

# Projet Sémantique de la bioDiversité GDR SémanDiv



## Constat de départ : Hétérogénéité sémantique

Rencontre Inist (cellule Terminologie) / CEFE / CESAB (janv. 2014)

**CEFE**: Centre d'Ecologie Fonctionnelle et Evolutive (CNRS)

**CESAB**: CEntre de Synthèse et d'Analyse sur la Biodiversité (FRB)

- Les sciences de la biodiversité et de l'écologie se caractérisent par leur interdisciplinarité, la multitude des bases de données d'observation ou d'expérimentation, l'hétérogénéité syntaxique et sémantique des données
  - → nécessité d'harmoniser les concepts et leurs définitions pour rendre ces données interopérables.

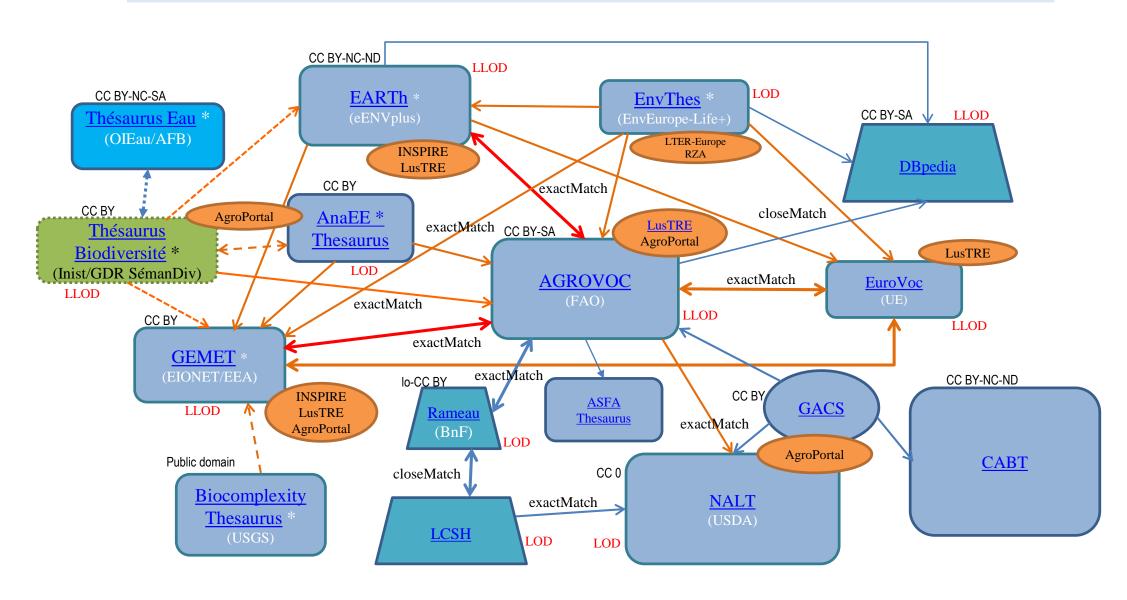
# Rampe de lancement (2014)

Création d'un groupe de travail (juin 2014) réunissant l'Inist, des chercheurs écologues et des informaticiens dans le cadre du réseau Bases de Données rBDD (groupe Interopérabilité): CNRS, INRA, FRB, AnaEE...

« Vers un thésaurus de la biodiversité »

- Réalisation à l'Inist d'un inventaire des thésaurus existants en Sciences de l'Environnement avec leurs complémentarités et leurs alignements
  - → absence de thésaurus dédié à la Biodiversité disponible dans le web sémantique
- Présentation de la méthodologie de constitution de ce thésaurus au Carrefour de l'IST (nov. 2014)
- Veille sur ces ressources terminologiques : liste évolutive et schéma relationnel au sein du Linked Open Data (LOD Cloud)

## Concept « biodiversité » dans les thésaurus sur l'Environnement et l'Agronomie au sein du web sémantique



# Premier étage de la fusée (2015-2016)

- Elaboration à l'Inist d'un thésaurus de 615 concepts de la composante écologique de la biodiversité en approche top-down à partir de nos propres vocabulaires d'écologie/agroécologie (référentiels terminologiques) et en optimisant la structuration des thésaurus généralistes préexistants
  - -Bilingue, polyhiérarchique, riche en relations associatives, en synonymes (850) et variantes lexicales (4500)
  - -Concepts regroupés en 3 collections (catégories, domaines, EBVs)
  - -Interopérable: conforme aux normes ISO 25964 et aux standards du W3C (SKOS/RDF)
- Présentation à plusieurs **séminaires** (<u>IndexMed</u>) et **ateliers** hébergés à la Fondation pour la Recherche sur la Biodiversité (FRB)
- Collaboration à la constitution d'un dossier de demande de création d'un GDR (Groupement de recherche) sur la Sémantique de la biodiversité (début 2016) initié par Eric Garnier (directeur de recherche au CEFE/Montpellier)

## Deuxième étage de la fusée (janv. 2017)

Démarrage du GDR SémanDiv (40 personnes, 17 structures ESR) pour 4 ans, visant à promouvoir une discipline émergente : l'éco-informatique

Organisé autour de 4 axes de travail dont 2 intéressent principalement l'Inist :

- Axe 1 : Délimitation du champ scientifique et concepts retenus
- Le Thésaurus de la biodiversité doit constituer le socle des concepts-clés à partir duquel sera constituée une ontologie de domaine, en relation avec des thésaurus plus spécialisés proposés par des chercheurs du GDR
- Axe 3 : Recensement des standards et référentiels terminologiques disponibles

Veille sur les ressources sémantiques et terminologiques (thésaurus, ontologies, glossaires, dictionnaires...) du domaine librement accessibles dans le web de données (LOD)

Axe 4 : Visualisation, interrogation, mise en correspondance avec les bases de données

# Troisième étage de la fusée (2018)

Simultanément au lancement du GDR, se mettent en place au niveau français des évolutions notables dans la réglementation, l'organisation de la recherche et la création de nouvelles structures autour de la biodiversité :

Loi pour la reconquête de la biodiversité, de la nature et des paysages (2016) Agence française pour la biodiversité (2017)

- Présentation du Thésaurus Biodiversité à l'Agence française pour la biodiversité / Office international de l'eau (février 2018) à leur demande (portail documentaire)
  - → pistes de collaboration pour élargir leur **Thésaurus Eau** à la biodiversité
- Exposition du thésaurus (licence libre CC BY) sur la plateforme Loterre
- Dépôt dans la plateforme Ortolang
- Alignements avec d'autres thésaurus généralistes sur l'Environnement
- Préparation d'un poster sur l'interopérabilité sémantique pour le Congrès LIBER (juillet 2018) « Semantics of Biodiversity : from Thesaurus to Linked Open Data (LOD) »



Biodiversité

Langue des données français - X Chercher

Accueil / Explorer / Naviguer / Vocabulaires / Biodiversité

Aide

in English

Liste					Hiérarchie					Groupes					
Α	В	С	D	Ε	É	F	G	Н	L	J	K	L	М	N	
0	Р	0	R	5	Т	U	V	Χ	7						

*abioseston* → tripton abondance relative des espèces → abondance écologique abondance écologique abri acaricide acarophage accroissement de population accumulation biologique aciculisylve → taïga acidification action des vagues activité animale activité anthropique → activité humaine activité humaine activité microbienne activité pollinisatrice → pollinisation adaptation au milieu adaptation ad infleti adaptation écologique → adaptation au milieu agglomération urbaine → zone urbaine agriculture biologique agriculture climato-intelligente agriculture conventionnelle agriculture conventionnelle
agriculture durable
agriculture intelligente face au climat → agriculture
climato-intelligente
agriculture organique → agriculture biologique
agriculture écologique → agroécologie
agrobiodiversité *agrosystème* → agroécosystème agroécologie agroécosystème

#### Description du vocabulaire

TITRE

Thésaurus Biodiversité

	Biodiversité							
SUJET	Sciences de l'environnement							
DESCRIPTION	Ce thésaurus bilingue structure les concepts-clefs de la biodiversité dans ses composantes écologiques fondamentales et appliquées. Il utilise la polyhiérarchie et comprend 400 relations associatives. Chaque descripteur présente un grand nombre de synonymes et de variantes interrogeables en français et en anglais. Les concepts sont regroupés par catégories sémantiques, par domaines et par classes d'EBV. Des définitions sont données avec leur source.							
CRÉATEUR	Institut de l'information scientifique et technique (Inist) - CNRS/UPS76							
LANGUE	http://lexvo.org/id/iso639-3/eng http://lexvo.org/id/iso639-3/fra							
VERSION	1.1							
NOM D'ATTRIBUTION	Institut de l'information scientifique et technique (Inist) - CNRS/UPS76							
CC:ATTRIBUTIONURL	http://www.inist.fr							
LICENCE	http://creativecommons.org/licenses/by/4.0/							
DC:ALTERNATIVE	Thésaurus de la biodiversité							
DATE DE CRÉATION	2015							
IDENTIFIANT	http://dx.doi.org/10.13143/lotr.4647							



environment quality environmental factor functional trait

Biodiversity

Langue des données anglais \* X Chercher

Accueil / Explorer / Naviguer / Vocabulaires / Biodiversity

Aide | in English

Hiérarchie Liste Groupes ecological parameter > biodiversity > species diversity natural resource > natural heritage > biological heritage > biodiversity > species diversity agricultural chemical product natural resource > natural biological resource > biological heritage > biodiversity > species diversity applied ecology -biocide species diversity TERME PRÉFÉRENTIEL -bioclimatology biogeography biological effect CONCEPT(S) GÉNÉRIQUE(S) biodiversity -biological marker CONCEPT(S) SPÉCIFIQUE(S) -biopesticide botanical diversity cultural heritage CONCEPT(S) ASSOCIÉ(S) community structure dynamic climatology ecological abundance ecological model ecological parameter species richness -aerobiosis SYNONYME(S) taxonomic diversity -anaerobiosis -anthropogenic factor APPARTIENT AU GROUPE Community composition -biocomplexity Ecology\_Miscellaneous biodiversity Property / Parameter / Characteristic\_Miscellaneous -agrobiodiversity ecosystem diversity **TRADUCTIONS** diversité spécifique français ecotopic diversity diversité taxinomique functional diversity diversité interspécifique genetic diversity species diversity biodiversité spécifique botanical diversity diversité taxonomique biological productivity biomass URI http://data.loterre.fr/ark:/67375/BLH-X0H2ZWNQ-G community structure RDF/XML TURTLE JSON-LD Télécharger ce concept: density dependence ecological connectivity ecological equilibrium ecological indicator ecological integrity CONTACT | MENTIONS LÉGALES | COOKIES ecological niche

### Dans le cadre du GDR

- Mise à disposition du thésaurus dans un outil collaboratif (<u>Thesauform</u>) pour la relecture des définitions par les experts du domaine
- Tests sur d'autres outils collaboratifs de gestion open-source (Opentheso, VocBench)
- Intégration du thésaurus dans la plateforme d'ontologie <u>AgroPortal</u> (LIRMM Montpellier) Group SEMANDIV
- Rédaction d'un guide « Terminologie des langages documentaires et du web sémantique » déposé sur le site CoRe du GDR
- **Enrichissement** du thésaurus par :
  - -extraction terminologique (littérature et rapports scientifiques): démarche contextuelle -transformation de synonymes/variantes en nouveaux concepts: démarche conceptuelle
- Présentation au <u>séminaire SIST</u> (28-29 juin 2018) : Séries Interopérables et Systèmes de Traitement / « Qualité des données » (utilisation de référentiels sémantiques permettant l'interopérabilité des données) « T-Semandiv - un nouveau thesaurus en biodiversité »
- Préparation d'un article pour une revue internationale en écologie à partir des travaux réalisés à l'Inist



#### **Thesauform Semandiv**

in progress

Hierarchy search

Index search

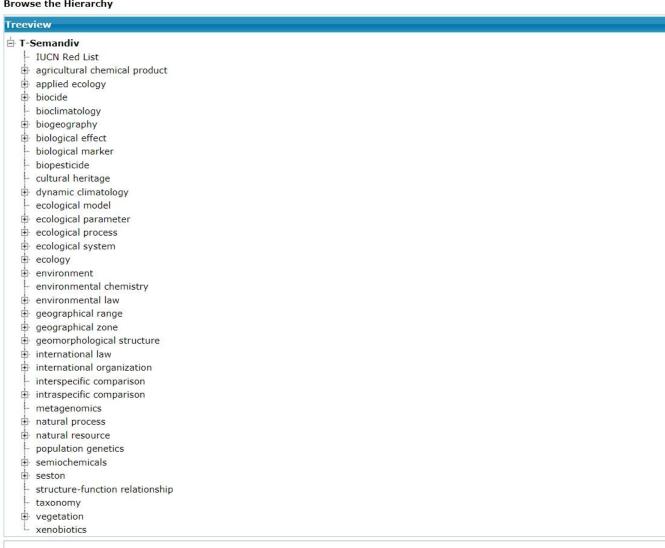
API

Annotation

Vote

Administration

#### Browse the Hierarchy



## Objectifs du GDR (2020)

- Mise en orbite du satellite de communication permettant d'interconnecter, d'annoter et relier sémantiquement les bases de données et de raisonner sur les données de biodiversité
- Formalisation du **thésaurus** des concepts-clés de la biodiversité vers une **ontologie** légère du domaine (affinement des relations sémantiques)
- Constitution de thésaurus spécialisés de sous-domaines de la biodiversité pour les combiner avec cette ontologie générale
- Mise en correspondance des bases de données avec ces ressources sémantiques

## Equipe projet à l'Inist

#### Membres du GDR:

- Isabelle Gomez (participant aux axes 2 et 3); conception du thésaurus
- Dominique Vachez (comité de pilotage du GDR / co-animateur de l'axe 3 avec Irstea / participant à l'axe 2); conception du thésaurus
- Nathalie Vedovotto (participant aux axes 2 et 3)

Conseils d'ordre technique : Majid Khayari

#### Acronymes:

**AnaEE**: **Ana**lyse et **E**xpérimentation sur les **E**cosystèmes

**CABT**: Centre for Agricultural Bioscience International Thesaurus

**EARTh:** Environmental Applications Reference Thesaurus

**EBV**: **E**ssential **B**iodiversity **V**ariable

**EnvThes: Environmental Thesaurus** 

GACS: Global Agricultural Concept Scheme

**GEMET**: **GEneral Multilingual Environmental Thesaurus** 

IMBE: Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale

**Irstea**: Institut national de recherche en sciences et technologies

pour l'environnement et l'agriculture

**LLOD**: Linguistic Linked Open Data

**Loterre**: Linked Open TERminology REsources

Lustre: Linked Thesaurus fRamework for Environment

**NALT**: National Agricultural Library Thesaurus

**SKOS**: Simple Knowledge Organization System

#### Liens:

**Site rBDD** (ateliers)

http://rbdd.cnrs.fr/spip.php?article170 http://rbdd.cnrs.fr/spip.php?article193

#### Site Inist

http://www.inist.fr/?Participation-de-l-Inist-au-GDR

**Site Loterre** 

https://www.loterre.fr/skosmos/BLH/fr/

Site GDR

http://www.semandiv.cnrs.fr/





#### **Inist-CNRS**





















Merci pour votre attention!



