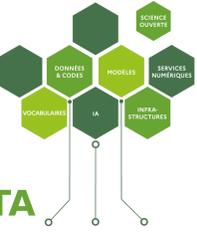


Vocabulaires thématiques dans le PEPR FORESTT Cas d'étude du multirisque



Peyratout Nolwenn, Pichot Christian, Rigolot Eric, Clastre Philippe

UR0629 URFM Écologie des forêts méditerranéennes, INRAE

NUM-DATA



CONTEXTE

Les projets du PEPR FORESTT rassemblent des communautés traitant de la résilience des forêts et pour lesquelles la pleine maîtrise du vocabulaire est un enjeu majeur.

La mobilisation et la consolidation de référentiels thématiques constituent une des actions structurantes de FORESTT.

Appuyée par le projet transversal NUM-DATA, ce travail est dans un premier temps réalisé au sein du projet X-RISKS.



OBJECTIF

Disposer de référentiels sémantiques fédérateurs des communautés et contribuer à l'interopérabilité des données partagées

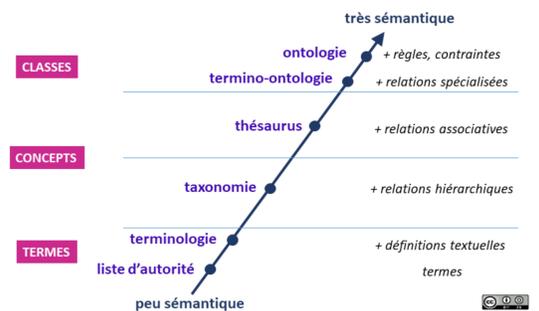
De la liste de termes à l'ontologie

Les vocabulaires définissent les concepts d'un domaine et les organisent et qualifient de façon plus ou moins avancée

Quelles pratiques dans FORESTT ?

L'identification de vocabulaires thématiques répond à l'analyse des besoins des communautés qui participent aux projets

- Raisonner**
Décrire le réel pour valider, simuler, générer de nouvelles connaissances
- Abstraire**
Indexer dans une ou plusieurs langues, documenter de manière non ambiguë
- Classer**
Organiser les objets numériques selon leurs propriétés ou leur thématique
- Normaliser**
S'affranchir de la variabilité de la langue, proposer une recherche à facettes



Typologie des vocabulaires

INRAE (2024). Les types de vocabulaires et leurs usages. Vocabulaires Ouverts@INRAE, <https://vocabulaires-ouverts.inrae.fr/types-de-vocabulaires-et-usages>

UN THÉSAURUS THÉMATIQUE DU RISQUE ET MULTIRISQUE EN FORÊT

1. Proposition d'une structure de thésaurus en se basant notamment sur l'état de l'art (ex. Simpson 2021) et l'approche du GIEC qui définit le risque comme le croisement de quatre composantes : **enjeu, vulnérabilité, aléa et réponse**.

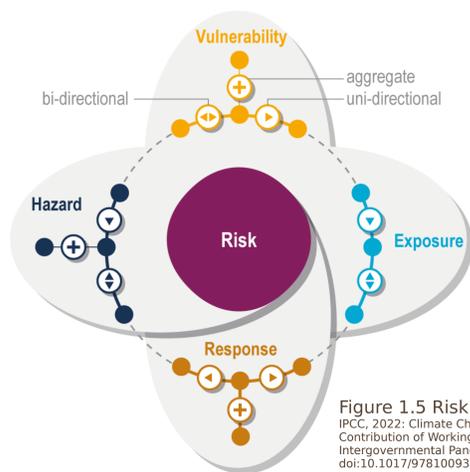


Figure 1.5 Risk in IPCC assessment through time
IPCC, 2022: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, 2023. doi:10.1017/9781009325844.

2. Appui d'un **groupe d'experts** qui intervient à toutes les étapes de construction du thésaurus : des **concepts** aux **définitions**. Les experts proposent des publications de référence et contribuent au cours de réunions individuelles.

3. Constitution d'un **corpus** d'une centaine de documents comme base bibliographique, source de concepts pour la création d'une première ébauche de thésaurus. Les réunions avec les experts permettent de **modifier, ajouter et hiérarchiser les concepts du thésaurus**. Elles sont gage de la bonne compréhension du vocabulaire utilisé et permettent d'affiner au mieux les concepts.

4. Rédaction ou récupération des **définitions** des concepts faite de différentes manières :

- Utilisation de **Gemini AI** en « deep search » pour proposer des définitions sourcées de concepts.
- Utilisation du **corpus** de texte pour retrouver des définitions écrites.
- **Exploration d'autres thésaurus** via le logiciel AML permettant l'alignement des concepts et la récupération de définitions d'intérêt. <https://forge.inrae.fr/christian.pichot/AML-Project>

5. La dernière étape portera sur la **validation des définitions** du thésaurus avec le groupe d'experts.

POINT D'AVANCEMENT

- Mise en place d'un travail collaboratif et fédérateur
- Thésaurus de 500 termes sur le multirisque
- Majorité des termes définis en français

ET PERSPECTIVES

- Finalisation et publication du thésaurus X-RISKS
- Analyse des besoins des autres communautés de FORESTT
- Exploitation des vocabulaires construits

