard O&M (Observation & Measurement), une « observation » peut être définie, comme l'action dont le résultat (**result**) est la valeur à un moment donné d'une propriété (**observedProperty**) d'un objet d'intérêt (**featureOfInterest**) obtenu grâce à une procédure (**procedure**).

Chaque élément de la collection **observations** est un objet JSON décrivant une observation composée des éléments suivants :

- **M observationId** : une chaîne de caractère composée d'un code formé par les 4 premières lettres du nom du fournisseur de données conforme à la liste fournie en annexe 1, suivi de la chaîne de caractère «_OBS_», suivi d'un identifiant unique de l'observation. L'identifiant unique de l'observation dans le jeu de données peut être défini de manière incrémental et **contrairement à l'identifiant du jeu de données**, il n'est pas nécessaire de le faire persister.

Ex : AGRH_OBS_1

- **M observedProperty** : identifie la variable décrivant le phénomène observé. Cette variable est liée à l'objet d'intérêt de l'observation (**featureOfInterest**). L'objet **observedProperty** est un objet JSON composé des champs suivants :
 - o **M name** : chaîne de caractères décrivant le nom de la variable tel qu'il est intitulé dans le système d'information du fournisseur de données. ; A charge du SI Theia/OZCAR de réaliser les associations avec le thésaurus Theia/OZCAR.
 - o **M unit** : chaîne de caractère décrivant l'unité de mesure de la variable
 - o **O description** : chaîne de caractère proposant une description de la variable. Cette description pourra faciliter l'association de la variable au thésaurus de variables Theia/OZCAR et faciliter la compréhension de la mesure effectuée.
 - o **O gcmdKeywords** : **déprécié**. Collection de d'objet JSON décrivant chacun un champ de la taxonomie GCMD pour les noms de variables (cf : Global Change Master Directory controlled vocabularies: [Earth Science and Earth Science Services](https://w3id.org/ozcar-theia/c_341ce66a)). Chacune des associations avec la taxonomie GCMD réalisées par l'équipe Theia/OZCAR lors de l'étape du travail sur le vocabulaire de variable, viendront enrichir ce champ. Chaque objet de la collection est composé des membres suivants :
 - **M category** : la catégorie du champ GCMD
 - **M topic** : le topic du champ GCMD
 - **O term** : le term du champ GCMD
 - **O variableLevel1** : la variable de niveau 1 du champ GCMD
 - **O variableLevel2** : la variable de niveau 2 du champ GCMD
 - **O variableLevel3** : la variable de niveau 3 du champ GCMD
 - **O uuid** : l'identifiant unique du champ GCMD

Le champ **gcmdKeywords** est déprécié car remplacé par le champ plus générique **keywords**.

- o **M theiaCategories**: Tableau JSON de chaîne de caractère. Chaque chaîne de caractère représente une URI identifiant un terme de type catégorie de variable (TYPE = Category of variable) dans le thésaurus Theia/OZCAR (https://w3id.org/ozcar-theia/c_341ce66a) et permet d'associer le nom de variable producteur à la taxonomie de catégorie de variable Theia/OZCAR. Il est possible de renseigner plusieurs URI dans le cas où l'on veut associer la variable producteur à plusieurs catégories. Chaque chaîne de caractère doit être une URI valide d'un niveau hiérarchique permettant à minima d'identifier le compartiment de la zone critique dans lequel est effectué la mesure (ex. pour une variable Water level, pouvoir distinguer si elle concerne les eaux karstiques ou

souterraines). Alternativement, il est désormais possible de référencer une URI de variable plutôt qu'une URI de catégorie si ce terme de type "Variable" (TYPE = Variable) existe déjà dans le thésaurus et décrit assez précisément la variable producteur. Dans ce cas le tableau JSON contiendra une et une seule URI représentant la variable Theia/OZCAR avec laquelle associer la variable producteur.

- **O keywords** : Tableau JSON d'Objet JSON. Chacun des objets JSON représentant un mot clé. L'Objet JSON représentant un mot clé est composé de 2 couples nom/valeur. Le premier couple avec nom = « keyword » est un champ de texte libre décrivant le mot-clé, et le second couple avec nom= « uri » est un champ de texte libre décrivant l'URI du mot clé dans un thésaurus publié
- **M featureOfInterest** : est une abstraction du monde réel pouvant être définie par des dimensions spatiales. Cette entité d'intérêt correspond à la cible de l'observation. L'entité d'intérêt est représentée ici comme un objet JSON composé du champ suivant :
 - **M samplingFeature** : c'est l'entité d'échantillonnage de la mesure. Ce champ permettra de localiser l'observation dans l'application. Cette entité est décrite selon des dimensions spatiales et peut être représenté par un objet GeoJSON composé des champs suivants :
 - **M name** : le nom de l'entité d'échantillonnage. Par exemple, le nom de la station de mesure pour une observation ponctuelle, le nom de la surface pour une observation concernant une surface. **Il est obligatoire que le nom choisi soit persistant dans le temps.**
 - **M geometry** : membre décrivant la géométrie de l'entité d'échantillonnage. L'objet **geometry** est composé des deux membres suivants :
 - **M type** : le type de la géométrie de la mesure. Les types d'emprise possible sont "Point", "MultiPoint", "LineStrings", "MultiLineStrings", "Polygon", et "MultiPolygon".
 - **M coordinates** : Pour un type "Point", le membre **coordinates** doit être une position, soit, un doublet de coordonnées longitude/latitude ou un triplet longitude/latitude/élévation. Les coordonnées sont exprimées en degré décimaux. Pour un type "MultiPoint" ou "LineString", le membre **coordinates** est un tableau JSON de positions. Pour un type "MultiLineString" ou "Polygon", le membre **coordinates** est un tableau JSON de tableau JSON de positions. Pour type "MultiPolygon" le membre **coordinates** est un tableau JSON de tableau JSON de tableau JSON de positions. Chaque position étant décrite par le doublet de coordonnées [longitude,latitude] ou un triplet [longitude,latitude,élévation].
 - **M** La constante **"type" : "Feature"** est attribué afin de faire correspondre l'objet au schéma GeoJSON.
 - **M** La constante **"properties" : {}** est attribué afin de faire correspondre l'objet au schéma GeoJSON.
 - **O procedure** : protocole d'observation décrivant le traitement utilisé pour produire un résultat sur une propriété observée de l'objet d'intérêt. Elle est souvent assurée par un instrument ou un capteur mais peut aussi être une chaîne de traitements, un observateur humain, une simulation. Cet objet permet de documenter le protocole d'acquisition de l'observation mais aussi les traitements effectués depuis l'acquisition. Cet objet est décrit par un objet JSON composé des champs suivants :
 - **O dataProduction** : objet JSON permettant de décrire l'acquisition de la donnée. Il est composé des champs suivants :
 - **O method** : chaîne de caractère décrivant la méthode d'acquisition.

- **R sensors** : collection d'objet JSON décrivant le dispositif ayant été utilisé pour la production de la donnée (capteur physique ou virtuel). Ce champ permet la mise en place de webservice standardisé d'échange de données (OGC SensorThings, OGC SOS). Les objets de la collection correspondent à un des 2 objets JSON suivants :

PhysicalSensor :

- **R model** : le modèle du capteur
 - **R manufacturer** : le fabricant du capteur
 - **O serialNumber** : le numéro de série du capteur
 - **O sensorType** : le type de capteur
 - **O operationalMode** : chaîne de caractère décrivant le mode opératoire du capteur
 - **O calibration** : chaîne de caractère décrivant la calibration du capteur.
 - **R activityPeriods** : collection d'objets JSON ayant les membres suivants :
 - **M dateBeg** : chaîne de caractère décrivant une date au format ISO 8601 "YYYY-MM-DDThh:mm:ssZ".
 - **M dateEnd** : chaîne de caractère décrivant une date au format ISO 8601 "YYYY-MM-DDThh:mm:ssZ".
 - **O documents** : collection des documents associés au capteur physique représentée par un tableau JSON. Les documents peuvent être de deux types : « Publication » ou « Manual ». L'objet **documents** est un tableau JSON. Chacun des éléments de ce tableau est un objet JSON comprenant deux membres. Un membre « type » pouvant prendre la valeur « Publication » ou « Manual », un membre « url » représentant l'url à laquelle est accessible le document. Cet élément optionnel permet de documenter les publications de références associées au capteur physique.
- VirtualSensor :**
- **M name** : le nom du capteur virtuel (Ex : nom du modèle)
 - **O parametrisationDescription** : la description des forçages et des paramètres appliqués au modèle
 - **O documents** : collection des documents associés au capteur virtuel représentée par un tableau JSON. Les documents peuvent être de deux type : « Publication » ou « Manual ». L'objet **documents** est un tableau JSON. Chacun des éléments de ce tableau est un objet JSON comprenant deux membres. Un membre « type » pouvant prendre la valeur « Publication » ou « Manual », un membre « url » représentant l'url à laquelle est accessible le document. Cet élément optionnel permet de documenter les publications de références associées au capteur virtuel.
- **O lineageInformations** : collection de d'objet JSON permettant de documenter les traitements post-acquisition subis par les données. Chacun des éléments de la collection sont composés des champs suivants :
 - **M processingDescription** : chaîne de caractères décrivant le traitement.
 - **O processingDate** : la date à laquelle le traitement a été effectué. Cette date est écrite au format ISO 8601 "YYYY-MM-DDThh:mm:ssZ".

- **M result** : Le résultat d'une observation est une estimation de la valeur d'une propriété d'un objet d'intérêt obtenu en utilisant une procédure spécifique. Pour des données de séries temporelles, l'objet result représente la série temporelle d'une variable donnée à une station donnée. L'objet result est décrit par l'objet JSON composé des membres suivants :
 - o **O missingValue** : code associé aux valeurs manquantes dans les fichiers de données.
 - o **R qualityFlag** : collection d'objet décrivant les flags de qualité et de leur signification associée aux mesures. Les objets JSON de la collection sont composés des membres suivants :
 - **M code**: chaîne de caractère décrivant le code flag utilisé dans les fichiers de données
 - **M description** : chaîne de caractères décrivant la signification du flag qualité représentant par le code.
 - o **R additionalValues** : Des valeurs additionnelles peuvent être ajoutées aux mesures de l'observation (exemples : incertitudes, erreurs, valeurs de paramètres liés à l'instrumentation...). Ces valeurs additionnelles sont présentes dans le fichier de données et sont décrites dans le format-pivot par une collection d'objet JSON. Les objets JSON de la collection sont composés des membres suivants :
 - **M columnName** : chaîne de caractère décrivant le nom de la colonne du fichier de données où sont détaillés les valeurs additionnelles de l'observation.
 - **M name** : chaîne de caractère décrivant le nom de la valeur additionnelle.
 - **R description** : chaîne de caractères décrivant la valeur additionnelle
 - **O unit** : chaîne de caractères décrivant l'unité de la valeur additionnelle.
 - o **M dataFile** : objet décrivant le fichier de données contenant les résultats de l'observation. Chaque objet est décrit par le champ suivant :
 - **M name** : le nom du fichier avec son extension. Ce champ permettra de retrouver le fichier de données dans le système. L'application pourra alors utiliser le fichier pour effectuer des opérations (par exemple, la visualisation).

La nomenclature des noms de fichier et son format est décrit dans le document description fichier de données.

- **M dataType** : le format du résultat de l'observation. Ce champ permettra à l'application d'effectuer des opérations sur l'observation comme par exemple de la visualisation. Ce champ peut prendre les valeurs suivantes : "Numeric", "Text", "Vector", "Raster", "Photo", "Video", "Audio", "Other".
 - **M timeSerie** : booléen égale à TRUE si l'observation est une série de données temporelles. Ce champ permettra à l'application d'effectuer des opérations sur l'observation comme par exemple de la visualisation.
 - **M temporalExtent** : décrit l'extension temporelle valide des données de l'observation. Si l'observation est une série temporelle : [Date de début de la série, date de fin de la série]. Si l'observation n'est pas une série temporelle, la période de validité de la mesure selon le scientifique (par exemple :
 - [Date de la mesure, Date de la mesure + 6 mois] pour une mesure que l'on peut considérer constante sur 6 mois.
 - [Date de la mesure, 9999-12-31T00:00:00Z] pour une mesure que l'on peut considérer atemporelle.). Si la date de la mesure est inconnue [9999-12-31T00:00:00Z, 9999-12-31T00:00:00Z]
- L'objet temporalExtent est composé des membres suivants :
- o **M dateBeg** : chaîne de caractère décrivant une date au format ISO 8601 "YYYY-MM-DDThh:mm:ssZ".
 - o **M dateEnd** : chaîne de caractère décrivant une date au format ISO 8601 "YYYY-MM-DDThh:mm:ssZ".
- **M processingLevel** : Niveau de traitement de l'observation. L'objet peut prendre une des valeurs suivantes : "Raw data", "Quality-controlled data", "Derived products".

Raw : les données n'ont pas subi de contrôle de qualité. **Quality-controlled data**: les données ont subi un contrôle de qualité comme une inspection visuelle ou statistique (ex : correction de la dérive du capteur, suppression des valeurs aberrantes). **Derived products**: les données sont le résultat d'une interprétation scientifique et/ou technique. Les données issues d'une interprétation scientifique ou basé sur un modèle qui utilise d'autres données et/ou est basé sur de fortes hypothèses. Ex : données construites à partir de données de plusieurs capteurs, données gap-filled, données issues de modèle...

Annexes 1: identifiants des observatoires

Observatoires	identifiants
AgrHyS	AGRH
AMMA-CATCH	CATC
Auradé	AURA
BVET	BVET
CRYOBS-CLIM	CRYO
Draix Bléone	DRAI
ERORUN	ERUN
H+	HPLU
HYBAM	HYBA
MSEC	MSEC
OBSERA	OBSE
OHGE	OHGE
OHM-CV	OHMC
OMERE	OMER

OPE	OPEA
ORACLE	ORAC
OSR	OSRC
Réal Collobrier	REAL
SNO Karst	KARS
SNO Tourbière	TOUR
Yzeron	YZER

Annexes 2: valeurs des énumérations

Valeurs possibles du membre type d'un élément de l'objet **fundings** :

<<Enumeration>> EnumFundingTypes
+ French research institutes
+ French universities and schools
+ Other universities and schools
+ Other research institutes
+ Research unit
+ French federative structure
+ Other federative structure
+ Research program
+ Other

Valeurs possibles du membre rôle d'un élément de l'objet **contacts**

<<Enumeration>> EnumContactPersonRoles
+ Principal investigator
+ Project leader
+ Project member
+ Data manager
+ Data collector

<<Enumeration>> EnumContactOrganisationRoles
+ Research Group

Valeurs possibles de l'objet **dataType** :

<<Enumeration>> EnumDataTypes
+ Numeric
+ Vector
+ Raster
+ Photo
+ Video
+ Text
+ Audio
+ Other

Valeurs possibles de l'objet **geologies** :

<<Enumeration>> EnumGeologies
+ Volcanic rocks
+ Plutonic rocks
+ Metamorphic rocks
+ Carbonate rocks
+ Quaternary soils
+ Other sedimentary rocks

Valeurs possibles de l'objet **climates** :

<<Enumeration>> EnumClimates
+ Oceanic climate
+ Tropical climate
+ Mediterranean climate
+ Mountain climate
+ Continental climate
+ Polar climate
+ Arid climate
+ Equatorial climate

Valeurs possibles de l'objet `inspireTheme` (cf: <http://inspire.ec.europa.eu/theme>)

<<Enumeration>> EnumInspireTheme
+ Addresses
+ Administrative units
+ Agricultural and aquaculture facilities
+ Area management/restriction/regulation zones and re
+ Atmospheric conditions
+ Bio-geographical regions
+ Buildings
+ Cadastral parcels
+ Coordinate reference systems
+ Elevation
+ Energy resources
+ Environmental monitoring facilities
+ Geographical grid systems
+ Geographical names
+ Geology
+ Habitats and biotopes
+ Human health and safety
+ Hydrography
+ Land cover
+ Land use
+ Meteorological geographical features
+ Mineral resources
+ Natural risk zones
+ Oceanographic geographical features
+ Orthoimagery
+ Population distribution - demography
+ Production and industrial facilities
+ Protected sites
+ Sea regions
+ Soil
+ Species distribution
+ Statistical units
+ Transport networks
+ Utility and governmental services

Valeurs possibles de l'objet `topicCategories` (cf : <http://inspire.ec.europa.eu/metadata-codelist/TopicCategory>)

<<Enumeration>> EnumTopicCategories
+ Boundaries
+ Biota
+ Climatology / Meteorology / Atmosphere
+ Economy
+ Elevation
+ Environment
+ Farming
+ Geoscientific Information
+ Health
+ Imagery / Base Maps / Earth Cover
+ Inland Waters
+ Intelligence / Military
+ Location
+ Oceans
+ Planning / Cadastre
+ Society
+ Structure
+ Transportation
+ Utilities / Communication

Valeurs possibles du membre type d'un élément de l'objet `document`:

<<Enumeration>> EnumDocumentTypes
+ Publication
+ Manual

Valeurs possibles de l'objet `processingLevel` :

<<Enumeration>> EnumProcessingLevels
+ Raw data
+ Quality-controlled data
+ Derived products