

Session « Définition et rédaction des métadonnées »

# Vocabulaires contrôlés et données FAIR

Source : Utiliser des vocabulaires contrôlés pour FAIRiser les données  
Séminaire SIST2022

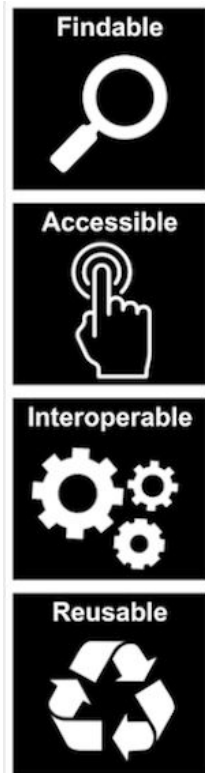
*GT Catalogue et Vocabulaires Data Terra/Gaia Data  
Victoria Agazzi - Data Terra*

*Ecole Thématique DATA SDUE - 12 Octobre 2022, Strasbourg.*

# Vocabulaires contrôlés et données FAIR

1. {Données, Metadonnées, Vocabulaires} FAIR
2. Panorama des thésaurus des sciences de la Terre et du vivant
3. Thésaurus IR Data Terra

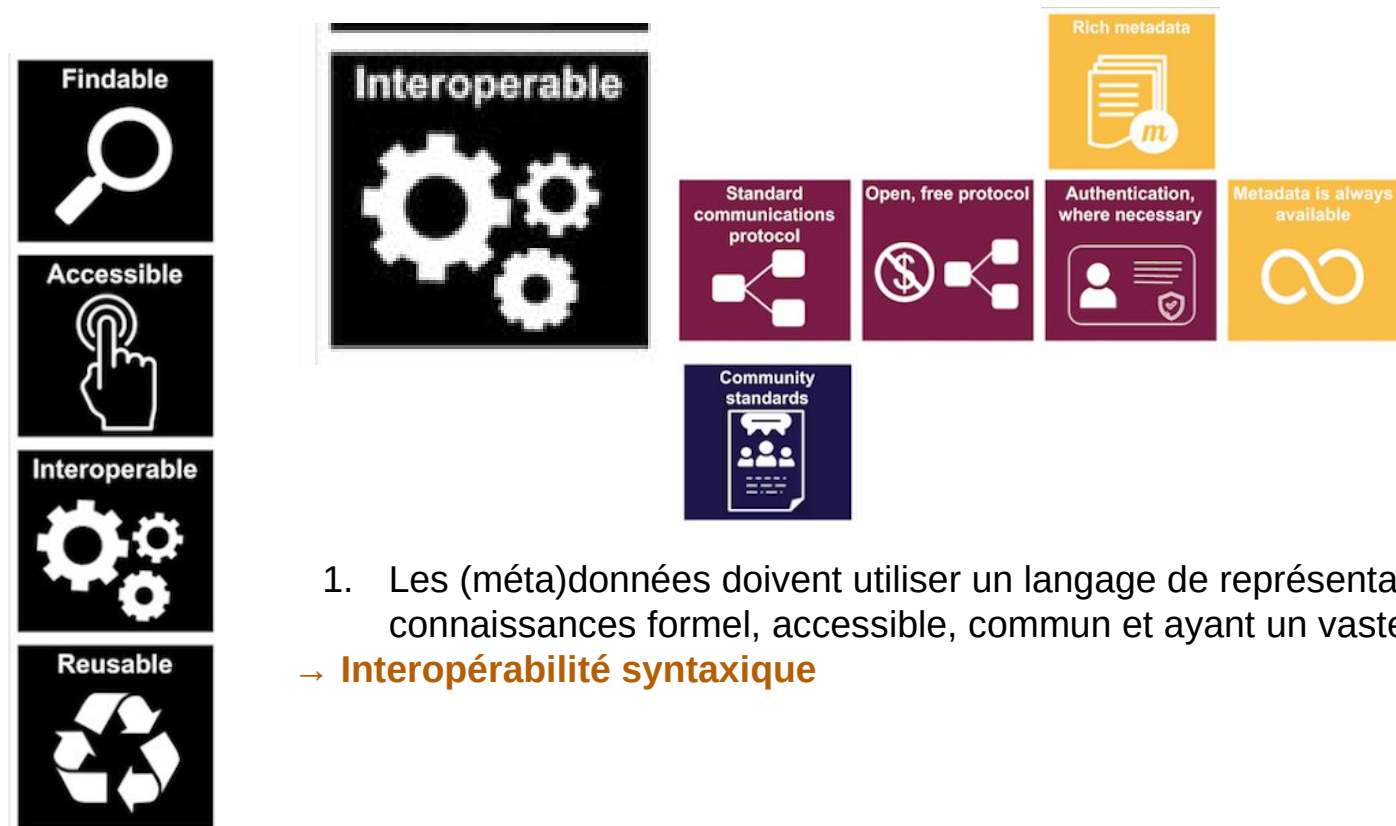
# Interêt des données FAIR



**Objectif** : les données doivent pouvoir être réutilisées et mises en relation avec d'autres données au delà de sa propre base de données

*“L’interopérabilité des données correspond à leur capacité à être intégrées avec d’autres données et à être utilisées et interprétées par des applications et des processus d’analyse et ce de manière automatique”*

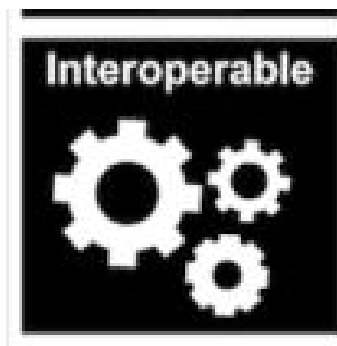
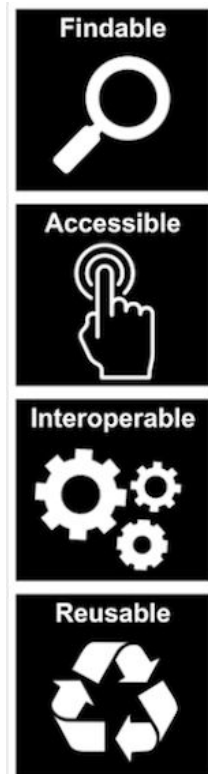
# Assurer l'interopérabilité des bases de données



1. Les (méta)données doivent utiliser un langage de représentation des connaissances formel, accessible, commun et ayant un vaste champ d'application

→ **Interopérabilité syntaxique**

# Assurer l'interopérabilité des bases de données



1. Les **(méta)données** doivent utiliser des **vocabulaires** qui suivent les principes **FAIR**
  2. Les **(méta)données** doivent **inclure des références** vers d'autres (méta)données
- **Interopérabilité sémantique**

# Assurer l'interopérabilité sémantique

Associer une signification aux données, les positionner dans un domaine de connaissance

Développer des **vocabulaires et de schémas standard** pour décrire les données et les liens entre les données.

Décrire les **données avec des métadonnées** en provenance des **vocabulaires formalisés et partagés**.



# Définitions : notion de vocabulaire

**Ensemble de termes définis, utilisés dans des métadonnées pour associer un sens aux données**

## **Vocabulaire thématique**

exemple : télédétection, agro pastoralisme

## **Vocabulaire sur l'observation** (variable, instrument...)

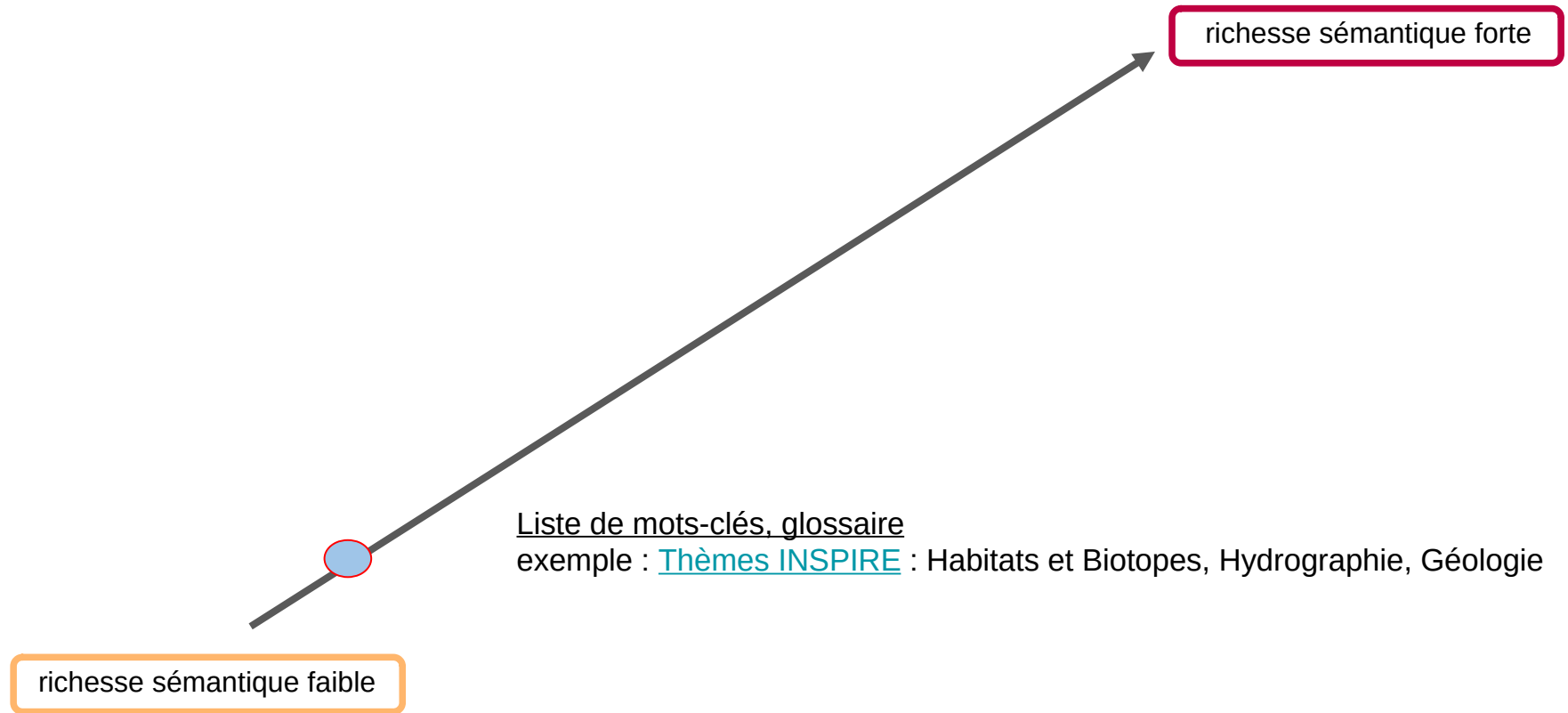
exemple : salinité, température de l'eau, cumul de précipitations ; bouée argo, pluviomètre

## **Vocabulaire géographique**

exemple : Sahel, amazonie, Brésil

**Objectifs** : décrire des données avec des vocabulaires communs permet de les connecter entre elles selon diverses optiques

# Divers modèles de formalisation des connaissances





# Divers modèles de formalisation des connaissances

INSPIRE  
Commission européenne  
Système de publication de registres

Commission Européenne > INSPIRE > Système de publication de registres pour INSPIRE > Registre de thème INSPIRE > Hydrographie

**Hydrographie**

Help us improving the **Re3gistry software!** Please fill our quick survey at <http://europa.eu/1Bn84Ct>

Identifiant: <http://inspire.ec.europa.eu/theme/hy>  
This version: <http://inspire.ec.europa.eu/theme/hy:3>  
Latest version: <http://inspire.ec.europa.eu/theme/hy>  
Previous versions: <http://inspire.ec.europa.eu/theme/hy:2>  
<http://inspire.ec.europa.eu/theme/hy:1>

Etiquette: **Hydrographie**  
Définition: Éléments hydrographiques, y compris les zones maritimes ainsi que toutes les autres masses d'eau et les éléments qui y sont liés, y compris les bassins et sous-bassins hydrographiques. Conformément, le cas échéant, aux définitions établies par la directive 2000/60/CE du Parlement européen et du Conseil du 23 octobre 2000 établissant un cadre pour une politique communautaire dans le domaine de l'eau [2] et sous forme de réseaux.

Description: The theme Hydrography is a basic reference component and, therefore, of interest for many users and uses.  
[Not available in French]  
For mapping purposes (to provide a map background for orientation and to understand place relationships), it includes

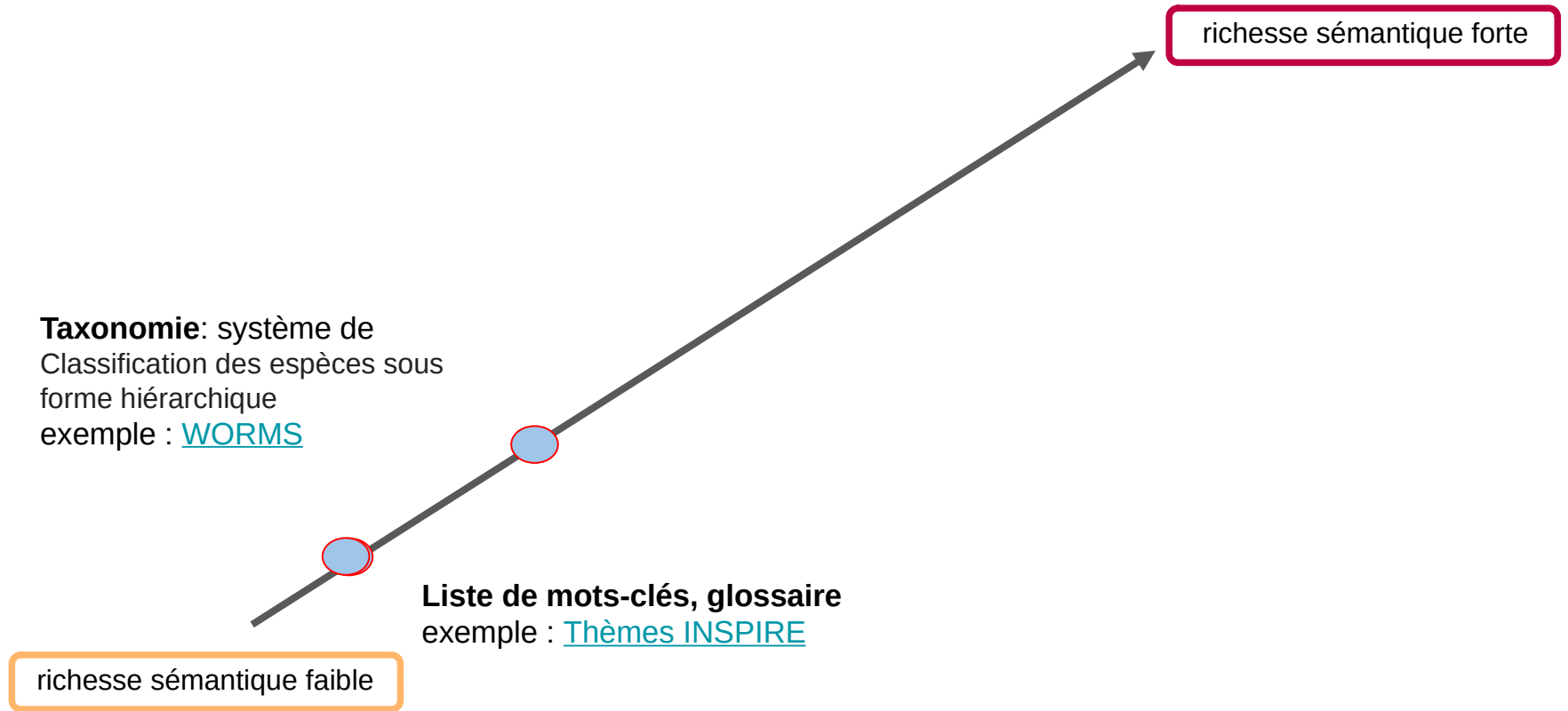
richesse sémantique forte

Liste de mots-clés, glossaire

exemple : [Thèmes INSPIRE](#) : Habitats et Biotopes, **Hydrographie**, Géologie

richesse sémantique faible

# Divers modèles de formalisation des connaissances



# Divers modèles de formalisation des connaissances

**Taxonomie:** système de Classification des espèces sous forme hiérarchique  
exemple : [WORMS](#)

The screenshot shows the WoRMS website interface. At the top, the logo 'WoRMS World Register of Marine Species' is visible. Below the logo is a navigation bar with links: Home, About, Subregisters, Users, Photogallery, Documents, LifeWatch. A search bar is also present. The main content area displays 'WoRMS taxon details' for *Licmophora abbreviata* C.Agardh, 1831. The details include:

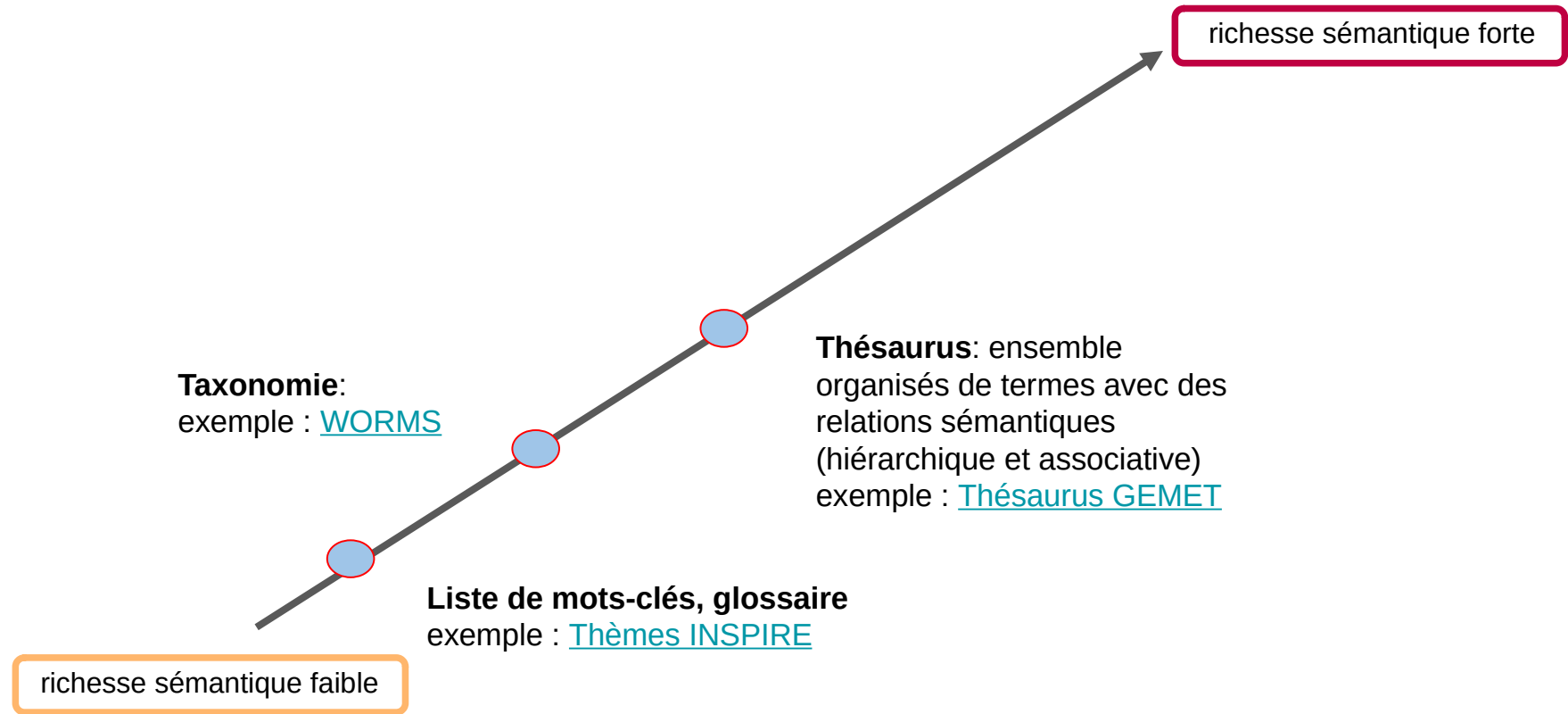
- AphiaID:** 157062 (um:lsid:marinespecies.org:taxname:157062)
- Classification:** Biota > Chromista (Kingdom) > Harosa (Subkingdom) > Heterokonta (Infrakingdom) > Ochrophyta (Phylum) > Khakista (Subphylum) > Bacillariophyceae (Class) > Fragilariophycidae (Subclass) > Fragilariophycanae (Superorder) > Licmophorales (Order) > Licmophoraceae (Family) > Licmophora (Genus) > Licmophora abbreviata (Species)
- Status:** accepted
- Rank:** Species
- Parent:** Licmophora C.A. Agardh, 1827
- Orig. name:** Licmophora abbreviata C.Agardh, 1831
- Environment:** marine, terrestrial
- Fossil range:** recent only
- Original description:** Not documented
- Descriptive notes:** Distribution geographic region not obtained; recorded as being found in the Bay of Fundy
- Taxonomic citation:** Kocolek, J.P.; Blanco, S.; Coete, M.; Ector, L.; Liu, Y.; Karthick, B.; Kulikovskiy, M.; Lundholm, N.; Ludwig, T.; Potapova, M.; Rimet, F.; Sabbe, K.; Sala, S.; Sar, E.; Taylor, J.; Van de Vijver, B.; Wetzel, C.E.; Williams, D.M.; Witkowski, A.; Witkowski, J. (2022). DiatomBase. *Licmophora abbreviata* C.Agardh, 1831. Accessed through: World Register of Marine Species at: <https://www.marinespecies.org/aphia.php?p=taxdetails&id=157062> on 2022-06-02

richesse sémantique forte

**Liste de mots-clés, glossaire**  
exemple : [Thèmes INSPIRE](#)

richesse sémantique faible

# Divers modèles de formalisation des connaissances



# Divers modèles de formalisation des connaissances

**hydrosphère**

**Definition**

The waters of the Earth, as distinguished from the rocks (lithosphere), living things (biosphere), and the air (atmosphere). Includes the waters of the ocean; rivers, lakes, and other bodies of surface water in liquid form on the continents; snow, ice, and glaciers; and liquid water, ice, and water vapour in both the unsaturated and saturated zones below the land surface. Included by some, but excluded by others, is water in the atmosphere, which includes water vapour, clouds, and all forms of precipitation while still in the atmosphere.

*Definition is not available for the current language.*

**Related terms**

**Narrower:**

[cours d'eau](#) | [cycle de l'eau](#) | [eaux \(géographie\)](#) | [entité hydrologique](#) | [lac](#) | [mer](#) | [réservoir d'eau](#)

**Themes:** [eau](#)

**Group:** [HYDROSPHERE](#) ([eau douce](#), [eau de mer](#), [eaux](#))

**Other relations**

**Has close match:** [UMTHES: Hydrosphère](#)

**Wikipedia article:** [Hydrosphère](#)

richesse sémantique forte

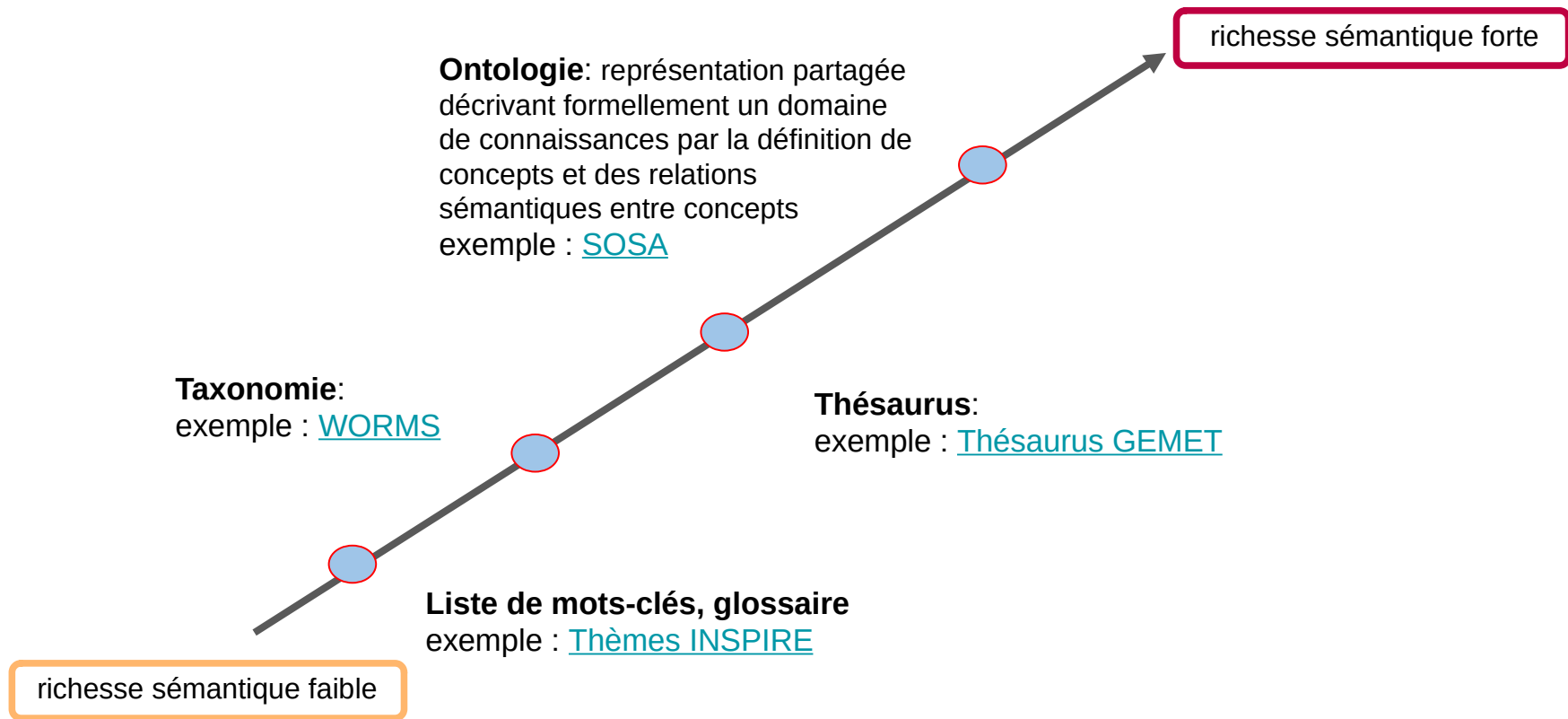
**Thésaurus:** ensemble organisés de termes avec des relations sémantiques (hiérarchique et associative)  
exemple : [Thésaurus GEMET](#)

lès, glossaire

exemple : [Thèmes INSPIRE](#)

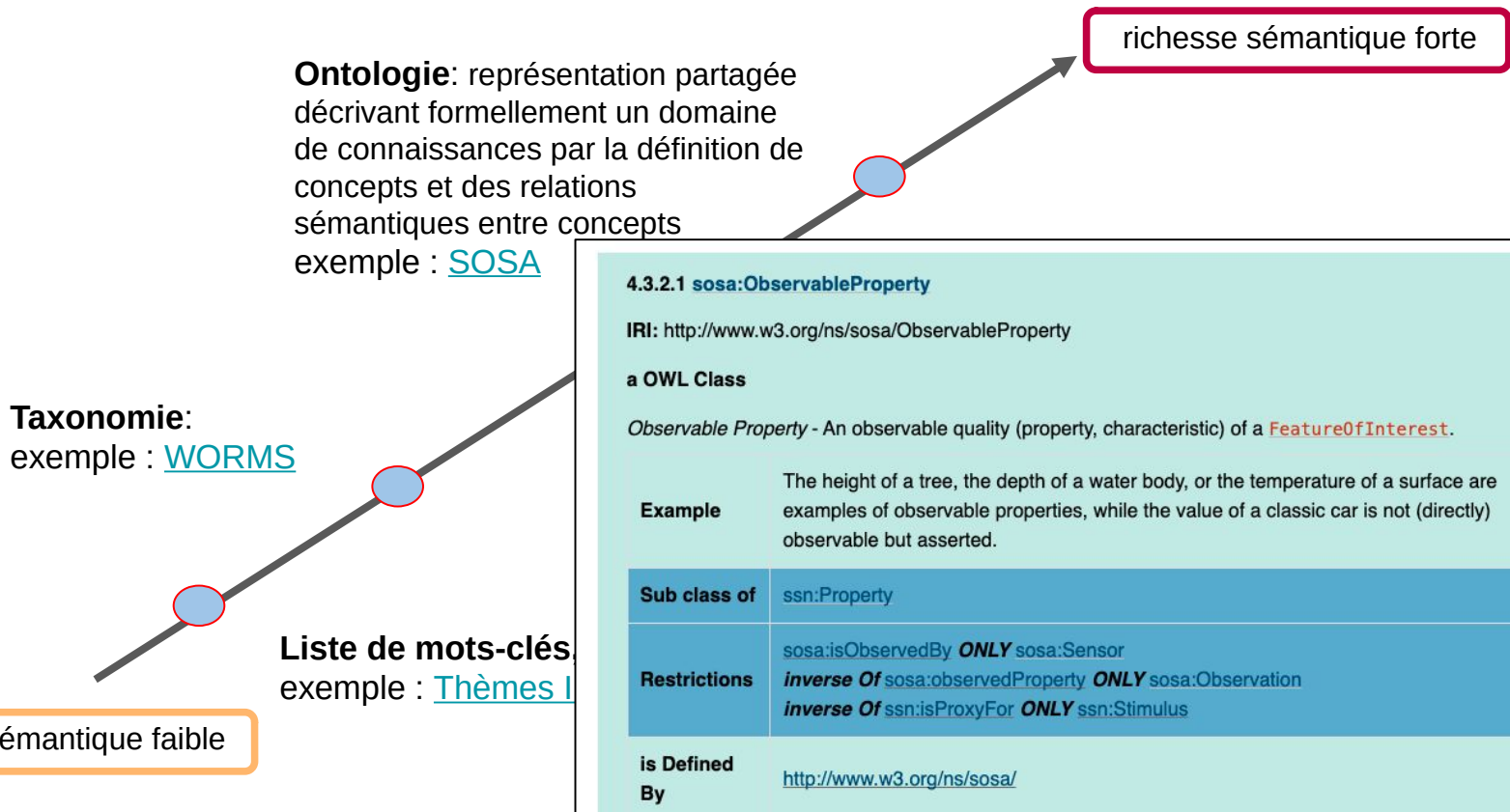
richesse sémantique faible

# Divers modèles de formalisation des connaissances



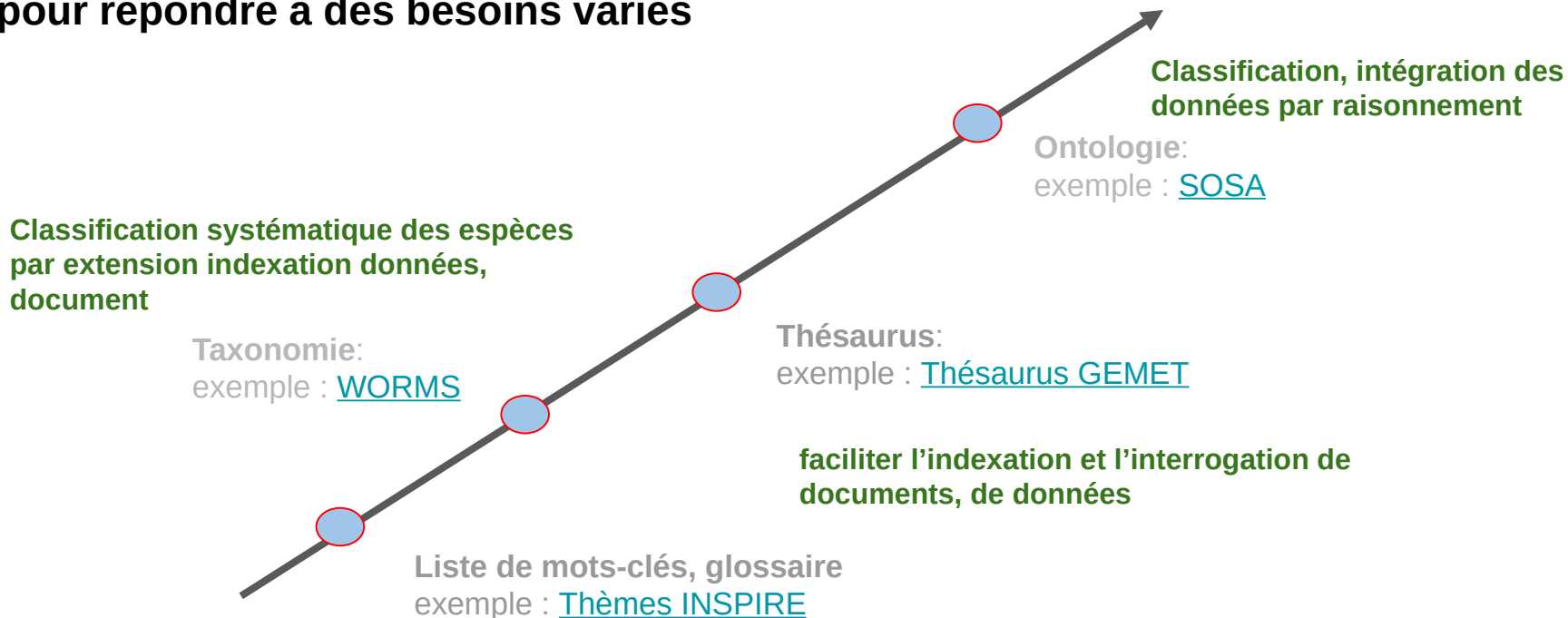
**Spectre des artefacts sémantiques** ou niveau **formalisation de la connaissance**

# Divers modèles de formalisation des connaissances



# Divers modèles de formalisation des connaissances

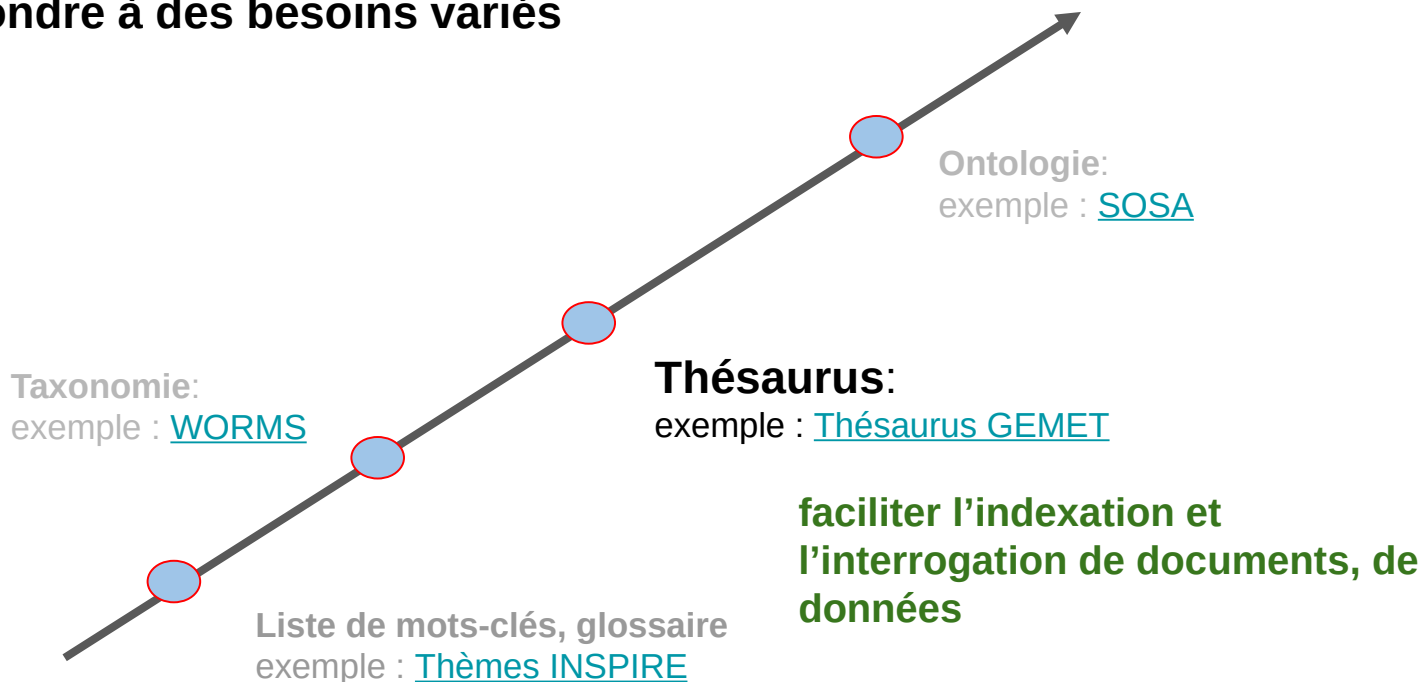
pour répondre à des besoins variés



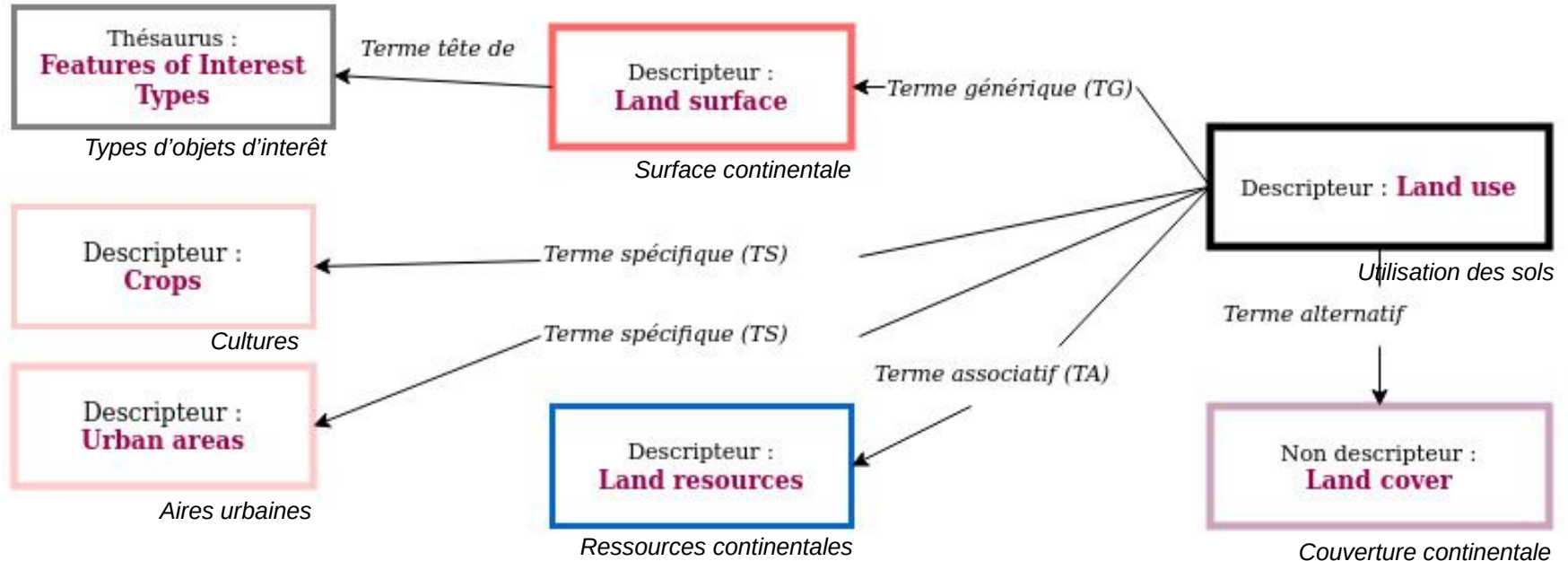


# Divers modèles d'artefact sémantiques

pour répondre à des besoins variés



# Notion de thésaurus : exemple pour *Land surface*



# Utilisation des termes issus des thésaurus dans les métadonnées

## Vision concrète Exemple : SENTINEL-2 MSI Level-1C User Product

```
## Image Dataset
:S2A_MSIL1C_20200709T105031_N0209_R051_T31TCJ_20200709T125257.SAFE
  a dcat:Dataset ;
  dct:description "This Dataset is a concatenation of differents observable bands by Sentinel-2.
                  Some processing can be use on the same product, here the product is level L1C."@en ;
  dct:title "SENTINEL-2 MSI Level-1C User Product : S2A_MSIL1C_20200709T105031_N0209_R051_T31TCJ_20200709T125257"@en ;
```

...

|# Lien vers concepts dans thésaurus Data Terra -----

```
dcat:theme <https://terra-vocabulary.org/ncl/EarthFeatureType/c_46fa2569> ; # 'Land surface'
dcat:theme <https://terra-vocabulary.org/ncl/earthvariables/c_40d601fb> ; # 'TOA reflectance'
dcat:theme <https://terra-vocabulary.org/ncl/sciencesensorstypes/c_ef837376> ; # 'MSI'
dcat:theme <https://terra-vocabulary.org/ncl/scienceplatforms/c_2bc97847>,
           <https://terra-vocabulary.org/ncl/scienceplatforms/c_1ccda463> ; # 'SENTINEL-2', 'SENTINEL-2A'
```

Propriété dcat portant des URI des vocabulaires contrôlés

URI vers les thésaurus colonne vertébrale Data Terra

Valeurs de propriété skos:prefLabel pour les concepts pointés par les URI

*Liens entre métadonnées et terminologies/ontologies*

# Etat des lieux des thésaurus des sciences de la Terre et du Vivant

(basé sur l'étude du GT Catalogue Data Terra et mis à jour en juin 2022, **non exhaustif**)

- Nature
- Maturité sémantique
- Niveau de Fairisation

## 1 - Panorama des thésaurus dit de “référence”

		Pôles de données								standard	API SPARQL (skosmos, JIRoad, D. MacRae, search4)	API (REST, autre)	UR
		Type	Odis	Aeris	Thela	Ozcar	Formater	Climeri	PNDB				
1	Last update : 2022-10-06												
2	Vocabulaires de référence												
3	WORMS	thésaurus								RDF/XML	ND	<a href="https://www.marinespecies.org/afsi/">https://www.marinespecies.org/afsi/</a>	NI
4	TaxRef	taxonomie								RDF/XML, N3, JSON-L	<a href="http://sparql.acroportal.lirmm.fr/">http://sparql.acroportal.lirmm.fr/</a>	<a href="http://data.acroportal.lirmm.fr/">http://data.acroportal.lirmm.fr/</a>	NI
5	NVS NERC Voc Server												
6	BODC collection -P01 (Voc général)	thésaurus								RDF/SKOS (RDF/XML)			ht
7	BODC collection -P02 (Découverte)	thésaurus								RDF/SKOS (RDF/XML)	<a href="https://vocab.nerc.ac.uk/spe/">https://vocab.nerc.ac.uk/spe/</a>	<a href="https://vocab.seadatanet.org/search">https://vocab.seadatanet.org/search</a>	ht
8	BODC collection -P07 (mapping to Convention CF)	thésaurus								RDF/SKOS (RDF/XML)			ht
9													
10	EPOS Vocabulary Registry	thésaurus								RDF, TTL, RDF/XML	<a href="https://registry.epos-eu.org/">https://registry.epos-eu.org/</a>	ND	ht
11	SANDRE groupe de paramètres	thésaurus								RDF/SKOS, OWL, JSK	<a href="https://www.sandre.eaufrance.fr/">https://www.sandre.eaufrance.fr/</a>	<a href="https://www.sandre.eaufrance.fr/sol/">https://www.sandre.eaufrance.fr/sol/</a>	NI
12	GCMD - NASA												
13	GCMD-Instruments	thésaurus								RDF/SKOS (RDF/XML)			ht
14	GCMD-Locations	thésaurus								RDF/SKOS (RDF/XML)	ND	<a href="https://gcmd.earthdata.nasa.gov/kms/">https://gcmd.earthdata.nasa.gov/kms/</a>	ht
15	GCMD-Platforms	thésaurus								RDF/SKOS (RDF/XML)			ht
16	GCMD-Science Keywords	thésaurus								RDF/SKOS (RDF/XML)			ht
17	INSPIRE												
18	GEMET INSPIRE	thésaurus								RDF/SKOS (RDF/XML)	ND	<a href="https://www.eionet.europa.eu/gemet/en/">https://www.eionet.europa.eu/gemet/en/</a>	
19	eLTER vocabulary												
20	EnvThes (Ecology)	Thésaurus								RDF/SKOS (RDF/XML)	ND	<a href="https://raw.githubusercontent.com/ILT/">https://raw.githubusercontent.com/ILT/</a>	ht
21	Climate and Forecast vocabulary												
22	CF Parameters (standard names)	thésaurus								RDF/XML, TTL, N3	<a href="https://mmisw.org/ont/sparql/">https://mmisw.org/ont/sparql/</a>	ND	ht
23	Essential Variables												NI
24	EOVs Oceans	liste									ND	ND	NI
25	EOVs Climate	liste									ND	ND	NI
26	EBVs Biodiversity	liste								JSON	ND	<a href="https://portal.ceobon.org/api/v1/">https://portal.ceobon.org/api/v1/</a>	ht
27	EGVs Geodiversity	liste											
28	EAVs Agriculture	liste											

# Etat des lieux des thésaurus des sciences de la Terre et du Vivant

## 2 - Panorama des thésaurus des dispositifs de diffusion des données (OSU, Pôles de données, IR)

Last update : 2022-10-06	Type	Odatis	Aeris	Theia	Ozcar	Formater	Climeri	PNDB	standard	API SPARQL (skosmos)	AI
<b>Vocabulaires pôles - IR</b>											
<b>ODATIS</b>											
BODC collection -OD1 Odatis	thésaurus								RDF/SKOS (RD	<a href="https://vocab.net">https://vocab.net</a>	N
odatis : variables	liste										
odatis : centre donnees	liste								RDF (RDF/XML)	ND	N
odatis : type_jeux_donnee	liste										
odatis : thematiques	liste										
<b>Theia</b>											
ozcar-TheiaThesaurus	thésaurus								RDF/SKOS (RD	<a href="https://in-situ.the">https://in-situ.the</a>	N
Theia spatial Thesaurus	thésaurus								RDF/SKOS (RD	SKOSMOS	N
<b>Form@ter</b>											
Formater : Disciplines et thèmes	thésaurus										
Formater : Objets d'intérêt	thésaurus										
Formater : Instruments	thésaurus								RDF ttl, RDF/XML	<a href="https://service.pr">https://service.pr</a>	N
Formater : Plateformes	thésaurus										
<b>Aeris</b>											
Aeris : Parameters	thésaurus										
Aeris : Instruments	thésaurus								RDF/SKOS (RD	<a href="https://skosmos">https://skosmos</a>	N
Aeris : Platforms	thésaurus										
Aeris : Projects	thésaurus										
Aeris : Actris	thésaurus								RDF/SKOS, RDI	ND	N
<b>CLIMERI IR</b>											
Variable CMIP6	liste								JSON	ND	N
<b>Data Terra IR</b>											
Data Terra : Types d'entités d'intérêt	thésaurus								RDF/SKOS SOS		N
Data Terra : Variables	thésaurus										N
Data Terra : Types de capteurs	thésaurus									<a href="https://terra-voc">https://terra-voc</a>	N
Data Terra : Plateformes	thésaurus										N
<b>AnaEE IR (Continental ecosystems / biodyn)</b>	Thésaurus								SKOS	ND	hi
	Utilisation										
ND	No Data										

# Etat des lieux des thésaurus des sciences de la Terre et du Vivant

## 3 - Autres thésaurus “généralistes”

Last update : 2022-06-15	Type	Theme	standard	API SPARQL (skosmos, UKgov/LD, VocPrez, vocabench)	API (REST, autre)	UI
Vocabulaires généralistes						
AgroVOC	Thésaurus	Alimentation / Agriculture	RDF/SKOS-XL	<a href="https://agrovoc.fao.org/sparc">https://agrovoc.fao.org/sparc</a>	<a href="https://agrovoc.fao.org/">https://agrovoc.fao.org/</a>	
EARTH	Thésaurus	Environmental Applications	SKOS	<a href="http://vocabularies.cnr.it/earth/">http://vocabularies.cnr.it/earth/</a>	<a href="https://vocabulary.cnr.it/">https://vocabulary.cnr.it/</a>	
UNESCO	Thésaurus	Education / culture / natural sciences / social and human sciences	SKOS * ( <a href="https://www.unesco.org/">https://www.unesco.org/</a> )	<a href="https://skos.um.es/sparql/">https://skos.um.es/sparql/</a>	<a href="http://skos.um.es/">http://skos.um.es/</a>	
USGS	Thésaurus	USGS activities	SKOS	ND	<a href="https://apps.usgs.gov/">https://apps.usgs.gov/</a>	

# Thésaurus de référence par compartiment

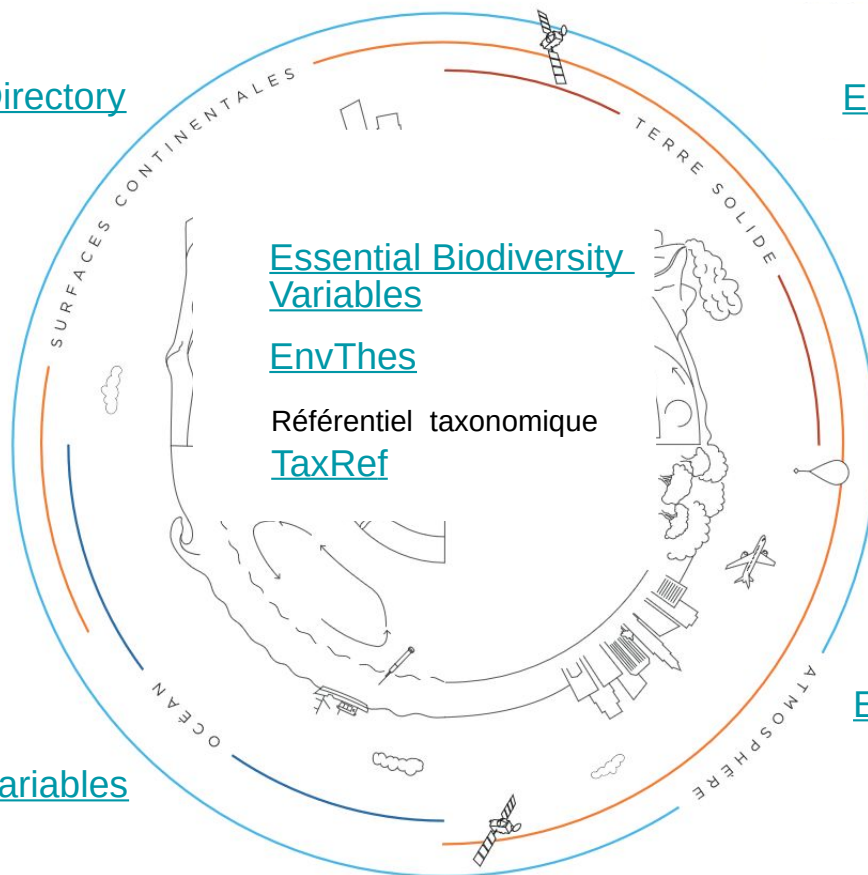
[Global Change Master Directory](#)

Référentiel sur l'eau  
[SANDRE](#)

[WORMS](#) : world  
register of marine species

[NERC Vocabulary Server](#)  
from British Oceanography Data  
Center

[Essential Oceans Variables](#)



[EPOS Vocabulary Registry](#)

[Convention Climate Forecast:](#)  
variable climatique

[Global Change Master Directory](#)

[Essential Climate Variables](#)

# Thésaurus des dispositifs de diffusion des données / IR

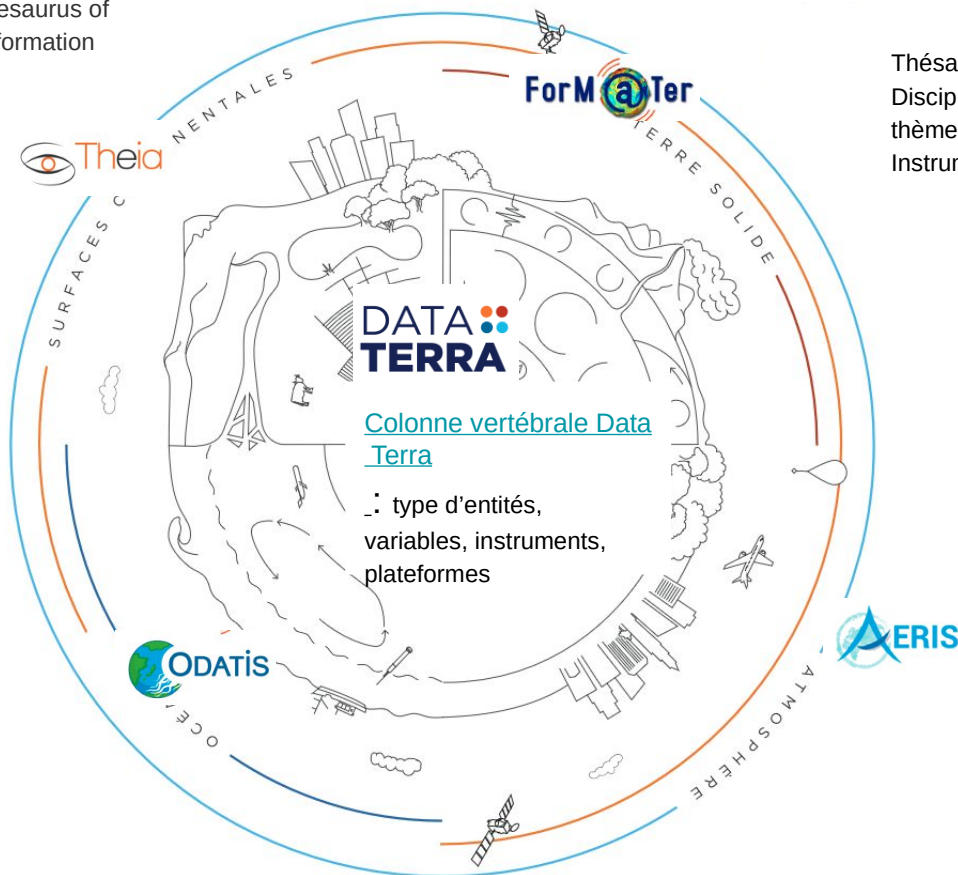
[OZCAR Theia](#) : Thesaurus of the Theia in-situ information system

[AnaEE](#): Thésaurus écologie et biodiversité continentales

[Theia spatial](#): Thesaurus des données spatiales Theia \_

Thesaurus [BODC Collection-OD1 Odatis](#) : Aggregation parameters and Essential Variable names

[Thesaurus Odatis](#) : variables, centre de données, type de jeu de données, thématiques



Thésaurus [Formater](#): Disciplines et thèmes, Objets d'intérêt, Instruments, Plateformes

[CLIMERI](#) : variables CMIP

Thésaurus [Aeris](#) : Parameters, Instruments, Platforms, Project

Thésaurus [ACTRIS](#): climat et qualité de l'air



# Thésaurus des dispositifs de diffusion des données / IR

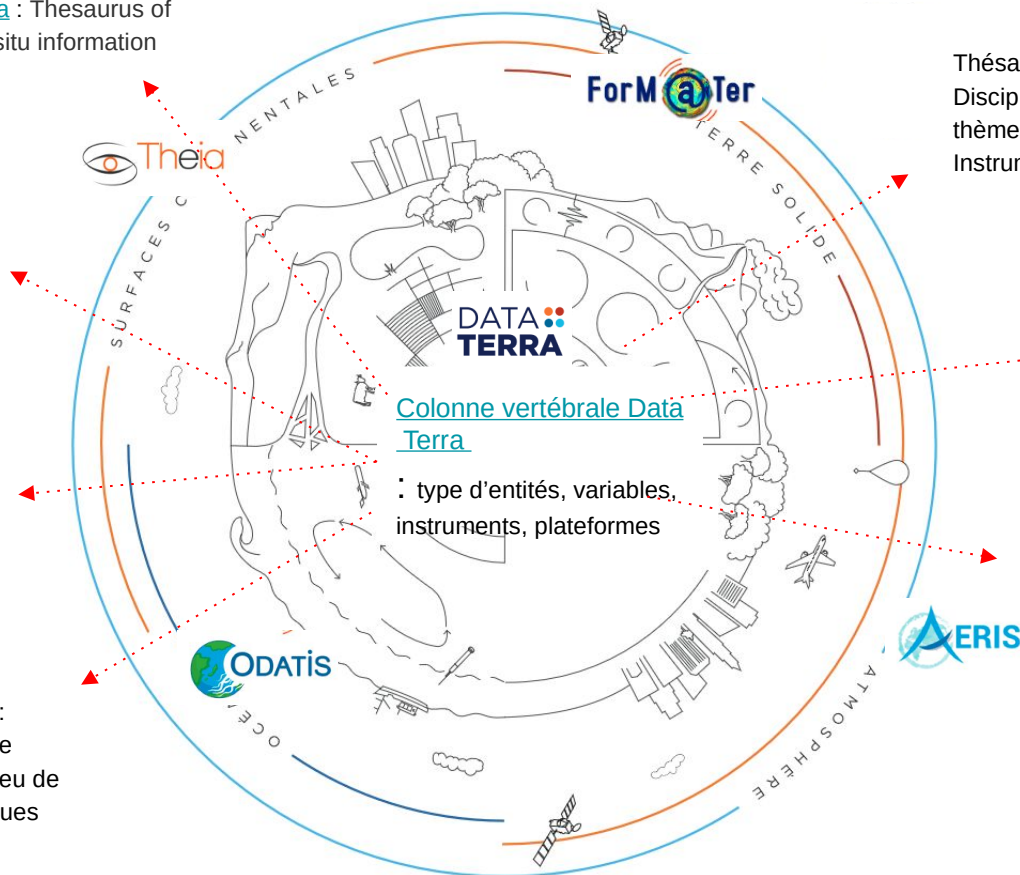
[OZCAR Theia](#) : Thésaurus of the Theia in-situ information system

[AnaEE](#) : Thésaurus écologie et biodiversité continentales

[Theia spatial](#) : Thésaurus des données spatiales Theia

Thésaurus [BODC Collection-OD1 Odatis](#) : Aggregation parameters and Essential Variable names

[Thésaurus Odatis](#) : variables, centre de données, type de jeu de données, thématiques



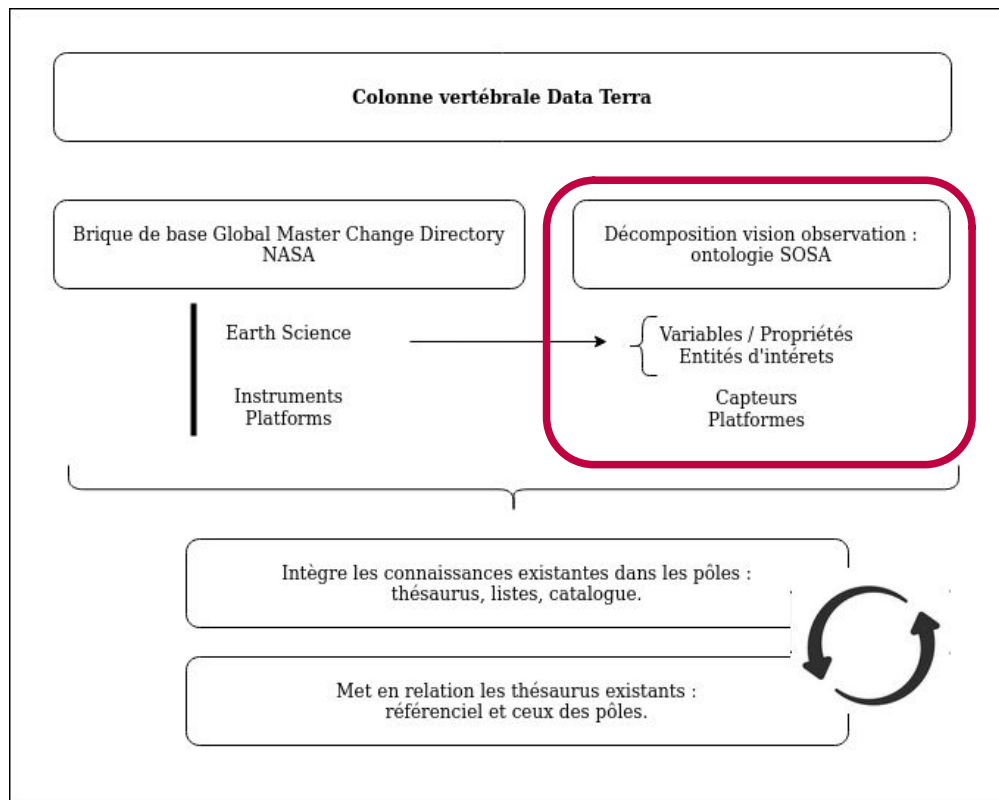
Cohérence d'ensemble : vision système

# Thésaurus IR Data Terra : quels besoins ?

- IR Data Terra :
    - observation intégrée du système Terre
  - Données multi sources et multi thématiques
  - Besoin des vocabulaires :
    - cohérents pour ce système,
    - avec une vision centrée sur l'observation
- 
- Colonne vertébrale : moyen d'articuler et d'harmoniser les thésaurus existants



# Thésaurus IR Data Terra : quelle démarche ?



# Thésaurus IR Data Terra : service FAIR !

- Service Incubator FAIR pour la colonne vertébrale des thésaurus

## 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 ▲
 <a href="#">Earth features types</a>	earthfeaturetype	The current thesaurus contains concepts related to the types ...	<a href="#">Container</a> , <a href="#">concept scheme</a> <a href="#">Register</a>	<a href="#">experimental</a>
 <a href="#">Earth science platforms</a>	earthscienceplatforms	The current thesaurus contains concepts related to data acqui...	<a href="#">Container</a> , <a href="#">concept scheme</a> <a href="#">Register</a>	<a href="#">experimental</a>
 <a href="#">Earth Science Sensors</a>	earthsciencesensors	The current thesaurus contains concepts related to data acqui...	<a href="#">Container</a> , <a href="#">Register</a> , <a href="#">concept scheme</a>	<a href="#">experimental</a>
 <a href="#">Earth science variables</a>	earthsciencevariables	The current thesaurus contains concepts related to Earth scie...	<a href="#">Container</a> , <a href="#">concept scheme</a> <a href="#">Register</a>	<a href="#">experimental</a>

Showing 1 to 4 of 4 entries

[Previous](#)

[1](#)

[Next](#)

# Idées clés à retenir

- Pour assurer **l'interopérabilité** des données les (méta)données doivent utiliser des **vocabulaires** qui suivent les **principes FAIR**
- Les **thésaurus** sont un ensemble de termes organisés avec des relations sémantiques dont le rôle est de **faciliter l'indexation et l'interrogation de documents, de données**
- Les **données sont décrites**, dans leur métadonnées, en faisant **référence aux termes** (libellé et identifiant unique) **d'un ou plusieurs thésaurus**
- Les **thésaurus de référence** sont les **briques de base** pour construire les thésaurus dédiés à la description de nos données

# Vocabulaires contrôlés et données FAIR

Questions ?

Merci !

*GT Catalogue et Vocabulaires Data Terra/Gaia Data*

*Victoria Agazzi – Data Terra*

*Ecole Thématique DATA SDUE – 12 Octobre 2022, Strasbourg.*