

# Panorama des thésaurus existants dans le domaine des sciences de l'environnement

1

**Ecole thématique e-ENVIR 2019 (28-31/10/2019):**

**Les données ouvertes en sciences environnementales : concepts et méthodes de qualification et de partage des données à l'ère du big data**

# Base de travail : Inventaire comparatif des thésaurus, ontologies et terminologies des sciences de l'environnement et de leurs alignements dans le web sémantique

- « *Etude comparative des principaux thésaurus sur l'Environnement et de leurs alignements. Propositions pour un thésaurus de la Biodiversité* » (2014)
- Dans le cadre du **GDR SémanDiv** (2017) **Axe 3- Recensement des standards et référentiels terminologiques disponibles** :
  - **Liste évolutive** de thésaurus généralistes ou spécialisés, glossaires, bases de données terminologiques, ontologies de domaine, ressources lexicales :  
« *Ressources sémantiques et terminologiques en sciences de l'environnement et de la biodiversité* »  
accessible par le site du GDR
  - **Base Zotero** de ressources catégorisées :  
« *Semantic and Terminology Resources in Environmental and Biodiversity Sciences* »  
accès public aux métadonnées : titre, auteurs, résumé, langues, date, URL, licence ...

# Critères de comparaison : qualité des thésaurus

- **Relations sémantiques** : **profondeur** (nombre de niveaux) **hiérarchique** (concepts spécifiques/génériques); nombre d'**associations** entre concepts (skos:broader/narrower; skos:related)
- Regroupements en **collections** thématiques (skos:Collection)
- **Types de hiérarchies** : mono ou **polyhiérarchies** ; sorte de / partie de / instance de ; **cohérence** et compatibilité
- **Richesse linguistique** : synonymes (skos:altLabel), termes cachés (hiddenLabel) permettant la **fouille de texte**
- **Equivalences linguistiques** - Multilinguisme
- Présence de **définitions terminologiques (intensionnelles)** : 1 caractère générique + des caractères distinctifs
- Utilisation dans des réseaux **européens** ou **internationaux**
- Respect des **standards RDF** / **Identifiant** pérenne (**URI**) pour chaque concept
- **URI déréréférencables**
- Interconnexion avec le **web sémantique (LOD / Linked Open Data cloud)**
- **Alignements** déjà réalisés avec d'autres thésaurus
- Liens **sortants** et/ou **entrants** (réciprocité)
- **Transitivité** des correspondances : **skos:exactMatch** (ou closeMatch ou mappingRelation)
- Ressource sous **licence libre**, **téléchargeable** et **interopérable (SKOS/RDF)**
- Affichage dans le **Linguistic Linked Open Data (LLOD) cloud**

# Conditions d'interopérabilité / réutilisabilité des thésaurus

- Outils de gestion **collaborative** (open-source de préférence)
- Exports **SKOS** disponibles et conformes aux **standards W3C**
- **Vérification** du **SKOS** (SKOS Play, qSKOS, Skosify, VocBench)
- Unicité et **pérennité** des **URI** pour les concepts
- Richesse **sémantique** (relations, définitions)
- Résolution de la **polysémie** et de l'**homonymie** (désambiguïsation);
  - **préférentiels identiques pour des concepts distincts** (ex. doublons multiples Gemet)
  - **préférentiel identique au synonyme d'un concept distinct**
  - **préférentiels identiques** au nombre près (**singulier/pluriel**) ex. pâturage(s) Agrovoc
- Cohérence de structure (norme **ISO 25964**)
- **Synonymies** et **multilinguisme** (facilitant les alignements)
- **Alignement** et exposition dans le **LOD Cloud** et le **LLOD Cloud**
- Formes variantes (termes cachés hiddenLabels) pour la **fouille de données** (TDM)

# Thésaurus généralistes avec une composante environnementale

- **AGROVOC thesaurus** (FAO) : multilingue [Linguistic Linked Open Data LLOD] SKOS [CC BY-IGO]  
<http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/>
- **NALT - National Agricultural Library's Agricultural Thesaurus** (USDA, IICA) [LOD] SKOS [CC 0] anglais/espagnol  
<http://agclass.nal.usda.gov/>
- **CAB Thesaurus** (CABI) : anglais/espagnol [CC BY-NC-ND] multilingue <http://www.cabi.org/cabthesaurus>
- **GACS - Global Agricultural Concept Space** = intersection {AGROVOC, CABT, NALT} [CC BY] : multilingue [LLOD]  
<http://browser.agrisemantics.org/gacs/en/>
- **EuroVoc Thesaurus : Domaine Environnement** (Union européenne) : multilingue [LLOD]  
<https://publications.europa.eu/fr/web/eu-vocabularies/th-concept/-/resource/eurovoc/632>
- **UNESCO microthesauri** : multilingues [LOD] [CC BY-SA] **Science: Environmental sciences and engineering & Natural resources** <http://vocabularies.unesco.org/browser/thesaurus/en/>
- **PLOS Thesaurus : Ecology and environmental sciences**  
<http://www.vocabularyserver.com/plos/index.php?tema=1609>
- **USAID Thesaurus : Natural resources and the environment**  
<http://www.vocabularyserver.com/usaidthesaurus/index.php?tema=4437&/natural-resources-and-environment>
- **UNBIS Thesaurus : NATURAL RESOURCES AND THE ENVIRONMENT**  
<https://vocabularyserver.com/unbis/en/index.php?tema=3794&/natural-resources-and-environment>

# Thésaurus du domaine Environnement

6

- **GEMET** - **General Multilingual Environmental Thesaurus** (Eionet/EEA European Environment Agency): multilingue (6500 concepts) [LLOD] [CC BY]
- **EARTH** - **Environmental Applications Reference Thesaurus** (CNR-IIA/EKOLab): anglais/italien (14350 concepts) [LLOD] [CC BY-NC-ND]
- **EnvThes** - **Environmental Thesaurus** (LTER-Europe: EnvEurope-Life+/ExpeER) (référentiel du réseau ILTER / DEIMS-SDR : International Long Term Ecological Research) anglais (3700 concepts) [LOD]
- **Thésaurus AnaEE** (Analyse et Expérimentation sur les Écosystèmes/AnaEE-France) français/anglais (3300 concepts) [LOD] [CC BY]
- **Thésaurus de la Biodiversité** (Inist-CNRS) : bilingue (français/anglais) (615 concepts) [LOD] [CC BY]
- **Thésaurus d'Irstea** (Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture) français (5800 concepts)
- **OZCAR-Theia thesaurus** (Observatoires de la Zone Critique: Application et Recherche - Pôle thématique surfaces continentales) anglais (400 concepts) [LOD]
- **GCMD Earth Science keywords Thesaurus** (Global Change Master Directory/NASA)
- **AFO - Natural resource and environment ontology** (National Library of Finland) [CC BY] anglais/finlandais (4400 concepts en anglais)
- **Earth Observation Thesaurus (EOStem)** (CNR-IIA-EKOLab) anglais/italien (2600 concepts)

# Thésaurus spécialisés

- ▶ **LifeWatch Italy thesauri** anglais/italien (LifeWatch ERIC) [LOD]  
(Alien species / Aquatic organisms / Fish traits / Phytoplankton / Zooplankton)
- ▶ **Thésaurus Eau** (OIEau/AFB) bilingue [CC BY-NC-SA] *téléchargeable sur inscription*
- ▶ **Law & Environment Ontology (LEO)** (InforMEA : United Nations Information portal on Multilateral Environmental Agreements) multilingue
- ▶ **Thesaurus Of Plant characteristics for ecology and evolution (TOP)** (CEFE-CNRS) OWL *anglais*
- ▶ **Thesaurus for Soil Invertebrate Trait-based Approaches (T-SITA)** (BETSI/CESAB) OWL *anglais*
- ▶ **Thésaurus d'Agroécologie** (INRA) *français* [CC BY]

# Thésaurus dans des formats incompatibles avec le LOD

8

- **Biocomplexity Thesaurus** (USGS/U.S. Geological Survey) anglais [*en libre consultation mais non disponible en SKOS*]
- **ThesBio - thesaurus em biodiversidade** (BHL SciELO/MZUSP, Brésil) interface *portugaise*, préférentiel monolingue, définitions trilingues, non téléchargeable, non aligné
- **ASFA/ASFIS Thesaurus** (FAO-Aquatic Sciences and Fisheries Information System Thesaurus) multilingue [*n'est plus accessible dans le LOD, le pdf (2009) reste disponible*]
- **Thésaurus MOTBIS (Microthésaurus nature et environnement)** et **THESAGRI** (Réseau Canopé)
- **Thésaurus Ecoplanète** (Portail du SIDE: Système d'Information du développement durable et de l'environnement / Ministère de la transition écologique et solidaire)
- **Thésaurus Environnement Thes'Envir** (Réseau Régional d'Education à l'Environnement Languedoc-Roussillon) 10000 entrées monolingues [*payant*]
- **EnVoc Multilingual Thesaurus of Environmental Terms** (United Nations Environment Programme / Infoterra)
- **"Thésaurus" FRB** (Fondation pour la Recherche sur la Biodiversité) classement thématique
- **Thésaurus francophone du développement durable** (Agora 21)

# Intégration entre différents thésaurus

## ► **LusTRE** - Linked Thesaurus fRamework for Environment

(eENVplus/ IMATI-CNR/ NatureSDIPlus) <http://showcase.eenvplus.eu/client/thesaurus.htm>

= **EARTH + GEMET + AGROVOC + EuroVoc** (+ **EUNIS-Habitats + EUNIS-Species** / European Nature Information System) (**multilingue** mais *interface + interrogation uniquement en anglais*, avec recherche de traductions)

## ► **GACS** - Global Agricultural Concept Space : intersection de 10 000 concepts

< **AGROVOC & NALT & CABT** (interface Skosmos **multilingue**)

## ► **AgroPortal** : AGROVOC, NALT, GACS, GEMET, AnaEE, Biodiversity, TOP ...

(*interrogeable uniquement par l'anglais*)

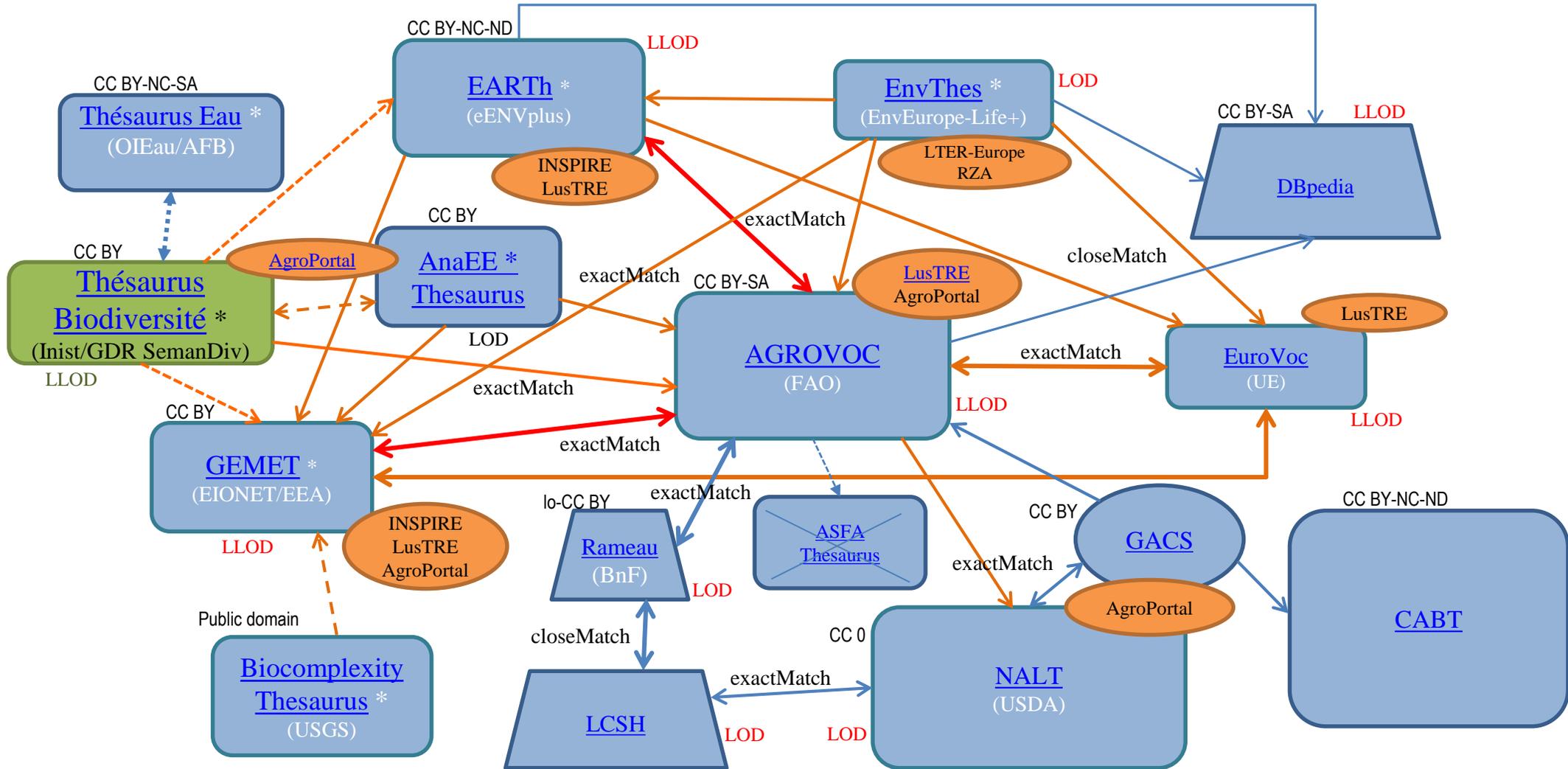
## ► **EcoPortal** : LifeWatch Italy thesauri, EnvThes ... (en cours de développement)

# Base Zotero : ressources sémantiques et terminologiques catégorisées (sciences de l'environnement)

[https://www.zotero.org/groups/2181298/semantic\\_and\\_terminology\\_resources\\_in\\_environmental\\_and\\_biodiversity\\_sciences/items](https://www.zotero.org/groups/2181298/semantic_and_terminology_resources_in_environmental_and_biodiversity_sciences/items)

- Library
  - **1-Thésaurus spécialisés environnement**
  - **2-Thésaurus généralistes**
  - 3-Ontologies de domaine
  - 4-Portails d'ontologies
  - 5-Terminologies (glossaires, nomenclatures, vocabulaires spécialisés)
  - 6-Bases de données terminologiques
  - 7-Ressources Linked Open Data
  - 8-Outils de gestion terminologiques
  - 9-Documents de référence

Concept « Biodiversity » (concept-URI)  
among **thesauri** on Environment and Agronomy in the **web of data**



# *Etude préalable*

Présélection (2014) de 10 thésaurus (pertinence, taille, audience) :

- **USGS Biocomplexity Thesaurus, AGROVOC, EnvThes, GEMET, EARTH, NALT, EuroVoc,**
- **UNESCO, UNBIS, THESAGRI**



Choix de 6 thésaurus :

**AGROVOC, GEMET, EARTH, EnvThes, USGS, NALT**

auxquels se sont ajoutés (2015-2017) :

- Thésaurus **Biodiversité**
- Thésaurus **AnaEE**
- Thésaurus **Irstea**
- Thésaurus **GACS**
- Thésaurus **Eau**

# Thésaurus AGROVOC

<http://agrovoc.uniroma2.it/agrovoc/agrovoc/en/>

- **FAO Food and Agriculture Organization** (ONU)
- **36 500 concepts** (agriculture, foresterie, alimentation, **environnement**)
- **multilingue** (35 langues) dont 33 800 termes **français**
- **termes sans équivalent français** : *antibiosis, environmental engineering, fishing, natural environment*
- riche en **synonymes** et **quasi-synonymes** (anglais 9800; français 7800) ; **absence** de **hiddenLabels**
- quelques **définitions** + des **scope-notes**
- regroupements sous **25 top-concepts** (**pas de collections**) ; profondeur hiérarchique : **7 niveaux**
- **schéma de concepts** formalisé en OWL puis en **SKOS-XL** (2009)
- relations **ontologiques** (onto-terminologie fondée sur [Agrontology](#))
- intègre le vocabulaire du thésaurus **LandVoc** (Linked Land Governance Thesaurus)
- nombreux **alignements** (en exact/ close ou broadMatch) sortants et entrants dans le **Linked Open Data** (18 vocabulaires)
- **téléchargeable** (**SKOS** Core, RDF/XML et NT)
- interface de **consultation** multilingue **Skosmos** + SPARQL **endpoint**
- présent sur [AgroPortal](#) (mode dégradé)
- **licence libre CC-BY** IGO 3.0
- présent dans le **Linguistic Linked Open Data** cloud (LLOD)

# AGROVOC Multilingual Thesaurus

Content language English  Search

14

Alphabetical Hierarchy

- resources
  - cultural heritage
  - economic resources
  - educational resources
  - energy sources
  - human resources
  - inputs
  - natural resources
    - agricultural resources
    - biological resources
      - animal resources
      - biodiversity**
        - agrobiodiversity
        - forestry biodiversity
        - genetic diversity (as resource)
        - natural heritage
      - genetic resources
      - plant resources
      - wildlife
    - commons
    - communal territory
    - energy resources
    - feed resources
    - fishery resources
    - food resources
    - forest resources
    - marine resources
    - nonrenewable resources
    - renewable resources
    - soil resources
    - raw materials

resources > natural resources > biological resources > biodiversity

PREFERRED TERM

**biodiversity**

DEFINITION

The variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems. Diversity indices are measures of richness (the number of species in a system); and to some extent, evenness (variances of species' local abundance). They are therefore indifferent to species substitutions which may, however, reflect ecosystem stresses (such as those due to high fishing intensity).  
(en)

BROADER CONCEPT

[biological resources \(en\)](#)

NARROWER CONCEPTS

[agrobiodiversity \(en\)](#)  
[forestry biodiversity \(en\)](#)  
[genetic diversity \(as resource\) \(en\)](#)  
[natural heritage \(en\)](#)

ENTRY TERMS

[biological diversity \(en\)](#)  
[ecosystem diversity \(en\)](#)  
[species diversity \(en\)](#)

SCOPE NOTE

The variety of life on earth at the genetic, species and ecosystems levels  
(en)

HAS COMPONENT

[species \(en\)](#)

INFLUENCES

[ecosystems \(en\)](#)  
[soil functions \(en\)](#)

IS INFLUENCED BY OR  
DEPENDS ON

[biosafety \(en\)](#)

IS OBJECT OF ACTIVITY

[biodiversity conservation \(en\)](#)  
[bioprospecting \(en\)](#)

IS PROPERTY OF

[gene pools \(en\)](#)  
[germplasm \(en\)](#)

IN OTHER LANGUAGES

تنوع حيوي (en) Arabic

[http://aims.fao.org/aos/agrovoc/c\\_33949](http://aims.fao.org/aos/agrovoc/c_33949)  
biodiversity

Property	Value
rdf:type	skos:Concept
skos:inScheme	<a href="http://aims.fao.org/aos/agrovoc">http://aims.fao.org/aos/agrovoc</a>
skos:broader	<a href="http://aims.fao.org/aos/agrovoc/c_92341">http://aims.fao.org/aos/agrovoc/c_92341</a>
skos:narrower	<a href="http://aims.fao.org/aos/agrovoc/c_37977">http://aims.fao.org/aos/agrovoc/c_37977</a> <a href="http://aims.fao.org/aos/agrovoc/c_37967">http://aims.fao.org/aos/agrovoc/c_37967</a> <a href="http://aims.fao.org/aos/agrovoc/c_0a269dfe">http://aims.fao.org/aos/agrovoc/c_0a269dfe</a> <a href="http://aims.fao.org/aos/agrovoc/c_33952">http://aims.fao.org/aos/agrovoc/c_33952</a>
skos:exactMatch	<a href="http://eurovoc.europa.eu/5463">http://eurovoc.europa.eu/5463</a> <a href="http://lod.gesis.org/thesoz/concept/10065031">http://lod.gesis.org/thesoz/concept/10065031</a> <a href="http://linkeddata.ge.imati.cnr.it:2020/resource/EARTH/27980">http://linkeddata.ge.imati.cnr.it:2020/resource/EARTH/27980</a> <a href="http://lod.nal.usda.gov/nalt/5559">http://lod.nal.usda.gov/nalt/5559</a> <a href="http://www.eionet.europa.eu/gemet/concept/827">http://www.eionet.europa.eu/gemet/concept/827</a> <a href="http://zbw.eu/stw/descriptor/19625-0">http://zbw.eu/stw/descriptor/19625-0</a>
skos:closeMatch	<a href="http://dbpedia.org/resource/Biodiversity">http://dbpedia.org/resource/Biodiversity</a>

Property	Value
dcterms:created	2012-02-03T16:29:11Z
dcterms:modified	2019-06-15T20:52:06.21+02:00
<a href="http://aims.fao.org/aos/agrovoc/c_7280">http://aims.fao.org/aos/agrovoc/c_7280</a>	<a href="http://aims.fao.org/aos/agrovoc/c_7280">http://aims.fao.org/aos/agrovoc/c_7280</a>
<a href="http://aims.fao.org/aos/agrovoc/c_2482">http://aims.fao.org/aos/agrovoc/c_2482</a>	<a href="http://aims.fao.org/aos/agrovoc/c_2482">http://aims.fao.org/aos/agrovoc/c_2482</a> <a href="http://aims.fao.org/aos/agrovoc/c_e5629cbd">http://aims.fao.org/aos/agrovoc/c_e5629cbd</a>
<a href="http://aims.fao.org/aos/agrovoc/c_35030">http://aims.fao.org/aos/agrovoc/c_35030</a>	<a href="http://aims.fao.org/aos/agrovoc/c_35030">http://aims.fao.org/aos/agrovoc/c_35030</a>
<a href="http://aims.fao.org/aos/agrovoc/c_50344">http://aims.fao.org/aos/agrovoc/c_50344</a>	<a href="http://aims.fao.org/aos/agrovoc/c_50344">http://aims.fao.org/aos/agrovoc/c_50344</a> <a href="http://aims.fao.org/aos/agrovoc/c_33955">http://aims.fao.org/aos/agrovoc/c_33955</a>
<a href="http://aims.fao.org/aos/agrovoc/c_3212">http://aims.fao.org/aos/agrovoc/c_3212</a>	<a href="http://aims.fao.org/aos/agrovoc/c_3212">http://aims.fao.org/aos/agrovoc/c_3212</a> <a href="http://aims.fao.org/aos/agrovoc/c_3249">http://aims.fao.org/aos/agrovoc/c_3249</a>
void:inDataset	<a href="http://aims.fao.org/aos/agrovoc/void.ttl#Agrovoc">http://aims.fao.org/aos/agrovoc/void.ttl#Agrovoc</a>
skos:definition	<a href="http://aims.fao.org/aos/agrovoc/xDef_13bbf236">http://aims.fao.org/aos/agrovoc/xDef_13bbf236</a>
skos:scopeNote	The variety of life on earth at the genetic, species and ecosystems levels A diversidade da vida na terra aos níveis da genética, da espécie e do ecossistema Genetyczne zróżnicowanie życia na ziemi, różnorodność gatunkowa ekosystemów A földi élet változatossága genetikai, faji szinten, ill. az ökoszisztéma szintjén Rozmanitost žívota na Zemi na úrovni genetické, druhové a ekosystémové Genetik açıdan, türler açısından ve ekosistem düzeyinde yeryüzündeki yaşam çeşitliliği ຄວາມຫຼາກຫຼາຍຂອງສິ່ງມີຊີວິດຢູ່ໃນໂລກ ນັບຕັ້ງແຕ່ລະດັບຂອງຜັງທຸກ ຫາ ລະດັບ ຂອງລະບົດຜັງ ຮອດ ລະດັບຂອງລະບົບນິເວດ. Varietà della vita sulla terra a livello genetico, delle specie e degli ecosistemi Die Vielfalt des Lebens auf der Erde auf der Genetik-, Spezies- und Ökosystemebene Variabilidad de especies y ecosistemas Variété et diversité de toutes les composantes du monde vivant aux niveaux génétiques, spécifiques et des écosystèmes تنوع زیستی بر روی زمین در سطوح ژنتیک، گونه‌ها و بوم‌سازگان‌ها ความหลากหลายของสิ่งมีชีวิตบนโลกตั้งแต่ระดับพันธุกรรม สปีชีส์และระบบนิเวศน์

prefLabel	altLabel	Lang
تنوع حيوي		ar
biodiverzita	biologická diverzita diverzita ekosystému druhová rozmanitost	cs
Biodiversität	Biodiversitaet Biologische Diversitaet Diversitaet im Oekosystem Diversitaet der Spezies Artenvielfalt Biologische Vielfalt Biologische Diversität Diversität im Ökosystem Diversität der Spezies	de
biodiversity	biological diversity ecosystem diversity species diversity	en
Biodiversidad	Diversidad biológica Diversidad de ecosistemas Diversidad de especies	es
زیست‌گوناگونی	گوناگونی زیستی گوناگونی بوم‌سازگان گوناگونی گونه(ها)	fa
Biodiversité	Diversité biologique Diversité d'écosystèmes Diversité des espèces	fr
जैव विविधता	जैविक विविधता पारिस्थितिक तंत्र विधिता जाति विविधता	hi
biodiverzitás	biológiai diverzitás ökoszisztéma-diverzitás fajdiverzitás	hu
Biodiversità	Diversità biologica Diversità degli ecosistemi Diversità di specie	it
生物多样性	生態系多様性 種多様性、種多様化	ja
ຄວາມຫຼາກຫຼາຍທາງຊີວະພາບ	ຄວາມຫຼາກຫຼາຍທາງດ້ານຊີວະວິທະຍາ ຄວາມຫຼາກຫຼາຍທາງລະບົດຜັງ ຄວາມຫຼາກຫຼາຍທາງລະບົດ	lo
Różnorodność biologiczna	Bioróżnorodność Różnorodność ekologiczna Zróżnicowanie gatunkowe	pl
Biodiversidade	Diversidade biológica Diversidade de ecossistema Diversidade de espécies	pt
биоразнообразие	биологическое разнообразие разнообразие экосистем	ru

# Thésaurus GEMET

<http://www.eionet.europa.eu/gemet>

- **General Multilingual Environmental Thesaurus** (Agence **Européenne** de l'Environnement / Eionet - EEA ) compatible thématiques **INSPIRE**
- **6500 concepts** (résultant de la fusion de plusieurs thésaurus/lexiques)
- **multilingue** (langues européennes)
- **définitions**
- téléchargeable en **SKOS**
- **regroupements thématiques** : GEMET utilise des groupes de concepts (35 **Groups**) : champs sémantiques regroupés en 4 grands **domaines** (ou super-groupes), ainsi que des thématiques distinctes (40 **Themes**) *mais dont certaines ont les mêmes intitulés que des groupes*
- *confusion entre les libellés de certains concepts, groupes et thématiques (ex. Biosphere)*
- profondeur hiérarchique : **8 niveaux**
- **alignements** multiples, compatible avec le **Linguistic Linked Open Data**
- présent dans **AgroPortal**, BARTOC
- **téléchargeable**
- licence **libre CC BY**
- **absence totale** de **synonymes**
- **nombreux préférentiels français identiques (anglais distincts)** : ex. « *aménagement du territoire* » (4 **URI** différents, dans des groupes ou hiérarchies différents → problème pour les alignements)
- **incohérences** hiérarchiques (ex. « biosphere » NT anatomy, ecology, organism)

# Thésaurus EnvThes

<http://vocabs.lter-europe.net/edg/tbl/EnvThes.editor>

- **Environmental Thesaurus** (EnvEurope-Life+ / ExpeER)
- **spécialisé** Environnement
- communauté : réseau **européen** / **LTER-Europe** (Long Term Ecological Research) + Réseau **Zones Ateliers** (SOERE RZA)
- **3700 concepts ; 7 super-concepts**
- *anglais* majoritaire
- **entièrement remanié** → **multilinguisme** (préférentiels, définitions); **hiérarchies**
- présent dans **EcoPortal** et **BARTOC**
- téléchargeable en **SKOS** ; compatible **Linked Open Data** ; (licence ?)
- basé sur **ontologies** OGC O&M, OBOE et SERONTO
- *interface monolingue anglais*
- **alignements uniquement sortants** (Agrovoc, Gemet, EARTH, US-LTER)
- *alt/hiddenLabel* peu ou pas utilisés; **altLabels non recherchables**
- **2900 concepts dépréciés interrogeables et conservés dans le thésaurus**

# Thésaurus EARTH

<http://linkeddata.ge.imati.cnr.it/resource/EARTH/>

- **Environmental Applications Reference Thesaurus**
- inclus dans la plateforme **européenne LuSTRE (eENVplus: eEnvironmental services for advanced applications within INSPIRE)**
- maintenu par **CNR-IIA-EKOLab** (Conseil national de la recherche, Italie)
- composé à partir de sources terminologiques multiples
- 14350 concepts ; bilingue **anglais**/italien (*pas de français*)
- **5 top-concepts** de haut niveau
- licence CC **BY-NC-ND**
- Linguistic Linked Open Data
- téléchargeable en **SKOS** - RDF/XML (version 2013)
- **alignements** avec Gemet, Agrovoc et EuroVoc
- *absent d'AgroPortal*
- se positionne comme concurrent de Gemet mais sans *le multilinguisme*

# Thésaurus GACS

<http://browser.agrisemantics.org/gacs/en/>

- **Global Agricultural Concept (Scheme) Space** (FAO/ USDA/ CABI)
- Projet d'**intégration** de 3 grand thésaurus en agronomie (démarré 2013): 10 000 concepts à l'intersection d'**AGROVOC**, **NALT** et **CABT**
- mise au point de technologies d'**alignements**
- plusieurs scénarios de structuration hiérarchique (**top-concepts**)
  - A- par **catégories sémantiques** de haut niveau (events&action/objects/properties) [ex. EARTH]
  - B- par **facettes** (activities/entities/features/organisms/phenomena/sites/subjects/substances...) [ex. Agrovoc]
  - C- par **classement thématique** (faible profondeur hiérarchique) [ex. CAB / Unesco thesaurus]
- GACS Core : **15 400 concepts**
- ambition : servir de **hub** pour d'autres thésaurus dans le LOD
- présent dans AgroPortal
- au cœur d'un projet d'alignements de thésaurus/ontologies : **D2KAB (Data to Knowledge in Agronomy and Biodiversity)**

# Thésaurus AnaEE (AnaEE-France Thesaurus)

<http://opendata.inra.fr/anaeeThes/>

- Analyse et Expérimentation sur les Écosystèmes / AnaEE-France)
- Infrastructure **européenne** spécialisée recherche expérimentale
- **3300** concepts; **anglais**/français; *1450 préférentiels français*
- **12 top-concepts**; *pas de collection*
- 360 concepts issus de l'ontologie OBOE
- **alignements** avec Agrovoc et Gemet
- présent dans **AgroPortal**, BARTOC
- **licence libre CC BY**
- à développer : préférentiels français

# Thésaurus d'Irstea

<https://voculaires.irstea.fr/skosmos/thesaurus/en/>

- IRSTEA fusionne en 2020 avec l'INRA → **INRAE** / Institut national de recherche pour l'agriculture, l'alimentation et l'**environnement**
- thésaurus récemment **skosifié** ; en cours de fusion avec **VocInra**
- interface de consultation **Skosmos**
- formalisation combinant des propriétés **SKOS** et **Dublin Core**
- **5800 concepts** (*majoritairement français* + quelques termes anglais)
- *peu structuré hiérarchiquement au niveau concept* : **3000 top-concepts**
- **regroupements** de concepts en domaines/microthésaurus, mais *ne sont pas gérés comme des collections*
- les **microthésaurus** sont considérés comme des sous-ConceptScheme <dc:isPartOf> du thésaurus et sont eux-mêmes classés en 17 grands domaines <dc:subject>
- les 17 **domaines** de regroupement sont les top-concepts d'un schéma de concepts « Domaines »  
→ importé dans VocBench : **autant de Schemes que de microthésaurus**
- *le même terme* (ex « **environnement** ») peut être utilisé comme **libellé** de domaine, de microthésaurus et de concept
- langue **anglaise** présente dans le fichier SKOS mais **n'apparaît pas** dans Skosmos
- **alignements** avec **Gemet** et **Agrovoc**, mais **aucun URI** (thésaurus, domaines ou concepts) n'est **déréférençable**
- *non téléchargeable*

# Thésaurus Eau (et biodiversité)

<http://thesaurus.oieau.fr/thesaurus/resource/ark:/99160/7af302a6-7518-4a8a-84a6-b8df7b595e14>

- Office International de l'Eau + Agence Française pour la Biodiversité
- 2700 concepts ; **biligue** français et anglais
- regroupements en 44 **thématiques** (iso-thes:ConceptGroup)
- tous les URI **déréférençables**
- **typographie** pauvre : majuscules sans accent
- **pas de hiddenLabel**
- **incohérences hiérarchiques** : « biodiversité » TS de « peuplement »; « habitat » TS de « biotope »
- **pas d'alignement**
- licence CC BY-NC-SA
- compatible SKOS Play, OnAGUI
- harmonisation à trouver avec le Glossaire sur l'eau (préférentiels, définitions)
- alignement envisagé avec le Thésaurus Biodiversité

# EcoLexicon

<http://ecolexicon.ugr.es/en/index.htm>

- ▶ Université de Grenade (projet Lexicon)
- ▶ **base de connaissances** terminologiques **multilingue** sur l'**environnement** (20 000 termes / 7 langues)
- ▶ **ontologie** de domaine (3500 concepts)
- ▶ **interface graphique** de consultation de **thésaurus**
- ▶ 5000 termes anglais / *seulement* 800 termes **français**
- ▶ objectif : affiner sémantiquement les relations **hiérarchiques** :
  - ▶ relations **générique**-spécifique : catégoriser plusieurs types d'**hyponymies** (10)
  - ▶ relations **partitives** : catégoriser plusieurs types de **méronymies** (6)
- ▶ *format actuellement non interopérable* : objectif d'intégration dans le LLOD **Linguistic Linked Open Data**
- ▶ alignements en cours avec DBpedia, GEMET, GeoNames

# Biocomplexity Thesaurus

[https://www1.usgs.gov/csas/biocomplexity\\_thesaurus/](https://www1.usgs.gov/csas/biocomplexity_thesaurus/)

- USGS (U.S. Geological Survey)
- thésaurus **orienté environnement**
- fusion de plusieurs thésaurus et glossaires d'organismes nord-américains
- **9500 concepts**
- **richesse** sémantique (hiérarchies, associations, synonymes)
- **librement** navigable
- *monolingue* **anglais**
- *absence d'URI* pour les concepts
- *absence de définition*
- *non téléchargeable*
- *non aligné*
- domaine public mais **inexploitable pour le LOD**

# Thésaurus Biodiversité

<http://data.loterre.fr/ark:/67375/BLH>

- Inist-Cnrs / GDR SémanDiv
- 615 concepts thématiques
- bilingue (français/anglais)
- 867 synonymes + 4600 termes cachés + 83 définitions  
+ 400 relations associatives
- profondeur hiérarchique à 10 niveaux
- 3 types de **regroupements** de concepts : par **catégories**, par **domaines** et par classes **EBV (Essential Biodiversity Variables)**
- interface **navigation Skosmos**
- **licence libre CC BY**
- **exposé** dans le **LOD : Loterre (Inist), Ortolang, Thesauform (Cesab) et AgroPortal**
- **téléchargeable** en **SKOS, CSV, PDF**
- **aligné** vers **Agrovoc** (+ GEMET, ENVO, EnvThes, Thésaurus EAU, **T-SITA**) avec **OnAGUI**



## Perspectives d'évolution du thésaurus Biodiversité

- Enrichissement **collaboratif** du thésaurus avec des experts écologues (**définitions**) du **GDR SémanDiv**
- **Interconnexion** avec des thésaurus internationaux (thésaurus pivots pour des **alignements**): GACS, GEMET, EnvThes, EARTH, AnaEE ...
- **Intégration** dans le **LLOD cloud** (Linguistic Linked Open Data), dans **EcoPortal** et dans **FAIRsharing**
- Affinement des relations hiérarchiques (**génériques, partitives**)
- Evolution vers une **ontologie** de la biodiversité (VocBench 3, AgroPortal) permettant d'effectuer des raisonnements, et alignable avec d'autres ontologies de domaine (ENVO, OBOE)
- Aide à la **fouille** de données textuelles (formes variantes hiddenLabel)
- Enrichissement par **extraction terminologique** d'un corpus bilingue d'articles et de rapports de synthèse sur les sciences de la biodiversité (outil de TDM / TermSuite)

# Référentiels terminologiques des portails de (méta)données

[Dat@OSU](#) – [OREME DATA](#) – [BBEES/InDoRES](#)

- **Agronomie**: thésaurus **AGROVOC**
- **Environnement, écologie**: thésaurus **GEMET & EnvThes**
  - **AGROVOC** : onto-terminologie spécialisée agriculture/alimentation (forte composante environnementale); **international** (FAO); **multilingue**
  - **GEMET** : cible **généraliste** environnementale; **européen** (AEE); compatible thématiques INSPIRE; **multilingue**; **absence de synonymes**
  - **EnvThes** : **spécialisé** environnement; communauté plus restreinte (**LTER** Europe); **européen** (EnvEurope); *anglais préférentiel*

→ *Complémentarités à trouver*

## Annexes : comparaison de Thésaurus

Exemples de traitements différenciés des **relations sémantiques** : hiérarchies, polysémies, équivalences, alignements

# Hiérarchie : concepts « Habitat / Biotope »

- **Agrovoc** : "habitats" TG "ecosystems" ; **Has Component** "biotopes"
- **NALT** : "habitats" TG "community ecology" ; **TS** "biotopes"  
*Déf. "The natural environment where an organism, population or community lives, including biotic and abiotic factors"*
- **GEMET** : "habitat" est placé au même niveau hiérarchique que "biocoenosis", "biotope" et "ecosystem" sous le TG "synecology"
- **EnvThes** : "habitat" et "biotope" étaient sous le même TG "organizational unit" → TG « ecosystem » et TG « environment » (l'inverse serait plus correct)
- **AnaEE** : "habitat structure" TG "ecosystem structure" ; **TS** "biotope"

→ **Thésaurus Biodiversité** : "habitat" TG "environnement" ;  
**TA** "biotope" (TG "écosystème" ), "milieu écologique" , "niche écologique"

# Polysémie : concept « Symbiose »

- **Agrovoc** : "symbiosis" TG "biological interaction" ; TS "commensalism", "mutualism", "mycorrhizal infection", "root nodulation"
- **CAB Thesaurus** : "symbiosis" TS "commensalism", "mutualism", "mycorrhizas", "parasitism", "phoresy"  
Déf. "The interaction of two organisms living together in more or less intimate association"
- **GEMET** : "symbiosis" TG "autoecology" ; TA "mycorrhiza"  
Déf. "A close and mutually beneficial association of organisms of different species"
- **EnvThes** : "symbiosis" avait pour TG "species interactions" → TG « biological system »  
2 définitions : sens large (close and obligatory association) et sens strict (mutualism)
- **AnaEE** : ignore le concept

→ **Thésaurus Biodiversité** : "symbiose" TG "mutualisme" ; TS "mycorhize"  
(pas d'équivalence exacte entre concepts pour les alignements)

# Equivalence : concept « Diversité spécifique »

- **Agrovoc** : "species diversity" renvoie sur le préférentiel "biodiversity"
- **GEMET** : ignore le concept
- **EnvThes** : "species diversity" TG "biodiversity" ; TA "species richness"
- **AnaEE** : ignore le concept

→ **Thésaurus Biodiversité** : "diversité spécifique" TG "biodiversité" ;  
TA "richesse spécifique", "abondance écologique"

# Comparaison de thésaurus sur l'Environnement et la Biodiversité

Relations sémantiques et alignements entre concepts (→)

Concept "biodiversity"

- **AGROVOC** [http://aims.fao.org/aos/agrovoc/c\\_33949](http://aims.fao.org/aos/agrovoc/c_33949)

1 TG, 4 TS, **6 SYN**, **8 RO** (agronomy) + 1 DEF [multilingue-**français**]

→ **EARTH**, EuroVoc, **GEMET**, NALT

- **GEMET** <http://www.eionet.europa.eu/gemet/concept/827>

1 TG, 4 TS, **0 SYN**, 2 TA + 2 DEF [multilingue-**français**]

→ **Agrovoc**, EuroVoc

- **EARTH** <http://linkeddata.ge.imati.cnr.it/resource/EARTH/27980>

1 TG, 2 TS, 1 SYN, **6 TA** + 1 DEF [bilingue anglais-italien]

→ **Agrovoc**, EuroVoc, GEMET

- **EnvThes** <http://vocabs.lter-europe.net/EnvThes/21673>

1 TG, **6 TS**, **0 SYN**, 1 TA + 1 DEF [**unilingue**]

→ Agrovoc, EARTH, EuroVoc, GEMET

TG: Terme générique

TS: Terme spécifique

SYN: Synonyme

TA: Terme associé

RO: Relation ontologique

DEF: Définition

→ liens réciproques en **gras**

# Comparaison de thésaurus sur l'Environnement et la Biodiversité

Concept «biodiversity»

- **EuroVoc** <http://eurovoc.europa.eu/5463>

1 TG, **0 TS**, 1 SYN, 1 TA + 1 DEF

[multilingue-**français**]

→ **Agrovoc, GEMET**

- **NALT** <http://lod.nal.usda.gov/nalt/5559>

2 TG, 3 TS, 2 SYN, **12 TA** + 1 DEF

[bilingue anglais-espagnol]

→ **GACS, LCSH** (*alignements non affichés*)

---

- **Thésaurus Biodiversité** <http://data.loterre.fr/ark:/67375/BLH-FHNG3BCR-H>

2 TG, **6 TS**, **11 SYN**, **7 TA** + 1 DEF + 20 **hiddenLabels**

[bilingue anglais-**français**]

→ **Agrovoc, GEMET**

## Comparaison de thésaurus sur l'Environnement et la Biodiversité

### Relations sémantiques entre concepts

#### Concepts “species diversity” et “species richness”

*absents de GEMET, EARTH et Agrovoc*

- **Agrovoc** : “species diversity” est synonyme de “biodiversity”
- **NALT** : “species richness” est synonyme de “species diversity”
- **AnaEE** : “species richness” (TG “biodiversity”)
- **EnvThes** : “species diversity” (TG “biodiversity”) / TA “species richness” (TG “ecosystem parameter”)
- **USGS** : “Species diversity” (TG “Biodiversity”) / “Species richness” (TG “Species-area relationship”)

→ **Thésaurus Biodiversité** : “diversité spécifique” (TG “biodiversité”)

TA “richesse spécifique”, “abondance écologique” (TG “indicateur de biodiversité”)

#### Concept “niche écologique”

*totalemment absent de EARTH, EnvThes, Agrovoc et AnaEE*

- **NALT** : “niches” (TG “habitats”)
- **GEMET** : “ecological niche” (TG “ecological parameter”)
- **USGS** : “Niches” TA “Biotopes”, “Ecosystems” et “Habitats”

→ **Thésaurus Biodiversité** : “niche écologique” (TG “paramètre écologique”) ; TA “habitat”

**Acronymes :****AnaEE** : **A**nalyse et **E**xpérimentation sur les **E**cosystèmes**CABT** : **C**entre for **A**gricultural **B**ioscience International **T**hesaurus**EARTH** : **E**nvironmental **A**pplications **R**eference **T**hesaurus**EBV** : **E**ssential **B**iodiversity **V**ariable**EnvThes** : **E**nvironmental **T**hesaurus**GACS** : **G**lobal **A**gricultural **C**oncept **S**pace**GEMET** : **G**eneral **M**ultilingual **E**nvironmental **T**hesaurus**IMBE** : Institut **M**éditerranéen de **B**iodiversité et d'**E**cologie marine et continentale**Irstea** : Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture**LLOD** : Linguistic **L**inked **O**pen **D**ata**Loterre** : **L**inked **O**pen **T**ERminology **R**Esources**LusTRE** : **L**inked **T**hesaurus **f**Ramework for **E**nvironment**NALT** : **N**ational **A**gricultural **L**ibrary **T**hesaurus**SKOS** : **S**imple **K**nowledge **O**rganization **S**ystem**Liens :****Site rBDD** (ateliers)<http://rbdd.cnrs.fr/spip.php?article170><http://rbdd.cnrs.fr/spip.php?article193>**Site Loterre**<https://www.loterre.fr/skosmos/BLH/fr/>**Site GDR SemanDiv**<http://www.semandiv.cnrs.fr/>**Inist-CNRS**