



**HAL**  
open science

# Étude comparative de thésaurus en Sciences de l'Environnement - Bonnes pratiques de conception et FAIRisation de thésaurus

Dominique Vachez

► **To cite this version:**

Dominique Vachez. Étude comparative de thésaurus en Sciences de l'Environnement - Bonnes pratiques de conception et FAIRisation de thésaurus. 2021. hal-03264803

**HAL Id: hal-03264803**

**<https://hal.science/hal-03264803>**

Preprint submitted on 18 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Etude comparative de thésaurus en Sciences de l'Environnement

## Bonnes pratiques de conception et FAIRisation de thésaurus

Dominique Vachez

*dominique.vachez@inist.fr*

*INIST-CNRS, 2 rue Jean Zay CS 10310 F-54519 Vandoeuvre-lès-Nancy, France*

**Résumé :** Nous avons réalisé une analyse comparative détaillée de la structure de plusieurs thésaurus du champ des sciences de l'environnement (EnvThes, GEMET, OZCAR-Theia) dans le but de mettre en évidence leurs limites et leurs complémentarités.

Les critères (qualitatifs et quantitatifs) retenus pour comparer ces ressources terminologiques sont énumérés et évalués de manière à proposer un guide de bonnes pratiques pour la conception et l'utilisation de thésaurus comme référentiels sémantiques ouverts et assurer ainsi l'interopérabilité des données de la recherche environnementale.

Afin de favoriser le partage et la réutilisation des (méta)données, les vocabulaires retenus pour représenter les connaissances d'un domaine doivent en effet respecter un ensemble de principes et de règles élaborés au niveau international (normes ISO, principes FAIR, recommandations du web de données, formats d'échange).

En complément de cette étude, nous avons également établi une comparaison avec 6 autres thésaurus spécialisés dans les sujets de l'agronomie, de l'écologie et de la biodiversité (Thésaurus Biodiversité, AGROVOC, GACS, EARTH, AnaEE-France, INRAE).

**Mots-clés :** thésaurus, sciences de l'environnement, ressource terminologique, interopérabilité sémantique, système d'organisation des connaissances, bonnes pratiques, web de données, vocabulaire FAIR

### A] Analyse comparative de trois thésaurus du domaine de l'Environnement

Trois thésaurus des sciences de l'environnement (**EnvThes**, **GEMET**, **OZCAR-Theia**) librement accessibles ont été comparés selon différents critères relatifs à leur qualité sémantique et à leur conformité avec :

- les recommandations du web de données liées et ouvertes (**Linked Open Data**) (W3C, 2006/2010)
- le modèle de données **SKOS-RDF** applicable aux systèmes d'organisation des connaissances (Simple Knowledge Organization System) (W3C, 2009)
- les normes [ISO 25964](#) relatives aux thésaurus et à leur interopérabilité (2011/2013)
- la norme **ISO 704** relative aux principes et méthodes du travail terminologique (2009)
- les **principes FAIR** (2016)

Pour rappel, l'ISO 25964 définit un thésaurus comme un vocabulaire contrôlé et structuré par une combinaison de relations hiérarchiques et associatives entre concepts, ainsi que des relations d'équivalence intra/inter-linguistiques entre les termes désignant ces concepts dans une même langue (préférentiel et synonymes) ou d'autres langues (traductions).

SKOS est un format standard faisant partie des technologies du Web sémantique, et dont le modèle repose sur un ensemble minimal d'axiomes et de conditions d'intégrité. Il peut être appliqué à divers systèmes simples d'organisation des connaissances, tels que les thésaurus.

Pour représenter chaque **concept** et dans chaque langue, on choisit comme désignation (ISO 704) un **terme préférentiel** (skos:prefLabel), dit aussi descripteur (ISO 25964), qui est donc normalisé mais se rapproche le plus possible de la langue de spécialité utilisée dans un champ de la connaissance.

Ce libellé préférentiel peut être accompagné d'un ou plusieurs synonymes/quasi-synonymes (libellés alternatifs skos:altLabel), encore appelés termes équivalents ou non-préférentiels.

Lorsqu'on crée des relations de correspondance entre les concepts appartenant à différents thésaurus, on parle d'alignements entre ressources sémantiques (skos:mappingRelation). Ces correspondances (ou équivalences conceptuelles) peuvent être exactes (skos:exactMatch) ou approchantes (skos:closeMatch/relatedMatch). En effet, dans les thésaurus, les définitions ne sont pas établies de manière aussi formelle que dans les ontologies.

Néanmoins, depuis la parution de ces normes et standards, et si ceux-ci sont correctement suivis, le thésaurus n'est plus vu uniquement comme un langage documentaire pour la recherche d'information, mais devient un véritable système de concepts qui peut être considéré en ingénierie des connaissances comme une ontologie légère. Moyennant quelques transformations basées sur une extension du modèle SKOS selon la norme ISO 25964 (iso-thes), il est d'ailleurs envisageable de faire de la réingénierie de thésaurus vers une ontologie simple.

A partir de cette comparaison, nous allons déduire un certain nombre de bonnes (ou moins bonnes) pratiques dans la conception de thésaurus, permettant d'orienter le choix des référentiels terminologiques destinés à enrichir les métadonnées dans le domaine de l'Environnement, et d'assurer une cohérence sémantique favorisant l'intégration et la réutilisation des données.

Précision terminologique : on voit que le terme d'équivalence peut être employé dans des sens variés  
→ équivalence entre termes d'un même vocabulaire  
(intralinguistique : synonymie / interlinguistique : traduction)  
→ équivalence entre concepts : alignement (correspondance entre différents vocabulaires)

## Principaux critères de comparaison des thésaurus

- **Richesse terminologique (termes) :**
  - nombre de **termes préférentiels** (skos:prefLabel)
  - nombre de **synonymes** (relations d'**équivalence**)  
(termes-libellés **alternatifs** / skos:altLabel / étiquettes **EM** : employer / **EP** : employé pour)
  - nombre de **variantes** orthographiques (termes-libellés **cachés** / skos:hiddenLabel)
  - langue de spécialité
  
- **Relations sémantiques (concepts) : hiérarchiques et/ou associatives**
  - nombre de concepts **génériques** (skos:broader / étiquette **TG** : terme générique)
  - nombre de concepts **spécifiques** (skos:narrower / étiquette **TS** : terme spécifique)
  - nombre de concepts de tête ou **top concepts** (étiquette **TT** : terme de tête)
  - nombre de niveaux hiérarchiques (**profondeur** hiérarchique)
  - nombre de concepts frères/adjacents (sous un même générique)
  - nombre de concepts **associés** (skos:related / étiquette **TA** : terme associé)

- **Hiérarchies** : 3 types (**générique** : sorte de/ **partitif** : partie de/ **instance** : exemple de) ; cohérence ; mono/**polyhiérarchie**
- **Regroupements de concepts** : collections, microthésaurus
- **Equivalences linguistiques** entre termes – **Multilinguisme**
- Présence de **définitions** (multilingues)
- Notoriété française, européenne ou **internationale**
- **Identifiant universel de ressource** (URI) pour chaque concept
- Liens avec le **web de données** (LOD - **Linked Open Data** cloud)
- **Alignements** réalisés avec d'autres thésaurus (**équivalences** conceptuelles)
  - Liens **entrants** et/ou **sortants** (**réciprocité**)
  - Affichage de la nature de ces liens (skos:**exact** / close / broad / narrow ou related**Match**)
- Ressource librement **consultable**, **téléchargeable** et **interopérable** (format **SKOS-RDF**)
- **Interfaces** de consultation / navigation (ergonomie) / interrogation / **SPARQL** endpoint / API
- **Licence** de (ré)utilisation (**libre/ouverte**)
- Maintenance et **versionnage** de la ressource

Note : dans la suite du texte, les expressions en **caractères gras latins** correspondent à des **qualités**, tandis que les expressions en **italiques** correspondent à des **limites**.

## 1] **EnvThes** (Environmental Thesaurus)

Thésaurus **européen** développé au sein du réseau **LTER-Europe** (Long-Term Ecosystem Research in Europe) (projet **EnvEurope** / **ExpeER**) depuis 2014

- généraliste et relié aux thématiques de la directive **INSPIRE**
- basé sur des ontologies (OGC O&M, OBOE et SERONTO) ainsi que sur le thésaurus **US LTER Controlled Vocabulary** et les mots-clés environnementaux de Microsoft Academic Research
- utilisé par **DEIMS-SDR** (**Dynamic Ecological Information Management System** - Site and Dataset Registry) à des fins d'annotation et d'interrogation des métadonnées, et mis en œuvre comme référentiel sémantique des jeux de données au sein du projet européen **ECOPOTENTIAL**
- préconisé notamment pour le géocatalogue des métadonnées du réseau des **Zones Ateliers** (RZA membre de **eLTER** / **ILTER**) et le portail des données de l'**OSU OREME** (Observatoire de REcherche Montpellierain de l'Environnement)
- URI de base du thésaurus : <http://vocabs.lter-europe.net/EnvThes/10000>

A noter : Ce thésaurus a beaucoup évolué depuis la première étude comparative réalisée en 2014 et contient de nombreux concepts dépréciés.

### Qualités / Avantages (EnvThes)

- richesse sémantique : **2805 concepts non dépréciés** (rassemblés sous 7 super-concepts)
- **multilinguisme** (26 langues) en cours de développement (anglais majoritaire)
- 1690 **termes alternatifs** (altLabel) anglais/français
- présence de **définitions** multilingues
- structure **polyhiérarchique**
- interface principale de navigation **TopBraid** en accès libre et **multilingue (préférentiels)**
- interface **EcoPortal** : n'affiche que les **concepts validés** ; accès par les **synonymes anglais**
- concepts **URI** (<http://vocabs.lter-europe.net/EnvThes/>) : identifiants **opaques** et **déréférençables**
- format **SKOS-RDF**
- thésaurus complet **téléchargeable** en RDF ttl ([GitHub](#), [EcoPortal](#))
- **nombreux alignements** sortants vers AGROVOC, GEMET, EARTH, US-LTER, EuroVoc, DBpedia
- [SPARQL endpoint](#) (CEH)
- présent dans le catalogue [FAIRsharing](#)

### Limites / Inconvénients (EnvThes)

- *pas de regroupements en collections*
- **mélange** de concepts **dépréciés** (2930) et de concepts validés (**difficulté pour les alignements**) parfois mis en relation *skos:related*
- les concepts **dépréciés** sont systématiquement mis en relation **skos:exactMatch** avec les *concepts remplaçants*, mais également avec les concepts d'*autres thésaurus*
- *défauts de gestion des synonymes* → concepts en **doublons** (ex. "rainforest" / "rain forest")
- *défauts de gestion des traductions* → **synonymes sans préférentiel** dans la même langue (ex. "endangered species" et "vascular plant species" ne comportent **aucune traduction du prefLabel** mais présentent des **altLabel** dans plusieurs autres langues)
- propriété **skos:altLabel** fréquemment utilisée pour des **traductions** (cf. ci-dessus) ou pour l'affichage de **variantes morpho-syntaxiques** (au lieu de *skos:hiddenLabel*, ex. "pine forested" "environmentally polluted")
- **erreurs orthographiques** : ex. *skos:prefLabel "permittiviy"@en*
- **termes français peu nombreux** (208 préférentiels en français)
- interface **EcoPortal** : - **interrogation unilingue (anglais)**, non interrogeable en français ;  
- termes **traduits sans étiquette de langue** ;  
- **bogues** à l'interrogation : **uniternes** ou **multitermes** parfois **introuvables** (ex. "biodiversity", "biomass", "genetic diversity"); réponses **incomplètes** et/ou ne suivant pas l'ordre **alphabétique** ;  
- **bogues** à l'affichage de la hiérarchie (ex. "object of interest") ;  
- certains termes **synonymes** restent **identiques au terme préférentiel** (ex. "arbre" /trees) ;  
- **non-affichage des concepts spécifiques** dans le détail de la notice
- interface **TopBraid** : - non interrogeable par les **synonymes** ;  
- affichage hiérarchique **mélangeant** préférentiels anglais et français (lorsque le terme est multilingue) ;  
- **affichage incomplet des résultats** (la recherche ne s'effectuant qu'en mode **Starts with**) ;

- **EcoPortal/TopBraid** : - aucune de ces interfaces ne permet de **retrouver la totalité** des termes **contenant** une chaîne de caractères (ex. \*diversit\*) [l'interface Skosmos de [BARTOC](#) n'étant **pas à jour (2017)**] ;  
- lors de l'affichage d'un **concept**, **l'arbre hiérarchique ne se déplie pas complètement** : ne **permet** donc **pas de visualiser** entièrement une **polyhiérarchie** (ex. event "algal bloom" ; parameter "plant nutrient")
- **anomalies SKOS-RDF** : thésaurus à la fois de type **Concept** et **ConceptScheme** ;  
- niveau **Concept** : **absence** de propriété **skos:inScheme** ;  
- alignements (**xMatch**) entre concepts **au sein du même thésaurus**
- **pas d'export possible SKOS-RDF au niveau concept individuel**
- introduction de **code html** dans des **notes textuelles** du fichier **rdf**
- **nombreuses définitions vides**
- **présence d'espaces excédentaires** dans des **libellés alternatifs**
- alignements **non réciproques (pas d'alignements entrants, excepté th. Biodiversité et OZCAR)**
- **pas de licence** d'utilisation

L'outil de vérification de conformité **qSKOS** ([SKOS Testing Tool](#)) a détecté un assez grand nombre d'anomalies ou de violations de la **spécification SKOS** (recommandation **SKOS Reference**) et/ou du guide utilisateur **SKOS Primer**.

Certaines erreurs présentes en 2018 ont été corrigées dans la dernière version GitHub, tels que des **synonymes identiques aux préférentiels** (qui apparaissent cependant toujours dans *EcoPortal*).

- **concepts sans aucun libellé** : <http://vocabs.lter-europe.net/EnvThes/22021>
- **concepts orphelins** : sans aucune relation sémantique avec d'autres concepts  
<http://vocabs.lter-europe.net/EnvThes/30020>

La plupart de ces "concepts" cumulent souvent plusieurs défauts (**concept orphelin, sans libellé préférentiel, URI non-opaque**) : [http://vocabs.lter-europe.net/EnvThes/water\\_pressure](http://vocabs.lter-europe.net/EnvThes/water_pressure)

- **codes langues absents** (libellés ou définitions)
- **libellés homonymes** : **préférentiels** identiques dans la même langue pour des concepts distincts. ex. "eau douce" ; "évaporation" ; "demography"
- concepts **reliés à la fois hiérarchiquement et associativement** ("aquatic fungi" TG/TA "fungi")
- **relations inverses absentes** (générique/spécifique)
- **relations symétriques absentes** (associative)
- **relations associatives** entre concepts adjacents du **même niveau**

## 2] **GEMET** (GEneral Multilingual Environmental Thesaurus)

Thésaurus **européen** développé par l'**Agence européenne pour l'environnement / Réseau européen d'information et d'observation pour l'environnement (AEE / Eionet)** depuis 1997

- généraliste et compatible avec toutes les thématiques de la directive **INSPIRE**

- issu de la compilation de plusieurs ressources sémantiques (lexiques, thésaurus)

- couramment préconisé pour l'enrichissement des métadonnées : OSU OREME, BBES/ [InDoRES](#) (Bases de données Biodiversité, Ecologie, Environnements, Sociétés / Inventaire des Données de la Recherche en Environnement et Sociétés)

- URI de base du thésaurus : <http://www.eionet.europa.eu/gemet/gemetThesaurus>

## Qualités / Avantages (GEMET)

- couvre l'ensemble des thématiques environnementales
- richesse sémantique : **5530 concepts**
- 4 super-groupes + 32 groupes + 40 thèmes (collections)
- interface [Eionet](#) (**multilingue**) : accès aux **top-concepts** à partir de chaque groupe ; interrogation réalisable avec n'importe quelle chaîne de caractères ex. "diversit"
- interface [AgroPortal](#) (en anglais) : navigation dans l'**arbre hiérarchique des concepts** à partir des **110 top-concepts**
- interface [LusTRE](#) (**Linked Thesaurus fRamework for Environment**) : recherche fédérée **multilingue** dans plusieurs thésaurus (GEMET, AGROVOC, EARTH, EuroVoc)
- interface [TemaTres](#) : liste **alphabétique** et arbre **hiérarchique**
- structure **polyhiérarchique**
- **multilinguisme** (37 langues)
- présence de **définitions**
- **alignements réciproques** avec AGROVOC, EuroVoc
- format **SKOS-RDF**
- concepts **URI** (<http://www.eionet.europa.eu/gemet/concept/892>) : identifiants **opaques** et **déréférencables**
- thésaurus complet **téléchargeable** en SKOS RDF ([Eionet Portal](#) ; [AgroPortal](#))
- **licence libre CC BY**
- [SPARQL endpoint](#) (EEA)
- historique de **versions**
- présent dans [FAIRsharing](#)

## Limites / Inconvénients (GEMET)

- **pas d'export SKOS-RDF au niveau concept individuel**
- **rarely synonymes anglais** (*alt* ou *hiddenLabel*) ; **aucun synonyme français**
- **préférentiels français identiques** pour des **concepts différents** (déf. + anglais distincts) : ex. "aménagement du territoire" (4 concepts) ; "pollution du sol" (2 concepts en relation TS/TG)
- **prefLabel français identique** pour un concept et son **générique** : "aménagement du territoire"
- quelques cas de **descripteurs unilingues** : "silviculture" (*en*)
- **confusion** entre les libellés de certains concepts, groupes et thématiques (ex. "Biosphere")
- **incohérence** sémantique au même niveau hiérarchique : ex. TG "biosphère"/TS "anatomie ; écologie ; évolution ; organisme ; processus biologique ; virus"
- **redondances hiérarchiques** : ex. "cours d'eau", "mer", "réservoir d'eau" sont à la fois des concepts spécifiques de "hydrosphère" et de "eaux (géographie)" (alors que ce dernier est lui-même spécifique de "hydrosphère")
- peu spécialisé (*profondeur hiérarchique* moyenne : 5 niveaux)
- **pas de relations inverses skos:topConceptOf**
- interface [Eionet](#) : **pas de navigation possible** dans un **arbre hiérarchique global**
- interface [AgroPortal](#) : - **pas de navigation possible par les groupes ou les thèmes** ;  
- **interrogation unilingue (anglais) ininterrogeable en français** ;  
- **à l'affichage d'un concept, l'arbre hiérarchique ne se déplie pas entièrement : ne permet donc pas d'identifier une polyhiérarchie** (ex. "watercourse")



- interface **LusTRE** : - **pas d'arbre hiérarchique** ; - **recherche ne s'effectuant qu'en mode Starts with** ; - **version 2012 non mise à jour**
- interface **TemaTres** : - **URI internes ne correspondant pas aux URI de base du thésaurus** ; - **interrogation monolingue (anglais)** ; - **regroupements absents** ; - **non mise à jour (2012)**

L'outil de vérification de la qualité du SKOS (**qSKOS**) permet de détecter plusieurs autres anomalies :

- **relation hiérarchique cyclique** : un concept est à la fois générique et spécifique du même concept. "ecosystem assessment" est **à la fois TG et TS** de "ecosystem boundary"
- **concepts orphelins** : "biological diversity"
- **libellés préférentiels homonymes** : "waste water treatment plant" (en) et autres concepts dans de nombreuses autres langues
- **collisions entre relations sémantiques hiérarchiques et associatives**
- **relations associatives entre concepts adjacents du même niveau**

### 3] **OZCAR-Theia thesaurus** (Observatoires de la Zone Critique-Application et Recherche)

Thésaurus **français** développé par le système d'information du Pôle de données **Theia** (IR Data Terra / **Système Terre – Surfaces continentales**) et l'infrastructure **OZCAR** (eLTER-RI) depuis 2019

- spécialisé sur certaines variables environnementales

- basé sur le vocabulaire **GCMD Earth Science** Keywords (Global Change Master Directory, NASA), et sur certains groupes de paramètres du référentiel de chimie du **Sandre** (Service d'administration nationale des données et référentiels sur l'eau)

- URI de base du thésaurus : <https://w3id.org/ozcar-theia/ozcarTheiaThesaurus>

#### **Qualités / Avantages (OZCAR-Theia)**

- richesse sémantique : **430 concepts** rassemblés sous 2 super-concepts + 2 collections
- **nombreux alignements** (exact, close ou relatedMatch) avec **EnvThes, AGROVOC, GACS, GCMD, EARTH, GEMET, AnaEE ...**
- interface de consultation **Skosmos** (**navigation** par l'**arbre hiérarchique** et par les **groupes**)
- **concepts** exportables individuellement en format **SKOS-RDF**
- concepts **URI** (<https://w3id.org/ozcar-theia/biosphere>): identifiants **permanents** et **déréférencables**
- ressource **téléchargeable** directement en rdf (ou par interrogation du triplestore)

#### **Limites / Inconvénients (OZCAR-Theia)**

- **monolinguisme** (anglais) : **français absent**
- confusion possible entre intitulés de collections et termes génériques
- absence de certains concepts généraux (*critical zone ; geosphere*)
- termes composites très éloignés du langage naturel ou recouvrant *plusieurs concepts* ex. "**COHV - Solvents - Freons (groundwater)**" (*complicant les alignements*)
- nombreux termes **homonymes** (différenciés par des parenthèses)
- graphies **hétérogènes**
- graphies **erronées**



- certains termes identiques se présentent plusieurs fois (avec le même URI concept) dans l'arbre hiérarchique, mais désignent des concepts différents : ex. "*precipitation amount*"
- aucune **relation associative** *skos:related*
- aucun **synonyme**
- aucune **définition**
- **absence** de **licence** d'utilisation
- concepts URI : identifiants **signifiants** (~prefLabel) → risque pour la pérennité à long terme
- **métadonnées** non renseignées (ressource, concepts) pas de **dates** de création/modification
- **alignements** : certains **exactMatch** sont en réalité des *closeMatch* ou *relatedMatch*.
- alignements avec des **concepts dépréciés** de [EnvThes](#) ou de [EARTh](#)
- **absent** de **FAIRsharing**

La qualité du SKOS a pu être testée avec le web service de [Loterre](#) (Linked open terminology resources), ainsi qu'avec le service en ligne de SKOS Play! [qSKOS](#), en décochant la case "Liens cassés" (pour éviter une erreur proxy, compte tenu des nombreux alignements). Ces services ne détectent pas toujours le même type de défauts et peuvent donc s'avérer utilement complémentaires.

Il est à noter qu'un grand nombre de ces anomalies étaient indétectables ou introuvables dans Skosmos.

- URI apparaissant comme membre de la collection Variable categories, mais **inexistant comme concept** (pas de *skos:Concept*, pas de libellé) et sans relation hiérarchique, donc **invisible dans Skosmos**
- URI apparaissant en *rdf:Description*, mais ne se présentant que par ses alignements, sans *skos:Concept*, ni *skos:prefLabel*, ni *skos:inScheme*, donc **invisible dans Skosmos**
- URI apparaissant comme *skos:narrower* d'un autre concept, mais absent en tant que *skos:Concept*. Terme **introuvable** dans la liste alphabétique et se présentant comme "**null**" dans la *hiérarchie Skosmos*
- **relation hiérarchique réflexive** : un concept est son *propre générique*
- **relations inverses absentes** (erreurs générique/spécifique) : en général, l'absence d'une des 2 relations inverses (*skos:broader/skos:narrower*) dans le fichier rdf, provoque des défauts d'affichage dans le dépliement de l'arbre hiérarchique de Skosmos (ex. "[Precipitation amount](#)" n'apparaît pas sous "*Liquid precipitation*")
- **libellés homonymes** apparaissant de manière **erronée** plusieurs fois dans la hiérarchie

➔ Après concertation avec l'équipe du système d'information Theia/OZCAR, la plupart de ces anomalies ont pu être corrigées dans une nouvelle version du thésaurus.

Remarque : Outre les outils de contrôle fournis par **qSKOS** et **Loterre**, si le thésaurus a été chargé dans l'éditeur open source **VocBench**, il est possible d'utiliser un module intégré destiné à tester la validité des contraintes d'**intégrité** du SKOS (ICV - [Integrity Constraint Validator](#)).

## B] Bonnes / Meilleures pratiques (structuration et description de thésaurus)

### ▪ Thésaurus (classe `skos:ConceptScheme`)

- Schéma de concepts = ressource sémantique ou système d'organisation des connaissances.

### ▪ Concepts (classe `skos:Concept`)

- Indiquer l'appartenance de chaque concept (= ressource conceptuelle) au thésaurus (= schéma conceptuel) avec la propriété `skos:inScheme`.

### ▪ Hiérarchies (propriétés `skos:broader/narrower`)

- **Homogénéité** dans les relations hiérarchiques entre concepts : éviter la confusion entre hiérarchie (`skos:broader/narrower`) et regroupement (`skos:Collection / skos:member`).

- Les relations `skos:broader` et `skos:narrower` sont **inverses** l'une de l'autre et il est préférable de les faire figurer toutes les deux dans le fichier rdf.

- **Cohérence** hiérarchique : les concepts spécifiques d'une même branche hiérarchique ne devraient pas appartenir à des **catégories sémantiques** fondamentalement différentes (propriété, discipline, organisme, substance, lieu, processus, action, méthode...).

- Ces catégories fondamentales doivent être traitées soit comme des concepts généraux de plus haut niveau, soit comme des collections de concepts.

- Selon les normes **ISO 25964** et **704**, on peut distinguer 3 types de hiérarchies :

- **générique** (est un / sorte de / genre de / type de) = relation de *subsumption*
- **partitive** (partie de / composant de / constituant de / localisé dans)
- **instance** (concept *individuel* / exemple de / entité nommée)

Note : un terme ou un concept sera dit "*générique*" quel que soit le type de hiérarchie, en précisant "terme *générique* générique" TGG, "terme *générique* partitif" TGP, "terme *générique* d'instance" TGI.

Ces relations peuvent être exprimées à l'aide des **sous-propriétés** de `skos:broader / narrower` issues de l'ontologie **iso-thes**, respectivement :

- **isothes:broader / narrowerGeneric** (étiquettes **TGG / TSG**)
- **isothes:broader / narrowerPartitive** (étiquettes **TGP / TSP**)
- **isothes:broader / narrowerInstantial** (étiquettes **TGI / TSI**)

- La **polyhiérarchie** est autorisée (mixant les divers types de hiérarchies).

Note : Des relations hiérarchiques de typologies différentes peuvent s'enchaîner mais elles ne sont pas obligatoirement transitives. Les propriétés `skos:broader/narrower` , qui établissent un lien hiérarchique *direct* entre deux concepts, ne sont donc pas déclarées transitives dans le modèle SKOS. En cas de besoin, il est néanmoins possible d'exprimer la *transitivité* d'une relation hiérarchique (*directe ou indirecte*) avec les propriétés `skos:broaderTransitive` ou `skos:narrowerTransitive`.

- Limiter le nombre de concepts de plus haut niveau (**top-concepts** / concepts de tête) pour faciliter la navigation dans le thésaurus.

- Les propriétés qui relient un thésaurus à tous ses top-concepts (**skos:hasTopConcept**) doivent figurer dans le fichier rdf sous skos:ConceptScheme ; les propriétés inverses correspondantes (**skos:topConceptOf**) se situent au niveau de chaque top skos:Concept.

#### ▪ **Associations** (propriété **skos:related**)

- La relation associative est **symétrique** et doit donc figurer dans les 2 sens.

- Cette propriété n'est **pas transitive**.

- La spécification SKOS ne permet pas de relier 2 concepts de manière *associative*, s'ils sont déjà reliés entre eux de manière *hiérarchique* (propriétés **disjointes**).

- Deux concepts peuvent généralement être associés s'ils sont de sens **voisins** ou **chevauchants** ("*zone côtière*" TA "*zone littorale*") ou bien si le premier concept participe de la définition du second, sans être en relation hiérarchique ("*écosystème*" TA "*service écosystémique*"; "*sols*" TA "*pédologie*").

- **Concepts adjacents** : On évitera d'associer systématiquement entre eux des concepts frères ou *adjacents* qui figurent au même niveau dans une *hiérarchie tout-partie*. Des associations trop fréquentes dans une même hiérarchie pourraient aussi dénoter une inhomogénéité hiérarchique ou une confusion entre hiérarchies et regroupements de concepts.

- Une relation associative peut surtout être l'occasion de qualifier la nature de cette relation et permettre ainsi d'aller vers une **ontologisation** du thésaurus.

Certains thésaurus (ex. AGROVOC) ont subdivisé la propriété **skos:related** en créant plusieurs **sous-propriétés** visant à **affiner** la relation associative en diverses **relations ontologiques** : cause/effet, agent/action, paramètre/processus, objet/propriété, procédé/produit ...

#### ▪ **Collections** (classe **skos:Collection**)

- Les **regroupements** de concepts par catégories, thématiques ou sous-domaines seront gérés comme des collections.

- Chaque collection est constituée d'une liste **non hiérarchisée** de concepts (**skos:member**). Il est cependant possible d'emboîter des collections et des sous-collections.

- On peut éventuellement gérer des groupes et sous-groupes en utilisant les sous-classes de skos:Collection tirées de l'ontologie **iso-thes** (**isothes:ConceptGroup** et/ou **isothes:ThesaurusArray**) et en les structurant en sous-collections (sous-groupes) avec les propriétés **isothes:superGroup** / **isothes:subGroup**.

#### ▪ **Définitions** (propriété **skos:definition**)

- Définition du concept : optionnelle mais largement recommandée pour enlever toute ambiguïté.

- Définition **non ambiguë** : si un concept comporte plusieurs définitions, cela peut signifier qu'on a affaire à différents concepts.

- Lors de la rédaction d'une définition, il convient de respecter certaines règles (norme **ISO 704** : Travail terminologique - Principes et méthodes) et d'éviter entre autres les définitions *circulaires*, *imprécises* ou *négatives*. Ainsi, un concept ne peut pas être défini au moyen d'un second concept qui est lui-même défini à l'aide du terme désignant le premier. Une définition ne doit pas non plus commencer en répétant la désignation (terme) du concept.

- La définition à privilégier est la **définition par intension** : elle consiste à reprendre un concept plus générique (*superordonné*) et à identifier un ensemble de caractères distinctifs permettant de différencier le concept à définir des autres concepts spécifiques (*subordonnés*).
- Donner la ou les **sources** de la définition.
- Attribuer un code **langue** à chaque définition.
- SKOS permet de choisir comme objet de la définition une valeur non littérale (ex. l'URI d'une ressource ou l'URL d'un document, au lieu d'une simple chaîne de caractères).
- Il est aussi possible d'utiliser la propriété **skos:scopeNote** de manière à clarifier et limiter le sens ou l'utilisation d'un terme polysémique pour un concept dans le domaine du thésaurus (= note explicative/note d'application).

**Remarques** : En toute rigueur la norme ISO 25964 indique dans son modèle que la classe **ScopeNote** est reliée (*definesScopeOf*) à **ThesaurusConcept**, tandis que **Definition** est reliée (*isDefinitionOf*) à **ThesaurusTerm**. En conséquence, le contenu de Definition pourrait d'une part dépasser le sens restreint que l'on attribue au terme au sein du thésaurus et autoriserait d'autre part à ajouter une définition à un terme non-préférentiel. Néanmoins, le modèle SKOS du W3C (2009) n'a pas été mis à jour après la parution de la norme ISO (2011) et les propriétés skos:scopeNote et skos:definition peuvent toutes deux servir à annoter un skos:Concept.

#### ▪ **Termes préférentiels** (propriété **skos:prefLabel**)

- Pas plus d'un **libellé lexical** (terme) **préférentiel** pour chaque concept et dans chaque langue (contrainte **SKOS Reference**).
- Eviter de désigner de manière ambiguë différents concepts avec le même terme préférentiel (recommandation **SKOS Primer**). De fait, les libellés **homographes** sont conflictuels, qu'il s'agisse d'homonymies ou de polysémies.
- Eviter le choix d'un terme préférentiel concaténant ou recouvrant **plusieurs** concepts à la fois, qu'ils soient séparés par des ponctuations (tiret, virgule) ou par des conjonctions (et, ou). En effet, contrairement à une classification, la *précoordination* de multiples notions dans un même terme *composé* (multiterme) devrait rester exceptionnelle dans un thésaurus (sauf s'il est ultra-spécialisé et destiné à une communauté restreinte d'utilisateurs).
- Attribuer un **code langue** à chaque terme.
- Le libellé lexical doit être suffisamment **précis** mais pas trop complexe afin de faciliter les **alignements**.
- L'usage des **parenthèses** doit être limité et sert surtout à désambiguïser ou qualifier le terme qui précède lorsque celui-ci est polysémique ou présente une homonymie avec un autre terme. Lorsque cela est possible, il est préférable d'employer un terme plus précis : ex. "*chaîne trophique*" (écologie) vs "*chaîne alimentaire*" (agroalimentaire).
- Les multitermes doivent être construits dans l'**ordre du langage naturel**. Un libellé exagérément long et éloigné du langage naturel peut s'apparenter davantage à une définition ou correspondre à une *combinaison* de plusieurs concepts.

- Le libellé préférentiel d'un concept peut être modifié dans une nouvelle version (utiliser la propriété `skos:changeNote`) tout en conservant son URI.

- Les termes préférentiels d'un vocabulaire ne doivent pas varier d'une plateforme terminologique à l'autre (en exposant toujours la dernière version).

#### ▪ Equivalences linguistiques

- Favoriser le **multilinguisme** (au min. le bilinguisme) avec un équivalent dans chaque langue pour chaque terme préférentiel, sauf si l'équivalence linguistique **exacte** n'existe pas.

- Dans certaines situations (entités nommées, noms latins, termes d'emprunt), le même terme pourra être employé à l'identique dans différentes langues.

#### ▪ Synonymies / termes alternatifs (propriété `skos:altLabel`) ou cachés (`skos:hiddenLabel`)

- Il est recommandé d'enrichir le thésaurus avec des termes équivalents ou synonymes (facilitant l'**interrogation** et/ou les **alignements**) :

- Préférer la synonymie exacte (équivalence stricte) à la synonymie approchée (ou quasi-synonymie)
- Les **quasi-synonymes** sont autorisés même s'ils ne sont interchangeable que dans certains contextes
- Les **acronymes** sont le plus souvent placés en **synonymes**, sauf si leur usage est largement répandu et si la terminologie est suffisamment spécialisée pour éviter toute ambiguïté (ex. ADN en biologie, PCB en chimie ...)
- Possibilité d'**inverser** synonyme et terme préférentiel (évolution des usages, dépréciation d'un terme, désambiguïsation). A préciser dans une note historique (`skos:historyNote`)
- Détecter les concepts identiques (*doublons*) qui pourraient être **dépréciés** et fournir de nouveaux synonymes.

Note : L'extension de SKOS pour les libellés ([SKOS-XL](#)) permet de traiter les libellés (entités lexicales) comme des ressources RDF (`skosxl:Label`) à part entière (et non pas simplement comme des suites de caractères) et de les relier entre elles par des propriétés `skosxl:labelRelation`.  
Il est possible par ce biais de créer des sous-propriétés de `skosxl:labelRelation` telle que l'**acronyme** d'une forme développée.

- La norme ISO 25964 autorise à placer en relation d'*équivalence* un terme plus *spécifique* en le désignant avec un terme plus *générique*, afin de ne pas multiplier le nombre de termes préférentiels, tout en conservant une entrée terminologique. Il faut toutefois considérer qu'un terme alternatif (`skos:altLabel`) plus précis que le terme préférentiel (`skos:prefLabel`) pourrait représenter un concept **émergent** plus spécifique (`skos:narrower`).

- Les résultats d'interrogation pourront être améliorés par la présence de multiples **libellés cachés** (formes variantes ou erreurs d'orthographe) recherchables mais non affichés (`skos:hiddenLabel`).

## ▪ Dépréciation de concepts

- Les thésaurus ne gèrent pas tous de la même manière les concepts lorsque ceux-ci deviennent obsolètes.

Ces concepts rejetés peuvent être entièrement supprimés ou bien signalés dans une nouvelle version du thésaurus à l'aide d'annotations sur les concepts qui les ont remplacés (*inversions* de synonymies, *scission* ou *fusion* de concepts).

- Plusieurs situations sont possibles :

- **Suppression** totale du concept déprécié, avec mention éventuelle de son existence passée dans l'**annotation** d'un autre concept (**skos:historyNote** ou **skos:changeNote**) ;
- **Maintien** du concept déprécié et de son URI mais avec un libellé préférentiel absent ou contenant (*deprecated*) ou (*obsolete*) dans le but de le démarquer nettement pour exclure les alignements en **exactMatch** avec des ressources externes ;
- **AGROVOC** : conservation de l'URI du concept déprécié, avec une annotation **skos:changeNote** mais sans aucun libellé ;
- **EnvThes** : maintien de l'URI et du libellé préférentiel du concept déprécié, sous le concept générique "*deprecated concept*" et mise en correspondance **exactMatch/closeMatch** avec l'URI du/des nouveau(x) concept(s).

- Si les libellés dépréciés restent présents dans le fichier SKOS, les outils d'exposition (Skosmos, EcoPortal...) peuvent être paramétrés de façon à les voir ou non s'afficher dans l'interface de visualisation alphabétique/hierarchique.

## ▪ Alignements (propriété **skos:mappingRelation**)

- Eviter l'emploi de relations d'alignement (**skos:xMatch**) pour relier des concepts appartenant au *même* thésaurus.

- Parmi les différentes relations **skos:exact / close / broad / narrow** ou **relatedMatch**, seule l'équivalence en **skos:exactMatch** est une propriété **transitive**. Elle peut donc être utilisée pour relier exactement des concepts dans plusieurs vocabulaires de manière indirecte, en s'assurant que cette équivalence ne soit pas approchée.

- Les alignements (semi) automatiques se font sur une base morphosyntaxique, mettant en jeu des algorithmes de **distance/similarité** entre les termes préférentiels et/ou synonymes. Une relecture s'avère donc indispensable si l'alignement se fait à travers les (quasi) synonymes.

- Peut poser problème en particulier le traitement des **unitermes** avec des définitions différentes d'un domaine à l'autre (que seuls les concepts génériques peuvent désambiguïser).

Ex. concept "*antagonisme*" : **Gemet TG** "*propriété chimique*" / **Agrovoc TG** "*compétition biologique*" sont **abusivement** mis en **exactMatch**, alors qu'il s'agit en fait de concepts distincts.

- Lorsqu'un thésaurus se réfère à un domaine très **spécialisé** et ne dispose pas de top concepts suffisamment généraux pour un alignement automatique, il est toujours possible dans une approche **ascendante** de réaliser manuellement des correspondances en **broadMatch** avec une ressource pivot plus généraliste.

## ▪ Identifiants (concept URI)

- Identifiant **pérenne** débutant avec un nom de domaine stable/permanent, et se terminant par une séquence de caractères non signifiante (idéalement **aléatoire** et/ou **opaque**) pour chaque concept.
- L'URI d'un concept doit normalement débiter par l'**URI de base** du thésaurus.
- Il importe de ne pas modifier ces URI au cours du temps. Les concepts URI ne devraient pas varier en fonction des versions ou des plateformes d'exposition. Il est par contre possible de les **déprécier**, mais ces URI ne seront pas réutilisables pour de nouveaux concepts.
- URI **HTTP déréférencables**. En cas de sécurisation du site, il faudra également s'assurer de la bonne redirection des URI http vers des URL en https.
- *Avertissement* : Certaines plateformes telles que TemaTres (recherchables mais non maintenues) imposent leur *propres identifiants URI* pour naviguer dans leurs ressources. Ces URI ne doivent pas être utilisés pour les alignements.

## ▪ Formats

- Choix du format **SKOS (RDF)** pour l'échange de données (**standard W3C**).
- Possibilité de téléchargement (niveau **concept** et niveau **ressource**) dans divers formats et sérialisations : SKOS (RDF/XML, Turtle...), JSON-LD, CSV, PDF ...

## ▪ Métadonnées du thésaurus

- Métadonnées concernant la ressource terminologique à placer dans la balise skos:ConceptScheme.
- Possibilité de télécharger séparément les **métadonnées** décrivant la ressource (ex. AgroPortal).
- Indications minimales (**Dublin Core / DCMI Metadata Terms**) :
  - description de la ressource (dc:description / rdfs:comment)
  - nom de la ressource (dc:title / rdfs:label)
  - sujet de la ressource (dc:subject)
  - version (owl:versionInfo)
  - créateur (dc:creator) ou contributeur (dc:contributor)
  - date de création (dct:created) et modification (dct:modified) (niveaux ressource + concept)
  - identifiants universel + pérenne au niveau ressource (URI http + DOI dc:identifier).
- Autres vocabulaires RDF applicables aux métadonnées et facilitant l'interopérabilité : **DCAT** (Data Catalog Vocabulary) et/ou en complémentarité avec **VOID** (Vocabulary of Interlinked Datasets) plus spécialement adapté aux jeux de données liées en RDF, et **PROV-O** pour les informations de provenance.
- Choix d'une **licence** d'utilisation **libre** :  
ex. type Creative Commons : CC 0 ou CC BY ou CC BY-SA (cc:license)  
avec un organisme d'attribution, responsable de la ressource (cc:attributionName).

## ▪ Référencement

- Amélioration du référencement du site hébergeant la ressource pour les moteurs de recherche.
- Optimisation des temps de réponse (un explorateur de thésaurus/lexiques comme **BARTOC** ayant un *timeout* très court peut pénaliser certains services terminologiques).



## ▪ Exposition

S'assurer que tous les répertoires/entrepôts terminologiques donnent accès à la **dernière version** du thésaurus.

Nous nous sommes efforcés d'appliquer cet ensemble de bonnes pratiques en constituant le **Thésaurus de la Biodiversité** à l'Inist-Cnrs (non repris dans cette étude, mais figurant dans le tableau comparatif en annexe), mis à disposition sur [Loterre](#) (exposition via **Skosmos** et son API) et sur [AgroPortal](#) (BIODIVTHES), et signalé dans [FAIRsharing](#).

En insistant plus particulièrement sur :

- le **bilinguisme** (exhaustif)
- la richesse en **synonymes** et **termes cachés**
- les regroupements en plusieurs types de **collections**
- la **polyhiérarchie** (selon ISO 25964)
- les **associations**
- les **identifiants pérennes** (URI de type **ARK**)
- les **alignements** dans le Linked Open Data cloud
- l'**ergonomie** de l'interface
- la **FAIRitude** (données et métadonnées) : interopérabilité (**SKOS**) et réutilisabilité (licence **CC BY**)
- les multiples formats de **téléchargement** : RDF/XML, Turtle, JSON-LD, CSV, PDF

Ce thésaurus a été mentionné dans le "[Guide de bonnes pratiques sur la gestion des données de la recherche](#)" (Atelier Données – groupe de travail inter réseaux de la MITI – CNRS, 2021).

En complément, on trouvera en Annexe un **tableau comparatif** selon 24 critères, illustrant la richesse sémantique et la réutilisabilité de 8 vocabulaires contrôlés en accès ouvert (Thésaurus Biodiversité, AGROVOC, GACS, EnvThes, GEMET, EARTH, AnaEE-France et INRAE) sur le concept de "**biodiversité**".

Ce tableau est la réactualisation 2021 d'un [inventaire comparatif](#) effectué en 2014 qui nous avait conduits à la décision d'élaborer un nouveau thésaurus bilingue rassemblant les concepts clefs généraux relatifs à la composante écologique de la diversité biologique, dans le cadre d'un Groupement de Recherche sur la sémantique de la biodiversité ([GDR SémanDiv](#)) : [Semantics of Biodiversity: from Thesaurus to Linked Open Data \(LOD\)](#)

Pour des exemples de comparaisons détaillées de la gestion des relations sémantiques concernant d'autres **concepts** et/ou d'autres **thésaurus environnementaux**, on peut aussi se référer à : [Panorama des thésaurus existants dans le domaine des sciences de l'environnement](#)

## **Conclusions :**

Les points forts et les points faibles qui viennent d'être mis en évidence pour chaque thésaurus indiquent que ces ressources terminologiques doivent pouvoir se **compléter** mutuellement, en tirant parti de leur spécialisation, du multilinguisme et des synonymies.

Leur accessibilité, leur ouverture et leur interopérabilité dans le **web de données** en font des outils sémantiques **standards** appropriés, dans un contexte de développement de la **science ouverte** et des **principes FAIR**.

Le choix de thésaurus internationaux comme **pivots**, suffisamment généralistes et riches en **alignements** exactMatch (tels que **EnvThes** ou **AGROVOC**), peut permettre par ailleurs d'établir des passerelles sémantiques, avec des **interconnexions** transitives entre plusieurs ressources **ontoterminologiques** du domaine interdisciplinaire de l'Environnement.

De plus, ces vocabulaires vont constituer des référentiels pour produire des **annotations sémantiques** de résumés scientifiques ou de jeux de données, en exploitant les synonymies, les alignements, ou l'expansion hiérarchique des termes rencontrés.

Ils serviront enfin à relier des portails de publications et des portails de (méta)données environnementales au moyen d'une **indexation** partagée.

## **Sources**

***Best Practice Recipes for Publishing RDF Vocabularies*** (2008)

<https://www.w3.org/TR/swbp-vocab-pub/>

***SKOS Simple Knowledge Organization System Reference*** (2009)

<https://www.w3.org/TR/skos-reference/>

***SKOS Simple Knowledge Organization System Primer*** (2009)

<https://www.w3.org/TR/skos-primer/>

***Linked data*** (T. Berners-Lee, 2009)

<https://www.w3.org/DesignIssues/LinkedData.html>

***Key choices in the design of Simple Knowledge Organization System (SKOS)*** (2013)

<https://arxiv.org/pdf/1302.1224>

***Best Practices for Publishing Linked Data*** (2014)

<https://www.w3.org/TR/ld-bp/>

***Towards the reuse of standardized thesauri into ontologies*** (WOP, 2014)

[https://www.researchgate.net/publication/272794487\\_Towards\\_the\\_reuse\\_of\\_standardized\\_thesauri\\_into\\_ontologies](https://www.researchgate.net/publication/272794487_Towards_the_reuse_of_standardized_thesauri_into_ontologies)

***Turning FAIR into reality*** (European Commission Expert Group on FAIR Data, 2018)

[https://ec.europa.eu/info/sites/default/files/turning\\_fair\\_into\\_reality\\_1.pdf](https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf)

***Technologies pour la gestion des thésaurus et de leur interopérabilité : standards, outils de gestion et d'alignement*** (2019)

[https://e-envir.sciencesconf.org/data/pages/J2\\_CM4\\_TechnologieThesaurus\\_VachezDominique.pdf](https://e-envir.sciencesconf.org/data/pages/J2_CM4_TechnologieThesaurus_VachezDominique.pdf)

***RESSOURCES SEMANTIQUES ET TERMINOLOGIQUES EN SCIENCES DE L'ENVIRONNEMENT*** (2020)

[https://hal.archives-ouvertes.fr/hal-02907484v2/file/RESSOURCES\\_SEMANTIQUES\\_ET\\_TERMINOLOGIQUES\\_EN\\_SCIENCES\\_DE\\_L'ENVIRONNEMENT\\_09.10.2020.pdf](https://hal.archives-ouvertes.fr/hal-02907484v2/file/RESSOURCES_SEMANTIQUES_ET_TERMINOLOGIQUES_EN_SCIENCES_DE_L'ENVIRONNEMENT_09.10.2020.pdf)

***Semantics of Biodiversity: from Thesaurus to Linked Open Data (LOD)*** (2020)

[https://hal.archives-ouvertes.fr/hal-02907484v2/file/Semantics\\_of\\_Biodiversity\\_from\\_Thesaurus\\_to\\_LOD\\_VachezDominique\\_GomezIsabelle\\_2020.pdf](https://hal.archives-ouvertes.fr/hal-02907484v2/file/Semantics_of_Biodiversity_from_Thesaurus_to_LOD_VachezDominique_GomezIsabelle_2020.pdf)

***D2.5 FAIR Semantics Recommendations - Second Iteration*** (Hugo, Le Franc et al., FAIRsFAIR 2020)

<https://doi.org/10.5281/zenodo.4314321>

***Ten Simple Rules for making a vocabulary FAIR*** (2020)

<https://arxiv.org/pdf/2012.02325.pdf>

***Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web*** (Garijo et al., 2020)

<https://arxiv.org/pdf/2003.13084.pdf>

***A Semantic Web methodological framework to evaluate the support of integrity in thesaurus tools***

<https://doi.org/10.1177%2F0165551519837195> (2020)

## ANNEXES

### ▪ Les quatre règles de base du web de données liées

([Linked Data](#), *Tim Berners-Lee*, 2006/2009)

1) **identifier** les ressources avec des **URI**

2) utiliser des URI **http (déréférencables)** de façon à pouvoir **accéder** à des informations sur les ressources

3) lorsqu'on déréférence un URI, renvoyer des données **structurées** respectant les **standards du web sémantique : RDF et SPARQL**

4) **relier** les URI entre eux pour créer un **réseau** de liens et découvrir de nouvelles informations

+ **Licences d'utilisation ouvertes / libres** (ex. CC BY) → **Linked Open Data (LOD) : web de données ouvertes et liées**

### ▪ Linked Open Data: 5 Stars (2010)

<https://5stardata.info/en/>

Les ressources sémantiques pourront obtenir de 1 à 5 étoiles dans l'échelle de gradation des Linked Open Data (LOD) telle que définie par T. Berners-Lee.

\* Données **librement** accessibles sur le **web** quel que soit le format (ex. PDF, JPG), avec mention d'une **licence ouverte (open data)** ;

\*\* Données dans un format **structuré** et lisibles par la machine (ex. XLS) ;

\*\*\* Format ouvert **non-propretaire** (CSV, XML, ODS ...) ;

\*\*\*\* **URI** pour **identifier** chaque ressource et **standards** ouverts du **W3C** (langages RDF\*) pour les représenter (RDFS, OWL, SKOS, JSON-LD...) ou les interroger (SPARQL) ;

\*\*\*\*\* Données reliées à d'autres données RDF via des **alignements** entre leurs URI (**LOD cloud**)

### ▪ Five Stars of Linked Data Vocabulary Use (2014)

<http://www.semantic-web-journal.net/content/five-stars-linked-data-vocabulary-use>

**Zéro** : Données liées mais sans mention du vocabulaire utilisé

\* **Information déréférencable** décrivant le vocabulaire

\*\* **Axiomatisation** du vocabulaire (langages du **W3C** : RDF\*, OWL...)

\*\*\* **Liens sortants** vers d'autres vocabulaires (**alignements**, equivalentClass)

\*\*\*\* **Métadonnées déréférencables** sur le vocabulaire (OMV, VOAF...)

\*\*\*\*\* **Liens entrants** à partir d'autres vocabulaires

## ▪ **Principes FAIR** (2016)

Les 4 **principes FAIR** fondamentaux (Facilement retrouvable, Accessible, Interopérable, Réutilisable) présentent des caractéristiques communes avec les 4 règles de base du Linked Data (2006) :

**identifiants pérennes, protocoles d'accès standards, formats et langages interopérables** avec d'autres ressources.

Les données FAIR doivent être correctement décrites avec des **métadonnées persistantes et standardisées**, ainsi qu'une **licence d'utilisation** clairement indiquée.

Dans le but de répondre au principe d'**interopérabilité** sémantique, les (méta)données FAIR doivent utiliser des terminologies / vocabulaires contrôlés **FAIRisés** pour la représentation des connaissances.

Les principes FAIR font eux-mêmes l'objet d'une ontologie : [FAIR Vocabulary](#)

**Tableau comparatif de 8 thésaurus sur l'Environnement et la Biodiversité (concept "biodiversity")**

CONCEPT BIODIVERSITY	Nb concepts spécifiques (narrower)	Nb concepts génériques (broader)	Niveau dans arbre hiérarchique	Nb synonymes altLabel fr/en	Nb termes cachés hiddenLabel fr/en	Nb concepts adjacents (même TG)	Nb concepts associés (related)	Multilingue (dont français)	Définitions	Groupes de concept (Collection)	Concept URI (LOD déréférenciation)	Mapping exactMatch	Liens entrants (réciprocité)	Plateforme	Linked Open Data (format SKOS/RDF)	Skosmos browser	SPARQL endpoint	LOD Cloud	Téléchargement niveau ressource	Téléchargement niveau concept	Licence	FAIRsharing	Origine	Termes (pref+syno) contenant "diversit"
<b>Thésaurus Biodiversity Thesaurus</b>	6	2	2 / 4	11	30	29	7	OUI (2)	1	3	<a href="http://data.loterre.fr/ark:/6737/5/BLH-FHNG3BCR-H">http://data.loterre.fr/ark:/6737/5/BLH-FHNG3BCR-H</a>	3 (agrovoc; gemet; envthes)	NON	AgroPortal	RDF/XML Turtle JSON-LD	<a href="https://www.loterre.fr/skosmos/BLH/">https://www.loterre.fr/skosmos/BLH/</a>	<a href="https://www.loterre.fr/sparklis/">https://www.loterre.fr/sparklis/</a>		<a href="https://www.loterre.fr/science-de-la-vie-sante/biodiversite-2/">https://www.loterre.fr/science-de-la-vie-sante/biodiversite-2/</a>	RDF/XML Turtle JSON-LD	CC BY	<a href="https://doi.org/10.25504/FAIRsharing.A29ckB">https://doi.org/10.25504/FAIRsharing.A29ckB</a>	français	24
<b>AGROVOC</b>	5	1	4 / 5	6	0	4	2	OUI (32)	1	0	<a href="http://aims.fao.org/aos/agrovoc/c/33949">http://aims.fao.org/aos/agrovoc/c/33949</a>	7 (earth; gemet; nalt; eurovoc)	earth; gemet; eurovoc	AgroPortal	RDF/XML NT	<a href="http://www.fao.org/agrovoc/fr/search">http://www.fao.org/agrovoc/fr/search</a>	<a href="https://agrovoc.uniroma2.it/sparql/">https://agrovoc.uniroma2.it/sparql/</a>	<a href="https://lod-cloud.net/dataset/agrovoc">https://lod-cloud.net/dataset/agrovoc</a>	<a href="http://www.fao.org/agrovoc/releases">http://www.fao.org/agrovoc/releases</a>	RDF/XML Turtle JSON-LD	CC BY SA	<a href="https://doi.org/10.25504/FAIRsharing.a9p9t">https://doi.org/10.25504/FAIRsharing.a9p9t</a>	international	9
<b>GACS</b>	3	1	5 / 6	3	0	29	8	OUI (22)	1	1	<a href="http://id.agrisemantics.org/gacs/C1918">http://id.agrisemantics.org/gacs/C1918</a>	3 (nalt; agrovoc; cabt)	nalt	AgroPortal	RDF/XML Turtle	<a href="http://browser.agrisemantics.org/gacs/en/">http://browser.agrisemantics.org/gacs/en/</a>	<a href="https://www.lod-cloud.net/dataset/GACS">https://www.lod-cloud.net/dataset/GACS</a>	<a href="https://github.com/gacs/gacs-scheme/raw/master/src/qacs-core-scheme.ttl">https://github.com/gacs/gacs-scheme/raw/master/src/qacs-core-scheme.ttl</a>	RDF/XML Turtle JSON-LD	CC BY	<a href="https://doi.org/10.25504/FAIRsharing.p1dodf">https://doi.org/10.25504/FAIRsharing.p1dodf</a>	international	9	
<b>EnvThes</b>	6	1	6 / 7	0	0	31	1	OUI (24)	1	0	<a href="http://vocabs.ter.europa.net/EnvThes/21673">http://vocabs.ter.europa.net/EnvThes/21673</a>	6 (agrovoc; earth; eurovoc; gemet; iter)	NON	EcoPortal	Turtle	<a href="https://bartoc.skosmos.unibas.ch/envthes/en/">https://bartoc.skosmos.unibas.ch/envthes/en/</a>	<a href="http://vocabs.ceh.ac.uk/edg/tbl/sparql/">http://vocabs.ceh.ac.uk/edg/tbl/sparql/</a>		<a href="https://github.com/LTER-Europe/EnvThes/tree/master/CurrentVersion">https://github.com/LTER-Europe/EnvThes/tree/master/CurrentVersion</a>	NON	inconnue	<a href="https://doi.org/10.25504/FAIRsharing.d52o69">https://doi.org/10.25504/FAIRsharing.d52o69</a>	européen	30
<b>GEMET</b>	4	1	6 / 8	0	0	22	2	OUI (37)	2	2	<a href="http://www.eionet.europa.eu/gemet/concept/827">http://www.eionet.europa.eu/gemet/concept/827</a>	2 (agrovoc; eurovoc)	agrovoc; earth; envthes; eurovoc	AgroPortal	RDF/XML	NON	<a href="https://sema.nic.eea.europa.eu/sparql/">https://sema.nic.eea.europa.eu/sparql/</a>	<a href="https://lod-cloud.net/dataset/gemet">https://lod-cloud.net/dataset/gemet</a>	<a href="https://www.eionet.europa.eu/gemet/en/exports/rdf/latest">https://www.eionet.europa.eu/gemet/en/exports/rdf/latest</a>	NON	CC BY	<a href="https://fairsharing.org/bsqs001444/">https://fairsharing.org/bsqs001444/</a>	européen	11
<b>EARTH</b>	2	1	5 / 6	1	0	8	6	NON (2 anglais / italien)	1	2	<a href="http://linkeddata.ge.imati.cnr.it/resource/EARTH/27980">http://linkeddata.ge.imati.cnr.it/resource/EARTH/27980</a>	3 (agrovoc; eurovoc; gemet)	agrovoc; envthes	TemaTres	RDF/XML	<a href="https://skosmos.dev.finto.fi/earth/en/">https://skosmos.dev.finto.fi/earth/en/</a>	<a href="http://linkeddata.ge.imati.cnr.it:8890/sparql">http://linkeddata.ge.imati.cnr.it:8890/sparql</a>	<a href="https://lod-cloud.net/dataset/environmental-applications-reference-thesaurus">https://lod-cloud.net/dataset/environmental-applications-reference-thesaurus</a>	<a href="https://old.datahub.io/datasets/environmental-applications-reference-thesaurus">https://old.datahub.io/datasets/environmental-applications-reference-thesaurus</a>	RDF/XML	CC BY NC ND	NON	européen	12
<b>AnaEE-France Thesaurus</b>	12	1	2 / 6	0	0	7	0	OUI (2)	2	0	<a href="http://opendata.inra.fr/anaeThes/c3_2393">http://opendata.inra.fr/anaeThes/c3_2393</a>	2 (agrovoc; gemet)	NON	AgroPortal	RDF/XML	<a href="https://bartoc.skosmos.unibas.ch/anaeThes/en/">https://bartoc.skosmos.unibas.ch/anaeThes/en/</a>	<a href="http://opendata.inra.fr:8080/openrdf-sesame/repositories/anaeThes">http://opendata.inra.fr:8080/openrdf-sesame/repositories/anaeThes</a>		<a href="http://data.agroportal.lirmm.fr/ontologies/ANAEETHES/download">http://data.agroportal.lirmm.fr/ontologies/ANAEETHES/download</a>	RDF/XML Turtle	CC BY	<a href="https://doi.org/10.25504/FAIRsharing.49bmk">https://doi.org/10.25504/FAIRsharing.49bmk</a>	français	11
<b>Thésaurus INRAE</b>	9	0	1 / 2	0	0	top concept	0	OUI (2)	0	2	<a href="http://opendata.inrae.fr/thesaurus/INRAE/c5836">http://opendata.inrae.fr/thesaurus/INRAE/c5836</a>	1 (agrovoc)	NON		RDF/XML	<a href="https://consultation.vocabaire-ouverts.inrae.fr/rest/v1/thesaurus-inrae/">https://consultation.vocabaire-ouverts.inrae.fr/rest/v1/thesaurus-inrae/</a>		<a href="https://consultation.vocabaire-ouverts.inrae.fr/rest/v1/thesaurus-inrae/data">https://consultation.vocabaire-ouverts.inrae.fr/rest/v1/thesaurus-inrae/data</a>	RDF/XML Turtle JSON-LD	Licence Ouverte		français	36	