

Le thésaurus COVADIS

Benoît DAVID

MEDDE/CGDD/DRI/MIG

Novembre 2014

Rappel des besoins à satisfaire

- Utilisation historique (GéoMAP/Géo-IDE) d'une arborescence, appelée arborescence Covadis, à 2 niveaux qui correspond à l'organisation physique des répertoires et fichiers dans les services (DDT, DRAAF, DREAL, ...)
- Cette approche est insuffisante pour certains utilisateurs
- En juin 2012, décision COVADIS/CCIG de mettre en œuvre un thésaurus
- Volonté de limiter l'ampleur du travail
- Choix de s'appuyer sur un fondement juridique
 - car de nombreux concepts d'entités sont fondés sur des textes juridiques (décret, loi, constitution, traité international, texte UE)
- Besoin d'intégrer l'affectation aux thèmes Inspire

Qu'est-ce qu'un thésaurus ?

- Ensemble organisé de mots-clés corresp. à des concepts et appelés termes
- 3 catégories de termes :
 - Descripteur ou terme préférentiel
 - Synonyme référençant un descripteur
 - Micro-thésaurus utilisé pour regrouper des descripteurs et non utilisé pour l'indexation des ressources

Organisation hiérarchique

- Les descripteurs sont organisés hiérarchiquement exprimant la **généricité** / **spécificité** du concept défini par le descripteur

- Notation:

- terme générique (TG) /
Broader term (BT)
- terme spécifique (TS) /
Narrower term (NT)

• Exemple :

–animal

•vertébré

–oiseau

–mammifère

»éléphant

»ours

»girafe

Représentation standardisée de la hiérarchie de genericité des concepts

- animal
 - vertébré
 - oiseau
 - mammifère
 - éléphant
 - ours
 - girafe

animal

TS: vertébré

éléphant

TG: mammifère

girafe

TG: mammifère

mammifère

TG: vertébré

TS: éléphant

TS: girafe

TS: ours

oiseau

TG: vertébré

ours

TG: mammifère

vertébré

TG: animal

TS: mammifère

TS: oiseau

Autres principes

- Relation d'association explicite une proximité de sens entre concepts voisins ou qui correspondent à des aspects différents d'une même réalité
(Voir aussi)
 - Exemple : Un **oiseau** est un vertébré ; L'**ornithologie** est une branche de la zoologie ; L'ornithologie a pour objet l'étude des oiseaux
- Des notes peuvent être associées pour expliciter la définition du concept ou sa règle d'utilisation et pour tracer les modifications du thésaurus
- Ajout de synonymes renvoyant vers un descripteur

Thésaurus Covadis (1/4)

3 ensembles de descripteurs

1. Concepts d'entités définis réglementairement et concepts d'entités géographiques statistiques
 - Exemples :
 - « réserve naturelle »
 - « petite région agricole »
2. Concepts d'entités relevant d'Inspire et n'appartenant pas au 1^{er} ensemble organisés par thème Inspire
 - Exemple : « occupation du sol sous la forme de polygones » dans « III.2. Occupation des terres »

Thésaurus Covadis (2/4)

3. Hiérarchie de politiques publiques

Les ressources qui ne correspondent à aucun descripteur des 2 premiers ensembles sont indexées par politique publique

- Exemples : un jeu de données « Poste fixe équipé pour chasser avec des filets » sera indexé par « chasse »
- Un jeu de données « Poste de guet d'intérêt DFCI » sera indexé par « défense de la forêt contre les incendies »

Thésaurus Covadis (3/4)

- Rattachement des entités géographiques par des relations hiérarchiques de généralité
 - au thème Inspire auquel il correspond
Exemple: « réserve naturelle » a pour terme générique « I.9. Sites protégés »
 - à la politique à laquelle il correspond
Exemple: « réserve naturelle » a pour terme générique « Nature, paysage, biodiversité »
- Utilisation de la poly-hiérarchie

Utilisation du std W3C SKOS

- Simple Knowledge Organization System
 - Recommandation du W3C pour représenter de manière standard des thésaurus, classifications ou tout autre type de vocabulaire contrôlé et structuré.
 - Déjà utilisé par exemple dans GéoSource
- Permet d'exprimer l'alignement entre vocabulaires différents
 - réserve naturelle (entité géographique COVADIS)
ExactMatch: réserve naturelle (Ecoplanète)
- Permet de définir des polyhiérarchies
 - réserve naturelle a pour concepts génériques :
 - entité géographique juridique
 - I.9. Sites protégés

Affinement de la méthode pour Inspire

- Le règlement interopérabilité définit des concepts d'entités (feature concepts) qui peuvent être utilisés comme terme
- Certains concepts d'entités sont précisés par certaines listes de codes
- Elaboration d'un vocabulaire dédié
VocInspire

Zone de gestion, de restriction ou de réglementation

chenal de navigation réglementé ...
district hydrographique
eaux de baignade
eaux désignées
masse d'eau au titre de la directive-cadre sur l'eau
région marine
unité de gestion pour les inondations
zone de gestion de la qualité de l'air
zone de gestion de zone côtière
zone de gestion forestière
zone de protection de l'eau potable
zone de protection phytosanitaire
zone de restriction pour des raisons de santé animale
zone faisant l'objet d'un permis d'exploration
et d'extraction minière
zone sensible
zone soumise à limitation du bruit
zone vulnérable aux nitrates
zones d'élimination des déchets
zones réglementées autour des sites contaminés

Proposition de démarche

- A. On part de 2 vocabulaires déjà définis
 - 1. Thèmes Inspire + descripteur « Hors Inspire »
 - 2. Concepts/sous-concepts d'entités Inspire
- B. On définit 3 vocabulaires
 - 1. Politiques publiques (hiérarchisé)
 - 2. Concepts d'entités règlementaires/statistiques (hiérarchisé)
 - Rattacher chacun à un A2 + un B1
 - 3. Concepts d'entités Inspire hors A2 et hors B2 (hiérarchisé)
 - Rattacher chacun à un A1
- C. Affecter à chaque série de données
 - A. Si dans Inspire à un A2, B2 ou B3
 - B. Si Hors Inspire à un B1 ou « Hors Inspire »