



# Technologies pour la gestion des thésaurus et de leur interopérabilité

standards, outils de  
gestion et d'alignement

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

# Introduction

- ▶ Nous distinguerons :
  - ▶ la « **terminologie** » en tant que **science** de la structure, de la formation et de l'usage des **termes** dans différents **domaines** de spécialité) et
  - ▶ les « **terminologies** » ensembles de désignations appartenant à une langue de spécialité (ISO 1087:2000) = termes d'un domaine
- ▶ Il faudra donc clarifier l'emploi de : terme, langage, vocabulaire, glossaire, thésaurus, ontologie
- ▶ **Domaine** (ISO 1087:2000) : « branche spécialisée de la connaissance »
- ▶ **Concept** : « unité de connaissance »
- ▶ **Terme** : **désignation** verbale d'un **concept général** dans un **domaine** spécifique
- ▶ Les liens renvoyant sur les **standards officiels** (normes ISO et recommandations W3C), les sources ou les documents déposés dans le cadre des travaux du **GDR SémanDiv** (sites du GDR et du rBDD) sont rassemblés en annexe

# Clarifier l'usage de certains termes : langage, vocabulaire, terminologie

L'utilisation du mot « **vocabulaire** » dans le cadre du Web sémantique prête à confusion dans la mesure où il peut désigner aussi bien le vocabulaire (**termes**) d'un **domaine spécialisé** (ou « *vocabulaire de valeurs* » au sens du W3C), que le « *vocabulaire **RDF*** » servant à décrire les classes et propriétés pour modéliser et structurer les **concepts** d'une terminologie ou d'une ontologie (dans ce dernier cas, on parle plutôt de **langage de représentation** des connaissances ex. RDFS, OWL).

- ▶ **Ressource lexicale** : ensemble d'**unités lexicales** (unités linguistiques) dont on donne une définition en **langage naturel** (ex. *dictionnaire, lexique*)
  - démarche **sémasiologique** (du mot à la signification)
- ▶ **Ressource terminologique** : ensemble de **termes** **ou appellations** (ISO 1087-1:2000 et ISO 704:2009) **désignant** les **concepts (généraux ou individuels)** d'un **domaine** de spécialité (ex. *terminologie, glossaire, nomenclature*) accompagné ou non de définitions
  - démarche **onomasiologique** (de la notion au terme)
- ▶ **Ressource sémantique** : ensemble de **concepts** structurés et reliés par des propriétés ou relations sémantiques (ex. *taxonomie, thésaurus, ontologie, base de connaissances*)

# Ressource sémantique

Ensemble de concepts structurés et reliés par des propriétés ou relations sémantiques, pouvant être enrichi de relations lexico-sémantiques

- ▶ **Relations logiques / ontologiques (entre concepts) :**
  - **hiérarchiques** (logiques)
    - ▶ génériques (taxinomiques)
    - ▶ partitives : composé/composant, ensemble/élément ...
  - **associatives** (ontologiques) : cause/effet , action/résultat , agent/patient ...
  
- ▶ **Relations lexico-sémantiques (entre termes ou unités lexicales) :**
  - relations **genre-espèce** (hyperonymie/hyponymie)
  - relations **tout-partie** (holonymie/méronymie)

**Labellisation** (lexicalisation, dénomination, désignation des concepts par des termes)

- relations d'**équivalence intralinguistique** (synonymies/quasi-synonymies)
- relations d'**équivalence interlinguistique** (traductions)

→ **Ressource termino-ontologique, thésaurus, ontoterminologie, ontologie**

# Vocabulaire contrôlé (vocabulaire d'autorité)

Liste de **termes normalisés** (approche normative) représentant des **concepts** :

- ▶ terme **préférentiel** (privilegié) : unique (dans chaque langue)  
traitement des homonymies et des polysémies (ex. « food chain » / « trophic chain »)
- ▶ termes **équivalents** (alternatifs) **intra**langues : optionnels
  - ▶ **synonymes** : **interchangeables** dans tous les contextes
  - ▶ **quasi-synonymes** : sens **approchant** (utilisables dans certains contextes)
- ▶ termes **équivalents inter**langues (traductions)

→ Liste d'autorité, nomenclature, thésaurus, catalogue de vedettes-matières

**Vocabulaire de référence, référentiel terminologique, standard terminologique**

# Qu'appelle-t-on « thésaurus » ?

- ▶ A l'origine il s'agit d'un langage **documentaire** destiné à indexer et retrouver les documents d'une base bibliographique
- ▶ C'est un vocabulaire **contrôlé**, normalisé (termes **préférentiels** et synonymes) et **structuré** avec des relations **hiérarchiques** et **associatives** (non formalisées) originellement selon normes ISO 2788:1986 (thésaurus monolingues) et ISO 5964:1985 (thésaurus multilingues)
- ▶ Avec les normes **ISO 25964-1:2011** et **25964-2:2013** (*Information et documentation - Thésaurus et interopérabilité avec d'autres vocabulaires*), on passe à un vocabulaire basé sur des **concepts** et optimisé pour couvrir la **terminologie** d'un **domaine**
- ▶ En format **SKOS** (2009), standard du web sémantique (**W3C**), on parle de **schéma de concepts** (skos:ConceptScheme) ou **système d'organisation des connaissances**
- ▶ Le **modèle SKOS** s'exprime lui-même en **langage RDF** (OWL Full)

# SKOS (W3C) versus ISO 25964

7

- Ce sont 2 **modèles de données** aux finalités voisines mais bien distinctes
- Le **standard W3C SKOS** (paru en 2004) s'attache à rendre différents systèmes simples d'organisation des connaissances (dont les thésaurus) compatibles et interconnectés dans le cadre du **web de données liées (LOD)**. Sa dernière révision date de **2009** et précède donc la parution de la norme **ISO**, ainsi que celle de **OWL 2** (dont il n'a pu tenir compte)
- La **norme ISO 25964 (2011-2013)** ne s'intéresse qu'aux **thésaurus** et à leur **interopérabilité** avec d'autres types de vocabulaires. Elle propose notamment une gestion (via des attributs ou rôles) des relations **hiérarchiques** entre concepts, proche des normes ISO 704 et 1087 :
  - **générique** (sorte de)
  - **partitive** (partie de)
  - **instance** (exemple de, individu)
- Le modèle ISO 25964 préconise SKOS parmi les formats d'échange et il existe une correspondance (non standardisée): **ISO 25964 SKOS extension (2015)** ou **iso-thes**  
<https://lov.linkeddata.es/dataset/lov/vocabs/iso-thes> <http://pub.tenforce.com/schemas/iso25964/skos-thes/>

# Définition officielle du thésaurus

- **Norme ISO 25964-1:** *Thésaurus et interopérabilité avec d'autres vocabulaires. Partie 1: Thésaurus pour la recherche documentaire (2011)*
- “Controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms.”
- **Vocabulaire contrôlé** et **structuré** dans lequel les **concepts** sont représentés par des **termes**, organisés de façon à ce que les relations entre les **concepts** soient rendues explicites, et les **termes préférentiels** soient accompagnés d'entrées pour les **synonymes** ou **quasi-synonymes**

# Référentiels sémantiques pour l'interopérabilité des données

- ➔ **Thésaurus** comme système d'**organisation des connaissances** (schémas conceptuels)
  - vocabulaire **contrôlé** : *concept* → 1 **terme** préférentiel + termes (quasi-)synonymes (compréhensible par l'**humain**)
  - relations **hiérarchiques** (TG/TS) et **associatives** (TA) entre concepts (hiérarchies souvent non différenciées donc ambiguës : sorte de/partie de/exemple de)
  
- ➔ **Ontologie** de domaine respectant un **formalisme** et une **logique** de description : concepts/**classes**, sous-classes et **instances** de classes, formalisés et reliés par des **propriétés** permettant d'effectuer des **raisonnements** par **inférence** (adapté à la **machine**)
  - « *spécification explicite formelle d'une conceptualisation partagée* » (Studer)
  
- ➔ **Ontoterminologie** :
  - « *terminologie dont le système notionnel est une ontologie formelle* » (Roche)
  - définitions par **intension/compréhension**  
(concept = un caractère générique + des caractères distinctifs) (Normes ISO 704 & 1087)

# Principes du web de données liées

## (Linked Open Data - LOD)

Les **quatre grands principes** du web de données liées :  
(**Linked Data**, *Tim Berners-Lee* 2006)

- **identifier** les ressources avec des **URI**
- utiliser des URI http (**déréférencables**) de façon à pouvoir utiliser ces URI pour **accéder** à des informations sur les ressources
- lorsqu'on déréférence un URI, renvoyer des données **structurées** respectant des **standards** RDF et SPARQL
- **relier** les URI entre eux pour créer un **réseau** de liens et découvrir de nouvelles informations

+ **Licences** d'utilisation **ouvertes** (lo) / **libres** (CC BY) → **Linked Open Data**

# Données FAIR et Linked Open Data

- Des données peuvent être **ouvertes sans être FAIR** (ou réciproquement) et il peut y avoir plusieurs **degrés** d'ouverture (openness) ou de FAIRness. Ainsi certaines données FAIR ne seront pas obligatoirement ouvertes à tous. <https://sparceurope.org/download/3585/>
- Les **ressources onto-terminologiques** (thésaurus, ontologies) sont vues comme des ensembles organisés de **concepts**, qui sont autant de jeux de **données RDF** accessibles via le web sémantique
- Les 4 **principes FAIR** fondamentaux (2016) présentent certaines caractéristiques communes avec les 4 règles de base du **Linked Data** (2006) : **identifiants** persistants, **protocoles** d'accès **standards**, **formats** et langages **interopérables** avec d'autres ressources
- Les données FAIR minimales, quel que soit leur niveau d'ouverture ou de partage, doivent être correctement décrites avec des **métadonnées** persistantes et standardisées, ainsi qu'une **licence** d'utilisation clairement définie. <https://doi.org/10.3233/ISU-170824>
- Les principes FAIR forment plutôt un guide de **bonnes pratiques**, mais ne constituent ni un standard, ni une spécification

# Données FAIR et Linked Open Data

- Les **ressources** sémantiques peuvent être dans le domaine **public** (licence ouverte, libre et gratuite), mais sont plus généralement diffusées sous **Licence Creative Commons** avec différents degrés de liberté de réutilisation pouvant aller de CC BY (*attribution*) à CC BY-NC-ND (*ni commercialisable, ni modifiable*).
- Strictement parlant, seules les données sous licence **CC 0**, **CC BY** ou **CC BY-SA** (*Share Alike: partageables dans les mêmes conditions*) sont considérées comme véritablement **libres** (les autres étant dites de « libre diffusion »)
- Par ailleurs, elles pourront obtenir de 1 à 5 **étoiles** dans l'échelle de gradation des **Linked Open Data** (LOD) telle que définie par T. Berners-Lee (2006)

# Linked Open Data : 5 Stars

- ▶ \* Données **disponibles** sur le **web**, avec une licence **ouverte**
- ▶ \*\* Données **structurées** et lisibles par l'ordinateur
- ▶ \*\*\* Format **non propriétaire** (CSV, ODS)
- ▶ \*\*\*\* **Identifiants** et **standards** ouverts **W3C** (URI, RDF\*, SPARQL)
- ▶ \*\*\*\*\* Données reliées par **alignements** dans le **LOD Cloud**  
(web de données ouvertes)

# Five Stars of Linked Data **Vocabulary Use**

- **Zero** : Données liées mais sans mention du vocabulaire utilisé
- \* **Information déréférençable** décrivant le vocabulaire utilisé
- \*\* **Axiomatisation** du vocabulaire (langages du **W3C**: RDF\*, OWL...)
- \*\*\* **Liens sortants** vers d'autres vocabulaires (**alignements**, equivalentClass)
- \*\*\*\* **Métadonnées déréférençables** sur le vocabulaire (OMV, VOAF...)
- \*\*\*\* **Liens entrants** à partir d'autres vocabulaires

# Conditions de l'interopérabilité syntaxique

- ▶ **Formats et protocoles** d'échanges **standardisés**, respectant les spécifications du **W3C** (World Wide Web Consortium) :
  - ▶ **Identifiants** pérennes **URI** (**Uniform Resource Identifier** - 2005) pour identifier chaque concept et aligner différents schémas conceptuels
  - ▶ **Modèle RDF** 1.1 (**Resource Description Framework** - 2014) pour représenter et échanger des données rassemblées dans des triplestores (bases de données RDF).  
Graphe de **triplets** (sujet-prédictat-objet) étiqueté et orienté
  - ▶ **SPARQL** 1.1 (**Simple Protocol And RDF Query Language** - 2013) : protocole et langage de **requête** pour les données RDF

# Conditions de l'interopérabilité sémantique

## ❖ Normes ISO :

- sur le travail **terminologique** (ISO **1087**:2000 et ISO **704**:2009)
- sur les **thésaurus** et leur interopérabilité avec différents vocabulaires/terminologies/ontologies (ISO **25964-1 & -2**: 2011 & 2013)

## ❖ Recommandations W3C permettant de structurer les connaissances et/ou de raisonner sur les données (**web sémantique**) : **vocabulaires RDF**

- RDFS 1.1 (RDF Schema, 2014) : pour ontologies légères
- OWL 2 (Web Ontology Language, 2012) : langage formalisé pour ontologies complexes
- SKOS (Simple Knowledge Organization System, 2009) : langage simple pour thésaurus

→ **Vocabulaires contrôlés (normalisés) et structurés en relations sémantiques** : utilisés comme **référentiels** permettant le **partage** et le **croisement** des données/métadonnées

# SKOS : Simple Knowledge Organization System

## Système **simple** d'organisation des connaissances

- **Standard** du web sémantique (**recommandation W3C 2009**) : Format destiné à faciliter le partage et l'**échange** de données
- **Langage RDF** (sérialisations RDF/XML, Turtle, JSON-LD...) donc compatible avec d'autres langages du **LOD / web de données** (RDFS, OWL, SPARQL)
- Respect de **conditions d'intégrité** favorisant l'**interopérabilité**
- Le **modèle** de données **SKOS** utilise une ontologie (OWL 1 Full) centrée sur le **concept**  
URI: <http://www.w3.org/2004/02/skos/core#>
- Ce n'est **pas** un langage formel de **représentation** de connaissances (peu de contraintes, mais des **conventions** d'usage); pas de distinction entre des concepts **généraux** (classes) et **individuels** (instances )
- Utilisable pour les **systèmes d'organisation des connaissances**, les schémas conceptuels de terminologies ou vocabulaires contrôlés tels que : thésaurus, taxonomie, classifications (mais **pas** les ontologies exprimées en OWL 2-DL)
- Formalisation en RDFS/**OWL 1 Full**: <https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html>
- Formalisation en **OWL 1 DL**: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>
- Les classes (**ConceptScheme/Collection/Concept**) sont formellement **disjointes** (owl:disjointWith)
- Certaines **propriétés** sont « déclarées » **disjointes** par la recommandation, mais ne le sont pas formellement (rdfs:comment)

# Notions de base en SKOS (Classes et propriétés)

18

## Schéma de concepts ([skos:ConceptScheme](#)) → Thésaurus

- Identifié par un **URI**
- propriété d'objet : [skos:hasTopConcept](#) → concepts de plus haut **niveau hiérarchique**

## ➤ Concepts ([skos:Concept](#))

- Chaque **instance** de concept est identifiée par un **URI**
- Désignés par des **libellés** lexicaux (langue naturelle) → **termes préférentiels, alternatifs** ou **cachés**
  - propriétés d'annotation : [skos:prefLabel](#), [skos:altLabel](#), [skos:hiddenLabel](#)
- Documentés par différents types de **notes** → **définitions, notes d'application...**
  - propriétés d'annotation : [skos:note](#), [skos:definition](#), [skos:example](#), [skos:scopeNote](#)
- Reliés **sémantiquement** → **hiérarchies** informelles et réseaux d'**associations**
  - propriétés d'objet : [skos:semanticRelation](#), [skos:broader](#), [skos:narrower](#), [skos:related](#)
- Intégrés à un **schéma** de concepts
  - propriétés d'objet : [skos:inScheme](#), [skos:topConceptOf](#)

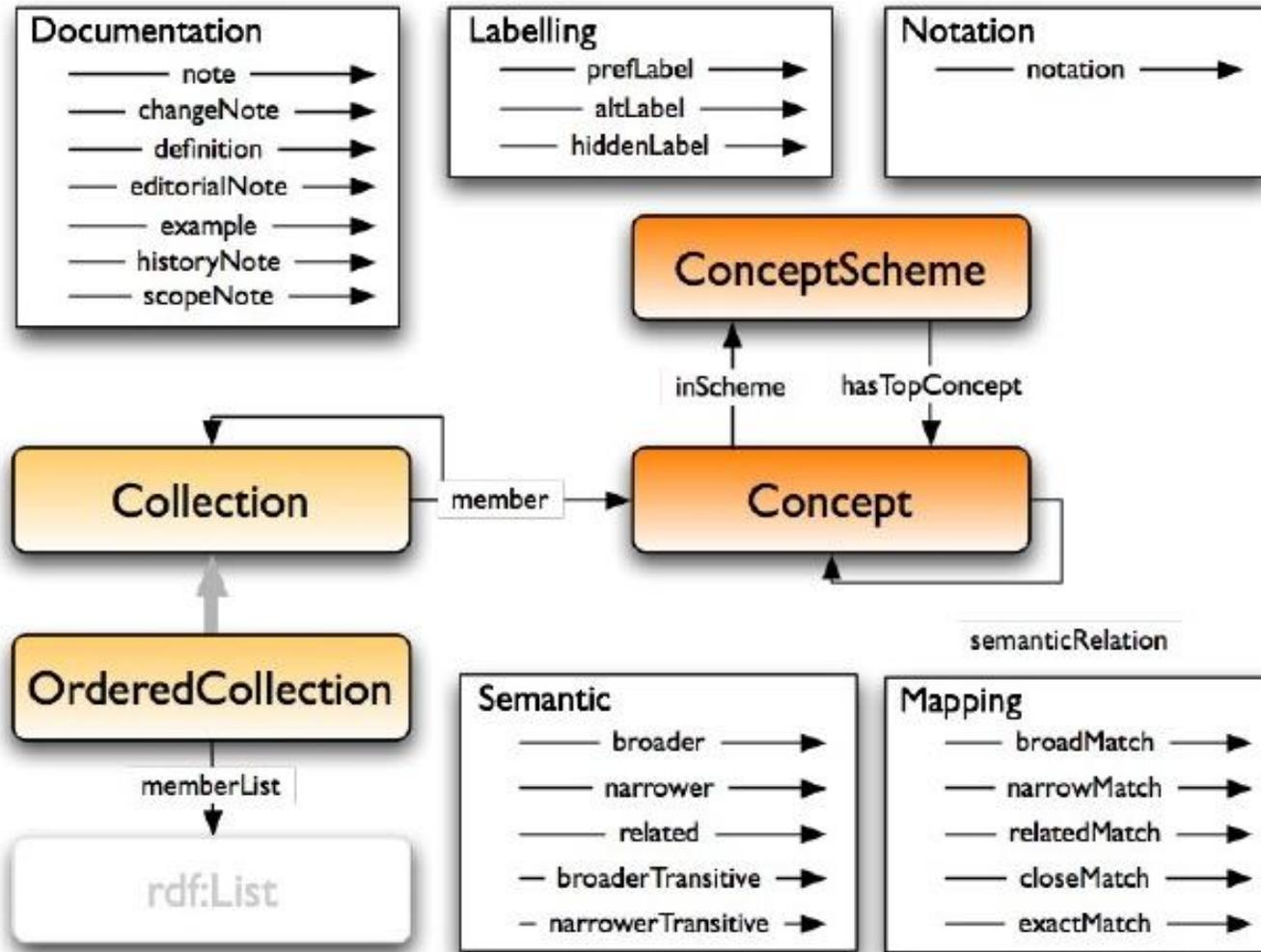
## → **Regroupables** dans des collections ([skos:Collection](#)) → **groupes** ou **facettes**

- Chaque collection est identifiée par un **URI**
- propriété d'objet : [skos:member](#)

## → **Reliables** aux concepts d'autres thésaurus par des alignements

- [skos:mappingRelation](#), [skos:exactMatch](#), [skos:closeMatch](#), [skos:broadMatch](#), [skos:narrowMatch](#), [skos:relatedMatch](#)

# SKOS Model



Source:  
Bechhofer (2010)

# Conditions d'intégrité de la spécification SKOS = Règles (axiomes) à respecter

- ▶ Pour un **concept** donné :
  - ▶ il existe au plus **un terme préférentiel** (libellé skos:prefLabel) pour chaque langue
  - ▶ un même terme ne peut pas être à la fois **préférentiel**, **alternatif** (skos:altLabel) ou **caché** (skos:hiddenLabel)
- ▶ Deux concepts ne peuvent pas être à la fois en relation **hiérarchique** (skos:broaderTransitive) et en relation **associative** (skos:related)
- ▶ Deux concepts de thésaurus différents ne peuvent pas être à la fois en **correspondance exacte** (skos:exactMatch), en correspondance **générique** (skos:broadMatch) ou en correspondance **associée** (skos:relatedMatch)
  - ▶ On dit que ces propriétés sont **disjointes**

# Conventions d'usage (bonnes pratiques) en SKOS facilitant l'interopérabilité

- Présence d'au moins un **terme préférentiel** par concept
- Absence de **doublons** entre concepts différents (termes préférentiels ou synonymes identiques)
- Présence d'un **code langue** pour chaque terme
- **Top-termes** du thésaurus sans terme générique
  
- Certaines **inférences** peuvent être traitées automatiquement :
  - relations **inverses** : broader/narrower; topConceptOf/hasTopConcept
  - relations **symétriques** : related
  - relations **transitives** : broaderTransitive, exactMatch

# SKOS-XL : SKOS eXtension pour les Libellés

- ▶ Partie intégrante de la recommandation **SKOS** (appendice)  
URI : <http://www.w3.org/2008/05/skos-xl#>
- ▶ Permet optionnellement de réifier les **libellés** des termes (valeurs lexicales attachées aux concepts) comme des instances de la **classe skosxl:Label** (les termes sont alors décrits comme des **ressources** (URI) et non plus seulement comme des **propriétés** d'annotation)
- ▶ Possibilité de mise en relation de ces libellés (ex. acronymes)
- ▶ Chaque instance possède 1 unique forme littérale **skosxl:literalForm** pour les propriétés d'objet : **skosxl:prefLabel**, **skosxl:altLabel** et **skosxl:hiddenLabel**

# Correspondances SKOS – norme ISO (étiquettes usuelles)

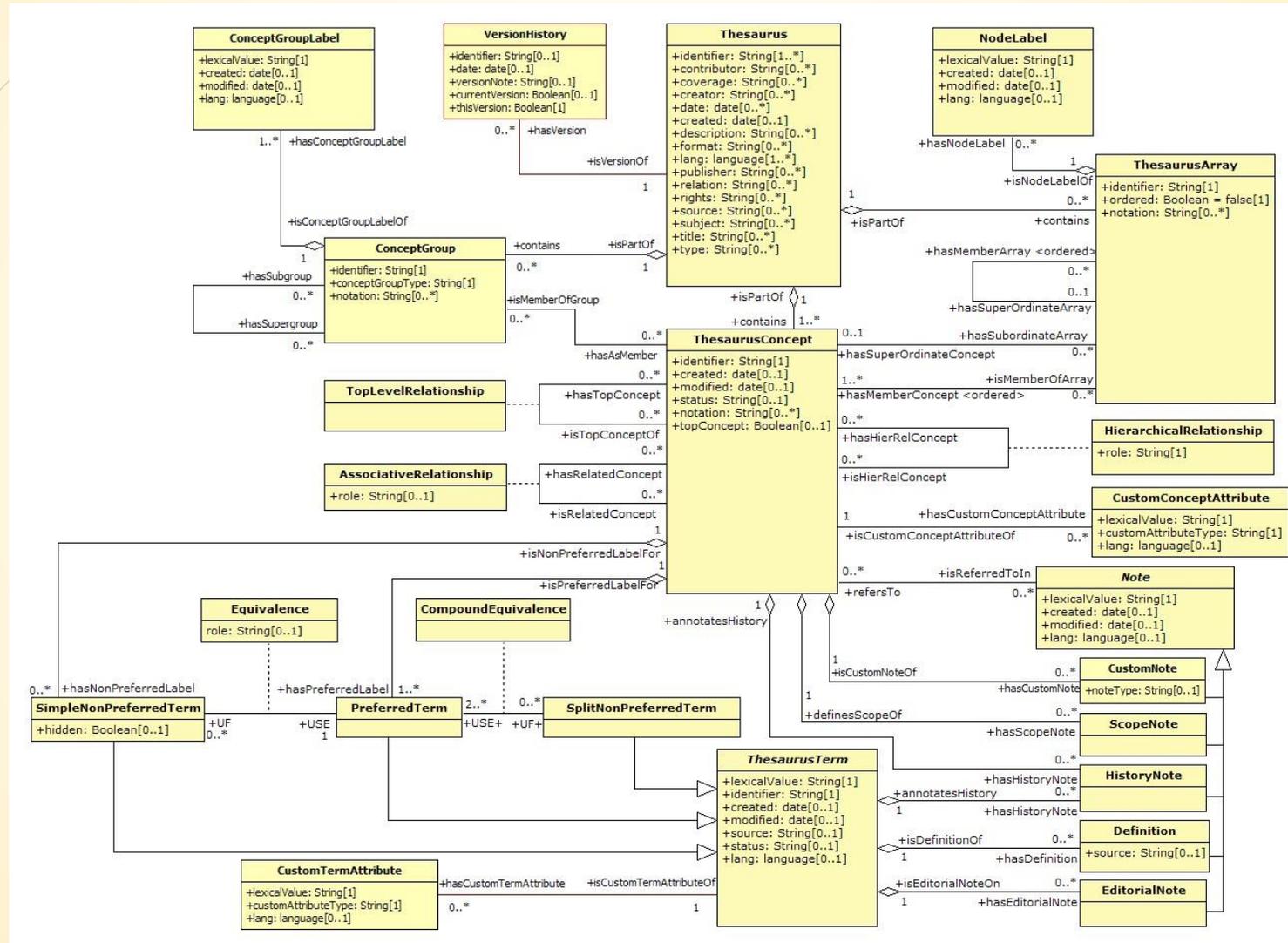
23

skos:**ConceptScheme** → Thésaurus (**Thesaurus**)

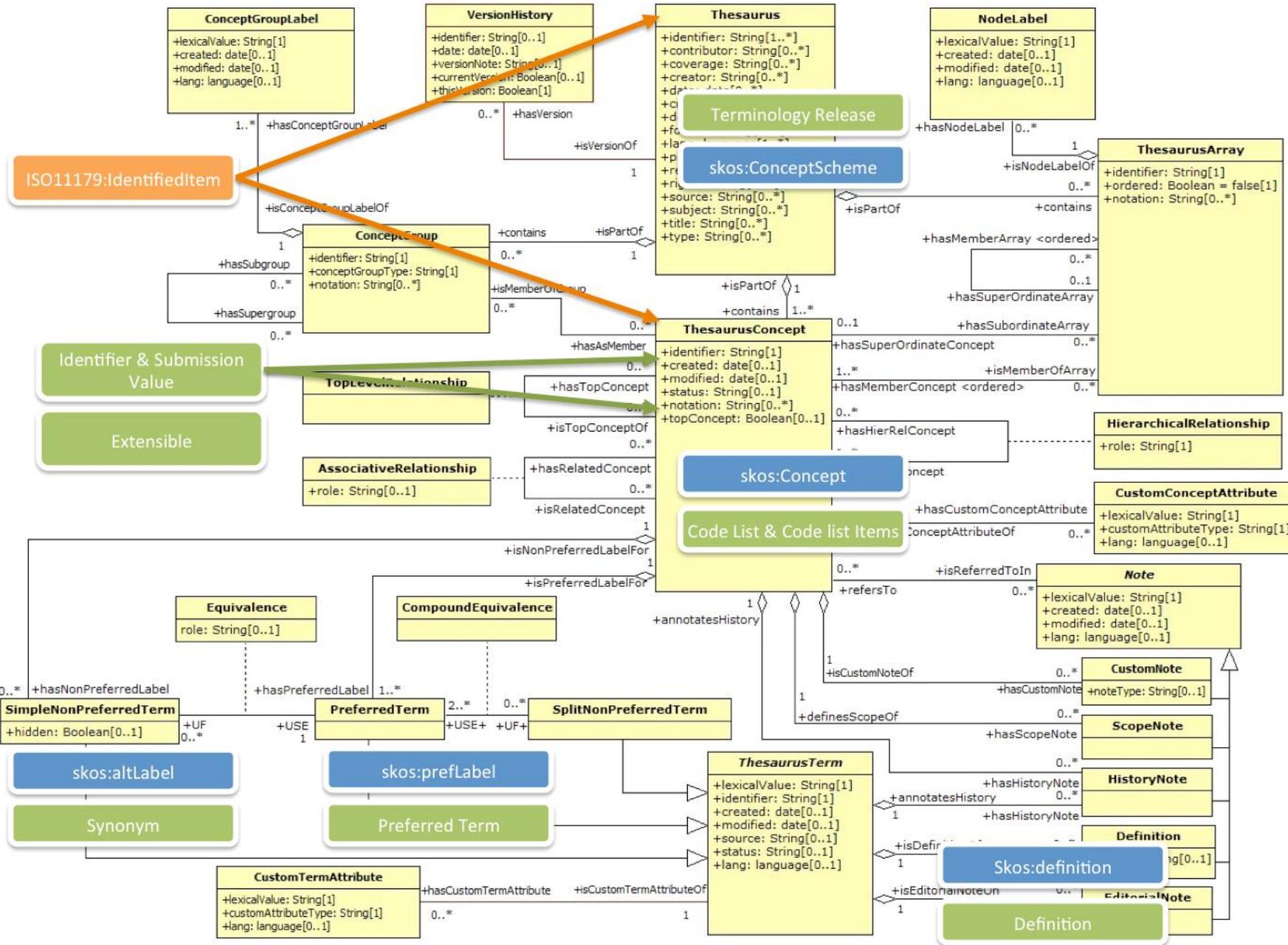
- ▶ skos:**Concept** → Concept (**ThesaurusConcept**)
- ▶ skos:**prefLabel** → Terme préférentiel (**hasPreferredLabel**)
- ▶ skos:**altLabel** → Terme alternatif (**hasNonPreferredLabel**) ; Employé Pour / Used For (**EP / UF**)
- ▶ skos:**hiddenLabel** → Terme caché, variante orthographique ou expression qu'on ne souhaite pas faire apparaître dans la publication mais utilisable pour la recherche
- ▶ skos:**broader** (lien hiérarchique direct) → (**HierarchicalRelationship**) Terme générique / Broader Term (**TG / BT**)
- ▶ skos:**narrower** (lien hiérarchique direct) → Terme spécifique / Narrower Term (**TS / NT**)
- ▶ skos:**broaderTransitive/narrowerTransitive** (liens hiérarchiques indirects) : super-propriétés de skos:broader/narrower
- ▶ skos:**related** → (**AssociativeRelationship**) Terme associé (**hasRelatedConcept**) / Related Term (**TA / RT**)
- ▶ skos:**definition** → Définition (**hasDefinition**) (**DEF**)
- ▶ skos:**Collection** → Regroupement de concepts (**ConceptGroup** ou **ThesaurusArray**) ; Subject Category (**SC**)
- ▶ skos:**member** → Membre d'un groupe ou collection (**hasAsMember** ou **hasMemberConcept**)
- ▶ skos:**inScheme** → Concept ou Collection appartenant (**isPartOf**) à un ConceptScheme
- ▶ skos:**scopeNote** → Note d'application (**hasScopeNote**) / Scope Note (**NA / SN**)
- ▶ skos:**hasTopConcept** → (**TopLevelRelationship**) Top terme (**hasTopConcept**) du thésaurus (**TT**)

# Modèle ISO 25964 (UML)

24



Source



Source

# Structuration d'un thésaurus

- Organisation centrée sur les **concepts**
- Structuration **hiérarchique**: concepts **superordonnés** (broader **TG**) et **subordonnés** (narrower **TS**)
- **Polyhiérarchie** autorisée et enchaînement possible dans la même hiérarchie de relations d'**instance (TGI)** et/ou relations **génériques (TGG)** et/ou relations **partitives (TGP)**  
*à noter : skos:broader, skos:narrower, isoches:broaderGeneric, isoches:broaderPartitive et isoches:broaderInstantial ne sont pas déclarées transitives → skos:broaderTransitive/narrowerTransitive*
- La relation d'**association** (skos:related) est une relation **symétrique** (réciproque) **TA**
- On distingue des thésaurus **généralistes** et des thésaurus **spécialisés**
- Les termes doivent être précis et proches du **langage naturel** (désignation ≠ dénomination)
- Un thésaurus généraliste peut être subdivisé en subthésaurus ou **microthésaurus** (**ConceptScheme** par **domaine** ou **thématique**)
- Au sein d'un ConceptScheme, une structuration en **collections** (optionnelle) peut être proposée en parallèle à la structuration hiérarchique mais sans se confondre avec elle

# Méthodologies de regroupements de concepts

Plusieurs choix possibles :

- création de plusieurs **ConceptScheme** (sous-thésaurus) disjoints au sein du même thésaurus : **microthésaurus** (sous-ensembles dont la réunion donne le thésaurus global) (ex. thésaurus Irstea, Unesco ou EuroVoc)
- création de **top-termes** comme termes de regroupement sans créer de collections ou de facettes, on considère alors que tous ces termes génériques désignent des **top-concepts** (ex. AnaEE Thesaurus) : <http://agroportal.lirmm.fr/ontologies/ANAEETHES?p=classes>
- création de **collections** (et sous-collections) avec skos:Collection, on peut éventuellement **imbriquer** entre elles différentes collections (**nested collections** : une collection pouvant être membre d'une autre collection)

# Gestion alternative des collections

- ▶ Pour gérer les **collections**, on peut faire appel à l'extension **iso-thes** suivant le modèle de données **ISO 25964** <http://purl.org/iso25964/skos-thes> selon 2 méthodes :
- **1)** répartition par **groupes (superGroup / subGroup)** de concepts (**isothes:ConceptGroup** sous-classe de **skos:Collection**) → classement thématique de concepts issus de différents arbres et différents niveaux hiérarchiques (broader/narrower), mais sans aucune indication de hiérarchies entre concepts (l'appartenance à un sous-groupe hérite l'appartenance au super-groupe, mais tous les membres du groupe apparaissent **au même niveau**, y compris les top-concepts du thésaurus)
- **2)** analyse par **facette (isothes:ThesaurusArray** sous-classe de **skos:Collection**) → subdivision des concepts (**superOrdinate**/superordonnés) selon certains critères distinctifs en sous-ensembles de concepts subordonnés (**isothes:subordinateArray**) de **même rang hiérarchique**, mais cela n'interdit pas de recréer une organisation par **sous-facettes** (indépendante de la structure hiérarchique globale du thésaurus ou de l'appartenance d'un concept à un groupe thématique). Cela permet aussi d'introduire une organisation proche d'une ontologie en créant des nœuds ou des concepts fondamentaux de plus haut niveau (organisme, objet, processus, activité, substance, lieu...)

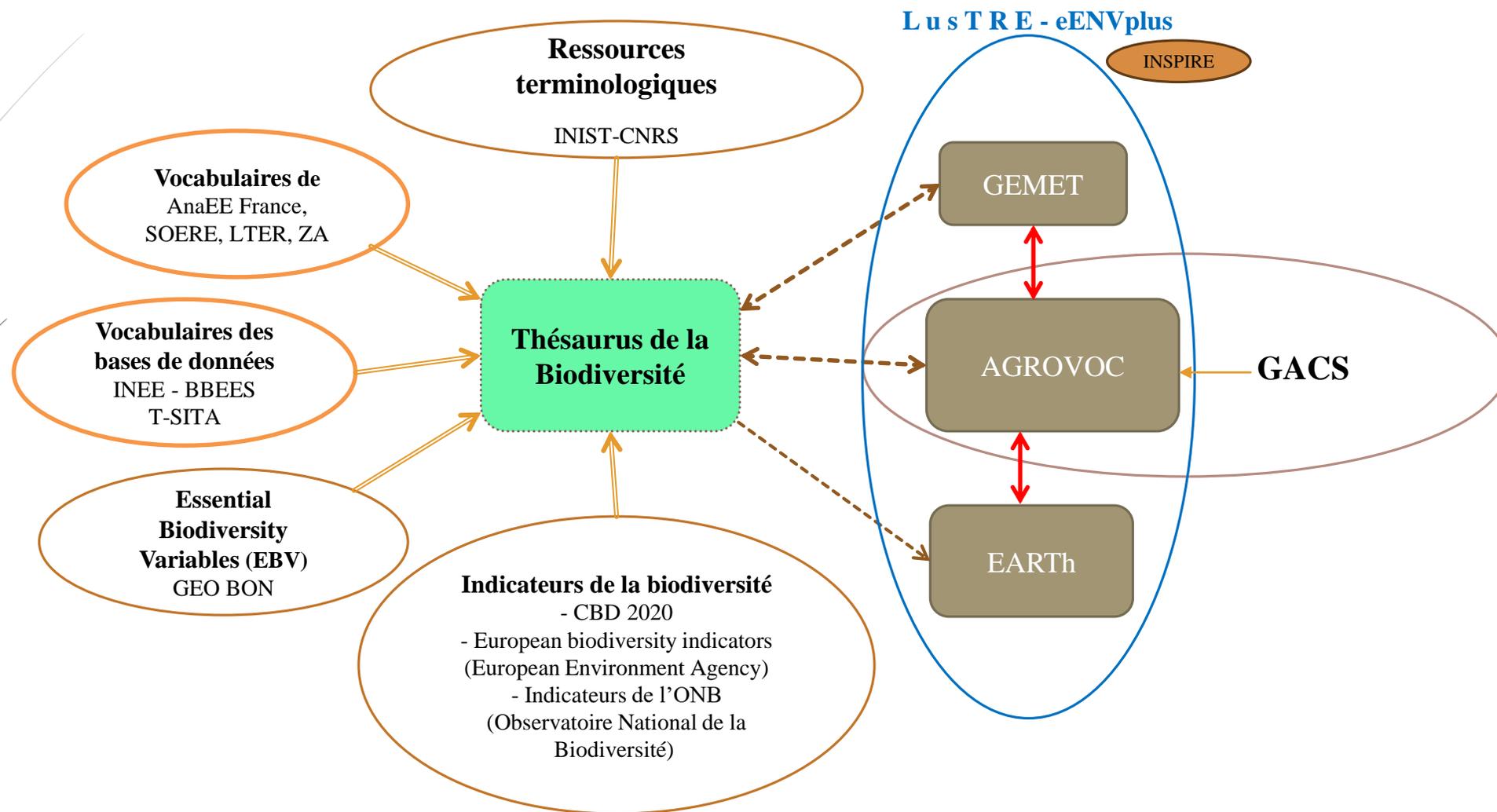
# Méthodologie de constitution du Thésaurus Biodiversité



- ▶ **Combinaison** de plusieurs approches *complémentaires* :
  - **descendante** (< concepts génériques : référentiels Inist + thésaurus généralistes en Environnement)
  - **ascendante** (< concepts + spécifiques : référentiels Inist + vocabulaires spécialisés + bases de données de recherche)
  - **comparative** (alignements entre ≠ thésaurus)
  
- ▶ **Travail terminologique** associant :
  - une démarche **conceptuelle** (du concept vers le terme) à partir des **ressources terminologiques** existantes
  - une démarche **contextuelle** (extraction terminologique) à partir de **corpus textuels** (anglais/français) du domaine de la biodiversité
  
- ▶ **Outils** : Excel, ITM (Mondeca), VocBench, TermSuite, OnAGUI, Skosmos ...

# Constitution du Thésaurus de la Biodiversité

30



# Conditions de l'interopérabilité 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)
- ▶ Unicité et pérennité des **URI** pour les concepts
- ▶ Richesse **sémantique** (relations, définitions)
- ▶ Résolution de la **polysémie** (désambiguïsation);
  - ▶ préférentiels identiques pour des concepts distincts (ex. doublons Gemet)
  - ▶ préférentiel identique au synonyme d'un concept distinct
  - ▶ préférentiels identiques au nombre près (singulier/pluriel)
- ▶ 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)

# Critères de comparaison de thésaurus

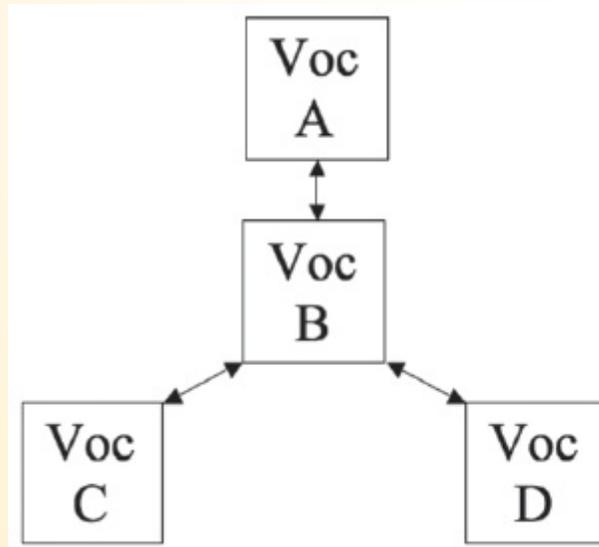
- **Relations sémantiques** : nombre de termes spécifiques ; nombre de termes génériques ; nombre de niveaux hiérarchiques ; nombre de termes associés
- Regroupements en **collections** thématiques
- **Types de hiérarchies** : mono ou poly-hiérarchies ; sorte de / partie de / instance ; cohérence et compatibilité
- **Richesse linguistique** : synonymes, variantes, acronymes, termes cachés pour la **fouille de texte**
- **Equivalences linguistiques** - Multilinguisme
- Présence et qualité des **définitions**
- Utilisation dans des réseaux **européens** ou **internationaux**
- Respect des **standards RDF** / Identifiant pérenne (URI) pour chaque concept
- **URI http déréférencables**
- Interconnexion avec le **web sémantique** (LOD Linked Open Data cloud)
- **Alignements** 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

# Méthodologies d'alignement entre thésaurus

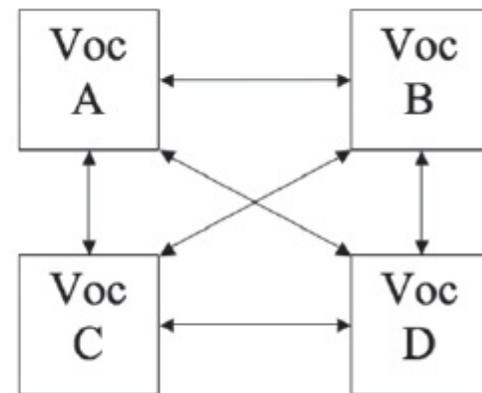
- Des **concepts** de différents thésaurus et/ou ontologies sont mis en **correspondance (équivalence)** en **exploitant** des similitudes de chaînes de caractères entre **préférentiels** et **synonymes** dans plusieurs **langues**, à l'aide d'algorithmes de comparaison
- Les méthodes sont **semi-automatiques** car elles nécessitent une relecture avant validation
- Les **désambiguïisations** sont facilitées par la richesse en **synonymes**, le **multilinguisme**, la **profondeur** des relations hiérarchiques, la présence de **définitions** ou de notes
- **SKOS** permet d'exprimer les équivalences entre concepts de manière **exacte** ou **approchante** (**exactMatch** est une sous-propriété de **closeMatch**)
- Pour les équivalences exactes indirectes (à travers un hub), on pourra exploiter la **transitivité** de la relation **skos:exactMatch**
- On peut aussi établir des correspondances de manière **hiérarchique (broad/narrowMatch)** ou **associative (relatedMatch)**
- Il est possible d'aligner un thésaurus avec une ontologie: `skos:Concept`  $\leftrightarrow$  `owl:Class`

# Mise en correspondance des vocabulaires (Norme ISO 25964-2)

- Correspondances **indirectes** via un vocabulaire pivot (hub)



- Correspondances **directes** 2 à 2



# Outils de gestion de thésaurus

- ▶ Outils d'**édition** (création, import/export)
  - ▶ open-source: ThManager, **TemaTres**, OpenSKOS, Thesauform, **VocBench**, Ginco, **Opentheso**
  - ▶ propriétaires: ITM, PoolParty, TopBraid Composer, iQvoc
- ▶ Outils de **vérification** (contrôle, transformation, validation)
  - ▶ RDF Validator
  - ▶ **Skosify**
  - ▶ **qSKOS**
  - ▶ SKOS Testing Tool
  - ▶ Loterre
- ▶ Outils d'**alignement**
  - ▶ **OnAGUI**
  - ▶ YAM++
  - ▶ **OpenRefine**
- ▶ Outils d'**exposition** (navigation, visualisation, export)
  - ▶ Skosmos (**Loterre**, BARTOC)
  - ▶ NCBO (**Agroportal**, EcoPortal)
  - ▶ **SKOS Play!** (5000 concepts max)

# Outils logiciels pour thésaurus en SKOS

## ► Gestion

- **Opentheso** : <https://github.com/miledrousset/opentheso/releases>
- **VocBench** : <http://vocbench.uniroma2.it/>

## ► Contrôle

- **Skosify** : <http://demo.seco.tkk.fi/skosify/skosify>
- **SKOS Testing Tool (q\$KOS)** : <http://labs.sparna.fr/skos-testing-tool>

## ► Conversion

- **Skos Play!** : <http://labs.sparna.fr/skos-play/?lang=fr>

## ► Alignement

- **OnAGUI** : <https://github.com/lmazuel/onagui>
- **YAM++ Online** : <http://yamplusplus.lirmm.fr/>

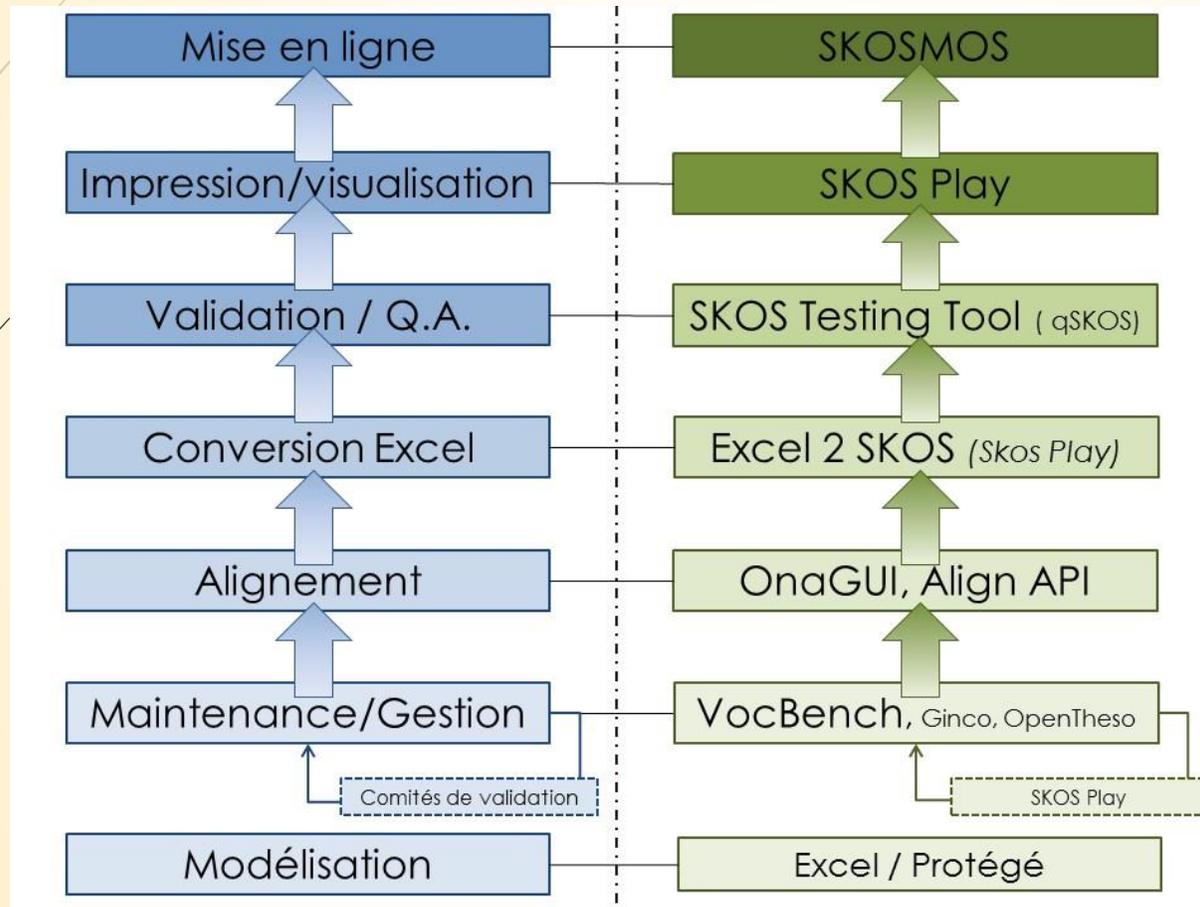
## ► Publication

- **Skosmos** : <http://skosmos.org/>

## ► Requêtage SPARQL

- **YASGUI** : <http://about.yasgui.org/>

# Outils de gestion/publication de thésaurus



Choix d'outils **open-source**

Source: [Sparna](#) (2018)

# Logiciel de gestion VocBench

Plateforme multilingue et collaborative de gestion d'ontologies (OWL), de thésaurus (SKOS/SKOS-XL) et de lexiques (OntoLex-lemon)

## AVANTAGES :

- version **VocBench3 (v5)** : manipulation possible de tous types de données RDF
- entièrement compatible avec les standards **W3C**
- gestion des **métadonnées** (vocabulaires standards)
- possibilité d'importer une ontologie extérieure dans une ontologie existante
- outil **Sheet2RDF**
- formats **import**: SKOS, SKOS-XL, OWL, RDFS / **export**: JSON, CSV, XLS, ODS, RDF\*
- maintenu par l'Office des publications de l'**Union européenne** et l'Univ. Rome
- **communauté** internationale d'utilisateurs : **FAO**, GACS, UNEP, EuroVoc, EEA, AnaEE (INRA)
- groupe d'utilisateurs (in English)
- gestion séparée des **schémas**, des **concepts** et des **collections**
- outil de **vérification d'intégrité**
- outil de validation **d'alignement**
- version standalone ou serveur
- **SPARQL** endpoint
- **licence** open source (BSD-3-Clause)

<http://vocbench.uniroma2.it/>  
<https://vocbench.inist.fr>

## LIMITES :

- *pas de génération des relations symétriques (skos:related)*
- *ne gère pas nativement iso-thes (ISO 25964)*
- *hiddenLabels non recherchables*
- *pas d'export PDF*

Navigation sidebar with search and class list:

Search:

- agricultural chemical product (en), produit agrochimique (fr)
- Anthropocene (en), Anthropocène (fr)
- applied ecology (en), écologie appliquée (fr)
- biocide (en), biocide (fr)
- bioclimatology (en), bioclimatologie (fr)
- biogeography (en), biogéographie (fr)
- biological effect (en), effet biologique (fr)
- biological marker (en), marqueur biologique (fr)
- biopesticide (en), biopesticide (fr)
- cultural heritage (en), patrimoine culturel (fr)
- dynamic climatology (en), climatologie dynamique (fr)
- ecological model (en), modèle écologique (fr)
- ecological parameter (en), paramètre écologique (fr)
  - aerobiosis (en), aérobiose (fr)
  - anaerobiosis (en), anaérobiose (fr)
- anthropogenic factor (en), facteur anthropique (fr)
- biocomplexity (en), biocomplexité (fr)
- biodiversity (en), biodiversité (fr)**
- biological productivity (en), productivité biologique (fr)
- biomass (en), biomasse (fr)
- community structure (en), structure de la communauté (fr)
- density dependence (en), dépendance de la densité (fr)
- ecological connectivity (en), connectivité écologique (fr)
- ecological equilibrium (en), équilibre écologique (fr)
- ecological indicator (en), indicateur écologique (fr)
- ecological integrity (en), intégrité écologique (fr)
- ecological niche (en), niche écologique (fr)

Class details for biodiversity (en), biodiversité (fr):

URI: <http://data.loterre.fr/ark:/67375/BLH-FHNG3BCR-H>

**Types:**

- skos:Concept

**Top Concept of:**

- (empty)

**Schemes:**

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

**Broaders:**

- biological heritage (en), patrimoine biologique (fr)
- ecological parameter (en), paramètre écologique (fr)

**Lexicalizations:**

- skos:prefLabel**
  - biodiversity
  - biodiversité
- skos:altLabel**
  - megabiodiversity
  - ecodiversity
  - megadiversity
  - ecological diversity
  - biotic diversity
  - biological diversity
  - écodiversité
  - diversité biotique
  - mégabiodiversité
  - diversité écologique

Status:

# Logiciel de gestion Opentheso

## Gestionnaire de thésaurus multilingue norme ISO 25964

### AVANTAGES :

- **licence** open source (CeCILL-C)
- gestion **collaborative** (candidats)
- bac à sable disponible en ligne
- développé par un organisme de recherche **français** (GDS CNRS/Univ. Lyon)
- utilisateurs: MOM, Frantiq, MASA, TGIR Huma-Num ...
- **réactivité** de l'équipe de conception/développement
- interface **conviviale** et multilingue
- compatibilité avec **SKOS** et avec **ISO 25964**
- gestion des identifiants **ARK**
- formats import: SKOS, Turtle, JsonLD, CSV / export: SKOS, Turtle, JsonLD, **CSV**, **PDF**
- **bilingue**
- Web service REST
- fonctionnalité d'**alignement** (REST, SPARQL, JSON-LD) avec une liste de sources (thésaurus, coordonnées GPS)

<https://github.com/miledrousset/opentheso/releases>

<https://opentheso4.mom.fr/opentheso4/index.xhtml>

### A RESOUDRE :

- gestion non séparée des **hiérarchies** (top-concepts) et des **groupes/collections**
- absence de serveur SPARQL



Pseudo  
Mot de passe  
Passe oublié

Thésaurus  
theso\_53(53) Ok français Ok



## Thésaurus

- E
- F
- index
- 1 Regroupement par Variable Essentielle de la Biodiv
  - 1.05 Populations d'espèces
  - 1.10 Fonctions des écosystèmes
  - 1.15 Traits des espèces
  - 1.20 Composition des communautés
  - 1.25 Structures des écosystèmes
  - 1.30 Composition génétique
- 2 Regroupement par catégorie sémantique
  - 2.05 Activité / Comportement / Société
  - 2.10 Espace
  - 2.15 Etre vivant
  - 2.20 Idée / Abstraction
  - 2.25 Matériau / Matière / Etat de la matière
  - 2.30 Objet
  - 2.35 Oeuvre / Document
  - 2.40 Phénomène / Processus / Propriété
  - 2.45 Structure du vivant
  - 2.50 Technique / Méthode
- 3 Regroupement par domaine
  - 3.05 Multidisciplinaire
  - 3.10 Sciences biologiques
    - 3.05 Ecologie
      - 3.05 Ecologie trophique
      - 3.10 Ecologie végétale
      - 3.15 Ecologie animale
      - 3.20 Ecologie aquatique
      - 3.25 Ecologie\_Divers
    - 3.10 Génétique
    - 3.15 Sciences agronomiques / Agriculture
    - 3.20 Sciences biologiques\_Divers
  - 3.15 Sciences de la Terre et de l'Univers
- 4 groupDefault
- Orphelins

### Terme(s) générique(s)

paramètre écologique (BT)  
patrimoine biologique (BT)

### biodiversité

id : BLH-FHNG3BCR-H  
ark: 67375/BLH-FHNG3BCR-H  
Domaine : Ecologie\_Divers, Propriété / Paramètre / Caractéristique\_Divers, Composition des communautés

Total de la branche  
Lien permanent

### Terme(s) spécifique(s)

agrobiodiversité (NT)  
diversité écosystémique (NT)  
diversité écotopique (NT)  
diversité fonctionnelle (NT)  
diversité génétique (NT)  
diversité spécifique (NT)

### Terme(s) associé(s)

conservation de la biodiversité (RT)  
équilibre écologique (RT)  
zone critique de biodiversité (RT)  
érosion de la biodiversité (RT)  
biosphère (RT)  
biocomplexité (RT)  
indicateur de biodiversité (RT)

### Terme(s) synonyme(s)

diversité biologique  
diversité biotique  
diversité écologique  
écodiversité  
mégabiodiversité

### Notes

**definition:** Diversité des organismes vivants, qui s'apprécie en considérant la diversité des espèces, celle des gènes au sein de chaque espèce, ainsi que l'organisation et la répartition des écosystèmes. Source: FranceTerme [http://www.culture.fr/franceterme/result?francetermeSearchTerme=biodive%C3%A9&francetermeSearchDomaine=0&francetermeSearchSubmit=rechercher&action=search] (fr)

### Alignement

exactMatch :  
- http://aims.fao.org/aos/agrovoc/c\_33949

### Coordonnées Gps

### Traduction

biodiversity (en)

Facette

### Image

Images externes



## Thesauform Semandiv

in progress

Home | Hierarchy search | Index search | API | Annotation | Vote | Administration

### Browse the Hierarchy

#### Treewiew

- [-] T-Semandiv
  - Liste rouge UICN
  - [-] aire géographique
  - [-] biocide
  - [-] bioclimatologie
  - [-] biogéographie
  - [-] biopesticide
  - [-] chimie environnementale
  - [-] climatologie dynamique
  - [-] comparaison interspécifique
  - [-] comparaison intraspécifique
  - [-] composé sémi chimique
  - [-] droit de l'environnement
  - [-] droit international
  - [-] effet biologique
  - [-] environnement
  - [-] génétique des populations
  - [-] marqueur biologique
  - [-] modèle écologique
  - [-] métagénomique
  - [-] organisation internationale
  - [-] paramètre écologique
  - [-] patrimoine culturel
  - [-] processus naturel
  - [-] processus écologique
  - [-] produit agrochimique
  - [-] relation structure-fonction
  - [-] ressource naturelle
  - [-] seston
  - [-] structure géomorphologique
  - [-] système écologique
  - [-] taxonomie
  - [-] végétation
  - [-] xénobiotique
  - [-] zone géographique
  - [-] écologie
  - [-] écologie appliquée

Search for your term here

# Logiciel d'alignement OnAGUI

*Ontology Alignment Graphical User Interface*

<https://github.com/Imazuel/onagui/wiki>

- ▶ Open source / **Licence libre** MIT
- ▶ Installation en local (Java8)
- ▶ **Bilingue** (français/anglais)
- ▶ Permet d'aligner des vocabulaires **RDF** (SKOS, OWL, RDFS) correctement lexicalisés (présence **synonymes**, **codes langues**) en exploitant tous les types de labels: skos, rdfs, dcterms, schema ...
- ▶ Compatibilité entre **thésaurus / ontologies**
- ▶ Utilise **3 algorithmes** différents (Exact match / I-Sub distance / Levenshtein distance) avec réglage possible du seuil de similarité
- ▶ Formats d'**export** : SKOS, RDF, CSV, **EDOAL**

OnAGUI - Ontology Alignment GUI

Fichier Rechercher Alignement Statistiques Application

Vocabulaire 1: file://vachez/Documents/Documents%20de%20travail/MAJ%20thésaurus%20biodiversité/Biodiversité\_SKOS.xml.rdf  
 Vocabulaire 2: file://vachez/Documents/Documents%20de%20travail/L.OTERRE/Ontologies/ENVO/lenvo\_owlapi\_AgroPorta%202019-03-18.xrdf

Filtrage de la table sur vocab 1
  Non alignable dans vocab 1
  Non alignable dans vocab 2
  Filtrage de la table sur vocab 2

Trouver le concept sélectionné

- SKOS Concept Schemes
  - http://data.loterre.fr/ark:/67375/BLH
    - aire géographique
    - biocde
    - bioclimatologie
    - biogéographie
    - biopesticide
    - chimie environnementale
    - climatologie dynamique
      - changement climatique
      - réchauffement climatique
      - variabilité climatique
    - comparaison interspécifique
    - comparaison intraspécifique
    - composé sémi chimique
    - droit de l'environnement
    - droit international
    - effet biologique
    - environnement
    - génétique des populations
    - Liste rouge UICN
    - marqueur biologique
    - modèle écologique
    - métagénomique
    - organisation internationale
    - paramètre écologique
    - patrimoine culturel
    - processus naturel
    - processus écologique
    - produit agrochimique
    - relation structure-fonction
    - ressource naturelle
    - seston
    - structure géomorphologique
    - système écologique
    - taxonomie
    - végétation
    - xénobiotique
    - zone géographique
    - écologie
    - écologie appliquée

Score: 1.0 Type: exactMatch

Ajouter Supprimer ligne Tout supprimer

Label	Label2	Type	Score	Validity	Method	Date
écossystème		exactMatch	1.0	OK	I-Sub distance	2019-10-24...
désert		exactMatch	1.0	OK	I-Sub distance	2019-10-24...
forêt prim	rest	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
changem	range	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
désertific	ion	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
milieu naturel	natural environment	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
communauté mi	!ERROR!	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
éclotone	ecotone	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
habitat	habitat	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
géodiversité	geodiversity	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
effet de serre	greenhouse effect	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
feu de végétation	bush fire	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
canopée	canopy	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
incendie de forêt	forest fire	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
cycle de fazote	nitrogen cycling	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
aire protégée	protected area	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
biosphère	biosphere	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
mer	!ERROR!	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
zone rurale	!ERROR!	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
corridor écologi	ecological corridor	exactMatch	1.0	OK	I-Sub distance	2019-10-24...
ressource alime	food source	exactMatch	0.9455	OK	I-Sub distance	2019-10-24...
pâtûre	pasturable land	exactMatch	0.928	OK	I-Sub distance	2019-10-24...
milieu terrestre	terrestrial environmental zone	exactMatch	0.928	OK	I-Sub distance	2019-10-24...
milieu terrestre	terrestrial natural environment	exactMatch	0.9176	OK	I-Sub distance	2019-10-24...
forêt	forestry	exactMatch	0.9143	OK	I-Sub distance	2019-10-24...
fragmentation é	habitat fragmentation process	exactMatch	0.9106	OK	I-Sub distance	2019-10-24...
milieu marin	marine environmental zone	exactMatch	0.91	OK	I-Sub distance	2019-10-24...
milieu aquatique	aquatic natural environment	exactMatch	0.9023	OK	I-Sub distance	2019-10-24...

Trouver le concept sélectionné

- Thing
  - entity
    - continuant
      - generically dependent continuant
      - information content entity
      - independent continuant
        - anatomical entity
        - immaterial entity
        - material entity
      - specifically dependent continuant
        - quality
          - physical object quality
          - process quality
          - qualitative
        - quality
          - information carrier
          - length
          - mass
          - morphology
          - organismal quality
          - quality of a single physical entity
          - quality of related physical entities
      - realizable entity
    - occurrent
      - process
        - biological\_process
          - cellular process
          - ecosystem fragmentation process
          - lithification
          - locomotion
          - metabolic process
          - multicellular organismal process
          - thermokarst formation process
        - cellular respiration
        - environmental system process
          - acidification of an aquatic environment
          - algal bloom process
          - algal bloom process phase
          - anthropogenic environmental process
          - atmospheric process
          - biogeochemical process
          - climate change
          - compaction process
          - coral bleaching process
          - deforestation
          - desertification
          - earthquake
          - ecosystem decay
          - ecosystem fragmentation process

Lexique Commentaires Annotations

<p>Lexique</p> <p>Libellés préférentiels: Pas d'informations</p> <p>Synonymes: Pas d'informations</p> <p>URI fragment: Pas d'informations</p>	<p>Lexique</p> <p>Libellés préférentiels: Pas d'informations</p> <p>Synonymes: Pas d'informations</p> <p>URI fragment: Pas d'informations</p>
---	---

# Services d'exposition et de publication de thésaurus

45

## ➤ Loterre (basé sur la technologie Skosmos)

- permet de naviguer en exploitant toutes les possibilités du **SKOS** (affichage des **collections**)
- interface trilingue; **interrogation multilingue** ou **choix de la langue** d'interrogation
- formes cachées (**hiddenLabel**) **interrogeables**
- service de téléchargement (SKOS, **PDF**, CSV)
- services de contrôle, de transformation et d'alignement (en développement)

## ➤ AgroPortal (basé sur la technologie NCBO)

- originellement destiné à l'exposition d'**ontologies**
- autorise l'importation du **SKOS** (mais *n'affiche pas les labels en SKOS-XL*)
- *n'affiche pas les collections*
- **skos:broader** est transformé en **rdfs:subClassOf** (= relation is\_A) → *pb pour les relations partitives*
- interface exclusivement en **anglais** → **interrogation monolingue**
- *hiddenLabels non interrogeables (anglais compris)*
- **alignements** automatiques (LOOM) entre ontologies (**en anglais**)
- service d'**annotation** (*en anglais*) et de **recommandation**
- **téléchargement** possible (SKOS, RDF/XML, CSV)

# LOTERRE

## (Linked Open TERminology REsources)

- Plateforme d'exposition de terminologies (format **SKOS**) lancée par l'**Inist-Cnrs** en mars 2018 : <https://www.loterre.fr/>
- Utilise la technologie **Skosmos** (Finto) pour naviguer dans les ressources
- Affichage et navigation **multilingue, multihierarchique,**
- Liste **alphabétique** / liste **hierarchique** / regroupements par **collections**
- Ressources exportables en **SKOS** (RDF/XML), **CSV** ou **PDF** et réutilisables sous **licence ouverte** (CC BY, Etalab ...)
- Ressources internes et externes à l'Inist
- Version 2 (nouveaux services fin 2019): **contrôle** du SKOS, **transformations,** **alignement, annotation**

Liste	Hiérarchie	Groupes
A B C D E É F G H I J K L M N		
O P Q R S T U V X Z		
baie <i>baie(côte)</i> → baie Bassin méditerranéen benthos berge <i>bioaccumulation</i> → accumulation biologique <i>bioagresseur</i> → organisme nuisible bioamplification biocide bioclimat bioclimatologie <i>biocoenose</i> → biocénose biocomplexité <i>biocontrôle</i> → lutte biologique <i>biocénologie</i> → synécologie biocénose <i>biocénotique</i> → synécologie <b>biodiversité</b> <i>biodiversité agricole</i> → agrobiodiversité <i>biodiversité fonctionnelle</i> → diversité fonctionnelle <i>biodiversité génétique</i> → diversité génétique <i>biodiversité spécifique</i> → diversité spécifique <i>biodiversité végétale</i> → diversité botanique <i>biodiversité écosystémique</i> → diversité écosystémique <i>biodécontamination</i> → bioremédiation biofilm biogéographie <i>bioindicateur</i> → indicateur biologique <i>biologie de la conservation</i> → conservation de la biodiversité <i>biomagnification</i> → bioamplification biomanipulation <i>biomarqueur</i> → marqueur biologique biomasse biome <i>biome anthropogénique</i> → anthrome <i>bionomie</i> → écologie biopesticide bioremédiation <i>bioregion</i> → écozone		

paramètre écologique &gt; biodiversité

ressource naturelle &gt; patrimoine naturel &gt; patrimoine biologique &gt; biodiversité

ressource naturelle &gt; ressource biologique naturelle &gt; patrimoine biologique &gt; biodiversité

TERME PRÉFÉRENTIEL

## biodiversité

DÉFINITION(S)

Diversité des organismes vivants, qui s'apprécie en considérant la diversité des espèces, celle des gènes au sein de chaque espèce, ainsi que l'organisation et la répartition des écosystèmes. Source: FranceTerme [<http://www.culture.fr/franceterme/result?francetermeSearchTerme=biodiversit%C3%A9&francetermeSearchDomaine=0&francetermeSearchSubmit=rechercher&action=search>]

CONCEPT(S) GÉNÉRIQUE(S)

paramètre écologique  
patrimoine biologique

CONCEPT(S) SPÉCIFIQUE(S)

agrobiodiversité  
diversité fonctionnelle  
diversité génétique  
diversité spécifique  
diversité écosystémique  
diversité écotopique

CONCEPT(S) ASSOCIÉ(S)

biocomplexité  
biosphère  
conservation de la biodiversité  
indicateur de biodiversité  
zone critique de biodiversité  
équilibre écologique  
érosion de la biodiversité

SYNONYME(S)

*diversité biologique*  
*diversité biotique*

# Biodiversity Thesaurus

Summary Classes Properties Notes Mappings Widgets Edit ontology information Add submission

Jump To:

- ⊕ agricultural chemical product
- ⊕ applied ecology
- ⊕ biocide
- ⊕ bioclimatology
- ⊕ biogeography
- ⊕ biological effect
- ⊕ biological marker
- ⊕ biopesticide
- ⊕ cultural heritage
- ⊕ dynamic climatology
- ⊕ ecological model
- ⊕ **ecological parameter**
- ⊕ ecological process
- ⊕ ecological system
- ⊕ ecology
- ⊕ environment
- ⊕ environmental chemistry
- ⊕ environmental law
- ⊕ geographical range
- ⊕ geographical zone
- ⊕ geomorphological structure
- ⊕ international law
- ⊕ international organization
- ⊕ interspecific comparison
- ⊕ intraspecific comparison
- ⊕ IUCN Red List
- ⊕ metagenomics
- ⊕ natural process
- ⊕ natural resource
- ⊕ population genetics
- ⊕ semiochemicals
- ⊕ seston
- ⊕ structure-function relationship

Details	Visualization	Notes ( 0 )	Class Mappings ( 1 )	Access Class JSON
Preferred Name	ecological parameter			
Synonyms	environmental parameter			
Definitions	Variable, measurable property whose value is a determinant of the characteristics of an ecosystem. Source: Service de terminologie environnementale (European Environment Agency) [ <a href="http://glossary.fr.eea.europa.eu/terminology/concept_html?term=param%C3%A8re%20%C3%A9cologique">http://glossary.fr.eea.europa.eu/terminology/concept_html?term=param%C3%A8re%20%C3%A9cologique</a> ]			
ID	<a href="http://data.loterre.fr/ark:/67375/BLH-ZKKF563R-D">http://data.loterre.fr/ark:/67375/BLH-ZKKF563R-D</a>			
altLabel	paramètre environnemental environmental parameter			
definition	Variable, measurable property whose value is a determinant of the characteristics of an ecosystem. Source: Service de terminologie environnementale (European Environment Agency) [ <a href="http://glossary.fr.eea.europa.eu/terminology/concept_html?term=param%C3%A8re%20%C3%A9cologique">http://glossary.fr.eea.europa.eu/terminology/concept_html?term=param%C3%A8re%20%C3%A9cologique</a> ]			
domain	Ecology Ecologie			
hiddenLabel	paramètres environnementaux			
identifiant	<a href="http://www.termssciences.fr/vocabs/Biodiversite/380">http://www.termssciences.fr/vocabs/Biodiversite/380</a>			
inScheme	<a href="http://data.loterre.fr/ark:/67375/BLH">http://data.loterre.fr/ark:/67375/BLH</a>			
	<a href="#">plant production</a> <a href="#">aerobiosis</a> <a href="#">geodiversity</a> <a href="#">anthropogenic factor</a> <a href="#">anaerobiosis</a> <a href="#">ecological connectivity</a> <a href="#">home range</a> <a href="#">functional trait</a>			

Powered by NCBO BioPortal

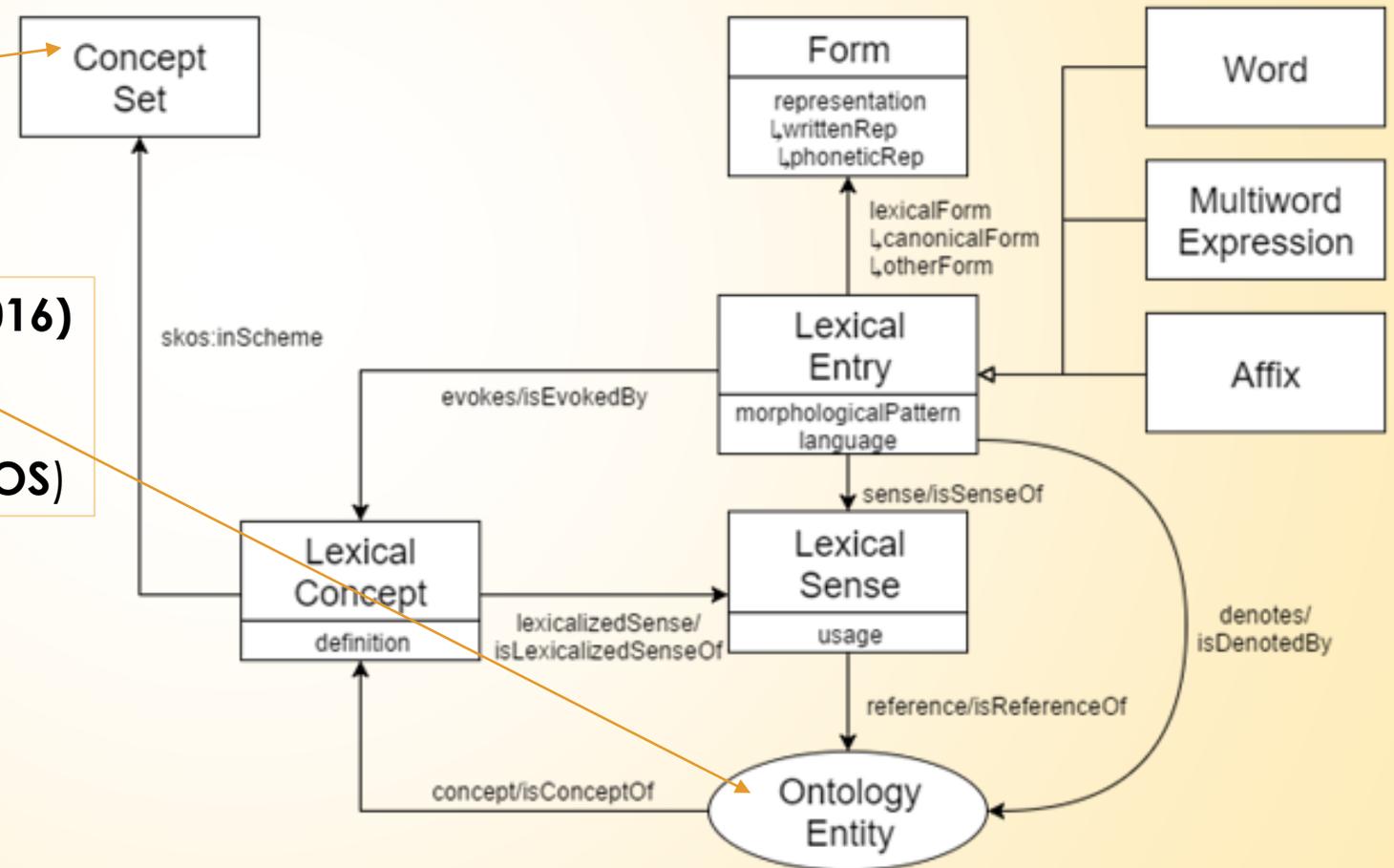
# Du thésaurus vers l'ontologie

- Une **ontoterminologie** est une « terminologie dont le système conceptuel est une ontologie formelle » (Roche, 2007)
- Une **ontologie** est une « spécification formelle et explicite d'une conceptualisation partagée » (Studer, 1998)
- Moyennant quelques transformations, un **thésaurus** bien construit pourra être considéré comme une ontoterminologie ou une « **ontologie légère** »
- **Définitions terminologiques** (par **intension** ou **compréhension**): pour chaque **concept**, on décrit un caractère **générique** et les caractères **distinctifs** (norme ISO 704)  
(Rappel: dans la norme **ISO 25964**, la définition est optionnelle et s'applique au terme)
- Il est également possible de **combiner** des assertions en **SKOS** et en **OWL**  
<MonConcept> est une instance de skos:Concept **et** une instance de owl:Class
- Possibilité d'exploiter les hiérarchies de classes/propriétés pour effectuer des inférences logiques

# Lexicalisation des ontologies

- On peut gérer de manière indépendante les **entités lexicales** et les **entités conceptuelles** d'un **thésaurus** ou d'une **ontologie** de domaine

→ Le **modèle lemon/OntoLex (W3C, 2016)** permet de mieux représenter les informations lexicales relatives à une ontologie (**OWL**) ou à un thésaurus (**SKOS**)



# Intérêt pour le GDR SémanDiv

- Utilisation d'un **thésaurus** des concepts-clefs de la **biodiversité** permettant de s'**aligner** avec d'autres ressources sémantiques interdisciplinaires sur les thématiques **environnementales** (Agrovoc, Gemet, EnvThes...) et de faciliter ainsi l'accès aux données par les mots-clés des métadonnées dans différents **portails** (OSU, PNDB, InDoRES...). Cf. Guides de saisie des métadonnées (BBEES, Zones Ateliers)
  
- **Objectifs Axe 2 du GDR SémanDiv :**
  - 1) élaborer un **thésaurus** bilingue français/anglais (définition, synonymes, concepts reliés sémantiquement) pour les concepts clefs les plus généraux relevant de la diversité biologique ;
  - 2) proposer une **ontologie simple** correspondant à ces concepts généraux, en utilisant des **relations formelles** entre concepts telles que « *est un* », « *fait partie de* », ou encore « *a pour caractéristique* » ;
  - 3) élaborer un ou des thésaurus qu'on appellera « **spécialisés** », correspondant à des domaines disciplinaires plus restreints et directement reliés aux bases de données (traits des plantes, interactions entre espèces), pour lesquels les compétences sont présentes au sein des membres du GDR ;
  - 4) évaluer les possibilités de combiner les informations issues de ces **thésaurus** spécialisés avec celles issues de l'**ontologie** plus générale de la biodiversité : confrontation d'approches **descendante** (à partir de concepts génériques) et **ascendante** (à partir des concepts issus de la terminologie de champs disciplinaires plus focalisés).

# Références et liens

52

- ▶ **Le tournant ontologique de la terminologie (J. ROCHE, 2018)** <https://tel.archives-ouvertes.fr/tel-02136227/>
- ▶ **SKOS : Système Simple d'Organisation de Connaissances – Référence** <http://www.sparna.fr/skos/SKOS-traduction-francais.html>
- ▶ **Agrisemantics Map of Data Standards** <https://vest.agrisemantics.org/about/structure>
- ▶ **Best Practices for Publishing Linked Data (W3C, 2014)** <https://www.w3.org/TR/ld-bp/>
- ▶ **Five Stars of Linked Data Vocabulary Use (2014)** <http://www.semantic-web-journal.net/system/files/swj653.pdf>
- ▶ **ISO 25964 – the international standard for thesauri and interoperability with other vocabularies**  
<https://www.iso.org/schemas/iso25964>
- ▶ **Key choices in the design of Simple Knowledge Organization System (SKOS) (2013)** <https://arxiv.org/pdf/1302.1224>
- ▶ **Correspondence between ISO and SKOS/SKOS XL Models**  
<http://storage.hinterland.nu/webdav/Documents/Information%20Architecture/Correspondence%20ISO25964-SKOSXL-MADS-2012-10-21.pdf>
- ▶ **Exchanging ISO 25964-1 thesauri data using RDF, SKOS and SKOS-XL**  
[https://nkos-eu.github.io/2012/presentations/ISO25964-mapping-to-SKOS-XL\\_TPD-2012-09-27v6.pdf](https://nkos-eu.github.io/2012/presentations/ISO25964-mapping-to-SKOS-XL_TPD-2012-09-27v6.pdf)
- ▶ **The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud**  
[http://svn.aksw.org/papers/2016/LREC\\_OWLG\\_LLOD\\_Cloud/public.pdf](http://svn.aksw.org/papers/2016/LREC_OWLG_LLOD_Cloud/public.pdf)
- ▶ **SKOS, Past, Present and Future (Bechhofer, 2010)** <https://www.slideshare.net/seanb/skos-past-present-and-future>
- ▶ **Terminologie des langages documentaires et du web sémantique** (Inist-Cnrs 2017)

# Normes ISO

- ▶ **ISO 1087-1:2000** : *Travaux terminologiques — Vocabulaire — Théorie et application*
- ▶ **ISO 704:2009** : *Travail terminologique – Principes et méthodes*
- ▶ **ISO 25964-1:2011** (*Information et documentation - Thésaurus et interopérabilité avec d'autres vocabulaires*) : **Thésaurus pour la recherche documentaire**
- ▶ **ISO 25964-2:2013** (*Information et documentation - Thésaurus et interopérabilité avec d'autres vocabulaires*) : **Interopérabilité avec d'autres vocabulaires**

# Recommandations et standards du W3C

- **URI (2005)** : <https://www.w3.org/TR/cooluris/>
- **SKOS (2009)** : <https://www.w3.org/TR/skos-reference/>
- **OWL 2 (2012)** : <https://www.w3.org/TR/owl2-overview/>
- **SPARQL 1.1 (2013)** : <https://www.w3.org/TR/sparql11-query/>
- **RDF 1.1 (2014)** : <https://www.w3.org/TR/rdf-syntax-grammar/>
- **RDF Schema 1.1 (2014)** : <https://www.w3.org/TR/rdf-schema/>
- **Lexicon Model for Ontologies (lemon 2016)** :  
<https://www.w3.org/2016/05/ontolex/>