



DATA
TERRA



Gaïa Data : modèle pivot de métadonnées et colonne vertébrale terminologique

Atelier thématique : Linked data / Web sémantique

Journée Interopérabilité et Innovation - J2I 23 janvier 2024

WP3 Services transversales

GT Vocabulaires

Jean-Christophe Desconnets - jean-christophe.desconnets@ird.fr

ViQui Agazzi - victoria.agazzi@ird.fr





01 – Modèle pivot. Données et services système Terre

- + Quels métadonnées pour un catalogue fédéré ?
- + Démarche de construction collaborative
- + Les bases du modèle
- + Guide de bonnes pratiques
- + évolutions

02 – Colonne vertébrale terminologique

- + Quelle gestion des connaissances ?
- + Conception des thésaurus Gaïa Data / Data Terra
- + Articulation des connaissances
- + Guide de bonnes pratiques
- + Service terminologique FAIR (incubator)
- + évolutions



01

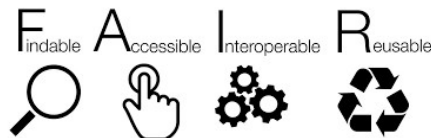
Modèle pivot de métadonnées



Modèle pivot : les besoins

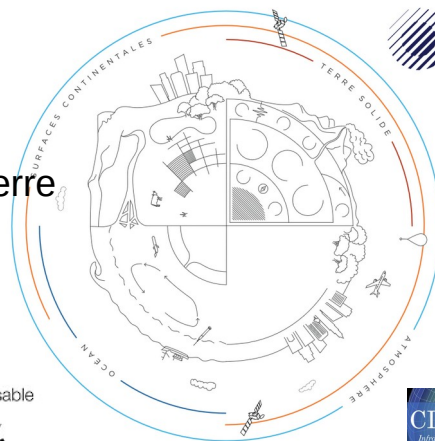
Modèle facilitant la fédération des catalogues existants du Système Terre

Modèle pour la découverte, la sélection et accès aux données / services facilitant la réutilisation



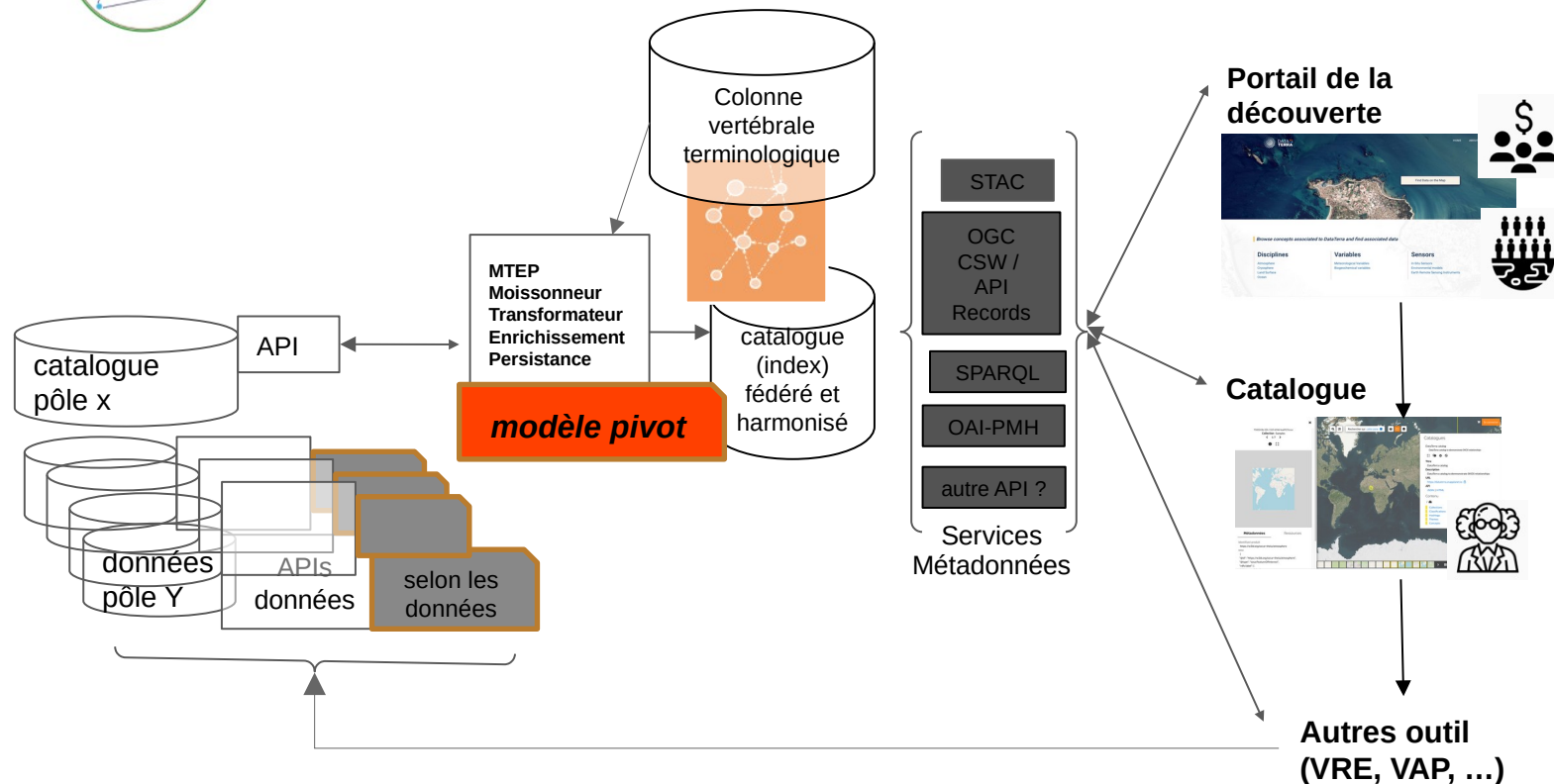
Modèle souple et extensible capable de gérer la diversité de pratiques de gestion { données, services, métadonnées } selon les domaines

Modèle permettant d'harmoniser les métadonnées (interopérabilité syntaxique) et d'enrichir sémantiquement les descriptions (interopérabilité sémantique).





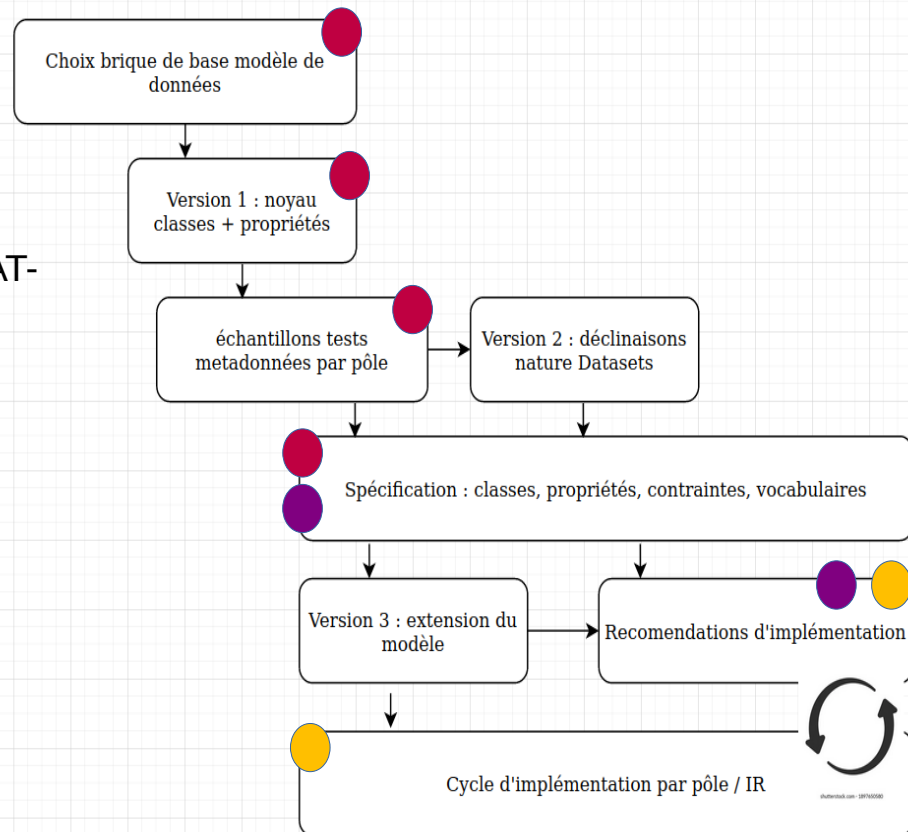
Place du modèle pivot dans l'architecture des services Gaïa Data





Démarche de construction collective : GT Vocabulaires

- Étape 1
 - choix briques de base : ontologies DCAT, GeoDCAT-AP.
 - confronter V1 avec des échantillons de métadonnées par pôle → [Exemples](#)
- Étape 2
 - partage des recommandations / spécifications dans le contexte Gaïa Data [Guide recommandations](#)
 - travail des correspondances (mappings) à partir des modèles des pôles / IR [Suivi des sujets traités](#)
- Étape 3 : amélioration des recommandations à partir des mappings conceptuels



Assemblage de classes et propriétés basées sur DCAT V2.0, complété par un ensemble de briques ontologiques pour assurer une description plus détaillée des données par domaine.

Le modèle pivot est un profil d'application de DCAT

Logique DCAT pour des descriptions en cascade :

→ **Dataset**

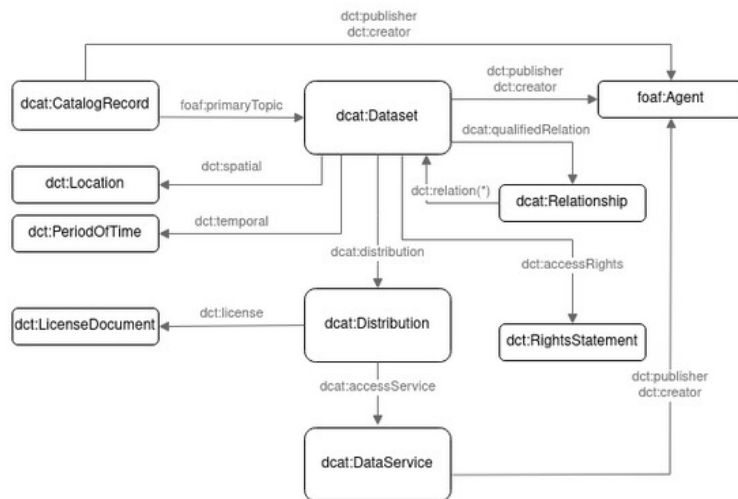
→ **Distribution(s)**

→ **Data Service(s)**

Permet découverte des datasets, et l'accès aux fichiers (descriptions distributions) et l'accès aux services (descriptions des API)

Descriptions harmonisés des contextes d'observation des Datasets

→ **Colonne vertébrale terminologique**



(*) dcat:Dataset is a sub-class of dcat:Resource

Diagramme UML simplifié des classes du cœur du profil d'application Data Terra



Les bonnes pratiques pour la FAIRisation des métadonnées

Guide de Recommandations :



https://gitlab.in2p3.fr/gaia-data/wp3-services/vocabulaires/dataterra_ap/guidelines_profil_application

- + *Spécifications techniques*
- + *Contextualisation Gaia Data*
- + *Exemples de code / descriptions*

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix : <http://example-dataterra.org/> .

:data_terra_dataset a dcat:Dataset ;
  dct:description "Description of the dataset."@en ;
  dct:title "Title of the dataset"@en ;
  dcat:contactPoint dataset_contactpoint ;
  dct:publisher dataset_publisher ;
  dcat:distribution dataset_distribution ;
  dcat:keyword "keyword1"@en, "keyword2"@en ;
  dct:spatial dataset_spatial ;
  dct:temporal dataset_temporal ;
  dcat:theme uri_dataset_variable, uri_dataset_foi, uri_
  dcat:accessRights dataset_access_rights ;
  dct:created dataset_creation_time ;
  dcat:qualifiedRelation relation_source_metadata, relat
```

Introduction

1. Terminologie
2. Motivations
3. Cas d'utilisation

Recommandations pour la création des métadonnées

1. Le profil d'application Data Terra
2. Spécification
 1. Classe : dcat:Dataset
 2. Classe : dcat:Distribution
 3. Classe : dcat:DataService
 4. Classe : foaf:Agent
 5. Classe : dct:Location
 6. Classe : dct:PeriodOfTime
 7. Classe : dcat:Relationship
 8. Classe : dct:LicenseDocument
 9. Classe : dct:RightsStatement
 10. Classe : dcat:CatalogRecord
3. Mappings
4. Références



Les évolutions (court terme)

MVP(v1) : Portail de la découverte



Stabiliser un sous-ensemble du modèle pivot avec une démarche sur 13 cas pratiques transversales (~60 datasets) : de la découverte aux traitements sous environnement virtuels de recherche (VRE).

		ir_poles_pepr id_use_case				
Datasets vs. Use cases		ODATIS	THEIA / OZCAR / AERIS	THEIA / OZCAR / AERIS	THEIA/FORM @TER	THEIA/AERIS/ FORM@TER
collection_name	dataset_name	3	4	5	6	7
Données de salinité GOSUD	2BPR6_2021_a_NRT_Gosud_V3.nc.gz					
Données SWOT KaRIn ocean	produits hauteur de mer	✓				
	courants de surface / drifters	✓				
Données de bouées dérivantes						
Produit: INSITU_GLO_PHYBGCWAV_DIS	cnems_obs-ins_glo_phybgcwav_mynrt_na_	✓				
CRETE_MYNRT_013_030	irr - reference					
Données in-situ météo	CATC_DAT_AL_Met_Od		✓			
Theia/OZCAR	OHGE_DAT_meteo			✓		
Données in-situ Météo France	à venir			✓		✓
MNT	MNT MERIT		✓			
	MNT Strengbach			✓		
Occupation sols	West Africa Land Use Land Cover 2013 - 2 km (layers=sb:swa_2013lulc_2km_geo) USGS		✓			
	ESA world cover - 10 m - 2021		✓			
	ESA Africa - 20m- 2015-2016		✓			
	OSO THEIA 10m raster		✓	✓	✓	✓
	OSO THEIA 10m vecteur			✓	✓	✓
Données de géophysique	à venir			✓		
Data cube Sentinel 2	à venir			✓		
Produits thermiques LST / THEIA	LST / THEIA					✓
Produits Réseaux Capteurs in-situ (thermique - observil/OZCAR, etc)	à venir					✓
Ground Deformation Monitoring with OPTICAL image time series	GDM-OPT/ForM@Ter				✓	
Interférogrammes à partir des données Sentinel-1, ainsi que des séries temporelles de déplacements	FLATSIM/ForM@Ter				✓	



02

Colonne vertébrale terminologique



01 – Modèle pivot. Données et services système Terre

- + Quels métadonnées pour un catalogue fédéré ?
- + Démarche de construction collaborative
- + Les bases du modèle
- + Guide de bonnes pratiques
- + évolutions

02 – Colonne vertébrale terminologique

- + Quelle gestion des connaissances ?
- + Conception des thésaurus
- + Articulation des connaissances
- + Guide de bonnes pratiques
- + Service FAIR (incubator)
- + évolutions

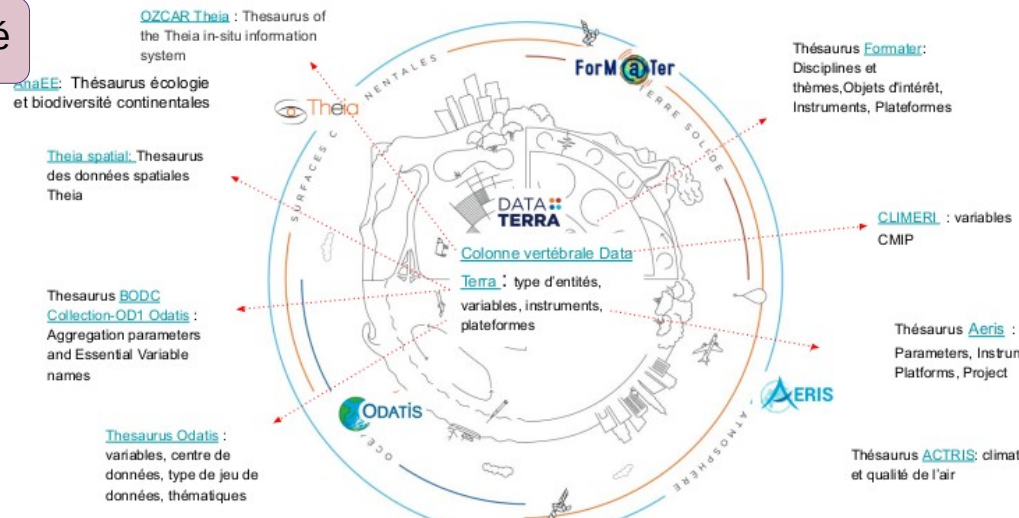
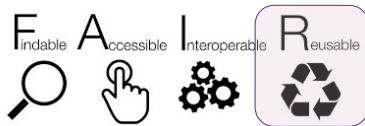


Gestion des connaissances : quels besoins ?

Vision de gestion en cohérence pour Gaïa Data
Catalogues pôles/IR → Catalogue fédéré
Vocabulaires pôles/IR → Vocabulaire fédéré

Formalisation des connaissances pour
l'harmonisation et l'articulation des vocabulaires
compartimentés dans les pôles de données

Interopérabilité sémantique : construire les liens
entre les vocabulaires dans les pôles/IR et des
référentiels dans les domaines du système Terre

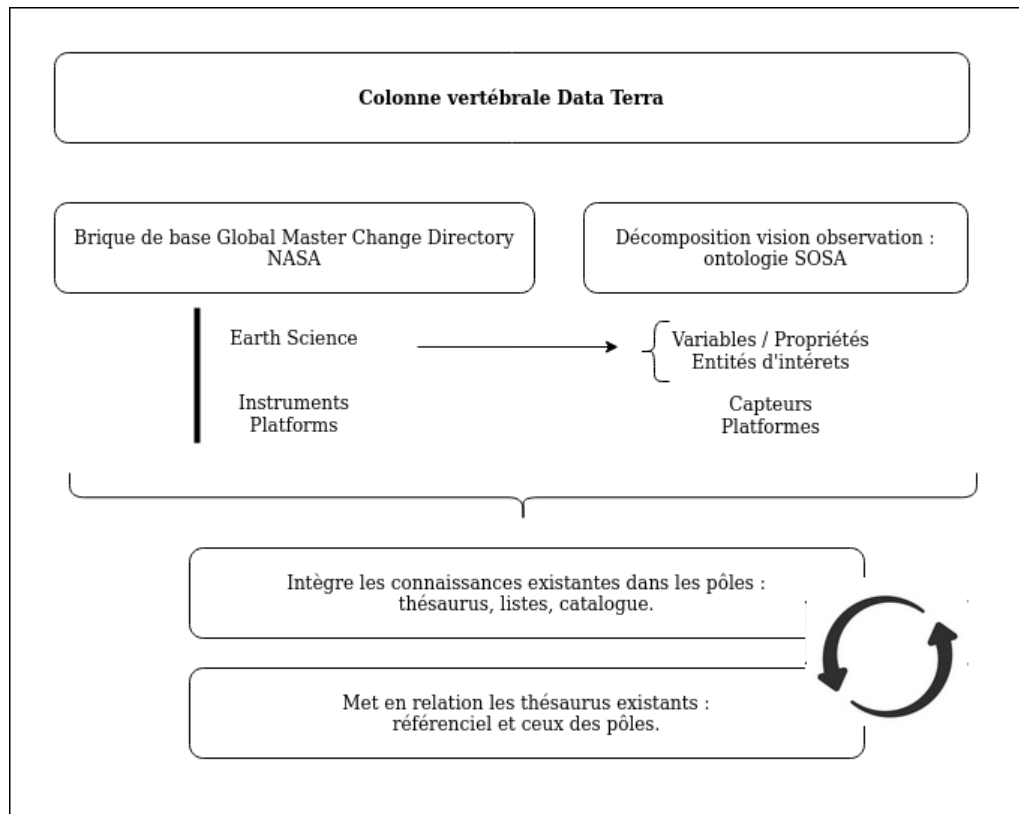




Conception colonne vertébrale terminologique

Ensemble de thésaurus conçus pour harmoniser et articuler la connaissance autours de l'observation intégrée du système Terre.

La colonne vertébrale alimente les métadonnées harmonisés du catalogue fédéré.





Structure des thésaurus

Ontologie SKOS

(Simple Knowledge Organization System)

Enrichie par une catégorisation selon les classes de l'ontologie SOSA

(Sensor, Observations, Sample and Actuator)

Variables
ce qui peut être
observé

4.3.2.1 **sosa:ObservableProperty**

IRI: <http://www.w3.org/ns/sosa/ObservableProperty>

a OWL Class

Observable Property - An observable quality (property, characteristic) of a *FeatureOfInterest*.

Objets d'intérêt (type)

4.6.2.1 **sosa:FeatureOfInterest**

IRI: <http://www.w3.org/ns/sosa/FeatureOfInterest>

a OWL Class

Feature Of Interest - The thing whose property is being estimated or calculated in the course of an *Observation* to arrive at a *Result*, or whose property is being manipulated by an *Actuator*, or which is being sampled or transformed in an act of *Sampling*.

Capteurs (type)

4.3.2.5 **sosa:Sensor**

IRI: <http://www.w3.org/ns/sosa/Sensor>

a OWL Class

Sensor - Device, agent (including humans), or software (simulation) involved in, or implementing, a *Procedure*. *Sensors* respond to a *Stimulus*, e.g., a change in the environment, or *Input* data composed from the *Results* of prior *Observations*, and generate a *Result*. *Sensors* can be hosted by *Platforms*.

Plateformes

4.9.2.1 **sosa:Platform**

IRI: <http://www.w3.org/ns/sosa/Platform>

a OWL Class

Platform - A *Platform* is an entity that *hosts* other entities, particularly *Sensors*, *Actuators*, *Samplers*, and other *Platforms*.

Structure qui permet de naviguer
dans la connaissance



Articulation des connaissances des pôles / IR

+ Mise en relation / articulation des connaissances avec des propriétés de l'ontologie skos : vers GCMD et les pôles.

skos:mappingRelation
skos:closeMatch
skos:exactMatch
skos:broadMatch
skos:narrowMatch
skos:relatedMatch

```
11 :c_46fa2569 a skos:Concept, <http://www.w3.org/ns/sosa/FeatureOfInterest>;
12   skos:prefLabel "Land surface"@en;
13   skos:exactMatch <https://w3id.org/ozcar-theia/c_3d4ce97d>,
14   <https://vocab.aeris-data.fr/parameter/0b47d139-cfde-c380-cd13-59ab6cc7ad32>,
15   <https://gcmd.earthdata.nasa.gov/kms/concept/6a426480-c58f-4b6b-8e35-0975b7f6edb5>;
16   skos:definition "Refers to the surface area and features on the surface of the Earth."@en;
```

Source : thésaurus types d'entités d'intérêts Data Terra

```
650
651 :c_b98112ca a skos:Concept, <http://www.w3.org/ns/sosa/ObservableProperty>;
652   skos:prefLabel "Dissolved organic carbon mass concentration"@en;
653   skos:inScheme <https://terra-vocabulary.org/ncl/earthvariables>;
654   skos:broader :c_499e75b6;
655   skos:narrowMatch <https://w3id.org/ozcar-theia/c_5c4e993c> .
```

Source : thésaurus variables Data Terra



Les bonnes pratiques pour la FAIRisation des thésaurus



Guide de Recommandations :

https://gitlab.in2p3.fr/gaia-data/wp3-services/vocabulaires/linked_data_registry/guidelines_thesaurus

- + *Modèle conceptuel*
- + *Spécifications techniques*
- + *Exemples de code*

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix sosa: <http://www.w3.org/ns/sosa/> .
@prefix ssn: <http://www.w3.org/ns/ssn/> .
@prefix : <https://terra-vocabulary.org/ncl/earthvariables/> .

:c_13c84e4c a skos:Concept, sosa:ObservableProperty ;
  ssn:isPropertyOf <https://terra-vocabulary.org/ncl/EarthFeatureType/c_fb12f2ac> ; #
  skos:prefLabel "Speed profile"@en ;
  skos:inScheme <https://terra-vocabulary.org/ncl/earthvariables> ;
  skos:definition "Ocean current velocity throughout the water column."@en ;
  skos:exactMatch <https://gcmd.earthdata.nasa.gov/kms/concept/7744f889-b25e-4d0e-bcf
```

Introduction

1. Terminologie
2. Motivations

Recommandations pour la création des thésaurus

1. Modèle conceptuel pour la construction des thésaurus
2. Spécification
 1. Classe : skos:ConceptScheme
 2. Classe : skos:Concept
 3. Classe : skos:Collection
 4. Classe : sosa:FeatureOfInterest
 5. Classe : sosa:ObservableProperty
 6. Classe : sosa:Sensor
 7. Classe : sosa:Platform
3. Bonnes pratiques FAIR
4. Références



Service FAIR (incubator) <https://terra-vocabulary.org/ncl/FAIR-Incubator>

Register: Incubateurs de concepts FAIR de Terra Vocabulary

URI: <https://terra-vocabulary.org/ncl/FAIR-Incubator>





Cette collection contient une liste préliminaire des registres et concepts à FAIRiser

[Core metadata](#)[All properties](#)[Download](#)

Contents

Show entries

Search:

Name ▲	Notation ⇅	Description ⇅	Types ⇅	Status ⇅
 Earth features types	earthfeaturetype	The current thesaurus contains concepts related to the types ...	Container , concept scheme , Register	experimental
 Earth science platforms	earthscienceplatforms	The current thesaurus contains concepts related to data acqui...	Container , concept scheme , Register	experimental
 Earth Science Sensors	earthsciencesensors	The current thesaurus contains concepts related to data acqui...	Container , Register , concept scheme	experimental
 Earth science variables	earthsciencevariables	The current thesaurus contains concepts related to Earth scie...	Container , concept scheme , Register	experimental

Showing 1 to 4 of 4 entries

[Previous](#)[1](#)[Next](#)



Les évolutions

Terminologies à intégrer :

- [Form@ter](#) / Sismologie (EaSy Data)
- Variables Theia Spatial
- Climeri : Variables prioritaires

Technique :

- Meilleure utilisation de l'API LDRegistry
- Brique pour faciliter VocBench vers LDRegistry

Register: Earth features types

URI: <https://terra-vocabulary.org/ncl/FAIR-Incubator/earthfeaturetype>

The current thesaurus contains concepts related to the types of physical Earth science features or entities.

Contents (tree view)

- [Solid Earth](#) Refers to "the Earth beneath our feet", the planet's solid su...
- [Oceans](#) Ocean is a very large expanse of sea, in particular, each of ...
- [Biosphere](#) The biosphere is made up of the parts of Earth where life exi...
- [Land surface](#) Refers to the surface area and features on the surface of the...
- [Atmosphere](#) The gaseous envelope surrounding the Earth in a several kilom...
- [Terrestrial hydrosphere](#) Terrestrial - related to the Earth, A hydrosphere is the tota...
 - [Coastline](#) The actual contour of the continents.
 - [Ground water](#) Ground water in its broadest sense includes all subsurface wa...
 - [Surface water](#) Surface water is water located on top of the Earth's surface...
 - [Basin](#) An area of land where surface water from rain, melting snow, ...
 - [Estuary](#) The place where animals or plants live in a semi-enclosed coa...
 - [Lake](#) An area of land where surface water from rain, melting snow, ...
 - [Pond](#) A body of standing water, either natural or man-made, that is...
 - [River](#) A natural watercourse, usually freshwater, flowing towards an...
 - [Spring](#) A natural flow of water from the sub-surface to the surface. ...
- [Cryosphere](#) The cryosphere is the frozen water part of the Earth system. ...



Idées clés

Le modèle pivot nous permet de décrire de façon harmonisée, la diversité de données et services liés à l'observation et compréhension du système Terre.

Les thésaurus de la colonne vertébrale Gaïa Data / Data Terra harmonisent et articulent les connaissances liés aux observations du système Terre.

Le modèle pivot et la colonne vertébrale appliquent les principes FAIR, en réutilisant des briques du web sémantique, profitant ainsi de sa souplesse dans les descriptions et la mise en relation des ressources à travers le Linked Data.



DATA
TERRA



WP3 Services transversales GT Vocabulaires

Jean-Christophe Desconnets - jean-christophe.desconnets@ird.fr
Viqui Agazzi - victoria.agazzi@ird.fr

www.gaia-data.org

Ce projet a bénéficié d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche au titre du programme Investissements d'Avenir Equipex+.

