



**HAL**  
open science

# Modèle de médiation pour la découverte de contenus géographiques structurés

Mathieu Zrhal

► **To cite this version:**

Mathieu Zrhal. Modèle de médiation pour la découverte de contenus géographiques structurés. Géographie. Université Gustave Eiffel, 2023. Français. NNT : 2023UEFL2037 . tel-04541505

**HAL Id: tel-04541505**

**<https://theses.hal.science/tel-04541505v1>**

Submitted on 10 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Modèle de médiation pour la découverte de contenus géographiques structurés

Thèse de doctorat de l'Université Gustave Eiffel

École doctorale n° 532, Mathématiques et Sciences et Technologies de l'Information et de la communication (MSTIC)

Spécialité de doctorat : Sciences et Technologies de l'Information Géographique

Unité de recherche : LaSTIG (IGN)

Thèse présentée et soutenue à l'Université Gustave Eiffel,  
le 26/01/2023, par :

**Mathieu Mehdi ZRHAL**

### Composition du Jury

<b>Bernd AMANN</b> Professeur, Sorbonne Université - LIP6	Examineur
<b>Elisabeth METAIS</b> Professeure, CNAM	Examineur
<b>Dimitris KOTZINOS</b> Professeur, CY Cergy Paris Université	Rapporteur
<b>Nathalie PERNELLE</b> Professeure, Université Sorbonne Paris Nord	Rapportrice

### Encadrement de la thèse

<b>Bénédicte BUCHER</b> Ingénieure des Ponts des Eaux et des Forêt, IGN - Université Gustave Eiffel	Directrice de thèse
<b>Fayçal HAMDI</b> Maître de conférences, CNAM	Co-Encadrant de thèse
<b>Marie-Dominique VAN-DAMME</b> Ingénieure, IGN - Université Gustave Eiffel	Co-Encadrante de thèse

---

## **Dédicaces**

Cette oeuvre est dédiée à ma famille, mes encadrants, mes amis et mes collègues qui, sans leur soutien indéfectible, n'aurait jamais vu le jour. Un grand merci du fond du coeur.

---

## Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude envers ma directrice de thèse, Bénédicte Bucher. Sa guidance éclairée, son soutien indéfectible et sa patience infinie ont été des piliers essentiels tout au long de ce parcours de recherche. Ses conseils précieux et son expertise ont grandement enrichi mon travail et ont été source d'inspiration.

Mes remerciements vont également à mes co-encadrants, Marie-Dominique Van-Damme et Fayçal Hamdi. Leur expertise complémentaire, leur engagement sans faille et leur encouragement ont été d'une importance capitale dans l'accomplissement de cette thèse. Leurs commentaires éclairés, leurs discussions stimulantes et leur soutien inconditionnel ont été des éléments déterminants dans la réalisation de ce travail de recherche.

Je souhaite également exprimer ma reconnaissance envers l'ensemble de l'équipe de recherche pour son soutien et son atmosphère collaborative, ainsi qu'à toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce projet.

Enfin, je voudrais adresser mes plus sincères remerciements à ma famille et à mes amis pour leur soutien constant, leurs encouragements et leur compréhension tout au long de cette aventure.

---

## Résumé

Depuis quelques années, la quantité de données géographiques disponibles sur le web n'a cessé d'augmenter. Cela est notamment dû non seulement aux spécificités propres aux données géographiques qui rendent leur utilisation indispensable dans plusieurs domaines d'application comme la météorologie, l'agriculture, prévention des risques naturels etc., mais aussi à l'adoption des principes d'ouverture des données par de plus en plus d'institutions et d'organisation publiques ou privés.

Traditionnellement, les données géographiques sont disponibles sur des portails en ligne qui permettent de les télécharger sous la forme de jeux de données en se basant sur des métadonnées qui décrivent ces derniers. La multiplicité de ces portails rend la tâche de retrouver le bon jeu de données difficile pour l'utilisateur.

Parallèlement, depuis quelques années, le Web se transforme d'un Web de documents à un Web de données. Cela a permis de développer de nouveaux outils puissants qui permettent de mieux structurer les données en ligne. Parmi ces outils, les graphes de connaissance ont permis une percée importante dans le domaine de la recherche d'information sur le web.

Cette thèse vise à étudier la possibilité de créer un graphe de connaissance qui permettrait d'améliorer la recherche de jeux de données géographiques sur le web. Pour cela, nous avons développé une approche basée sur un graphe de connaissances dédié au métadonnées de jeux de données géographiques. Nous avons ensuite implémenté un prototype de ce graphe de connaissance pour finalement évaluer avec l'aide d'experts l'apport de ce dernier dans l'amélioration de la recherche de jeux de données géographiques.

---

## **Abstract**

In recent years, the amount of spatial data available on the web has increased steadily. This is not only due to the specificities of spatial data that make their use essential in several application domains such as meteorology, agriculture, natural risk prevention, etc., but also to the embracing of open data principles by more and more public and private institutions and organizations.

Usually, spatial data are available on online portals that allow to download them in the form of datasets based on metadata that describe them. The multiplicity of these portals makes it difficult for the user to find the right dataset. This task is even more complex if the user needs to evaluate the costs and benefits of exploiting or reusing the spatial data.

At the same time, in the last few years, the Web is transforming from a Web of documents to a Web of data. This has led to the development of powerful new tools that allow better structuring of online data. Among these tools, knowledge graphs have allowed an important breakthrough in the field of information retrieval on the Web.

This thesis aims to study the possibility of creating a knowledge graph that would allow to improve the search of spatial datasets on the web. To do so, we have developed an approach based on a knowledge graph dedicated to spatial data metadata. We then implemented a prototype of this knowledge graph to finally evaluate with the help of experts the contribution of the latter in the improvement of spatial data sets search.

---

# Table des figures

1.1	Nombre de jeux de données sur data.gouv.fr et geo.data.gouv.fr . . . . .	5
1.2	Évolution du nombre de jeux de données sur data.europa.eu . . . . .	6
1.3	Exemple des résultats de recherche obtenus dans le catalogue Sandre en utilisant le mot-clé "Estuaire" . . . . .	7
2.1	Composants d'une infrastructure de données géographiques [78] . . . . .	12
2.2	Relation entre ISO 19115-1 et ISO 19115-2 . . . . .	17
2.3	Composants d'un moteur de recherche classique . . . . .	23
2.4	Pyramide des langages et standards du Web sémantique tels que présentés par Tim Berners-Lee en 2006 . . . . .	27
2.5	Le LOD cloud . . . . .	29
2.6	Diagramme UML des classes DCAT . . . . .	33
2.7	Diagramme UML du standard DCAT 2 . . . . .	34
2.8	Alignement entre DCAT 2 et ISO 19115 . . . . .	35
2.9	Les profils DCAT . . . . .	36
3.1	Processus global de Recherche . . . . .	41
3.2	Champs de métadonnées dans le standard DCAT-AP . . . . .	42
3.3	Processus de recherche des jeux de données géographiques grâce à un KG dédié . . . . .	43
3.4	Knowledge Graph Components . . . . .	45
3.5	Comparaison de Records . . . . .	46
3.6	Schéma d'ensemble du KG . . . . .	47
3.7	Exemple du concept "Cours d'eau" dans Gemet . . . . .	51
3.8	Exemple du concept "Cours d'eau" dans le Glossaire eau et biodiversité . . . . .	52
3.9	Exemple du concept "Cours d'eau" dans le Thésaurus Eau . . . . .	53
3.10	Exemple d'un Élément Wikidata . . . . .	55
3.11	Exemple d'alignements qui existent entre DBpedia et d'autres graphes de connaissances du LOD . . . . .	56

---

3.12 Exemple de scores de similarité sémantique calculés par l'outil KG semantic similarity . . . . .	57
4.1 Processus général de la création et de l'exploitation du KG . . . . .	61
4.2 Requêtes pour calculer l'alignement entre GEMET et Wikidata . . . . .	63
4.3 Exemple de requête pour annoter les mots-clés et thèmes présents dans les records RDF . . . . .	65
4.4 Exemple de résultats de recherche d'un concept à travers une interface utilisateur . . . . .	66
4.5 Requête pour trouver un concept Wikidata à travers son label . . . . .	67
4.6 Requête pour extraire les records depuis le KG . . . . .	68
4.7 Exemple de clusters de métadonnées du KG représentés sous la forme d'un graphe . . . . .	70
4.8 Extrait des métadonnées associées au jeu de données :Réseau de mesures de houle du Centre d'Archivage National des Données de Houles In-Situ (CANDHIS) - France métropole et outre-mer . . . . .	75
4.9 Résultats de recherche pour la requête "plan de prévention du risque inondation" sur le catalogue Sandre . . . . .	85
4.10 Résultats de recherche pour la requête "plan de prévention du risque inondation" sur le catalogue Cerema . . . . .	85
4.11 Résultats de recherche pour la requête "nappe phréatique" sur le catalogue Cerema . . . . .	88
4.12 Résultats de recherche pour la requête "nappe phréatique" sur le catalogue Sandre . . . . .	89

# Liste des tableaux

2.1	Liste des principaux Standards ISO pour les métadonnées et catalogues géographiques . . . . .	15
2.2	Liste des principales classes du standard ISO 19115-1 . . . . .	17
2.3	Alignement entre les classes de DCAT et les classes de Schema.org . . .	37
4.1	Liste des concepts GEMET bannis . . . . .	64
4.2	Les 10 meilleurs résultats obtenus pour la requête "plan de prévention du risque inondation" grâce au KG . . . . .	87
4.3	Les 11 meilleurs résultats obtenus pour la requête "nappe phréatique" grâce au KG . . . . .	90

---

# Table des matières

Dédicaces	iii
Remerciements	v
Résumés	vii
Liste des Figures	xii
Liste des tableaux	xiii
<b>1 Contexte, problématique et objectifs</b>	<b>1</b>
1.1 La découverte de données géographiques : enjeux applicatifs . . . . .	1
1.2 Problématique de la recherche de jeux de données géographiques . . . . .	6
1.3 Objectifs . . . . .	8
<b>2 État de l’art</b>	<b>11</b>
2.1 Introduction . . . . .	11
2.2 Métadonnées et catalogues géographiques . . . . .	12
2.2.1 Infrastructures de données géographiques . . . . .	12
2.2.2 Métadonnées géographiques . . . . .	14
2.2.2.1 ISO 19110 . . . . .	15
2.2.2.2 ISO 19115 . . . . .	15
2.2.2.3 ISO 19119 . . . . .	18
2.2.2.4 ISO 19139 . . . . .	18
2.2.2.5 ISO 19150 . . . . .	19
2.2.2.6 Directive INSPIRE . . . . .	19
2.2.3 Remplissage des métadonnées . . . . .	21
2.3 Web sémantique et recherche d’information sur le Web . . . . .	22
2.3.1 Moteurs de recherche sur le Web . . . . .	22
2.3.1.1 Mesures de similarité . . . . .	24
2.3.2 Web sémantique et Graphes de connaissances . . . . .	27
2.3.2.1 Les bases du Web sémantique . . . . .	27

---

2.3.2.2	Les graphes de connaissances . . . . .	28
2.3.3	Standards des Métadonnées sur le Web . . . . .	32
2.3.3.1	DCAT . . . . .	32
2.3.3.2	DCAT-AP . . . . .	35
2.3.3.3	GeoDCAT-AP . . . . .	36
2.3.3.4	Schema.org . . . . .	36
2.4	Conclusion . . . . .	37
<b>3</b>	<b>Conception d'un Knowledge Graph dédié à la recherche de jeux de données géographiques</b>	<b>39</b>
3.1	Introduction . . . . .	39
3.2	Présenter l'approche . . . . .	39
3.3	Analyser les exigences fonctionnelles du KG . . . . .	40
3.3.1	Processus Hors-Session . . . . .	41
3.3.2	Session de Recherche . . . . .	42
3.3.2.1	Supporter la spécification d'un concept d'intérêt utilisateur . . . . .	43
3.3.2.2	Reformuler la requête utilisateur en une requête sur les métadonnées . . . . .	44
3.3.2.3	Évaluer la pertinence des records candidats . . . . .	45
3.3.2.4	Regrouper les records et les renvoyer à l'utilisateur . . . . .	46
3.4	Analyse des ressources à intégrer au KG . . . . .	48
3.4.1	Les catalogues à moissonner . . . . .	48
3.4.1.1	Sandre . . . . .	48
3.4.1.2	Cerema . . . . .	49
3.4.2	Les vocabulaires thématiques . . . . .	49
3.4.2.1	GEMET . . . . .	50
3.4.2.2	Glossaire eau et biodiversité . . . . .	51
3.4.2.3	Thésaurus EAU . . . . .	52
3.4.3	Les vocabulaires de sens commun . . . . .	53
3.4.3.1	Wikidata . . . . .	54
3.4.3.2	DBpedia . . . . .	54
3.4.4	Mesures de similarité . . . . .	55
3.5	Conclusion . . . . .	57
<b>4</b>	<b>Implémentation et évaluation</b>	<b>59</b>
4.1	Introduction . . . . .	59
4.2	Méthodologie d'évaluation . . . . .	60
4.3	Implémentation et production du KG . . . . .	61

---

---

4.3.1	Moissonner les records . . . . .	61
4.3.2	Intégration des records et de GEMET au KG . . . . .	62
4.3.3	Enrichissement des records et alignement des vocabulaires . . . . .	63
4.4	Exploitation du KG lors d'une session de recherche . . . . .	66
4.4.1	Identification du concept utilisateur . . . . .	66
4.4.2	Sélection et classement des records potentiellement pertinents . . . . .	67
4.4.3	Regrouper des records de jeux de données géographiques . . . . .	68
4.5	Expérimentations . . . . .	69
4.5.1	première évaluation des fonctionnalités de session . . . . .	69
4.5.1.1	Tâche 1 : Pertinence de Wikipédia-Wikidata pour identifier des concepts utilisateur . . . . .	71
4.5.1.2	évaluer la capacité du moteur à pointer vers des records pertinents en lien avec un concept utilisateur . . . . .	73
4.5.1.3	Évaluer le regroupement des records . . . . .	77
4.5.2	Expérimentation : Fixer un seuil de similarité entre deux concepts . . . . .	78
4.5.2.1	Requête 1 : Nappe phréatique . . . . .	79
4.5.2.2	Requête 2 : Phytoépuration . . . . .	80
4.5.2.3	Requête 3 : Plan de prévention du risque inondation . . . . .	81
4.5.2.4	Requête 4 : Station de traitement des eaux usées . . . . .	82
4.5.2.5	Synthèse . . . . .	82
4.5.3	Expérimentation 3 : Identifier des records pertinents . . . . .	83
4.5.3.1	Requête 1 : plan de prévention du risque inondation . . . . .	84
4.5.3.2	Requête 2 : nappe phréatique . . . . .	88
4.5.3.3	Synthèse . . . . .	91
4.6	Conclusion . . . . .	91
<b>5</b>	<b>Conclusion et Perspectives</b>	<b>93</b>
<b>A</b>	<b>Questionnaire Experts</b>	<b>105</b>
<b>B</b>	<b>Réponses des experts au questionnaire</b>	<b>107</b>
<b>C</b>	<b>Questionnaire pour la similarité sémantique entre deux concepts du domaine de l'eau</b>	<b>109</b>
<b>D</b>	<b>Réponses au questionnaire pour la similarité sémantique entre deux concepts du domaine de l'eau</b>	<b>111</b>

---

# Chapitre 1

## Contexte, problématique et objectifs

### 1.1 La découverte de données géographiques : enjeux applicatifs

La donnée géographique ne cesse de jouer un rôle crucial dans un monde où les gouvernements et les entreprises se fondent de plus en plus sur les données pour leur prise de décision. En effet, la donnée géographique est une donnée qui combine, tout d'abord, une donnée associée à une localisation relative à la planète Terre que ce soit implicitement ou explicitement, et ensuite un ensemble d'attributs décrivant des phénomènes ou des objets.

Cette dualité permet à la donnée géographique d'être indispensable dans plusieurs domaines comme la météorologie, l'agriculture, prévisions des risques, gestion du changement climatique, l'aménagement du territoire, etc. De plus, la donnée géographique est un vecteur d'intégration de données provenant de sources différentes. En effet, il est possible de rassembler et combiner des collections de données grâce à leurs composantes géographiques. Cela rend l'utilisation des données géographiques pertinente même dans le cas d'une prise de décision non localisée. Cela pousse plusieurs institutions publiques et privées à produire des données géographiques relatives à leurs domaines de compétence respectifs et de les mettre à disposition de potentiels utilisateurs et organismes qui pourraient s'y intéresser.

La production de données géographiques est une activité coûteuse qui nécessite d'importantes ressources humaines et matérielles. Cela se manifeste tout au long du cycle de vie de la donnée géographique, en commençant par son acquisition, la création d'une première version du produit, sa qualification, son maintien à jour, sa mise à

disposition, etc.

Par exemple, la collecte des données géographiques nécessite l'utilisation de différentes ressources matérielles comme des avions ou drones pour les prises de vues aériennes, des radars pour la radio-détection, des lasers pour les nuages de points Lidar, etc., ainsi que les savoir-faire nécessaires pour manipuler ces ressources et collecter les données géographiques.

Dès les années 1990, l'enjeu de faciliter le partage et la réutilisation des données géographiques pour diviser les coûts d'acquisition et de production a été identifié et a donné naissance au domaine des infrastructures de données géographiques (*IDG*).

En parallèle de la mise en place des IDG, dès la fin des années 1990 et au début des années 2000, le développement et la popularisation du Web ont été vus par la communauté comme une opportunité de faciliter l'accès et l'utilisation des données géographiques. Ainsi, plusieurs portails en ligne et outils Web ont été mis en place par les producteurs des données géographiques pour permettre la mise à disposition des données géographiques à un maximum d'utilisateurs. Cela a permis une plus grande diffusion de ces données auprès d'un public plus large, notamment grâce à des applications Web comme Google Maps qui permettent à un utilisateur de consulter des données géographiques directement sur son navigateur sans passer par des logiciels et outils spécialisés qui nécessitent une expertise pour leur prise en main.

Les données géographiques ont aussi profité de l'essor du principe de l'ouverture des données qui a accompagné la popularisation et l'utilisation du Web au début des années 2000. En effet, le Web a été un outil pratique pour les producteurs de données pour publier leurs données plus facilement et les partager au plus grand nombre. L'utilisation d'internet a aussi permis aux chercheurs de publier plus facilement leurs jeux de données de recherche et à d'autres utilisateurs de publier leurs propres données. Cela s'inscrit plus globalement dans le paradigme d'ouverture des données sur le Web.

Le besoin d'outils en ligne a aussi mené à des projets collaboratifs de production des données géographiques comme OpenStreetMap qui se base sur le volontariat. Ce projet vise à fournir une cartographie en ligne et à la constitution d'une base de données géographiques du monde entier qui est mise sous licence libre. Le projet a connu un grand succès et un engouement certains de la part d'un grand nombre de contributeurs (7.000.000 en 2020). Cela souligne l'intérêt et l'utilité des données géographiques auprès du grand public notamment.

La production collaborative des données géographiques est à rapprocher plus largement au domaine des sciences ouvertes et des données FAIR (*FAIR data : Findable, Accessible, Interoperable, Reusable*). Les données FAIR impliquent que les données publiées sur internet répondent aux principes qui suivent. Tout d'abord, les données doivent être faciles à retrouver grâce à l'utilisation d'un identifiant unique et propre à chaque ensemble de données. Ensuite, les données FAIR ne sont pas forcément des données ouvertes, mais doivent être accessibles en fournissant assez d'informations permettant de facilement y accéder. Cela se fait grâce à des protocoles standard de communication qui sont ouverts, libres et d'usage universel. L'interopérabilité des données géographiques implique l'utilisation de métadonnées contextuelles et standardisées. Enfin, le dernier principe FAIR est la réutilisation des données géographiques qui se fonde sur des standards de données communs et l'élaboration de bases de données rassemblant des données claires et décrites convenablement.

Les principes FAIR ne concernent pas uniquement les données géographiques, ils ont été introduits comme une solution permettant la découverte et la réutilisation de données diverses et provenant de sources multiples et disponibles sur le Web. En effet, un nombre important de données ouvertes, qu'elles proviennent du citoyen, d'une autorité, du domaine de la recherche ou de la science ouverte ont été publiés. Ces différentes sources de données ont nécessité l'introduction de règles et de normes permettant de mieux découvrir, partager et réutiliser les données disponibles sur le Web.

L'Europe, à travers le parlement européen, a fait passer la directive INSPIRE en 2007 qui vise à créer une IDG au niveau de l'Union européenne. Celle-ci ambitionne à mieux homogénéiser le partage des données géographiques environnementales dans toute la communauté européenne et à fournir les données nécessaires à l'élaboration des politiques transfrontalières au sein de l'union. La directive INSPIRE définit un ensemble de règles de mise en œuvre qui sont directement applicables dans toutes les entités et instituts publics producteurs de données géographiques dans les 27 pays de l'UE.

La quantité des données et ressources géographiques disponibles sur le Web a inévitablement mené à l'enjeu de la découverte de ces dernières. Avec de plus en plus d'entités publiques et privées, ainsi que les principes de la recherche ouverte, un utilisateur aura plus de mal à rechercher les données dont il a besoin. La Figure 1.1 présente le nombre de jeux de données disponibles sur deux portails français, data.gouv (données publiques de l'état français) et geo.data.gouv (données géographiques publiques

de l'état français). Au niveau européen, la Figure 1.2 montre l'évolution du nombre de jeux de données publiés par des entités publiques européennes entre août 2019 (870000 jeux de données indexés) et juillet 2022 (1.500.000 jeux de données indexés).

En effet, il existe différents portails et outils permettant d'accéder aux données, chacune ayant ses spécificités et nécessitant une expertise préalable pour identifier la source où pourraient se trouver les données géographiques dont il a besoin. Une fois le bon portail ou outil identifié, l'utilisateur a encore besoin d'évaluer les coûts et les bénéfices nécessaires à l'exploitation de ces données que ce soit pour les formats dans lesquels les données sont disponibles ou les outils nécessaires à leur exploitation et à leur visualisation. De plus, il est rare qu'un unique ensemble de données soit suffisant pour satisfaire le besoin d'un utilisateur. Certaines applications nécessitent d'utiliser des données provenant de sources différentes et de les intégrer pour créer un nouvel ensemble de données.

Du point de vue des producteurs de données, le défi qui se présente est celui de la visibilité de ses données. La question qui se pose est de comment rendre des données géographiques plus visibles auprès d'utilisateurs qui s'y intéressent, mais dont le profil peut varier entre novices et experts et dont le comportement n'est pas clairement défini.

Cela rapproche le problème de la découverte et de réutilisation des données géographiques du domaine de recherche d'information (*IR*) sur le Web. L'*IR* est le domaine qui s'intéresse à la recherche des documents et des pages Web disponibles sur internet et a notamment donné naissance à la création des moteurs de recherche comme Google et Bing. De la même manière, il existe aussi des moteurs de recherches dédiés à d'autres types d'objets, comme les articles scientifiques (e.g., Google Scholar, DBLP, etc.) ou les données d'apprentissage (e.g., Kaggle), qui se basent sur des technologies similaires, et qui ont été adaptés au domaine d'application visé.

Depuis quelques années, le Web se transforme d'un Web de documents (c.-à-d. Web 2.0) à un Web de données (c.-à-d. Web 3.0). Cela est dû à l'importance grandissante des données disponibles dans le Web. En effet, de grands efforts ont été fournis pour mieux structurer et rendre plus accessibles les données sur le Web, notamment grâce aux principes du Web sémantique et aux outils qui y sont dédiés. Ainsi, les graphes de connaissances, aussi appelés Knowledge Graphs (*KG*), ont été largement utilisés pour améliorer la recherche d'information. En effet, plusieurs moteurs de recherche classiques comme Google ou Bing utilisent des *KGs* pour améliorer les résultats de recherche obtenus. D'autres grandes entreprises comme Facebook ou Amazon ont créé

The screenshot shows the 'Jeux de données' (Data Games) section of data.gouv.fr. At the top, there's a search bar with 'Recherche' and a magnifying glass icon. Below the search bar, there are navigation links: 'Données', 'Réutilisations', 'Organisations', 'Commencer sur data.gouv.fr', 'Actualités', and 'Nous contacter'. The main heading is 'Jeux de données' with a sub-heading 'Rechercher parmi les 41 848 jeux de données sur data.gouv.fr'. A search bar contains 'Ex : élections présidentielles 2022' and a 'Recherche' button. Below this, there are filters for 'Organisations' (set to 'Toutes les organ...') and 'Mots clés' (set to 'Tous les mots-clés'). The results section shows '10000 résultats' and a 'Trier par : Pertinence' dropdown. Two results are visible: 'Base Sirene des entreprises et de leurs établissements (SIREN, SIRET)' by Insee, and 'Hébergements équestre - Roi Morvan Communauté - Licence Ouverte'. A sidebar on the right shows filters for 'PUBLIÉ SUR DATA.GOUV.FR', 'DONNÉE OUVERTE', 'FORMAT DE DISTRIBUTION', 'LICENCE', and 'TYPE'.

FIGURE 1.1 – Nombre de jeux de données sur data.gouv.fr et geo.data.gouv.fr

des KGs dédiés à leurs secteurs d'activité respectifs et sont notamment utilisés pour améliorer les recommandations faites à leurs clients.

En parallèle des KGs construits par ces grands acteurs du Web, plusieurs KGs collaboratifs ont été développés et maintenus par une communauté de volontaires. On peut notamment citer Wikidata et DBpedia qui sont deux graphes de connaissances étroitement liés à l'encyclopédie en ligne Wikipédia. Ils fournissent un ensemble de données sur le monde réel sous une forme structurée et lisible à la fois par un utilisateur humain et par des machines. Toutes les données qui y sont contenues sont disponibles sous licence libre et sont accessibles librement.

Les graphes de connaissances ont déjà été utilisés pour améliorer la recherche de données de manière générale. Par exemple, Google Dataset Search, le moteur de recherche développé par Google, réutilise son KG pour améliorer les résultats de recherche et permet ainsi aux utilisateurs de retrouver différents jeux de données. Au niveau européen, le portail data.europa.eu emploie aussi les technologies du Web sémantique pour indexer et retrouver des jeux de données publiés par des établissements publics

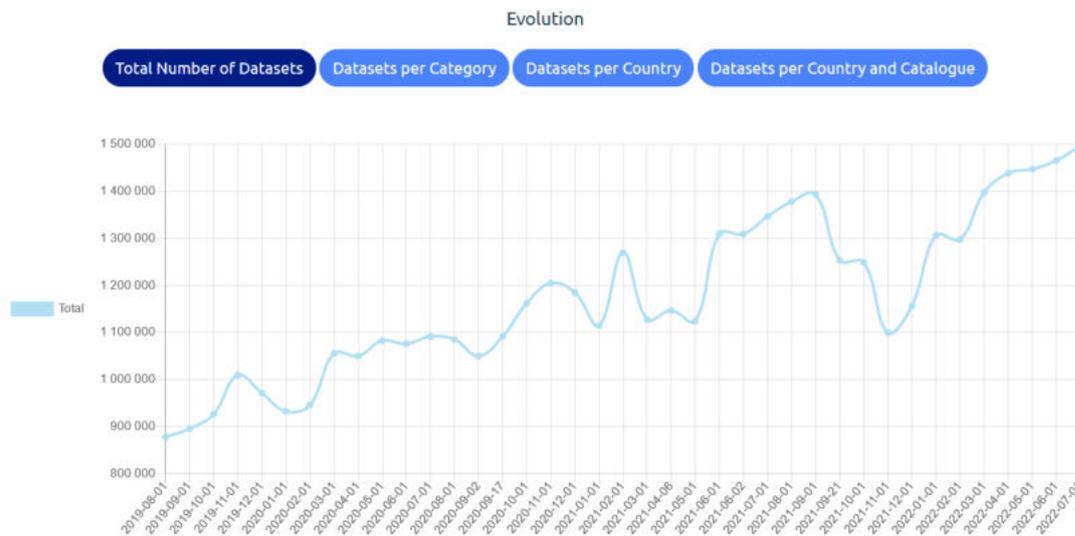


FIGURE 1.2 – Évolution du nombre de jeux de données sur data.europa.eu

de toute l'UE.

## 1.2 Problématique de la recherche de jeux de données géographiques

Pour en faciliter la découverte, les jeux de données géographiques sont accompagnés de métadonnées fournies par les producteurs de données. Les métadonnées géographiques peuvent être définies comme des données à propos des données géographiques. Leur but premier est de décrire un jeu de données géographiques et de fournir des informations telles que la disponibilité, l'accessibilité et l'interopérabilité des données.

La recherche de jeux de données géographiques est essentiellement fondée sur l'exploitation des métadonnées accompagnant les jeux de données qui sont rassemblées sur des portails dédiés qu'on appelle catalogues géographiques. Ces derniers collectent et indexent les métadonnées et permettent à l'utilisateur d'effectuer des recherches, le plus souvent, en utilisant des mots-clés. La Figure 1.3 présente l'interface de recherche du catalogue français Sandre<sup>1</sup>.

Les catalogues géographiques indexent des jeux de données en fonction de critères tels que les thématiques abordées, les standards utilisés, le type des données, la source du financement, etc. Ainsi, il existe de nombreux catalogues géographiques qui communiquent très peu entre eux (c.-à-d. chaque catalogue se spécialise dans un type de

1. <https://www.sandre.eaufrance.fr/>

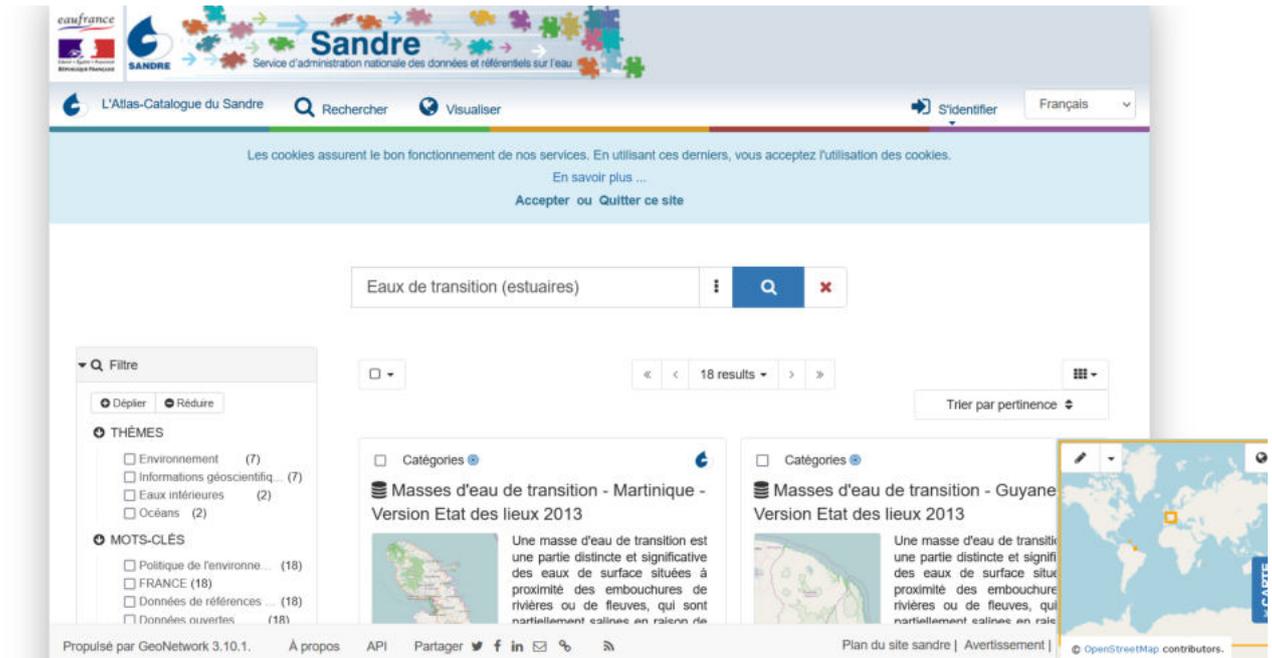


FIGURE 1.3 – Exemple des résultats de recherche obtenus dans le catalogue Sandre en utilisant le mot-clé "Estuaire"

données géographiques spécifique ou bien encore un producteur propose son catalogue), ce qui crée des silos de données qui évoluent indépendamment les uns des autres.

Rien qu'en France, un grand nombre de catalogues géographiques existent. Ainsi, en plus du Sandre qui rassemble les jeux de données de référence sur l'eau en France, ou encore le Géocatalogue<sup>2</sup> qui se spécialise dans les jeux de données géographiques INSPIRE, le catalogue Cerema<sup>3</sup> qui s'intéresse aux jeux de données autour des thématiques de l'environnement, du transport, des risques, etc. Si certains jeux de données peuvent être présents dans un ou plus de ces catalogues, rien ne garantit qu'un jeu de données spécifique peut être trouvé dans l'un d'entre eux.

Bien qu'il existe plusieurs standards permettant d'établir des métadonnées géographiques, ces derniers ont été développés pour laisser un maximum de liberté aux producteurs de données afin de décrire leurs jeux de données géographiques. Par exemple, on peut notamment citer la possibilité d'utiliser des vocabulaires métier ou thématiques dans les métadonnées ce qui a l'avantage de permettre aux producteurs des données de mieux décrire leurs données, mais ça implique par la même occasion, une hétérogénéité dans la manière dont les métadonnées sont spécifiées.

De plus, ces catalogues implémentent différents moteurs de recherche (p. ex., Geo-

2. <https://www.geocatalogue.fr/>

3. <https://www.cdata.cerema.fr/geonetwork/srv/fre/catalog.search#/search>

Network, Lucene, moteur développé en interne, etc.), adaptés aux jeux de données qui y sont indexés. Cela implique que l'utilisateur devra potentiellement adopter différentes stratégies de recherche pour trouver le résultat qui lui convient. Cette tâche n'est pas une tâche aisée et nécessite une certaine expertise.

Bien que les technologies du Web sémantique ont été utilisées pour améliorer la recherche de jeux de données de manière générale et plus précisément pour prendre en compte des jeux de données provenant de sources multiples et présents dans des portails différents. Cependant, à notre connaissance, aucune approche dédiée spécifiquement aux jeux de données géographiques n'a été développée à ce jour.

La problématique de la recherche de jeux de données géographiques se situe à l'intersection de trois domaines de recherche. Tout d'abord, les catalogues géographiques, les métadonnées associées aux jeux de données géographiques et les standards qui les structurent. Viennent ensuite la recherche d'information et les moteurs de recherche avec leur capacité à retrouver et classer les résultats pertinents pour l'utilisateur. Enfin, le Web sémantique et notamment les graphes de connaissances qui apportent les technologies nécessaires.

### 1.3 Objectifs

L'objectif principal de cette thèse est d'apporter une contribution au domaine de la recherche de jeux de données géographiques en appliquant les technologies du Web sémantique et de la recherche d'information. Pour cela, nous adopterons une approche basée sur les graphes de connaissance. Nous nous concentrerons sur un domaine d'application spécifique, à savoir, l'eau et l'environnement en France. Cela nous permettra de mieux comprendre le besoin de l'utilisateur et de prendre en compte son contexte, son approche, etc.

La suite de ce manuscrit s'organisera comme suit. Tout d'abord, le chapitre 2 se concentrera sur l'état de l'art des trois domaines de recherche adjacents, à savoir les métadonnées géographiques, la recherche d'information et le Web sémantique. Le chapitre 3 se concentrera sur la conception et la création d'un graphe de connaissances dédié à la recherche de jeux de données géographiques. On y abordera une analyse détaillée des exigences fonctionnelles d'un tel graphe de connaissances ainsi qu'une architecture qui y répond.

---

Le chapitre 4 détaillera la méthodologie suivie pour évaluer notre approche basée sur les graphes de connaissance et présentera et analysera les expérimentations qui ont été effectuées. Finalement, le chapitre 5 conclura ce manuscrit et présentera différentes perspectives au travail de recherche effectué tout au long de cette thèse.



## Chapitre 2

# État de l'art

### 2.1 Introduction

La recherche de jeux de données géographiques est un enjeu très important dans la découverte, le partage et la réutilisation des données géographiques.

La problématique de la recherche de jeux de données géographiques est à la croisée de deux domaines de recherches. D'un côté, le domaine des métadonnées et des catalogues géographiques, de l'autre, la recherche d'information sur le Web et le rôle que joue l'utilisation du Web sémantique et des données liées dans l'amélioration de la découverte des ressources sur le Web.

Ce chapitre aura pour rôle d'étudier et de présenter l'état de l'art des deux domaines connexes à la problématique de la recherche de jeux de données géographiques. La Section 2.2 présentera les infrastructures de données géographiques et notamment le domaine des catalogues géographiques avant de nous intéresser aux métadonnées géographiques, leurs standards et quelques freins à leur exploitation.

Ensuite, dans la Section 2.3 nous présenterons les domaines de la recherche d'information et notamment les moteurs de recherche et leur fonctionnement, pour ensuite nous intéresser au domaine du Web sémantique et plus précisément aux graphes de connaissance et leur apport à la recherche d'information sur le Web.

La Section 2.4 clôture le Chapitre en récapitulant les éléments essentiels et en définissant quelques hypothèses de recherche ainsi le cadre dans lequel s'inscrit ce travail de thèse.



FIGURE 2.1 – Composants d’une infrastructure de données géographiques [78]

## 2.2 Métadonnées et catalogues géographiques

### 2.2.1 Infrastructures de données géographiques

La géomatique (contraction entre les termes “géographie” et “informatique”) est le domaine qui s’intéresse au traitement informatique de la donnée géographique. Cela englobe différents outils et méthodes permettant l’acquisition, la représentation, l’analyse et l’intégration des données géographiques. Les infrastructures de données géographiques (*IDG*) sont la base sur laquelle s’appuie le partage et la réutilisation de données géographiques. En effet, le terme *IDG* est souvent utilisé pour indiquer l’ensemble des politiques, des technologies et des directives institutionnelles qui facilitent la disponibilité et l’accès à la donnée géographique [57]. Les *IDG* fournissent un socle pour la découverte, l’évaluation et le traitement des données géographiques qui est mis à disposition des utilisateurs et des fournisseurs de données.

Dans le cadre de cette thèse, nous nous intéresserons essentiellement aux composants suivants : métadonnées et catalogues de métadonnées, leurs standards et les services qui y sont associés.

Les infrastructures de données géographiques sont conçues pour répondre, entre autres, au besoin des utilisateurs d’accéder à des données géographiques provenant de

différentes sources et de les utiliser pour soutenir leurs prises de décisions. [56] préconise que l'utilisation de normes communes devrait augmenter les ressources totales disponibles pour la création d'infrastructures de données spatiales. Il identifie également les métadonnées standardisées et les catalogues de données géographiques comme des éléments clés pour soutenir la découverte de données.

[28] définit un catalogue de données comme une collection de records de métadonnées combinée à des outils de gestion et de recherche de données. Les catalogues de données sont un moyen populaire pour découvrir des données géographiques [49]. Le standard CSW (*Catalog Service for the Web*) permet la mise à disposition sur le Web de l'information des catalogues géographiques. Il se base sur des records sous format XML et nécessite de faire des requêtes au service pour accéder aux records de métadonnées [54]. Les différentes opérations qu'offrent le standard pour qu'un catalogue soit conforme sont :

- **GetCapabilities** : permet à un utilisateur de récupérer les métadonnées de service d'un serveur.
- **DescribeRecord** : permet à l'utilisateur de découvrir les éléments du modèle utilisé par le service du catalogue.
- **GetRecords** : permet d'extraire les records et envoie leurs identifiants.
- **GetRecordById** : permet d'extraire un record correspondant à un certain identifiant.

Plus récemment, plusieurs outils implémentent et améliorent le standard CSW. Ces améliorations ont été nécessaires pour prendre en compte les nouveaux formats d'échange de données sur le Web (JSON, JSON-LD, HTML, etc.) et aussi pour prendre en compte les principes FAIR pour les données de recherche [74]. Selon [48] cela pourrait permettre une meilleure intégration des IDG avec des outils de recherche d'information sur le Web et la découverte des données géographiques à travers des moteurs de recherche comme Bing ou Google. C'est pourquoi l'OGC a commencé à publier de nouveaux standards nommés OGC APIs [59, 60, 61, 62, 63] à partir de 2020.

Plusieurs outils permettent de créer des catalogues géographiques tout en respectant les standards de l'OGC. On peut notamment citer GeoNetwork, CKAN ou encore

OpenDataSoft. Ces derniers permettent de publier, éditer et rechercher des métadonnées de jeux de données géographiques [47]. Ils fournissent aussi des outils nécessaires pour créer un moteur de recherche pour les catalogues géographiques en se basant sur des recherches sous la forme de mots-clés donnés par l'utilisateur ou grâce à des filtres prédéfinis comme le type de ressources, les organisations, etc.

D'autres plate-formes de gestion des données et de publication de données permettent de partager des jeux de données en ligne. Toutes ne sont pas compatibles avec les principes et les standards OGC et certaines sont dédiées à des domaines d'applications spécifiques. [2] présente et compare plusieurs plateformes de gestion de données. Au niveau européen, la plateforme data.europa.eu utilise CKAN munis d'extensions qui ont été développées spécifiquement pour l'application de recherche en ligne [47]. En France, des catalogues comme Sandre Atlas catalogue, dédié aux données de référence sur l'eau, utilisent GeoNetwork.

Dans [32], les auteurs ont effectué une analyse approfondie des différents moteurs de recherche utilisés dans différents portails de données et ont constaté que ces derniers sont basés sur une recherche verticale en texte intégral (*full-text search*) associée à des filtres sur certains champs des métadonnées.

### 2.2.2 Métadonnées géographiques

De manière littérale, "métadonnée" signifie "donnée à propos de données". Ainsi, les métadonnées géographiques sont toutes les données qui accompagnent les jeux de données pour les définir ou les décrire. Dans son sens le plus large, les métadonnées peuvent être fournies sur un support papier ou électronique. Dans leur version informatique, les métadonnées peuvent se présenter sous la forme d'un simple fichier PDF contenant un texte détaillant le contenu des données, ou encore sous forme de données structurées. Cette dernière présente l'avantage de permettre un traitement automatique des métadonnées. Dans la suite de ce manuscrit de thèse, on définira les métadonnées géographiques comme des données structurées accompagnant des jeux de données géographiques.

Les métadonnées géographiques jouent un rôle clé dans la découverte et la réutilisation des jeux de données géographiques [56]. En effet, les métadonnées associées aux jeux de données géographiques sont indexées par les catalogues géographiques pour permettre à l'utilisateur de retrouver ou télécharger ces derniers. Ainsi, pour assurer un minimum d'interopérabilité entre les différents jeux de données, plusieurs standards

TABLE 2.1 – Liste des principaux Standards ISO pour les métadonnées et catalogues géographiques

Standard	Titre
ISO 19110	Méthodologie pour les feature catalogues
ISO 19115	Métadonnées
ISO 19115-1	Métadonnées - Partie 1 : Les Fondamentales
ISO 19115-2	Métadonnées - Partie 2 : Extensions pour l'acquisition et le traitement
ISO 19115-3	Métadonnées - Partie 3 : Schéma d'implémentation XML des concepts Fondamentaux
ISO 19119	Services
ISO 19139	Métadonnées - Implémentation du modèle XML
ISO 19150-1	Ontologie - Partie 1 : Framework
ISO 19150-2	Ontologie - Partie 2 : Règles pour développer des ontologies OWL
ISO 19157	Qualité des données

de métadonnées et de catalogues géographiques ont été développés aux niveaux international et européen.

Au niveau international, l'organisation internationale de la standardisation (*ISO*) a créé un comité technique de standardisation, ISO/TC211, dédié aux systèmes d'information géographiques. Depuis le début des années 2000, ce comité a développé plus de 90 standards relatifs aux données, métadonnées et catalogues géographiques. La Table 2.1 présente une liste des principaux standards ISO/TC211 qui traitent des métadonnées et catalogues géographiques.

### 2.2.2.1 ISO 19110

L'ISO 19110 :2016, qui succède à la version 2005, définit la méthodologie pour la création des "Feature Catalogs". La norme spécifie comment les "Feature Types" peuvent être organisés dans un Feature Catalogue et présentés aux utilisateurs d'un ensemble de données géographiques.

### 2.2.2.2 ISO 19115

L'ISO 19115 :2003 est le premier standard international, publié en 2003 par l'organisation internationale de standardisation (*ISO*). Il est dédié à la description des jeux de données géographiques et des services associés à travers des métadonnées structurées. Cela comprend le contenu des jeux de données, la qualité des données, l'accès et les droits d'utilisation, etc. Dans un premier temps, la description de la qualité des données

faisait partie de l'ISO 19115 :2003 avant de donner lieu à un standard indépendant, à savoir, ISO 19157 [42].

L'objectif de l'ISO 19115 :2003 est de fournir une procédure claire pour la description des jeux de données et des services géographiques en ligne afin que les utilisateurs puissent déterminer si ces données leur seront utiles et comment y accéder. En établissant un ensemble commun de terminologie, de définitions et de procédures d'extension des métadonnées, la norme ISO 19115 :2003 favorise l'utilisation appropriée et l'extraction efficace des données géographiques.

La norme ISO 19115 :2003 a eu droit à une mise à jour majeure publiée à partir de 2009 dans le but de mieux tenir compte de l'utilisation croissante du Web pour la gestion et le partage des métadonnées et données géographiques, ainsi que pour ajouter de nombreux autres éléments de métadonnées et de nombreuses nouvelles catégories d'éléments de métadonnées, appelées listes de codes (*code lists*). C'est ainsi que furent publiées les normes ISO 19115-1, ISO 19115-2 et ISO 19115-3.

Le standard ISO 19115-1 :2014 est la nouvelle génération des standards ISO concernant les métadonnées géographiques. Il introduit de nouveaux éléments de métadonnées en plus de ceux déjà présents dans ISO 19115 :2003. De plus, c'est à partir de cette version que les éléments concernant la qualité des données ont été transférés dans la norme ISO 19157. La Table 2.2 présente les classes principales définies dans la norme ISO 19115-1.

Le standard ISO 19115-2 :2019 [40] établit les éléments de description des métadonnées de jeux de données géographiques d'imagerie et de données d'acquisition (i.e., données raster). Le standard réutilise les éléments de métadonnées présents dans ISO19115-1 et ajoute notamment les éléments de métadonnées pour décrire le processus d'acquisition des données. La Figure 2.2 montre la relation entre les éléments de l'ISO 19115-1 et l'ISO 19115-2. Il est important de remarquer que dans le Standard ISO 19115-2, la classe principale des métadonnées devient "MI\_Metadata" et qu'une nouvelle classe "MI\_AcquisitionInformation" a été ajoutée.

Finalement, le standard ISO 19115-3 [41] définit la nouvelle version d'encodage XML des métadonnées géographiques qui était définie auparavant dans le standard ISO 19139 :2007, et lui inclut l'encodage des nouveaux éléments de métadonnées introduits dans les nouvelles versions de ISO 19115-1 et ISO 19115-2.

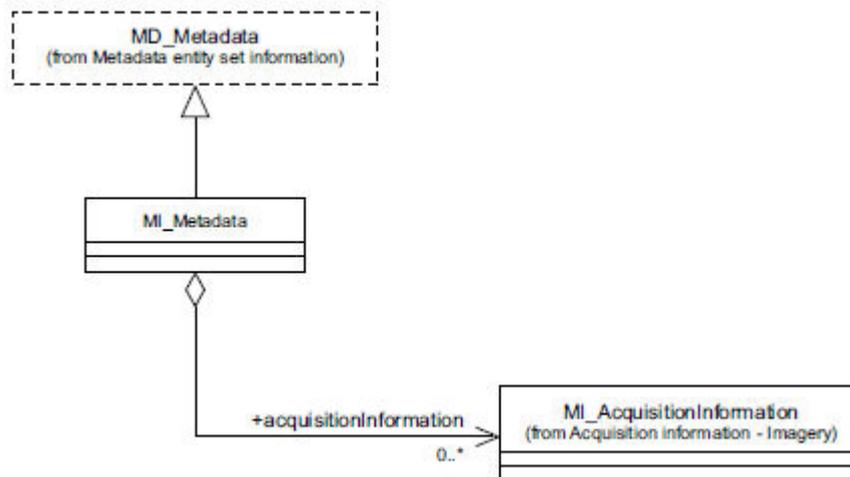


FIGURE 2.2 – Relation entre ISO 19115-1 et ISO 19115-2

TABLE 2.2 – Liste des principales classes du standard ISO 19115-1

Classe de Métadonnées	Description
MD_Metadata	Rassemble les éléments pour la gestion de la fiche des métadonnées comme la date d'édition de la fiche, l'agent en charge, etc.
MD_Identification	Rassemble les éléments nécessaire à l'identification d'un jeu de données tels que le titre, résumé, etc.
MD_Constraints	Représente les contraintes d'utilisation et de diffusion autour du jeu de données comme une licence d'utilisation, copyright, etc.
MD_Distribution	Décrit les modalités de distributions du jeu de données, à savoir, le format des données, moyen d'accès en ligne, etc.
MD_MaintenanceInformation	Spécifie, entre autres, la fréquence de mise à jour des données.
MD_ReferenceSystem	Décrit le système de référence spatiale utilisé dans le jeu de données
MD_SpatialRepresentation	Détaille les représentations vectorielles ou matricielles des données ainsi que la résolution utilisée, etc.
MD_ContentInformation	Précise le contenu avec des caractéristiques plus techniques, notamment quand il s'agit de jeux de données raster
LI_Lineage	Décrit la source du jeu de données et le processus utilisé pour la production des données
DQ_DataQuality	Rassemble les informations concernant la qualité des données comme définie dans la norme ISO 19157 :2013

[9] précise que la famille de normes ISO 19115 contient toutes les métadonnées pour la description des ensembles de données géographiques, que ça soit des données vectorielles ou d'imagerie, et que les éléments de qualité ont été déplacés dans la norme [42]. À partir des normes décrivant un jeu de données, alors que la description des métadonnées résume le contexte spatio-temporel des jeux de données géographiques, les métadonnées de qualité indiquent dans quelle mesure les données géographiques correspondent au monde réel, conformément aux spécifications des données.

### 2.2.2.3 ISO 19119

Le standard ISO 19119 :2005 [36] identifie et définit les modèles d'architecture pour les interfaces de service utilisées pour les informations géographiques, définit sa relation avec les systèmes de données ouverts, présente une taxonomie des services géographiques et une liste d'exemples de services géographiques placés dans la taxonomie des services. Elle indique également comment créer une spécification de service indépendamment de la plateforme utilisée, comment dériver des spécifications de service conformes et spécifiques à la plateforme, et fournit des lignes directrices pour la sélection et la spécification des services géographiques, tant du point de vue de la neutralité que de la spécificité de la plateforme.

La norme ISO 19119 :2005 a été révisée et a donné lieu à la norme ISO 19119 :2016. La version 2016 définit les exigences relatives à la création de spécifications de services neutres et spécifiques à une plateforme services, afin de permettre la spécification d'un service indépendamment d'une ou de plusieurs plateformes informatiques distribuées sous-jacentes. Le modèle de métadonnées de service et les exigences ont été déplacés vers la norme ISO 19115-1 :2014. Cette nouvelle version ne parle donc qu'en termes très généraux sur la nécessité des métadonnées de service.

Les spécifications ou les règles de mise en œuvre des métadonnées INSPIRE et les "directives techniques INSPIRE Metadata Technical Guidelines" sont basées sur les normes ISO 19115 et ISO 19119 :2005 pour la description des métadonnées pour les ensembles de données, les séries d'ensembles de données et les services.

### 2.2.2.4 ISO 19139

ISO/TS 19139 :2007 [37] définit l'encodage XML des métadonnées géographiques (gmd), une implémentation XML dérivée de la norme ISO 19115. Les métadonnées géo-

graphiques sont représentées dans la norme ISO 19115 comme un ensemble de paquets UML contenant une ou plusieurs classes UML. L'ISO 19115 fournit une vue universelle, indépendante de l'encodage, des métadonnées géographiques alors que l'ISO/TS 19139 fournit l'encodage de l'ISO 19115 sous format XML conforme aux règles décrites dans l'ISO 19118 (Encodage).

L'ISO/TS 19139 est actuellement en cours de révision et sera remplacée par l'ISO/NP TS 19139-1. Là encore, l'encodage XML spécifique aux métadonnées sera supprimé et fait désormais partie de l'ISO/TS 19115-3 : 2016 qui définit une implémentation XML intégrée des normes ISO 19115-1 et ISO 19115-2 (pour les données raster).

#### **2.2.2.5 ISO 19150**

L'ISO/TS 19150-1 :2012 [38] définit le cadre de l'interopérabilité sémantique de l'information géographique. Ce cadre définit un modèle de haut niveau des composants nécessaires pour traiter la sémantique dans les normes ISO d'information géographique avec l'utilisation d'ontologies.

La norme ISO 19150-2 :2015 [39] définit des règles et des lignes directrices pour le développement d'ontologies afin de soutenir mieux l'interopérabilité des informations géographiques sur le Web sémantique à travers le Web Ontology Language (*OWL*). Il définit la conversion des éléments de modélisation de la vue statique UML utilisés dans les normes ISO en OWL. Il définit en outre les règles de conversion pour la description des schémas d'application basés sur le modèle général des fonctionnalités définies dans la norme ISO 19109 en OWL.

#### **2.2.2.6 Directive INSPIRE**

Au niveau européen, la directive INSPIRE [21] introduite en 2007 par la Commission européenne, vise à créer une infrastructure de données géographiques de l'Union européenne. Cette infrastructure européenne de données géographiques devrait permettre le partage d'informations spatiales environnementales entre les organisations du secteur public, facilitera l'accès du public aux informations spatiales dans toute l'Europe et aidera à l'élaboration de politiques par-delà les frontières.

La directive INSPIRE souligne également les priorités pour le développement de l'infrastructure européenne de données géographiques : "les infrastructures d'informa-

tion géographique dans les États membres devraient être conçues de manière à ce que [...] il soit facile de découvrir les données spatiales disponibles, d'évaluer leur adéquation à la finalité et de connaître les conditions applicables à leur utilisation."

INSPIRE est basé sur les infrastructures d'information géographique établies et exploitées par les États membres de l'Union européenne. La directive aborde 34 thématiques de données géographiques nécessaires aux applications environnementales.

Pour rendre les sources d'informations thématiques provenant des 28 états membres de l'UE interopérables au niveau des métadonnées, INSPIRE a établi, sous la forme d'une législation européenne, des règles de mise en œuvre des métadonnées et des lignes directrices techniques pour les métadonnées.

Les "Règles de Mise en œuvre des Métadonnées INSPIRE" [19] (*INSPIRE Metadata Implementing Rules*) sont une réglementation de l'UE qui établit les obligations en vertu desquelles les organismes du secteur public de l'UE doivent publier des métadonnées descriptives sur les ensembles (séries) et services de données géographiques. Il décrit un ensemble de vingt et un éléments de métadonnées à utiliser pour décrire les jeux de données géographiques, les séries de données et les services. Cet ensemble d'éléments de métadonnées INSPIRE a été complété par cinq éléments supplémentaires en 2010 et un autre en 2013.

Afin d'harmoniser davantage la mise en œuvre de ces obligations, les "lignes directrices techniques INSPIRE pour les métadonnées" [20] (*INSPIRE Metadata Technical Guidelines*), non obligatoires, ont été définies. Ces lignes directrices techniques montrent comment les éléments de métadonnées du règlement correspondent aux éléments de métadonnées des normes ISO 19115 :2003 et ISO 19119 :2007.

La spécification contient également des exemples de la manière dont les métadonnées peuvent être transformées en XML (ISO 19139). Bien que les lignes directrices techniques ne soient pas contraignantes, il est fortement conseillé de les suivre. Dans le cas contraire, il est très difficile de démontrer que les métadonnées fournies dans un modèle/format différent sont conformes aux règles de mise en œuvre des métadonnées INSPIRE.

De plus, la directive INSPIRE, à travers une loi européenne promulguée en 2007, rend l'utilisation des standards ISO 19115 :2003 et ISO 19139 :2007 obligatoires pour toutes les entités publiques productrices de données en relation avec les thématiques

environnementales dans tous les pays membres de l'UE et notamment en France. Cela rend les versions de 2003 et de 2007 de ces standards incontournables dans la mesure où les entités publiques sont parmi les plus grands producteurs de données et sont généralement responsables de produire des données de référence.

### 2.2.3 Remplissage des métadonnées

Les métadonnées géographiques sont d'une grande importance dans le processus de recherche de jeux de données géographiques sur le Web. La création et l'édition des métadonnées restent une tâche longue et fastidieuse. Bien que plusieurs outils proposent des moyens pour créer des métadonnées compatibles avec les standards OGC et ISO, [81] précise qu'une grande majorité des métadonnées sont créées manuellement. Certaines alternatives ont été développées en se basant sur des communautés de crowdsourcing comme expliqué par [46] qui propose un prototype où des volontaires peuvent créer des métadonnées en partageant leurs notes et en analysant les mots-clés utilisés dans la recherche de l'utilisateur.

En regardant les métadonnées existantes de plus près, on peut se rendre compte que le remplissage des métadonnées manuellement crée des hétérogénéités malgré l'utilisation des standards vus précédemment. Si les standards rendent l'utilisation d'au moins un mot-clé obligatoire, il n'y a aucune recommandation sur la manière de le remplir. Par exemple, le champ des mots-clés est un champ à texte libre qui permet à un producteur de données de décrire son jeu de données avec une totale liberté. C'est pourquoi plusieurs travaux de recherche proposent différentes approches pour enrichir les champs de mots-clés dans les métadonnées en utilisant des vocabulaires contrôlés [83], ou encore en intégrant des mesures de similarité sémantiques [12].

Minimiser les hétérogénéités des mots-clés dans les métadonnées et d'autant plus importants que [11] a souligné que la recherche des jeux de données est principalement basée sur des mots-clés dans les métadonnées disponibles et que les résultats de la recherche sont calculés sur la base de filtres et d'expériences qui ont fonctionné pour la recherche d'information sur le Web. C'est particulièrement vrai en ce qui concerne la recherche de jeux de données géographiques.

De plus, [44] se sont intéressés au comportement des utilisateurs qui recherchent des jeux de données au Royaume-Uni. Pour ce faire, ils ont mené une étude sur les journaux de recherche de quatre portails gouvernementaux de données ouvertes. Ils ont découvert que les requêtes de jeux de données sont généralement décrites en uti-

lisant des limites et des restrictions concernant l'emplacement, la temporalité, le type de données spécifiques et/ou la granularité spécifique. Ils suggèrent que les critères les plus importants sont la couverture temporelle et spatiale avec une granularité variable. Ces résultats sont proches des critères identifiés par [70] et [50] pour les données géographiques, à savoir, la spatio-temporalité et les thématiques des données.

Le manque de métadonnées est un frein essentiel à la découverte et le partage des données géographiques et les rend non compatibles avec les bonnes pratiques recommandées pour les données FAIR [8]. Un autre point important pour la découverte des données géographiques et la qualité des métadonnées elles-mêmes. Ainsi, [82] ont analysé les métadonnées disponibles sur le portail de données espagnoles. Ils ont constaté que tous les fichiers de métadonnées portent sur la description des données, sans que les utilisateurs ne disposent d'informations sur la qualité des métadonnées. Ainsi, pour 3640 fichiers de métadonnées géographiques, 97% ont une bonne boîte de délimitation et 41% présentent des contradictions topologiques dues à des erreurs de saisie humaine. C'est pourquoi ils proposent une méthode pour mesurer la qualité des métadonnées géographiques.

De plus, dans [17], les auteurs ont étudié la qualité des thésaurus et vocabulaires contrôlés utilisés dans les métadonnées géographiques. Pour cela, ils se sont intéressés aux thèmes présents dans les fichiers de métadonnées disponibles dans le portail de données espagnoles. Ils ont constaté que même si l'utilisation d'au moins un thème provenant du vocabulaire GEMET est obligatoire, seuls 85% d'entre eux en utilisaient et que 56% d'entre eux utilisaient d'autres thésaurus ou vocabulaires contrôlés.

Sur son portail, data.europa.eu [47] propose aussi un indicateur pour évaluer la qualité des métadonnées qui y sont indexées et se base pour ça sur l'algorithme TLSH [64]<sup>1</sup>.

## 2.3 Web sémantique et recherche d'information sur le Web

### 2.3.1 Moteurs de recherche sur le Web

La recherche d'information (IR) est le domaine de recherche qui s'intéresse à la recherche d'une information dans un corpus donné. Le corpus peut correspondre à des documents, une ou plusieurs bases de données, etc. qui sont décrites par des métadonnées ou un contenu associé. La recherche d'information sur le Web se base essentiel-

---

1. <https://gitlab.com/european-data-portal/metrics/edp-metrics-dataset-similarities/-/tree/master/src/main/java/io/piveau/metrics/similarities>

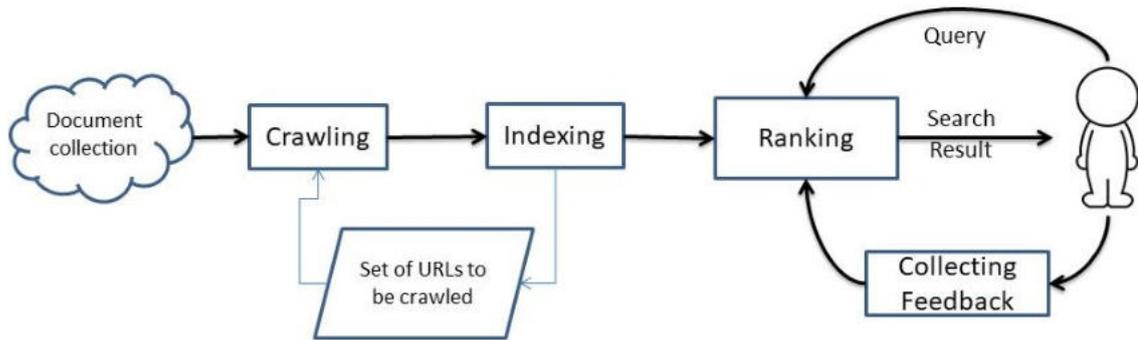


FIGURE 2.3 – Composants d'un moteur de recherche classique

lement sur des moteurs de recherche qui permettent d'indexer des documents et des pages Web et permet ensuite à un utilisateur d'effectuer une recherche, le plus souvent, sous la forme de mots-clés.

La Figure 2.3 présente le processus de recherche d'un moteur de recherche classique. A partir d'une collection de documents Web, le moteur de recherche parcourt et indexe celles-ci. Pour chaque page parcourue, il l'indexe en fonction de différents critères, il identifie les liens hypertextes qui y sont contenus pour les parcourir à leur tour et ainsi de suite. Cette opération est exécutée de manière automatique et continue par le moteur de recherche pour indexer un maximum de pages Web.

Ensuite, quand l'utilisateur veut retrouver une page, il fournit en entrée une requête qui va être utilisée et intégrée dans un modèle de pertinence (*Ranking Model*) pour retrouver les documents les plus pertinents par rapport à cette requête. Les résultats de la recherche sont ensuite renvoyés à l'utilisateur qui a la possibilité de fournir un retour (*feedback*) qui pourra être pris en compte pour améliorer le modèle de pertinence.

Le modèle de pertinence (la fonction de pertinence) est un élément central d'un moteur de recherche, qui va associer à une paire  $(R, D)$  un score qui exprime la pertinence du document  $D$  par rapport à la requête  $R$ . Par exemple, l'algorithme PageRank [7] a permis au moteur de recherche Google de se démarquer des autres moteurs de recherche concurrents et a initié l'ascension fulgurante de ce dernier en tant que géant du Web. Il prend en compte d'autres éléments que les mots contenus dans les pages comme par exemple les liens entre pages ou encore les comportements de navigation des usagers du moteur.

TF-IDF [77] est un modèle de pertinence très répandu qui permet d'associer un

score pour un mot-clé par rapport à un corpus de documents. Pour calculer la pertinence, ce dernier se base sur la fréquence du mot recherché (*TF - term frequency*) dans les documents indexés pour classer les documents [87]. Le modèle utilise aussi la fréquence inverse des documents (*IDF - inverse document frequency*) pour ignorer les mots trop fréquents dans le corpus de documents (c.-à-d. le, la, un, etc.) et qui ne sont pas assez discriminants. [11] précise qu'une grande majorité des moteurs de recherche utilisés par les catalogues géographiques utilisent des modèles de pertinence semblables à TF-IDF ou des variantes [53, 68]. Il indique aussi que cela a un grand impact sur les résultats de recherche des jeux de données géographiques et que ces modèles ne capturent pas toute la complexité qu'il peut y avoir dans les métadonnées associées aux jeux de données.

La recherche à facette (*Faceted search*) [80] est une autre technique de recherche d'information qui est populaire dans les portails Web et notamment dans les catalogues géographiques [51]. Cette technologie se base sur des facettes qui sont des critères à partir desquels les données ont été classifiées au préalable. [84] présente plusieurs techniques et méthodes pour la mise en place et l'amélioration de la recherche à facette en fonction du domaine d'application.

La recherche à facettes a permis de grandement améliorer la recherche d'objets plus complexes sur le Web et notamment les données géographiques. Dans [32], les auteurs ont étudié plusieurs catalogues géographiques qui proposent à leurs utilisateurs une recherche à facettes. Ils ont tout d'abord remarqué que la barre de recherche suggérait aux utilisateurs des mots-clés pour les aider, et que les sélectionner améliore les résultats obtenus. La majorité des catalogues étudiés proposent plusieurs facettes, à savoir, les mots-clés, les thèmes, les licences d'utilisation, les producteurs de données, les types de données, etc. Ces facettes correspondent généralement aux champs de métadonnées présents dans les standards décrits dans la Section 2.2.2. Sur la Figure 1.3, on peut voir qu'à gauche de l'interface utilisateur, plusieurs facettes sont utilisées dans le catalogue du Sandre.

### 2.3.1.1 Mesures de similarité

Les mesures de similarité jouent un rôle très important dans la recherche d'information sur le Web. En effet, de manière générale, un utilisateur fournit à un moteur de recherche une chaîne de caractères qui sera comparée aux documents qui y sont indexés. Différentes manières existent pour calculer la distance ou la similarité entre un mot ou un ensemble de mots et un corpus de documents.

### Mesures de similarité textuelles

Les mesures de similarité textuelles sont des mesures de similarité permettant de comparer des mots ou des ensembles de mots en se basant sur les chaînes de caractères qui les composent. Dans ce cas, le sens des mots à comparer n'est pas pris en compte.

La mesure de similarité textuelle la plus simple est la distance de Levenshtein [86] (c.-à-d. distance d'édition) qui se base sur le nombre minimal d'éditions (c.-à-d. substitutions, insertions et suppression) pour transformer une chaîne de caractère en une autre.

D'autres mesures de similarité textuelles permettent de comparer deux chaînes de caractères en se basant sur un vecteur binaire associé à chaque chaîne de caractère. Ainsi si  $X$  et  $Y$  sont deux vecteurs associés à deux chaînes de caractères, le coefficient de Jaccard [43] entre  $X$  et  $Y$  par :

$$jaccard(X, Y) = \frac{|X \cup Y|}{|X \cap Y|}$$

La valeur du coefficient de Jaccard est un réel compris entre 0 et 1 où 0 indique qu'il n'existe aucun recouvrement entre  $X$  et  $Y$  et 1 indique un recouvrement parfait entre  $X$  et  $Y$ .

Une autre mesure de similarité textuelle populaire est la mesure Cosinus [71]. Pour deux vecteurs  $X$  et  $Y$  la mesure entre  $X$  et  $Y$  est défini par :

$$cosinus(X, Y) = \frac{|X \cup Y|}{\sqrt{|X||Y|}}$$

Le résultat obtenu est un réel compris entre 0 et 1 où 0 indique qu'il n'existe aucun recouvrement entre  $X$  et  $Y$  et 1 indique un recouvrement parfait entre  $X$  et  $Y$ .

### Mesures de similarité sémantiques

À la différence des mesures de similarité textuelles, les mesures de similarité sémantique ne se contentent pas de comparer les chaînes de caractères, mais comparent le sens des mots. Pour cela, une mesure de similarité sémantique se base sur une ontologie, un thésaurus, un corpus de documents, etc.

La mesure de similarité sémantique la plus simple est basée sur le plus court chemin

entre deux concepts d'un même thésaurus. Pour deux concepts  $C_1$  et  $C_2$  appartenant à un même thésaurus :

$$sim_{pcc}(C_1, C_2) = 2 \times Max(C_1, C_2) - pcc(C_1, C_2)$$

où  $Max(C_1, C_2)$  est le chemin le plus long entre  $C_1$  et  $C_2$  dans le thésaurus utilisé et  $pcc(C_1, C_2)$  le plus court chemin entre  $C_1$  et  $C_2$  dans le thésaurus utilisé.

Il est possible d'adapter cette mesure de similarité sémantique en associant un poids lors du calcul du plus court chemin  $pcc(C_1, C_2)$  et du plus long chemin  $Max(C_1, C_2)$  pour obtenir une mesure de similarité sémantique avec des liens pondérés.

C'est sur ce principe qu'a été définie la mesure de similarité *Class* qui est une implémentation basée sur Wikidata. Comme Wikidata n'est pas un thésaurus, les calculs se font uniquement sur un ensemble de propriétés qui représentent des liens hiérarchiques, à savoir, *instanceOf* (P31), *subClassOf* (P279), *positionHeld* (P39) et *occupation* (P106). La mesure *Class* permet de calculer la similarité entre deux concepts Wikidata si ces deux derniers sont reliés par un chemin constitué par les quatre liens hiérarchiques considérés auparavant.

Une autre mesure de similarité sémantique répandue est la mesure de Wu-Palmer [85]. Pour deux concepts  $C_1$  et  $C_2$ , on note  $C$  le concept qui est le plus proche parent commun entre  $C_1$  et  $C_2$ . Alors

$$sim_{wp} = \frac{2 \times d}{d_1 + d_2 + 2 \times d}$$

où  $d_1$  est la distance entre  $C_1$  et  $C$ ,  $d_2$  la distance entre  $C_2$  et  $C$  et  $d$  la distance entre  $C$  et la racine du thésaurus.

La mesure de Wu-Palmer a inspiré plusieurs travaux de recherche pour développer des mesures de similarité sémantiques [75, 25, 73] basées sur le comptage du nombre d'arêtes qui séparent deux concepts qui appartiennent à un même thésaurus ou qui appartiennent à la même ontologie [45].

### Mesures de similarité multi-critères

Historiquement, les méthodes d'analyse multicritère (*MCDM*) [88] visent à résoudre des problèmes d'aide à la décision multicritères. Leur but est de trouver une ou plusieurs solutions optimales en fonction de plusieurs critères prédéfinis auxquels on associe un poids qui représente le coût ou l'utilité d'un certain critère.

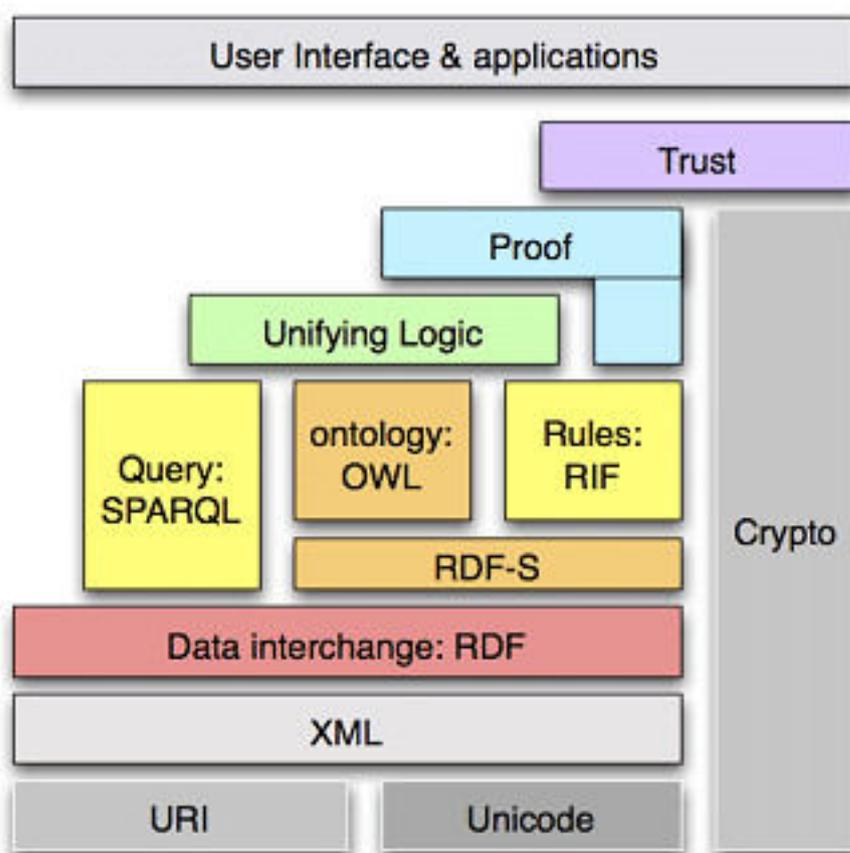


FIGURE 2.4 – Pyramide des langages et standards du Web sémantique tels que présentés par Tim Berners-Lee en 2006

Parmi ces méthodes, l'algorithme TOPSIS [52] est un algorithme qui permet de comparer plusieurs alternatives pour retrouver la meilleure possible. Chaque alternative est évaluée par rapport aux critères définis par l'utilisateur et aux poids associés à chacun des critères. L'algorithme TOPSIS et les mesures de similarité sémantiques ou textuelles ont été combinés pour créer une mesure de similarité multicritères qui permet la recherche d'information dans un ensemble d'images historiques [26].

## 2.3.2 Web sémantique et Graphes de connaissances

### 2.3.2.1 Les bases du Web sémantique

Le Web sémantique est une extension du Web classique introduite par Tim Berners-Lee en 1998 et standardisée par le World Wide Web Consortium (W3C). Dans son article fondateur (Berners-Lee, 1998), Tim Berners-Lee expose son ambition de transformer le Web d'un réseau de documents, qui sont essentiellement destinés aux humains, en un réseau de données interconnectées lisibles par les machines.

La Figure 2.4 présente les différents standards établis par le W3C et qui sont utili-

sés sur le Web sémantique. Je présente ici les standards directement liés à mon sujet de thèse. Le modèle RDF qui est la base du Web sémantique représente un domaine sous forme de triplets de la forme (Sujet, Propriété, Objet) où le Sujet est la ressource concernée par la représentation, le Prédicat une propriété qui est applicable à cette ressource et l'objet représente une valeur ou une autre ressource. Dans RDF, chaque ressource est identifiée par un identifiant unique (*Unique Resource Identifier - URI*). Un jeu de données RDF se présente donc sous la forme de triplets qui forment un graphe de données.

De plus, il existe deux vocabulaires permettant de représenter et décrire des ontologies et qui sont des recommandations du W3C : RDFS (i.e RDF Schema ) et OWL (i.e., Web Ontology Language). Ces standards permettent de représenter des connaissances en respectant le modèle de données RDF. Les ontologies sont utilisées pour décrire les liens (i.e prédicats) qui existent entre le sujet et l'objet d'un triplet RDF. De plus, comme les ontologies OWL et RDFS sont elles-mêmes décrites en RDF, les propriétés qu'elles contiennent sont définies par des URI, ce qui permet de connaître la définition exacte d'un lien qui existe entre des données et de parfois inférer de nouvelles connaissances [33]. Les ontologies décrites par OWL et RDFS sont utilisées pour décrire des données qui sont rassemblées sous la forme de jeux de données RDF publiées sur le Web.

Le langage SPARQL permet d'exécuter des requêtes dans un jeu de données RDF pour en extraire un sous-ensemble de triplets qui correspond au besoin de l'utilisateur. Il existe plusieurs extensions de SPARQL qui permettent d'outrepasser les limitations du langage quand il faut requêter des données plus spécialisées. Ainsi, le langage de requêtes GeoSPARQL est une extension de SPARQL qui permet de représenter et requêter des données géographiques RDF. C'est aussi un standard de l'OGC.

Les données liées (*Linked data - LD*) sont la concrétisation des principes du Web sémantique. La LOD [3] (*Linked Open Data*) représente la partie ouverte des données liées et rassemble un grand nombre de jeux de données RDF comme on peut le constater dans la Figure 2.5 où chaque noeud représente un jeu de données et chaque couleur représente le domaine des données.

### 2.3.2.2 Les graphes de connaissances

Le terme "Knowledge Graph" ou graphe de connaissances a été popularisé par Google<sup>2</sup> en 2012. Bien que la communauté scientifique ne s'accorde pas sur une dé-

---

2. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>

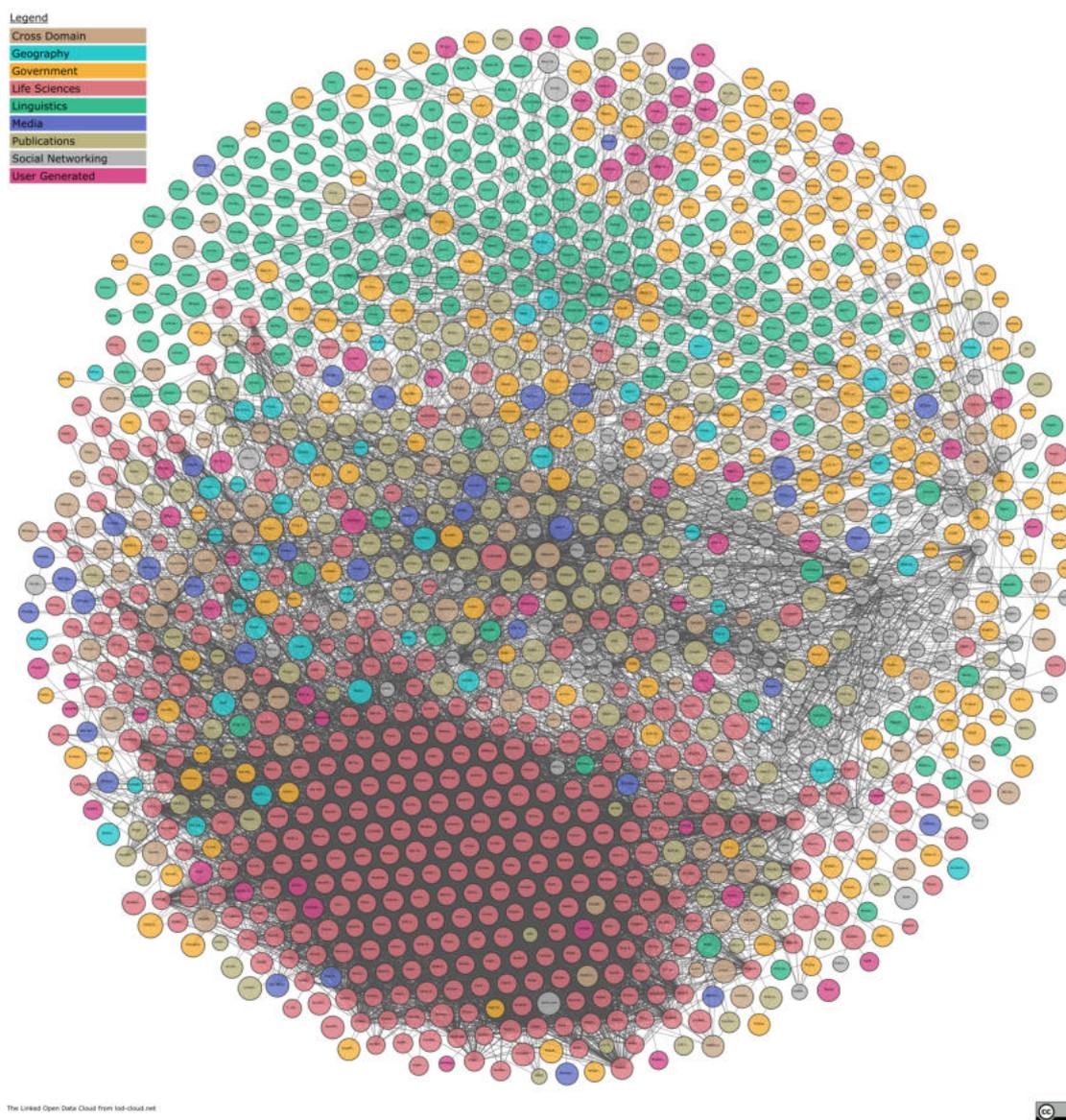


FIGURE 2.5 – Le LOD cloud

finition unique, les différentes définitions disponibles pour les KG sont d'accord sur certaines propriétés qu'un KG devrait satisfaire [18] : (1) des données structurées sous la forme d'un graphe, (2) des données qui décrivent des entités du monde réel ou pas et (3) les relations qui existent entre ces entités.

Bien que le terme "Knowledge Graph" a été utilisé par Google, on peut remarquer que les bases de données RDF satisfont ces conditions et peuvent donc être considérées comme des Knowledge Graphs. Parmi ces graphes de connaissances, plusieurs sortent du lot. On peut notamment citer Wikidata et DBpedia qui sont open source et dont les données sont disponibles pour tous et peuvent être réutilisées.

DBpedia est un projet communautaire qui propose une version de l'encyclopédie Wikipedia au format du Web sémantique. De son côté, Wikidata est une initiative de la fondation Wikimedia qui vise à constituer une source de données objectives, fiables et multilingues qui pourront ensuite être utilisées dans tous les projets Wikimedia dont Wikipedia. Les données contenues dans ces deux KG portent essentiellement sur un savoir de sens commun (dates de naissance, films, célébrités, géographie, etc.). Wikidata et DBpedia sont au centre de plusieurs applications du Web sémantique comme DBpedia Spotlight [55] utilisé pour annoter des textes sémantiquement, enrichir un document [23], enrichir la requête d'un utilisateur [16, 15, 14], classifier des documents [24] ou encore utiliser un assistant vocal comme Alexa pour obtenir des réponses [29].

Devant les différentes applications possibles à l'utilisation des graphes de connaissance, plusieurs entreprises ont développé leurs propres graphes de connaissance adaptés à leurs besoins. Par exemple, le KG de Google s'intègre à son moteur de recherche qui affiche une information structurée centrée sur l'entité de la requête de l'utilisateur et permet à ce dernier de naviguer facilement vers d'autres entités qui sont en lien avec sa requête. Le Knowledge Graph de Google rassemble non seulement un très grand nombre de connaissances factuelles sur le monde réel et la vie de tous les jours (cinéma, histoire, sports, géographie, musique . . .), mais inclut aussi différents services (disponibilités et réservations dans des restaurants, achats en ligne . . .) [58]. De manière similaire, Amazon a développé un graphe de connaissance qui rassemble les produits disponibles dans sa market place et les relations qui existent entre eux et s'en sert pour proposer à ses clients des produits qui pourraient les intéresser.

Assez rapidement, plusieurs graphes de connaissance dédiés à des domaines d'applications spécifiques ont commencé à émerger. On peut par exemple citer le domaine de la santé où un KG dédié aux maladies existantes a été proposé par [69] et qui permet

de rassembler un grand nombre de maladies et de symptômes. Dans [1], l'auteur fournit une liste de graphes de connaissance dédiés à des domaines très variés.

C'est tout naturellement que la communauté SIG s'est aussi intéressée aux applications possibles du Web des données et des KG au domaine de l'information géographique. On peut notamment citer Ordnance Survey qui a transformé ses données administratives sous la forme de données liées [27] ou encore l'IGN qui a proposé la plateforme DataLift qui permet de transformer et de publier des données géographiques sous la forme de données liées [72, 31]. Dans [10] les auteurs proposent de créer un KG européen dédié aux données géographiques et soutiennent que cela permettrait de grandement diminuer les hétérogénéités des données géographiques au niveau européen.

Un principe important porté par le Web sémantique et le LOD est la connexion entités des différents jeux de données RDF. En effet, relier les entités RDF avec d'autres entités du Web des données est aussi important que publier les données [4]. C'est pourquoi, très rapidement, la problématique de comparer et d'identifier automatiquement des entités RDF s'est rapidement posée avec la grande quantité des données disponibles sur le Web des données. En effet, deux KG peuvent représenter la même entité de deux manières différentes.

La problématique de l'alignement des entités de deux KGs vise à identifier la même entité qui est représentée dans deux KGs différents [89]. Depuis quelques années, plusieurs approches basées sur les plongements de KG (*KG Embeddings - KGE*) ont été proposées pour résoudre cette problématique. Un KGE est le plongement d'un KG dans espace vectoriel où il est possible de calculer la similarité entre deux entités [76].

Il existe plusieurs modèles permettant de créer un KGE, comme par exemple, TransE [79] ou encore ComplEx [5]. Dans [22], les auteurs comparent plusieurs méthodes de KGE pour l'alignement d'entités. Ces modèles peuvent être adaptés pour mesurer la similarité sémantique entre concepts [13, 35, 34].

Une autre problématique intéressante autour de l'alignement des données RDF est la détection des liens erronés qui peuvent relier deux jeux de données RDF. En effet, il est fréquent de relier des entités appartenant à deux jeux de données différents grâce à la propriété "owl:sameAs" [65]. Néanmoins, dans la mesure où les données du LOD sont majoritairement produites de manière collaborative, il n'est pas rare que des erreurs se glissent parmi ces liens [30].

Plusieurs approches ont été développées pour détecter les liens d'identification erronés. Par exemple [67] propose une méthode basée sur l'utilisation de métriques de graphes pour détecter les liens erronés.

### 2.3.3 Standards des Métadonnées sur le Web

Plusieurs vocabulaires permettent la description, sous format RDF, des métadonnées des jeux de données géographiques et sont destinés à la diffusion sur le Web.

#### 2.3.3.1 DCAT

DCAT est un vocabulaire développé pour faciliter l'interopérabilité entre les jeux de données publiés sur le Web et compte à ce jour 3 versions. A l'origine, le vocabulaire DCAT a été développé et hébergé par le Digital Enterprise Research Institute (*DERI*), puis affiné par le groupe d'intérêt eGov, et enfin normalisé en 2014 par le groupe de travail Government Linked Data (*GLD*) et est ensuite devenu une recommandation du W3C. Une deuxième révision recommandée de DCAT, nommée DCAT 2, a été élaborée par le groupe de travail sur l'échange de données en réponse à un nouvel ensemble de cas d'utilisation et d'exigences recueillis à partir de retours d'expérience d'utilisateurs de la première version de DCAT, et des nouvelles applications qui n'avaient pas été prises en compte.

La première version de DCAT introduit quatre classes. Ces classes constituent le noyau de base permettant de décrire un catalogue, une distribution, un record et un jeu de données. Ces classes resteront inchangées dans les versions ultérieures de DCAT :

- Catalog : représente un catalogue de métadonnées
- Dataset : qui représente un jeu de données disponible dans un catalogue
- Distribution : représente un format accessible d'un jeu de données
- CatalogRecord : qui décrit une représentation d'un jeu de données dans un catalogue. Ainsi un même jeu de données peut avoir deux représentations dans deux catalogues différents

La Figure 2.6 représente le diagramme UML des classes définies par DCAT, les propriétés de chacune d'entre elles ainsi que les différentes propriétés qui relient ces

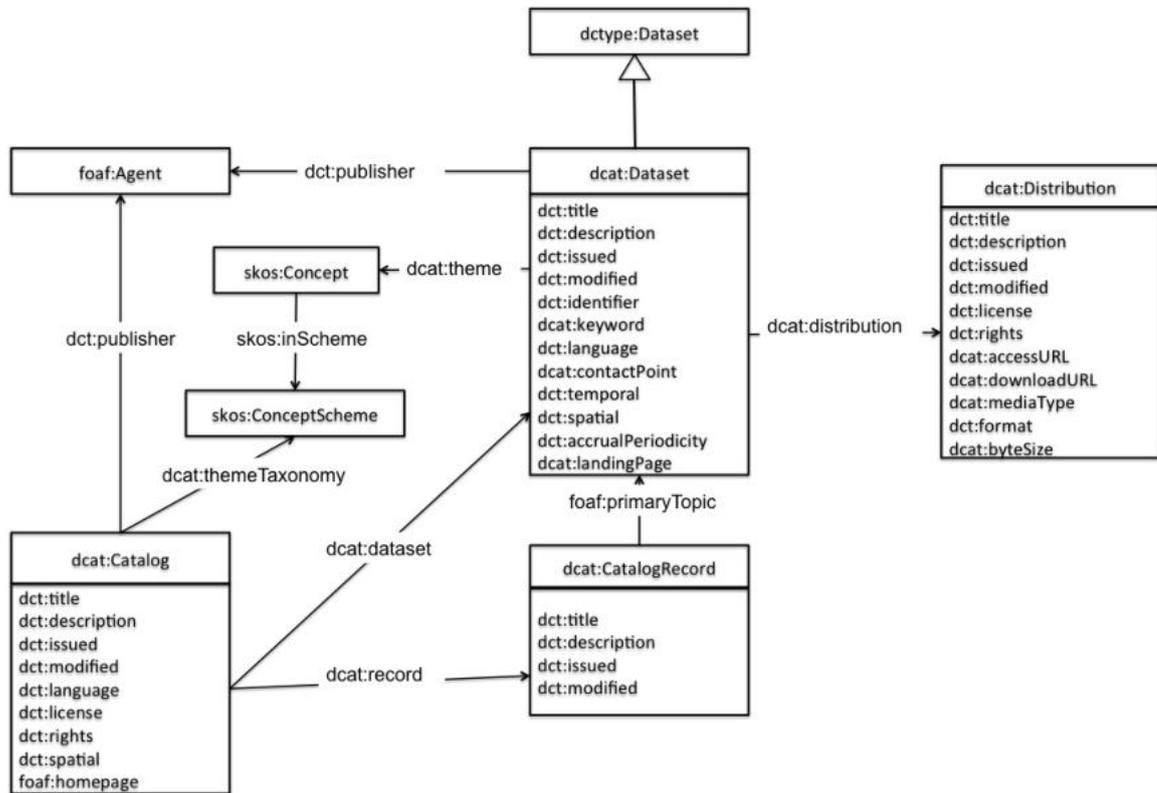


FIGURE 2.6 – Diagramme UML des classes DCAT

classes. On peut remarquer que DCAT reprend des classes de l'ontologie SKOS et notamment la classe `skos:Concept` qui permet de décrire un concept donné et la classe `skos:ConceptScheme` qui permet de décrire une taxonomie ou un vocabulaire contrôlé contenant des concepts.

DCAT 2, publié en 2020, apporte plusieurs modifications par rapport à DCAT en introduisant plusieurs nouvelles classes tout en gardant les anciennes. DCAT 2 définit notamment la classe «`dcat:Resource`» qui décrit toutes les ressources d'un catalogue et la classe «`dcat:DataService`» qui décrit les services de données tels que les services d'accès aux données. De plus, une classe «`dct:Location`» a été ajoutée à DCAT 2, accompagnée de trois propriétés «`locn:geometry`», «`dcat:bbox`» et «`dcat:centroid`» qui permettent de spécifier la couverture spatiale d'une ressource.

La dernière version en date de DCAT, DCAT 3, a été développée par le groupe de travail sur l'échange de données, en tenant compte de certains des cas d'utilisation et des demandes les plus pressants parmi ceux qui n'ont pas été pris en compte lors du cycle de normalisation précédent. DCAT 3 introduit la notion de séries de jeux de données, qui sont représentées par la classe «`dcat:DatasetSeries`», et qui correspond à un jeu de données représentant une collection de jeux de données, publiés séparément, mais qui partagent des caractéristiques communes qui les regroupent.

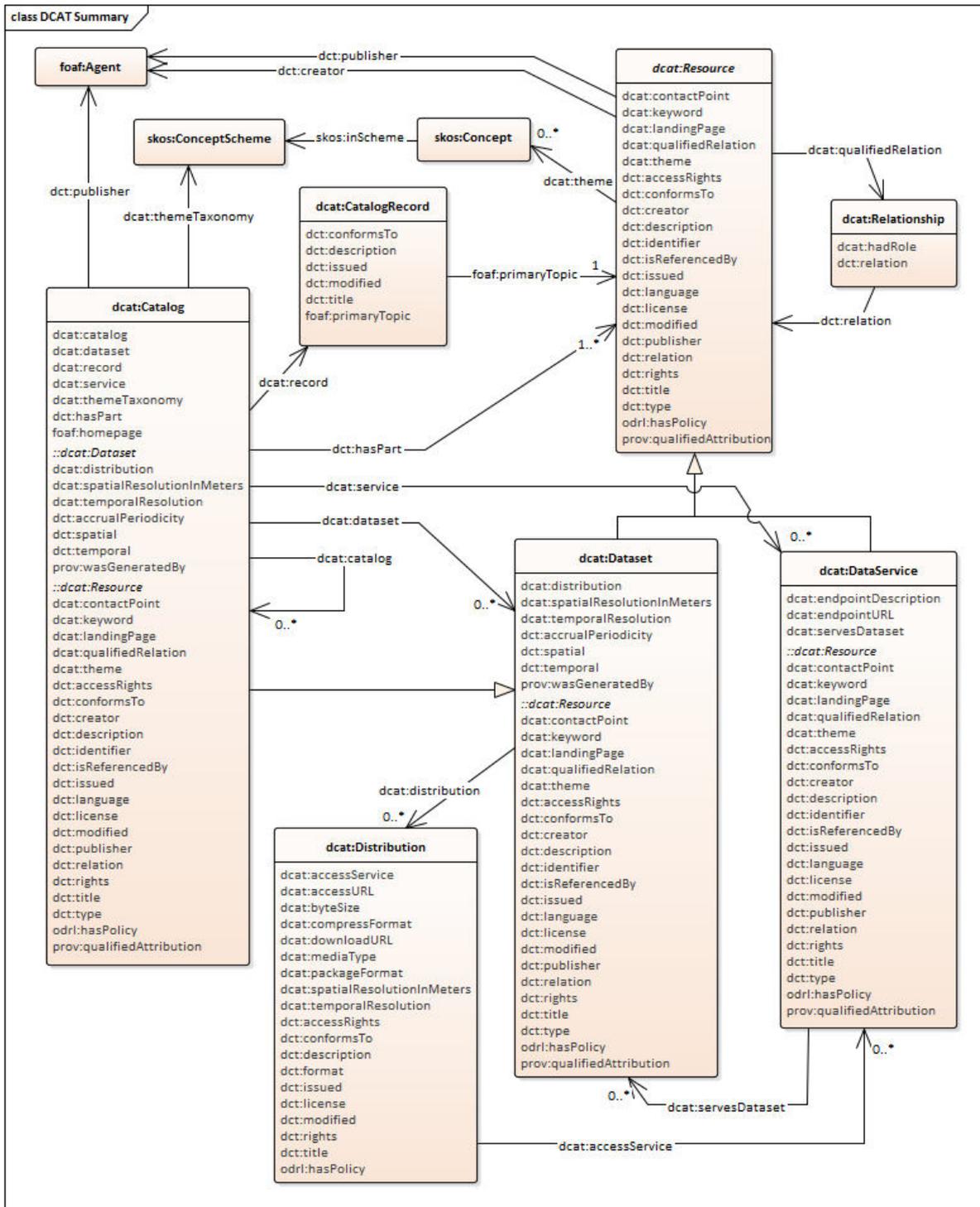


FIGURE 2.7 – Diagramme UML du standard DCAT 2

Mapping entre ISO 19115 et DCAT

The screenshot shows a dataset page for 'Scottish Road Accident Statistics' on data.gov.uk. Red annotations map DCAT properties to ISO 19115 classes. For example, 'dot:title (Dataset)' points to the dataset title, 'dot:description' points to the description, 'dot:accessURL' points to download links, and 'dot:issued' points to the 'Added to data.gov.uk' date. A table on the right lists the mapping between ISO 19115 and DCAT entities.

ISO 19115	DCAT
Service (catalog service)	Catalog
Metadata	Catalog record
Dataset	Dataset
Distribution / Online resource	Distribution
Originating controlled vocabulary	Concept scheme
Descriptive keyword / ISO topic category	Concept
Responsible party	Organization / person

Src : DCAT Application Profile for Data Portals in Europe, Nikolaos Loutas, Michiel De Keyzer, Stijn Goedertier

Src: [https://www.w3.org/2015/spatial/wiki/ISO\\_19115\\_-\\_DCAT\\_-\\_Schema.org\\_mapping](https://www.w3.org/2015/spatial/wiki/ISO_19115_-_DCAT_-_Schema.org_mapping)

FIGURE 2.8 – Alignement entre DCAT 2 et ISO 19115

A noter qu'il existe un alignement entre les standards DCAT 2 et ISO 19115, comme on peut le voir dans la Figure 2.8.

2.3.3.2 DCAT-AP

Toutes les versions de DCAT permettent la création de profils. Cela s'explique par la volonté de permettre à DCAT d'être adapté à des besoins spécifiques à certaines utilisations. Un profil DCAT est une spécification qui ajoute des contraintes supplémentaires à DCAT. Ces contraintes supplémentaires dans un profil peuvent inclure un nombre minimal de champs de métadonnées requis, de nouvelles classes ou propriétés non couvertes par DCAT, l'utilisation de vocabulaires comme valeurs acceptables pour des propriétés, etc.

C'est ainsi que le Profil d'Application DCAT pour les portails de données en Europe (i.e., DCAT-AP) a été développé par la Communauté européenne de l'interopérabilité sémantique (SEMIC) en 2015. DCAT-AP rend l'utilisation des classes `dcat:Dataset`, `dcat:Catalog` et `foaf:agent` obligatoire, tandis qu'il recommande l'utilisation des classes `dcat:Distribution` et `dct:LicenseDocument`. Les autres classes restent optionnelles. De plus, le standard précise pour chaque classe quelles propriétés sont obligatoires, lesquelles sont recommandées et lesquelles sont optionnelles.

Plusieurs pays européens (Italie, Norvège, Belgique, Suède, Allemagne) ont aussi

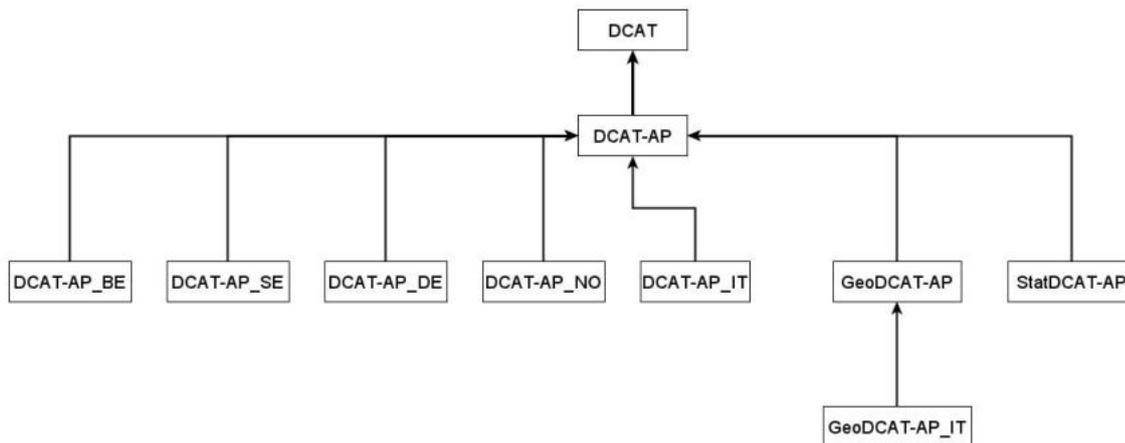


FIGURE 2.9 – Les profils DCAT

développé leurs propres profils DCAT qui ajoutent des contraintes supplémentaires à DCAT-AP. La Figure 2.9 présente les différents profils DCAT et les relations qui existent entre eux.

### 2.3.3.3 GeoDCAT-AP

GeoDCAT-AP [66] est une extension de DCAT-AP et de DCAT qui permet de rendre les métadonnées compatibles avec les standards ISO/OGC et notamment le standard ISO 19115. GeoDCAT-AP permet notamment de décrire la couverture spatiale d'un jeu de données (`dcate:bbox`) sous la forme de données RDF. Cela permet notamment d'utiliser le langage GeoSPARQL pour requêter les données.

### 2.3.3.4 Schema.org

Plus récemment, en 2011, le vocabulaire Schema.org a été introduit par des géants du Web comme Bing, Google, Yahoo et Yandex. Le but premier de Schema.org est d'améliorer l'indexation des pages Web en utilisant les micro-données qui ont été préalablement ajoutées par les propriétaires des pages en question. Son utilisation permet de saisir plus précisément le sens d'une page indexée. Schema.org permet, entre autres, de décrire des jeux de données qu'ils soient géographiques ou non.

Google Dataset Search (GDS) a été publié pour la première fois dans sa version Bêta en septembre 2018. Ce n'est plus le cas depuis janvier 2020. Le moteur de recherche permet de retrouver un jeu de données parmi ceux qui ont été indexés grâce à une requête utilisant des mots-clés. GDS moissonne le Web à la recherche de jeux de données à indexer. Dans [6] les auteurs expliquent qu'il est capable d'identifier les jeux

TABLE 2.3 – Alignement entre les classes de DCAT et les classes de Schema.org

Classe DCAT	Élément Schema.org
dcat :Resource	schema :Thing
dcat :Catalog	schema :DataCatalog
dcat :Dataset	schema :Dataset
dcat :Dataset	schema :DataFeed
dcat :DatasetSeries	schema :CreativeWorkSeries
dcat :Distribution	schema :DataDownload
dcat :DataService	schema :WebAPI
dcat :Relationship	schema :Role

de données dont les métadonnées sont décrites en DCAT et Schema.org. Une fois les métadonnées brutes collectées, celles-ci sont normalisées et nettoyées. Ensuite, les doublons potentiels sont identifiés et agrégés. Finalement, les métadonnées sont indexées. Quand l'utilisateur veut effectuer une recherche, il formule sa requête, et les résultats sont classés en fonction de leur pertinence. Le modèle de pertinence utilisé est un modèle de pertinence Web classique (c.-à-d. les modèles utilisés par les moteurs de recherche de documents classiques). Cela est dû à l'inexistence d'un modèle de pertinence dédié aux jeux de données.

Contrairement à Google Dataset Search, data.europa.eu (anciennement European Data Portal) n'a pas pour vocation de rassembler tous types de jeux de données, mais est dédié à la recherche de jeux de données institutionnels ouverts. Ainsi, dans [47], les auteurs affirment que EDP ne moissonne pas le Web pour retrouver les métadonnées, mais seulement un nombre prédéfini de catalogues appartenant essentiellement à des entités publiques productrices de données de référence ou d'intérêt pour l'UE. Ensuite, les métadonnées sont transformées en données liées et indexées. Les métadonnées sont décrites en utilisant l'ontologie DCAT-AP. EDP fournit à l'utilisateur la possibilité de formuler sa requête directement en SPARQL en plus de la possibilité de formuler sa requête grâce à des mots-clés.

## 2.4 Conclusion

Ce Chapitre a présenté les principaux travaux de recherche nécessaires à la compréhension de la complexité de la recherche de jeux de données géographiques sur le Web. En effet, la littérature scientifique identifie les métadonnées géographiques comme l'élément central permettant d'adresser la recherche de jeux de données géographiques.

Si plusieurs standards de métadonnées géographiques ont été introduits dans le but d'augmenter l'interopérabilité des jeux de données et de diminuer les hétérogénéités dans les métadonnées géographiques, plusieurs facteurs font que cet objectif n'est toujours pas atteint. En effet, la qualité des métadonnées et la manière dont ils sont saisis font que ces derniers ne sont pas parfaitement homogènes que ce soit au niveau national, européen ou mondial.

Les Standards ISO et OGC ont été pensés pour être assez permissifs pour permettre aux producteurs de données de les adapter à leurs propres données. Si ça a permis une large adoption de ces standards dans la communauté des producteurs de données, mais en contrepartie, cela n'a pas permis de gommer les hétérogénéités qui peuvent exister entre différents catalogues géographiques.

Le Web des données et les technologies du Web sémantiques ont permis d'améliorer de manière significative la recherche d'information dans différents domaines, grâce à l'utilisation des KGs. Bien que plusieurs standards pour décrire des métadonnées sous format RDF, ces derniers n'ont pas encore été complètement adoptés notamment par les entités publiques productrices des données géographiques européennes. Cela est notamment dû à la directive INSPIRE qui les obligent à utiliser le standard ISO 19115 :2003 et ISO 19139 :2007.

Dans la suite de cette thèse, et pour adresser le problème de la recherche de jeux de données géographiques, nous formulons les hypothèses de recherches suivantes :

- L'utilisateur exprimera sa requête sous la forme d'un concept d'intérêt provenant d'un vocabulaire contrôlé ou d'un graphe de connaissance à définir.
- Nous utiliserons par la suite des métadonnées géographiques conformes au profile INSPIRE.
- Nous nous concentrons sur un domaine d'application spécifique à savoir le domaine de l'eau et de l'environnement en France.
- Quand cela est possible, nous réutiliserons des ressources pré-existantes et disponibles sur le Web.

## Chapitre 3

# Conception d'un Knowledge Graph dédié à la recherche de jeux de données géographiques

### 3.1 Introduction

Ce chapitre a pour but de décrire la conception d'un KG dédié à la recherche de jeux de données géographiques. Il se basera essentiellement sur les travaux suivants [91] et [90].

Pour cela, nous allons tout d'abord présenter dans la Section 3.2 notre approche globale basée sur l'exploitation d'un KG. Ensuite, dans la Section 3.3 nous analyserons les exigences fonctionnelles du KG pour identifier les fonctionnalités nécessaires qui lui permettront de supporter la recherche de jeux de données géographiques. Finalement, la Section 3.4 s'intéressera aux ressources nécessaires à la mise en oeuvre de notre approche et qui permettent la mise en oeuvre du KG avant de conclure le chapitre à la Section 3.5.

### 3.2 Présenter l'approche

Pour adresser la problématique de la recherche de jeux de données géographiques, et en prenant en compte les enjeux applicatifs des utilisateurs et des producteurs de données géographiques présentés dans le Chapitre 1, nous développons une approche qui reprend les principes de recherche d'information basés sur le Web sémantique et les catalogues de métadonnées géographiques présentés dans le Chapitre 2. En d'autres termes, nous visons la conception d'un KG qui s'intégrerait et améliorerait le processus

de recherche de jeux de données géographiques. Le KG se fera sur le modèle de ce qui se fait plus généralement dans le domaine de la recherche de ressources sur le Web.

Pour y parvenir, et dans le cadre de cette thèse, nous faisons les hypothèses de recherche suivantes. Tout d'abord, (1) on peut améliorer la recherche de jeux de données géographiques en permettant à l'utilisateur d'exprimer son besoin avec des concepts de sens commun. Ensuite, (2) nous nous concentrons sur des métadonnées compatibles avec le standard DCAT-AP et finalement, (3) nous nous focaliserons sur un domaine applicatif, à savoir le domaine de l'eau et de l'environnement en France.

Pour développer cette approche, nous allons tout d'abord analyser les exigences fonctionnelles permettant de concevoir un tel KG ; c'est-à-dire le besoin en connaissances et ressources pour supporter la recherche de jeux de données géographiques. Pour cela, nous analyserons des pour et des contres des différentes ressources Web nécessaires pour la construction du KG et pour ensuite l'implémenter et l'évaluer. Cette dernière étape est particulièrement délicate dans la mesure où l'état de l'art ne fournit aucun moteur de recherche dédié à la recherche de jeux de données géographiques semblable pour comparer les performances de notre approche. La méthodologie d'évaluation sera traitée et détaillée dans la Section 4.2.

### 3.3 Analyser les exigences fonctionnelles du KG

Cette section a pour but premier d'analyser les fonctionnalités de recherche qui s'appuieront sur le KG pour à chaque fois identifier les éléments nécessaires et qui devront être intégrés au KG ainsi que les requêtes et rationnements nécessaires à leur exploitation. Pour cela, nous faisons la distinction entre deux processus schématisés dans la Figure 3.1 :

1. ***Le processus hors-session*** : décrit les étapes et les traitements nécessaires au bon fonctionnement de l'approche et que ne nécessite aucune interaction de l'utilisateur. Ce processus concerne notamment la collecte et l'indexation des ressources nécessaires au KG, comme par exemple le moissonnage des catalogues géographiques, la maintenance du KG ou sa mise à jour.
2. ***Session de recherche*** : décrit le processus où l'utilisateur interagit directement avec le KG à travers le moteur de recherche. Durant la session utilisateur, ce dernier peut exprimer son besoin et le moteur de recherche lui renvoie les résultats

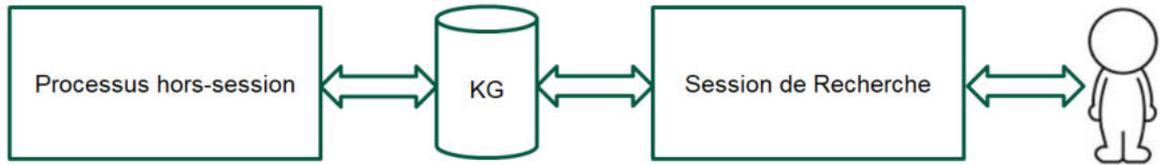


FIGURE 3.1 – Processus global de Recherche

qui correspondent au mieux à sa requête.

### 3.3.1 Processus Hors-Session

Le processus hors-session décrit les étapes et traitements nécessaires au bon fonctionnement d'un moteur de recherche, n'impliquant pas l'intervention de l'utilisateur, et qui doivent être mis en place avant la session utilisateur. Comme présenté dans la Section 2.3.1, un moteur de recherche doit au préalable moissonner et indexer des ressources avant de permettre à un utilisateur d'exécuter une requête. En effet, un moteur de recherche est incapable de retrouver des ressources qu'il n'a pas indexées préalablement.

Le KG devra donc être doté des moyens lui permettant de moissonner et indexer les ressources dont il a besoin, à savoir les métadonnées de description des jeux de données géographiques. Plusieurs descriptions peuvent exister pour un même jeu de données. Cela peut être dû à plusieurs facteurs comme, par exemple, la présence d'un même jeu de données dans des catalogues différents, les différents formats dans lesquels les métadonnées de description sont fournies (XML, RDF, PDF, etc.) ou encore le standard utilisé. C'est pourquoi nous introduisons la notion de *record*.

Dans la suite de ce manuscrit de thèse, nous définissons un record comme un fichier de description d'un jeu de données géographiques. Quand cette description est conforme au standard DCAT-AP, on parlera de records RDF et de manière similaire, nous appelons un record brut, un fichier de description d'un jeu de données géographiques conforme au standard ISO 19115.

Les records sont les ressources de base que le KG doit indexer et exploiter pour prendre en charge la recherche de jeux de données géographiques. Pour cela, le KG devra récupérer des records directement depuis des catalogues géographiques. Il est ainsi possible de se baser sur une liste prédéfinie de catalogues et de mettre à jour régulièrement la liste des records indexés.

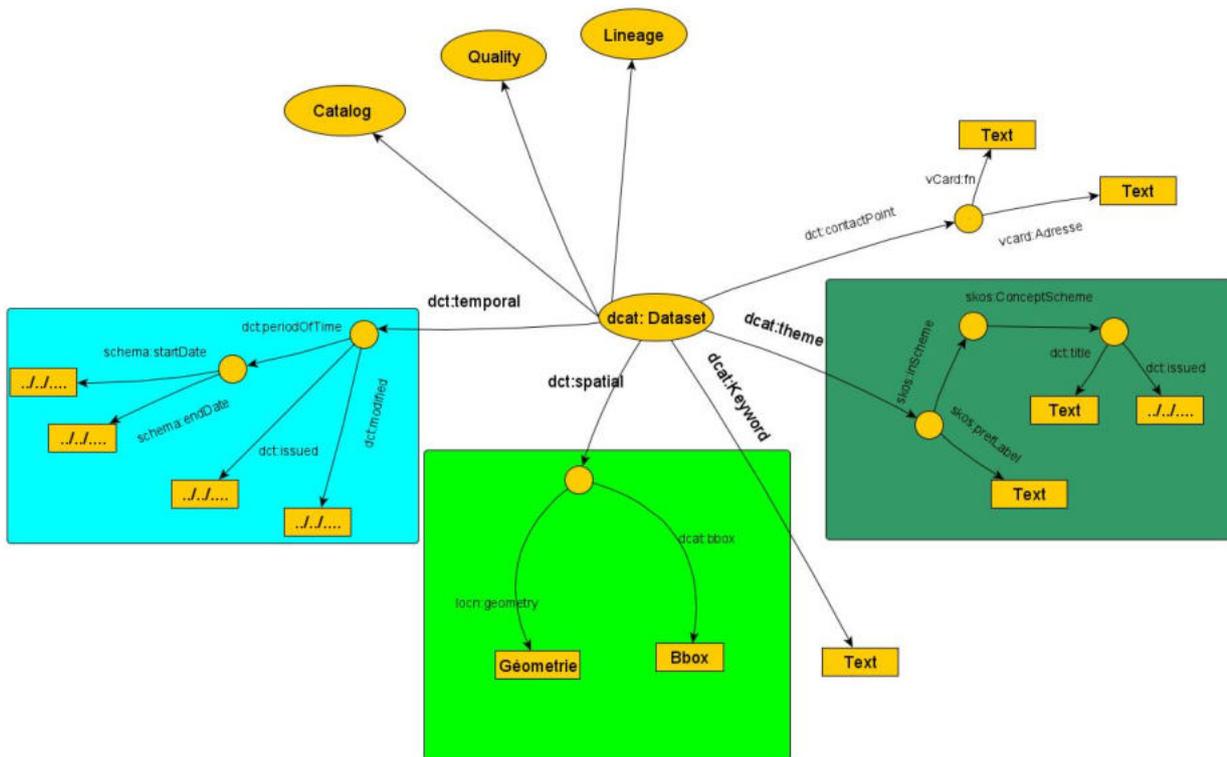


FIGURE 3.2 – Champs de métadonnées dans le standard DCAT-AP

Chaque record de métadonnées se présente sous la forme d'un ensemble de champs de métadonnées décrivant les jeux de données géographiques et contenant des informations comme le titre, la description, la couverture spatiale, les thèmes, les keywords, etc. Les champs de métadonnées utilisés dans un record et la façon de les remplir dépendent fortement du standard utilisé et de la façon dont les producteurs remplissent ces champs. La Figure 3.2 représente les champs de métadonnées les plus importants dans le standard DCAT-AP.

Dans la mesure où en Europe, la directive INSPIRE impose l'utilisation du standard ISO 19115 pour décrire les métadonnées géographiques à toutes les entités publiques européennes, en moissonnant les catalogues géographiques, le KG récupérera essentiellement des records bruts. Il faudra ensuite que le KG soit capable de transformer les records bruts en records RDF pour pouvoir les stocker.

### 3.3.2 Session de Recherche

La session de recherche est la partie du processus qui implique directement l'utilisateur. Ce dernier interagit avec le moteur de recherche grâce à une requête utilisateur pour obtenir les résultats de recherche. En nous inspirant des étapes élémentaires du

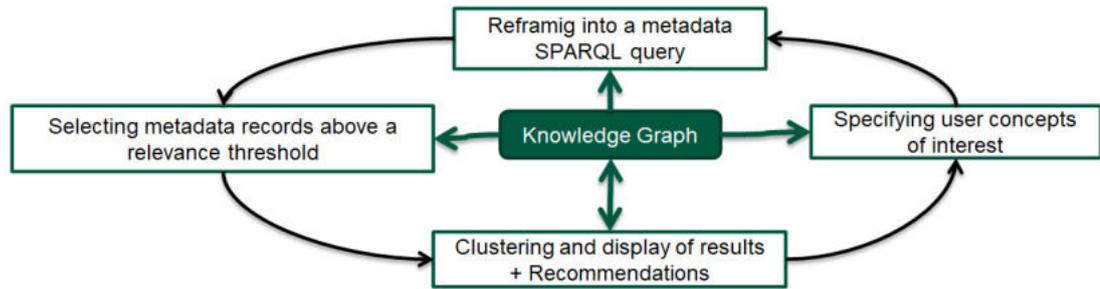


FIGURE 3.3 – Processus de recherche des jeux de données géographiques grâce à un KG dédié fonctionnant d’un moteur de recherche classique présentées dans la Section 2.3.1. Nous les adapterons à la recherche de jeux de données géographiques en mettant le KG au centre du processus et en essayant de définir pour chaque étape, quelles ressources et informations doivent y être intégrées. Figure 3.3 présente le fonctionnement d’un moteur de recherche adapté à la recherche de jeux de données géographiques.

Ainsi, une session de recherche peut être divisée en 4 étapes :

- **Étape 1** : L’utilisateur exprime sa requête sous la forme d’un concept d’intérêt.
- **Étape 2** : Le KG traduit le concept d’intérêt en une requête interne sur les métadonnées.
- **Étape 3** : Le KG sélectionne les records au-dessus d’un seuil de pertinence.
- **Étape 4** : Les résultats de recherche les plus pertinents sont ensuite regroupés et présentés à l’utilisateur sous la forme de groupes de résultats.

### 3.3.2.1 Supporter la spécification d’un concept d’intérêt utilisateur

La première étape d’une session de recherche implique que l’utilisateur ait la possibilité d’exprimer sa requête sous la forme d’un concept d’intérêt. Un concept d’intérêt, ou concept utilisateur est un concept choisi par l’utilisateur lors d’une session de recherche pour exprimer son besoin et provenant d’une ontologie, d’un thésaurus ou d’un vocabulaire contrôlé.

Le KG doit permettre à l’utilisateur de choisir parmi un ensemble de concepts relatifs au domaine d’application, mais aussi de sens commun. Prenons l’exemple d’un

utilisateur qui s'intéresse aux jeux de données géographiques concernant les rivières. L'utilisateur pourra donc choisir des concepts d'intérêt relatifs aux rivières comme "rivière", "affluent" ou encore "bassin versant", mais aussi des entités nommées comme "Seine", "Loire" ou "Rhône" qui sont des instances de rivières.

Il est donc important que le KG intègre des vocabulaires permettant à l'utilisateur de choisir un concept d'intérêt. Ces vocabulaires doivent proposer suffisamment de concepts pour couvrir le besoin de l'utilisateur et couvrir non seulement le domaine de l'eau et de l'environnement en France, mais aussi d'autres domaines. En effet, quand l'utilisateur exprime sa requête, il peut notamment choisir un concept d'un domaine plus ou moins éloigné du domaine d'application. C'est pourquoi le KG doit inclure au moins un vocabulaire de sens commun qui contient à la fois des concepts relatifs au domaine de l'eau et de l'environnement, mais aussi des concepts de sens commun. A cela doivent s'ajouter un ou plusieurs vocabulaires thématiques reliés au domaine de l'eau et de l'environnement et qui contiennent des concepts plus spécialisés.

### 3.3.2.2 Reformuler la requête utilisateur en une requête sur les métadonnées

L'étape 2 consiste à transformer la requête utilisateur en une requête interne sur les records de métadonnées. Ainsi, l'utilisateur choisit un ou plusieurs concepts d'intérêt à partir d'un vocabulaire pris en compte par le KG, et crée une requête SPARQL pour récupérer un ensemble de records candidats. Si le nombre de records est insuffisant, la requête devra être étendue pour fournir à l'utilisateur des records qui ne correspondent pas entièrement à sa requête, mais qui pourraient l'intéresser.

Dans la mesure où l'utilisateur exprime sa requête sous la forme d'un concept d'intérêt provenant d'un ensemble de vocabulaires prédéfinis, il est donc indispensable d'aligner les thématiques et les mots-clés présents dans les métadonnées aux vocabulaires permettant à l'utilisateur d'exprimer sa requête.

De manière générale, les producteurs de métadonnées utilisent des vocabulaires pour décrire les thématiques présentes dans un record. Comme présenté dans la Section 2.2.2, l'utilisation de certains vocabulaires est rendue obligatoire par la directive INSPIRE, mais en fonction du catalogue géographique en question et de son domaine d'application, d'autres vocabulaires sont utilisés pour décrire des thématiques plus spécifiques. Les vocabulaires utilisés dans les records sont généralement disponibles sous forme de données liées, et leurs thèmes peuvent donc être identifiés par des URIs.

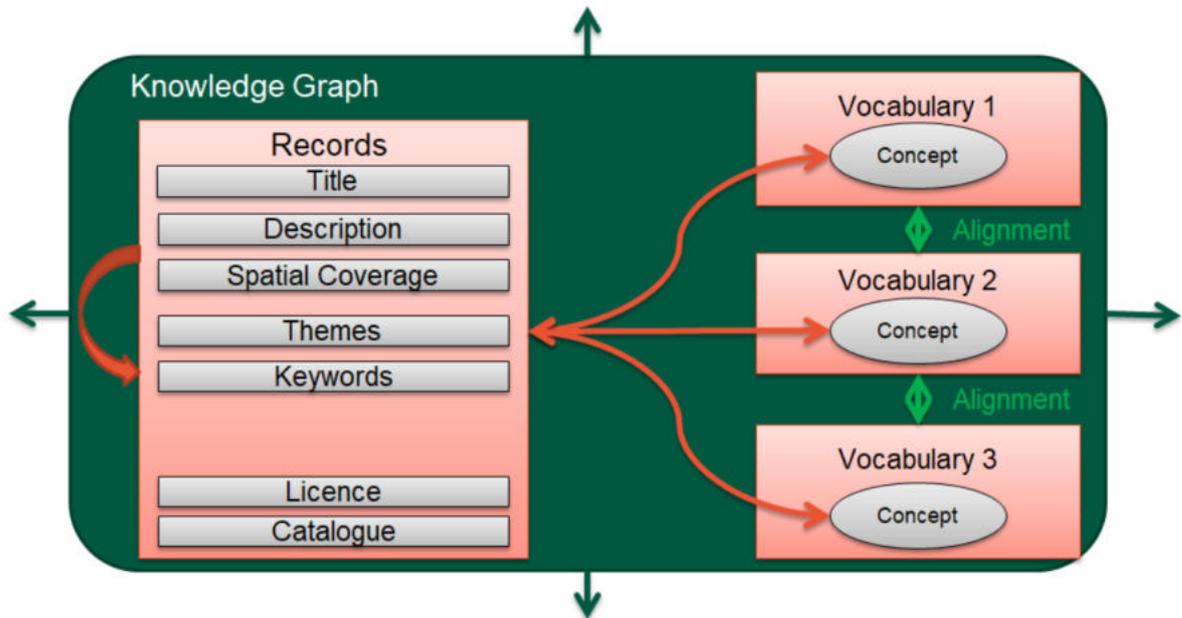


FIGURE 3.4 – Knowledge Graph Components

Les mots-clés sont de simples chaînes de caractères qui permettent de mieux décrire un jeu de données et d'améliorer la visibilité de ce dernier pour un moteur de recherche classique. Il est indispensable d'aligner les mots-clés présents dans les records de métadonnées avec des concepts déjà présents dans le KG. Autrement dit, il faudra munir chaque mot-clé d'un URI provenant d'un vocabulaire du KG.

La Figure 3.4 présente un schéma conceptuel des principaux composants du KG à ce stade. En vert sont représentés les alignements entre vocabulaires et en rouge les alignements entre les records et les vocabulaires.

### 3.3.2.3 Évaluer la pertinence des records candidats

L'étape 3 permet, à partir d'un ensemble de records candidats fourni, d'évaluer la pertinence de chaque record par rapport au concept d'intérêt de l'utilisateur. Pour réaliser cette étape, le KG devra être capable d'évaluer la similarité entre le ou les concepts d'intérêt choisis par l'utilisateur et les concepts présents dans les records.

À partir des similarités deux à deux, il faudra déterminer un score de pertinence global pour chaque record présent dans le KG. Ensuite, seuls les records ayant un score supérieur à un seuil de pertinence satisfaisant seront sélectionnés et retournés à l'utilisateur comme résultats de sa recherche. Plusieurs mesures de similarité sémantiques existent et seront présentées en détail dans la Section 3.4.

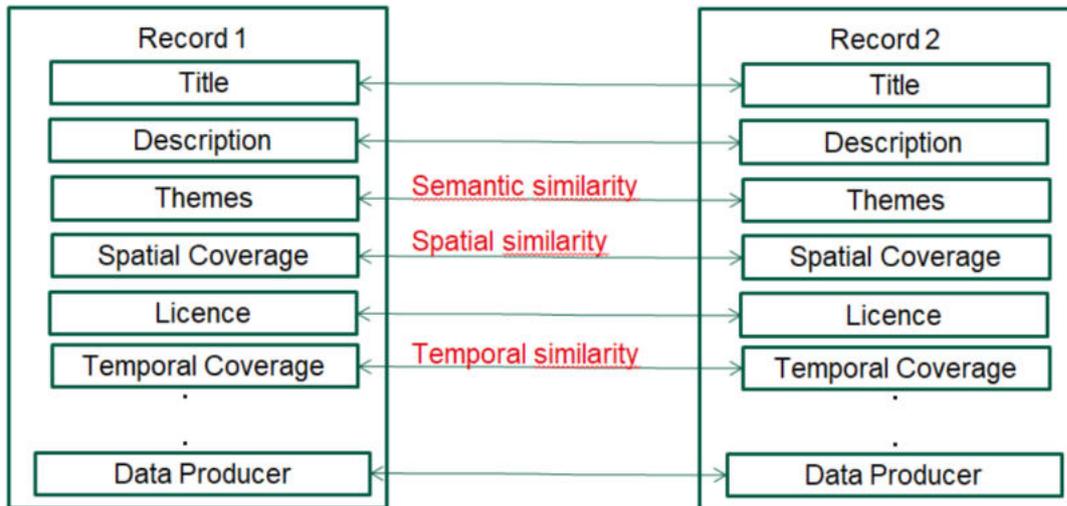


FIGURE 3.5 – Comparaison de Records

### 3.3.2.4 Regrouper les records et les renvoyer à l'utilisateur

Pour compléter l'étape 4, le KG doit être en mesure de créer des clusters de records de métadonnées. Cela permettra à l'utilisateur d'obtenir des résultats de recherche agrégés en fonction de leur similarité et donc de simplifier la tâche à l'utilisateur en ce qui concerne le choix du jeu de données qui l'intéresse.

Regrouper des records implique de disposer d'une méthode pour les comparer et évaluer à quel point deux records sont semblables. Ainsi, une première étape serait d'identifier quels critères sont pertinents pour comparer de tels objets pour ensuite utiliser une mesure multi-critères et en déduire une comparaison globale entre chaque paire de records dans le KG.

Pour chaque critère identifié, on utilisera une mesure de similarité ou distance qui permet de comparer les champs de métadonnées concernés. Par exemple, pour comparer les emprises géographiques de deux records, on privilégiera une mesure de similarité spatiale, et une mesure de similarité sémantique pour les thèmes et mots-clés. Ainsi, chaque champ de métadonnées dans un record sera comparé aux champs de métadonnées correspondant dans le deuxième record. La Figure 3.5 présente un exemple de champs de métadonnées à comparer.

La Figure 3.6 donne un schéma détaillé des composants du KG, à partir de l'analyse des exigences fonctionnelles qui a été faite. Différents éléments des métadonnées ont été représentés, notamment le producteur de données (`dct :contactPoint`), la couverture spatiale, les thèmes, les mots-clés, etc. En rouge sont représentés les alignements entre les thèmes et mots-clés présents dans les records et les vocabulaires (thématique et

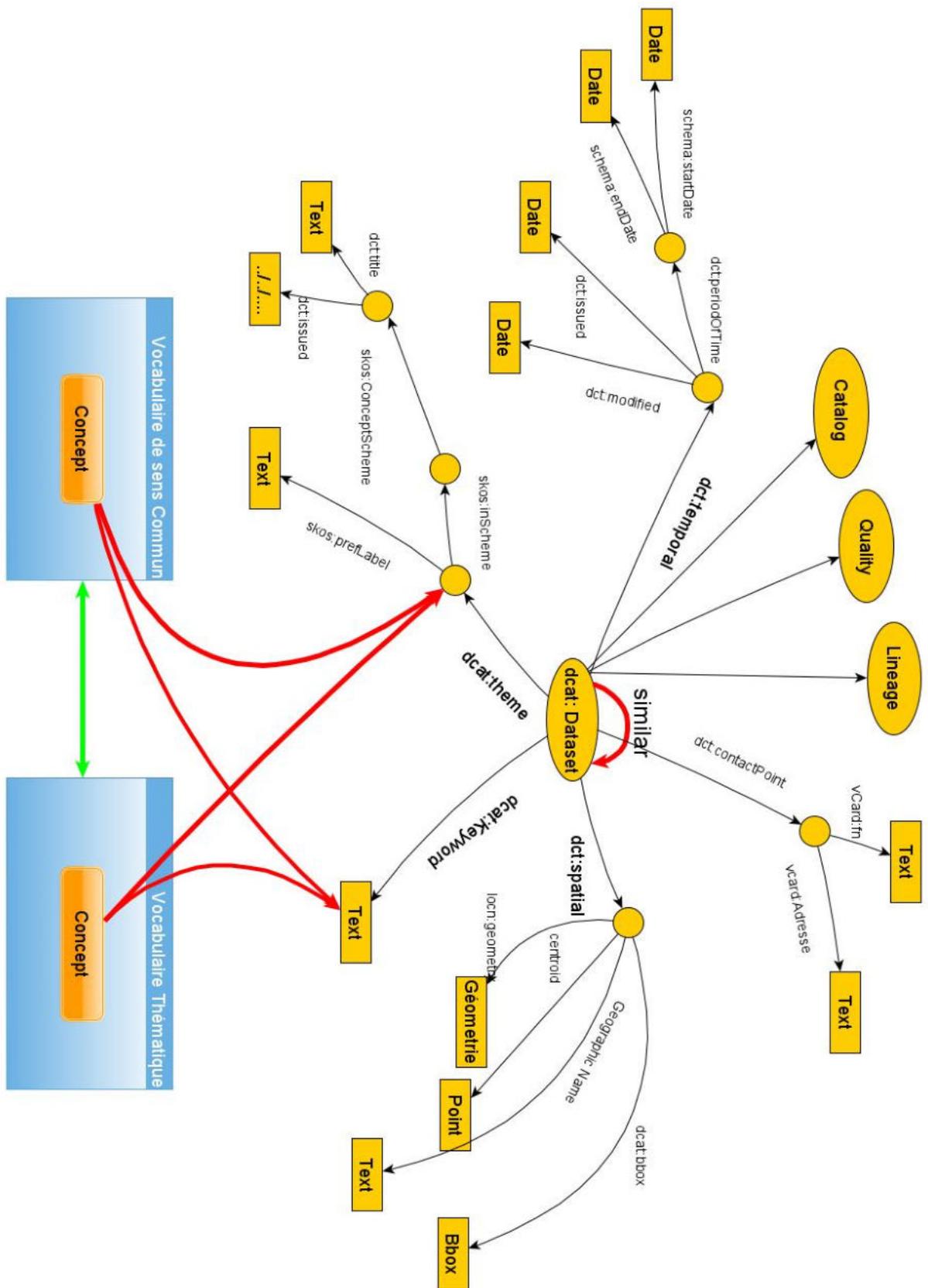


FIGURE 3.6 – Schéma d'ensemble du KG

de sens commun) utilisés dans le KG. En vert est représenté l'alignement entre les alignements entre les vocabulaires utilisés.

## 3.4 Analyse des ressources à intégrer au KG

Cette section s'intéressera aux différentes ressources disponibles sur le Web et qui pourraient être utilisées pour créer et exploiter un KG en réponse à l'analyse faite à la section précédente. Le but est d'évaluer les coûts et bénéfices associés à chaque ressource correspondant aux exigences fonctionnelles identifiées dans la Section 3.3 pour choisir celles qui s'intégreront le mieux au KG.

### 3.4.1 Les catalogues à moissonner

Comme présenté dans la Section 3.3, la première étape pour constituer le KG est de collecter les métadonnées géographiques et donc de moissonner des catalogues géographiques en relation avec le domaine de l'eau et de l'environnement en France.

Comme présenté dans la Section 2.2, les catalogues géographiques rassemblent les métadonnées associées aux jeux de données géographiques et sont généralement organisés en fonction des institutions productrices des données ou encore en fonction des thématiques abordées dans les jeux de données. Rappelons aussi qu'en France, la directive INSPIRE impose aux entités publiques productrices des données géographiques d'utiliser les standards ISO et OGC dans leurs catalogues géographiques et notamment le standard CSW qui standardise le moissonnage des métadonnées géographiques.

#### 3.4.1.1 Sandre

Un premier catalogue d'intérêt est le catalogue du Sandre<sup>1</sup> (i.e. *Service d'Administration Nationale des Données et Référentiels sur l'Eau*). La mission du Sandre est d'établir, mettre à disposition et entretenir un système d'information sur l'eau (i.e. *SIE*) au niveau national. Le SIE se compose de plusieurs éléments, dont des jeux de données de références autour de l'eau en France.

Le catalogue du Sandre rassemble 276 jeux de données de référence accessibles par une API respectant le standard CSW. Le catalogue indexe aussi des records de jeux de données produits par d'autres organismes, comme par exemple, les limites administratives publiées par l'IGN, et qui sont indispensables pour compléter le SIE développé

---

1. <https://www.sandre.eaufrance.fr/>

par le Sandre.

Les records sont disponibles sous format XML et sont conformes au standard ISO 19115.

#### 3.4.1.2 Cerema

Le Cerema est le centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement. C'est un établissement public dont les missions concernent l'ensemble des thématiques de l'aménagement et développement durable (c.-à-d. urbanisme, environnement, gestions des risques, transport, mer et littoral, etc.). Dans le cadre de ses missions, le Cerema est amené à publier plusieurs jeux de données concernant chacune de ces thématiques.

Parmi les missions du Cerema, certaines sont directement liées au domaine de l'eau et de l'environnement en France, notamment l'environnement, la gestion des risques ou encore la mer et le littoral. De plus, le Cerema travaille en étroite relation avec les collectivités locales et produit des jeux de données au niveau régional, départemental ou encore communal. Ainsi, les jeux de données du Cerema et du Sandre sont complémentaires et proposent un large panel d'applications possibles pour l'utilisateur.

Le catalogue CeremData est le catalogue CSW qui rassemble les jeux de données produits par le Cerema. Il indexe 191 records disponibles sous format XML et conformes au standard ISO 19115. Grâce au standard CSW, il est possible de moissonner les records en utilisant l'API mise à disposition.

#### 3.4.2 Les vocabulaires thématiques

Comme discuté dans la Section 3.3, il est indispensable d'inclure un ou plusieurs vocabulaires thématiques dans le KG pour permettre à la fois à l'utilisateur d'exprimer sa requête sous la forme d'un concept d'intérêt, mais aussi pour identifier les thèmes et mots-clés présents dans les métadonnées.

Nous choisirons les vocabulaires à intégrer au KG en fonction de ces trois critères :

- **L'apport au KG** : quels bénéfices y a-t-il à intégrer le vocabulaire au KG ?
- **Facilité d'intégration au KG** : quels sont les coûts associés à l'intégration du vocabulaire dans le KG ?

— **Alignements connus** : existe-t-il un ou plusieurs alignements permettant de relier le vocabulaire à d'autres vocabulaires ou ontologies ?

### 3.4.2.1 GEMET

Un premier vocabulaire d'intérêt est GEMET (i.e. *GEneral Multilingual Environmental Thesaurus*). GEMET est un thésaurus qui a été développé par Eionet (i.e. *European Environment information and observation network*) dans le but de hiérarchiser et catégoriser plusieurs concepts relatifs au domaine de l'environnement et donc de définir une terminologie générale du domaine de l'environnement. GEMET contient plus de 6500 concepts décrits en 27 langues européennes et extra européennes. Il est régulièrement mis à jour. Ainsi, la version la plus récente de GEMET (v4.1) date de 2021 alors que la première version date quant à elle de 2001.

GEMET est disponible sous format RDF. Chaque concept est défini de manière unique par un URI et la relation entre concepts est décrite grâce au vocabulaire SKOS. GEMET est disponible sous la forme de données ouvertes et peut être facilement téléchargé sous la forme d'un fichier, consulté sous format HTML ou à travers une API, ou encore interrogé grâce un SPARQL endpoint. Il met à disposition, dans ses données, plusieurs alignements vers d'autres vocabulaires largement utilisés comme DBpedia, les thèmes géographiques d'INSPIRE, Agrovoc, EuroVoc, etc. D'autres vocabulaires comme Wikidata fournissent aussi des alignements avec GEMET. Il est aussi important de noter que GEMET est très utilisé dans les métadonnées géographiques au niveau Européen.

GEMET contient uniquement des concepts, mais aucune instance de ces concepts. Un utilisateur peut trouver des entités comme "cours d'eau", "ville", "pays" ou "hydrographie", mais il lui sera impossible de trouver des instances comme "Garonne", "Paris", ou "France". C'est pourquoi GEMET n'est pas suffisant pour permettre à l'utilisateur d'exprimer pleinement sa requête ni pour décrire tous les thèmes et mots-clés disponibles dans les records de métadonnées.

La Figure 3.7 présente la description associée au concept "cours d'eau" dans GEMET.

**GEMET** Version 4.0  
Multilingual  
Environmental  
Thesaurus

Thematic Listings    Hierarchical Listings    INSPIRE Spatial Data Themes    Alphabetic Listings

Search concepts by name (French)

**cours d'eau**

**Definition**  
A natural stream arising in a given drainage basin but not wholly dependent for its flow on surface drainage in its immediate area, flowing in a channel with a well-defined bed between visible banks or through a definite depression in the land, having a definite and permanent or periodic supply of water, and usually, but not necessarily, having a perceptible current in a particular direction and discharging at a fixed point into another body of water.

*Definition is not available for the current language.*

**Related terms**  
**Broader:** [eaux \(géographie\)](#) | [hydrosphère](#)  
**Related:** [bassin hydrographique](#) | [cours d'eau international](#)  
**Narrower:** [fleuve](#) | [ruisseau](#) | [source](#)  
**Themes:** [eau](#) | [zones naturelles, paysages, écosystèmes](#)

**Translations**

Arabic:	مسار المياه
Armenian:	գրունի, ցրի ճակատագրի
Azerbaijani:	su yolu
Basque:	ur-ibilgu
Bulgarian:	Воден поток
Catalan:	curs d'aigua
Chinese:	水道
Croatian:	vodotok
Czech:	tok vodní
Danish:	vandløb
Dutch:	waterloop
English:	watercourse
English (US):	watercourse
Estonian:	vooluveekogu
Finnish:	vesistö, vesitie, väylä
French:	cours d'eau
Georgian:	წყლის ხეობა
German:	Wasserlauf
Greek:	υδάνην οδό
Hungarian:	vízfolyás
Icelandic:	vátnfall
Irish:	cúrsa uisce
Italian:	corso d'acqua
Latvian:	ūdenskrātuve

FIGURE 3.7 – Exemple du concept "Cours d'eau" dans Gemet

### 3.4.2.2 Glossaire eau et biodiversité

Le Glossaire sur l'Eau, les milieux marins et la biodiversité est un vocabulaire qui rassemble un ensemble de définitions sur les thématiques de l'eau, des milieux aquatiques et de la biodiversité. Le glossaire est piloté par l'office international de l'eau (i.e. *OiEAU*) avec le soutien de l'office français de la biodiversité (i.e. *OFB*).

Le glossaire contient plus de 1899 termes associés à des définitions associées à un sens de lecture particulier (commun, technique ou réglementaire). Il prend en charge 3 langues d'utilisation, à savoir le français, l'espagnol et l'anglais et est structuré autour de 17 thématiques comme (risque, climat, gestion de l'eau et des milieux, eau potable et assainissement, etc.).

Le glossaire est collaboratif et permet de soumettre des ajouts ou des modifications à un groupe de relecteurs composés de techniciens et de spécialistes des domaines de l'eau et de la biodiversité qui garantissent la fiabilité des informations diffusées. Le glossaire est disponible selon les standards du Web sémantique sous formats RDF et utilise SKOS pour la description des concepts qu'il contient.

Il faut néanmoins noter que le glossaire eau et biodiversité ne fournit pas beaucoup d'alignements externes. Ainsi, il existe 4 liens avec GEMET, 312 liens avec Wikipedia

The screenshot shows the website interface for the 'Glossaire Eau, Milieu marin et Biodiversité'. At the top, there are logos for 'eaufrance', 'milieumarinefrance', and 'naturefrance'. A search bar is visible with the text 'Rechercher' and a search icon. Below the search bar, there are navigation links: 'A propos du glossaire', 'Consulter le glossaire', 'Télécharger le glossaire', and 'Espace perso'. The main content area is titled 'Cours d'eau' and includes a dropdown menu for 'Accéder directement à un terme'. Below the title, there is a 'Mis à jour le 05/03/2019' date and social media icons for Facebook, Twitter, LinkedIn, and Email. The 'Synonymes' section lists 'Masse d'eau cours d'eau, Rivière.'. The 'Sens commun' section provides a detailed definition: 'Caractérisé par la permanence du lit, le caractère naturel du cours d'eau ou son affectation à l'écoulement normal des eaux (par exemple, un canal offrant à la rivière, dans un intérêt collectif, un débouché supplémentaire ou remplaçant le lit naturel) et une alimentation suffisante, ne se limitant pas à des rejets ou à des eaux de pluies (l'existence d'une source est nécessaire). Source: d'après Agence de l'eau Rhône-Méditerranée et Corse'. The 'Thématiques associées' section lists 'Hydromorphologie, Milieu et habitat, Cours d'eau.'. The 'Autres termes utiles à la compréhension' section lists 'Eau de surface, Eaux courantes.'. The 'Termes associés à des définitions du Sandre' section lists 'COURS D'EAU, ENTITE HYDROGRAPHIQUE.'. The 'Traductions' section lists 'Watercourse [en], Curso de agua [es]'. The 'Statut de relecture' section lists 'Soumis au comité de relecture.'.

FIGURE 3.8 – Exemple du concept "Cours d'eau" dans le Glossaire eau et biodiversité

(dont seulement 4 sont des liens d'équivalence "owl :sameAs") ou encore 284 avec Thésaurus Eau.

La Figure 3.8 présente la description associée au concept "cours d'eau" dans le Glossaire eau et biodiversité.

### 3.4.2.3 Thésaurus EAU

Le thésaurus EAU est un vocabulaire créé par les six agences de l'eau françaises, l'Onema, l'OiEAU et la direction de l'eau et de la biodiversité du ministère de l'écologie et du développement durable. Il rassemble une liste de termes contrôlés et normalisés autour de la thématique de l'eau. Il s'adresse aux centres de documentation dont l'activité, principale ou partielle, concerne l'environnement et plus spécifiquement l'eau.

Le thésaurus EAU agrège 2 687 définitions de concepts disponibles en français et en anglais organisés autour de 44 champs sémantiques et destinés à être utilisés lors de l'indexation et la recherche de ressources sur le Web. Le thésaurus traite de plusieurs sujets comme l'agriculture, l'environnement, la faune, la qualité de l'eau, les risques, les milieux, etc.

Il est à noter que le thésaurus EAU est disponible sous format RDF et fournit un SPARQL Endpoint pour envoyer des requêtes, mais ne fournit aucun alignement externe avec d'autres vocabulaires ou thésaurus.

Descripteur	<ul style="list-style-type: none"> <li>▪ (fr-FR) cours d'eau</li> <li>▪ (en-US) stream</li> </ul>	<p>Terme(s) de premier niveau</p> <ul style="list-style-type: none"> <li>▪ milieu naturel</li> </ul>
Employé pour	<ul style="list-style-type: none"> <li>▪ (fr-FR) eaux fluviales</li> <li>▪ (fr-FR) fleuve</li> <li>▪ (fr-FR) rivière</li> </ul>	<p>Groupes de concepts</p> <ul style="list-style-type: none"> <li>▪ 0026 milieu naturel</li> </ul>
Terme(s) générique(s)	<ul style="list-style-type: none"> <li>▪ <a href="#">ecosysteme aquatique</a></li> </ul>	
Terme(s) spécifique(s)	<ul style="list-style-type: none"> <li>▪ <a href="#">annexe hydraulique</a></li> <li>▪ <a href="#">atterrissement</a></li> <li>▪ <a href="#">berge</a></li> <li>▪ <a href="#">continuite ecologique</a></li> <li>▪ <a href="#">evolution du lit</a></li> <li>▪ <a href="#">hydromorphologie</a></li> <li>▪ <a href="#">lit majeur</a></li> <li>▪ <a href="#">lit mineur</a></li> <li>▪ <a href="#">lone</a></li> <li>▪ <a href="#">meandre</a></li> <li>▪ <a href="#">riviere phreatique</a></li> <li>▪ <a href="#">ruisseau</a></li> <li>▪ <a href="#">torrent</a></li> </ul>	
Terme(s) associé(s)	<ul style="list-style-type: none"> <li>▪ <a href="#">acide humique</a></li> <li>▪ <a href="#">adaptation paysagere</a></li> <li>▪ <a href="#">baignade</a></li> <li>▪ <a href="#">boue de curage</a></li> <li>▪ <a href="#">eau de baignade</a></li> </ul>	

FIGURE 3.9 – Exemple du concept "Cours d'eau" dans le Thésaurus Eau

La Figure 3.9 présente la description associée au concept "cours d'eau" dans le thésaurus EAU.

### 3.4.3 Les vocabulaires de sens commun

En se basant sur la Section 3.3, en plus des vocabulaires thématiques, il est nécessaire d'inclure des vocabulaires de sens commun au KG. Cela permettra à l'utilisateur d'exprimer sa requête sous la forme d'un concept d'intérêt parmi un large panel de concepts disponibles. Pour cela, ces vocabulaires doivent contenir un grand nombre de concepts et d'informations structurées et facilement exploitables.

Comme pour les vocabulaires thématiques, nous choisirons les vocabulaires de sens commun à utiliser dans le KG en fonction des trois critères :

- **L'apport au KG** : quels bénéfices y a-t-il à intégrer le vocabulaire au KG ?
- **Facilité d'intégration au KG** : quels sont les coûts associés à l'intégration du vocabulaire dans le KG ?
- **Alignements connus** : existe-t-il un ou plusieurs alignements permettant de relier le vocabulaire à d'autres vocabulaires ou ontologies ?

Deux alternatives semblent correspondre aux besoins du KG : Wikidata et DBpedia.

### 3.4.3.1 Wikidata

Tout d'abord, Wikidata est un KG ouvert contenant à la fois des données conceptuelles et des instances du monde réel. Il est né d'un effort communautaire qui vise à rassembler les différents projets Wikimedia (c.-à-d. Wikipédia, Wikivoyage, Wiki commons etc.) sous la forme de données liées. Wikidata contient plus de 99.000.000 de concepts différents et 12.500.000.000 triplets.

Wikidata est un élément central des données liées ouvertes (i.e. *LOD*) et aligné avec un grand nombre de vocabulaires traitant de différentes thématiques. Wikidata est notamment aligné avec GEMET, AgroVoc, GeoNames, Wikipedia, schema.org, DCAT et un grand nombre de vocabulaires et ontologies très utilisés dans plusieurs domaines très variés.

Pour retrouver un concept sur Wikidata, il existe différentes méthodes. En effet, il est possible de faire une recherche directement sur le site Web de Wikidata grâce au moteur de recherche qui y est fourni. Ensuite, Wikidata fournit un SPARQL Endpoint qui permet de requêter les données que ce soit sur l'interface graphique fournie par le site Web ou à travers une API RestFul. Il est aussi possible de télécharger toutes les données de Wikidata, à un moment précis, sous la forme d'un fichier RDF appelé un *dump*.

Une dernière alternative est d'utiliser une page Wikipedia pour retrouver le concept Wikidata correspondant en utilisant un lien dédié "Élément Wikidata", comme on peut le voir sur la Figure 3.10. Ce lien entre Wikipedia et Wikidata peut s'avérer très intéressant pour un utilisateur qui n'a aucune expérience d'utilisation de Wikidata, mais qui est habitué à utiliser Wikipédia.

### 3.4.3.2 DBpedia

DBpedia est un KG ouvert qui se base sur l'exploration et l'extraction automatique des données des Infobox des pages Wikipedia. Il propose des données structurées et normalisées au format RDF des contenus Wikipedia. DBpedia a été développé dans l'optique d'en faire un élément central de la LOD et donc un Hub d'interconnexion et d'alignements avec un maximum d'autres vocabulaires et KG du Web des données.

The image shows a screenshot of the French Wikipedia article titled "Rivière". The page layout includes a search bar at the top, a navigation menu on the left, and the main article content. The article text defines a river in hydrography, mentions examples like the Saône and Rhône, and discusses its flow characteristics. There are two images: one showing a river in a mountainous area and another showing a riverbed with a large log jam. The page also includes a "Présentation" section and a list of interlanguage links.

FIGURE 3.10 – Exemple d'un Élément Wikidata

DBpedia contient plus de 4.500.000 de concepts différents et environ 9.500.000.000 triplets. DBpedia fournit plus de 38 millions de labels et résumés en 125 langues différentes, 25 millions d'images et plus de 80 millions de catégories Wikipédia. Il est aussi aligné avec Wikidata.

Comme avec Wikidata, il existe plusieurs méthodes pour retrouver un concept DBpedia. Il est possible d'utiliser le site Web de DBpedia ou de passer par le SPARQL endpoint ou l'API dédiée. Mais contrairement à Wikidata, il n'est pas possible de trouver l'élément DBpedia correspondant à une page Wikipédia directement depuis celle-ci.

### 3.4.4 Mesures de similarité

En plus des catalogues et des vocabulaires à inclure dans le KG, un autre type de ressources indispensables à l'exploitation du KG dans le cadre de la recherche de jeux de données géographiques sont les mesures de similarité. En effet, les étapes 3 et 4 du processus présenté dans la Section 3.3 dépendent essentiellement des mesures de similarité sémantiques, géographiques ou textuelles qui existent et qui sont exploitables par le KG. Cette partie s'intéresse aux différentes mesures de similarité qui pourraient être exploitées par le KG et de leurs potentielles implémentations.

Comme discuté dans le Chapitre 2, plusieurs manières de comparer des concepts existent. En se basant sur l'analyse des exigences fonctionnelles du KG, et dans la mesure où l'utilisateur exprime sa requête sous la forme d'un concept d'intérêt dans Wikidata, et que les records de métadonnées sont alignés avec des concepts Wikidata,



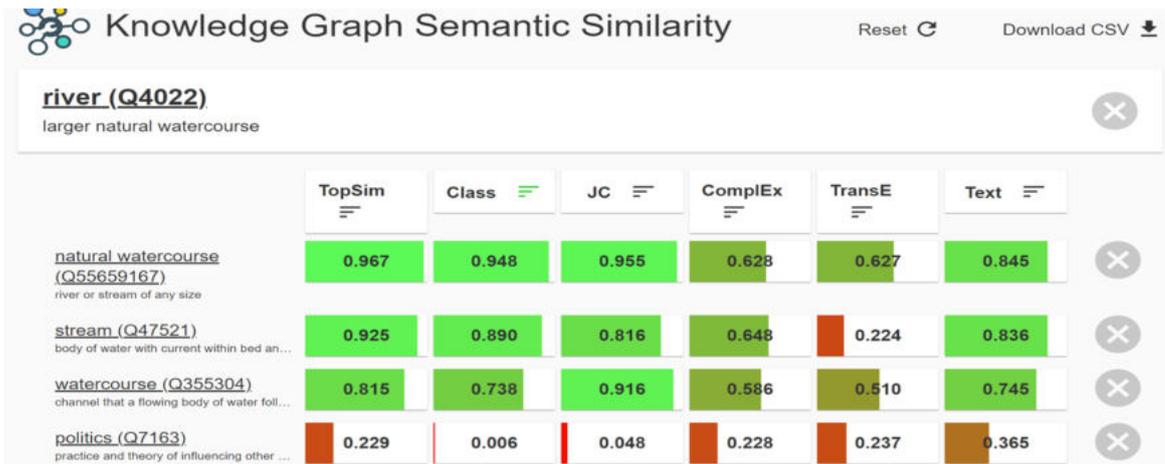


FIGURE 3.12 – Exemple de scores de similarité sémantique calculés par l’outil KG semantic similarity

notre but est de comparer la similarité sémantique entre deux concepts Wikidata.

Pour cela, il existe l’outil KGTK-Similarity<sup>2</sup> (*Knowledge Graph ToolKit - Similarity*) qui permet à travers une API Rest de comparer deux concepts de Wikidata en utilisant leurs URIs. KGTK-Similarity implémente six mesures de similarité sémantiques dédiées aux concepts de Wikidata, à savoir, la mesure "Class", "JC", "TransE", "ComplEx", "Text" et "TopSim". La Figure 3.12 montre un exemple de scores de similarités calculés pour le concept "rivière" dans Wikidata grâce à l’outil KGTK-Similarity.

### 3.5 Conclusion

Dans ce Chapitre nous avons présenté notre approche basée sur un graphe de connaissances pour adresser la problématique de la recherche de jeux de données géographiques. Pour développer notre approche, nous nous sommes inspirés des étapes classiques de la recherche d’information grâce à un moteur de recherche et nous les avons adaptées à la recherche de jeux de données géographiques et en tenant compte du rôle du KG.

Ensuite, nous avons détaillé les exigences fonctionnelles du KG pour lui permettre de supporter la recherche des jeux de données géographiques ce qui nous a permis de définir les composants essentiels et les ressources Web nécessaires à son fonctionnement. Cela nous a amenés à analyser parmi ces ressources, lesquelles étaient disponibles et lesquelles étaient manquantes.

2. <https://github.com/usc-isi-i2/kgtk-similarity>

S'il existe plusieurs catalogues que l'on peut moissonner grâce aux outils CSW, il est aussi nécessaire de les transformer sous format RDF avant de les intégrer au KG. Pour le domaine d'application choisi, nous avons identifié 2 catalogues intéressants : Cerema et Sandre.

Pour mieux enrichir les métadonnées et notamment les mots-clés et les thèmes présents dans les records, nous avons investigué les vocabulaires à intégrer au KG. Notre choix s'est porté pour deux vocabulaires, à savoir, GEMET qui est un thésaurus dédié au domaine de l'environnement et Wikidata qui est un KG de sens commun qui permettra à l'utilisateur d'exprimer sa requête sous la forme d'un concept d'intérêt. Nous avons aussi identifié l'outil KGTK-Similarity qui permettra de comparer le concept utilisateur et les concepts qui ont été utilisés pour enrichir les records de métadonnées.

Si plusieurs ressources nécessaires au fonctionnement du KG sont disponibles et sont facilement intégrables au KG, d'autres ressources comme certains vocabulaires du domaine de l'eau ne sont pas disponibles en RDF. De plus, l'alignement entre Wikidata et GEMET n'existe pas et doit donc être calculé. Même si une propriété "GEMET ID" a été introduite dans Wikidata en 2022, celle-ci ne concerne que quelques concepts (52 concepts seulement en octobre 2022) et ne peut donc pas être utilisée.

## Chapitre 4

# Implémentation et évaluation

### 4.1 Introduction

Ce chapitre présente l'implémentation et l'évaluation de notre approche. L'évaluation porte sur le processus hors session de production et d'alimentation du KG d'une part, et sur les fonctionnalités clés d'une session de recherche de jeux de données géographiques en exploitant ce KG.

Avant de détailler notre méthodologie, il faut souligner plusieurs spécificités de l'évaluation dans ce domaine de recherche. L'évaluation d'une session de recherche ne peut porter que sur des fonctionnalités prises isolément, car construire un moteur qui supporte le processus complet implique du travail d'intégration et cela sort du cadre d'une thèse. Par ailleurs, concevoir les expérimentations pour évaluer ces fonctionnalités pose des défis méthodologiques, car ces fonctionnalités sont difficilement isolables des experts.

Le choix des experts est primordial pour la validation de notre approche. En effet, l'idéal est de trouver des personnes qui ont une bonne maîtrise des différents jeux de données géographiques relatifs au domaine de l'eau et de l'environnement en France, et aussi une expérience approfondie concernant leurs métadonnées et leurs processus de production.

Il est donc nécessaire d'identifier de tels experts, qui ne sont pas forcément dans le domaine de la recherche, et de construire avec eux un échange qui permet de bien se comprendre. Cela nécessite un dialogue avec eux et non pas juste les laisser répondre à un sondage.

Pour cela, nous avons pris la décision de diviser notre approche sous la forme de briques fonctionnelles qui seront évaluées indépendamment. Cela nous permettra

d'identifier pour chaque brique les experts qui correspondent le mieux à notre besoin. Chaque brique fonctionnelle sera évaluée grâce à une expérimentation conduite avec l'aide d'experts. Chaque expérimentation est suivie d'un entretien pour échanger avec eux et consolider leurs retours. Suite aux entretiens, des modifications sont parfois apportées au prototype pour intégrer au mieux leurs retours.

Ainsi, ce chapitre s'organise comme suit. Tout d'abord, la Section 4.2 présente en détail la méthodologie d'évaluation de notre approche. Ensuite, la section 4.3 décrira en détail comment nous avons produit un prototype du KG, et nous présenterons les algorithmes et les méthodes utilisées pour permettre d'exploiter le KG pour la recherche de jeux de données géographiques. La Section 4.5 s'attardera sur les différentes expérimentations qui ont été entreprises pour évaluer les différentes briques fonctionnelles qui ont été identifiées. La Section 4.6 conclura ce chapitre.

## 4.2 Méthodologie d'évaluation

La première phase de l'évaluation est l'implémentation d'un prototype du KG qui constituera une première base qui nous permettra d'échanger au mieux avec des experts.

Pour la deuxième phase de l'évaluation, nous avons contacté des experts des jeux de données du domaine de l'eau et de l'environnement en France et nous leur avons proposé un questionnaire pour évaluer plusieurs points de notre approche. Tout d'abord, nous avons évalué l'expressivité et l'ergonomie de Wikipédia combiné à Wikidata pour permettre à l'utilisateur de désigner un concept d'intérêt et formuler une requête liée au domaine de l'eau. Ensuite, nous avons évalué la capacité de notre approche à identifier les records pertinents et à organiser la présentation des résultats en groupes pour faciliter leur exploration. Finalement, pour évaluer les clusters de records, nous devons d'abord identifier les critères de comparaison des records.

Ensuite, à partir des retours de la deuxième phase, nous nous sommes intéressés à la tâche de fixer un seuil de similarité à partir duquel deux concepts de Wikidata sont trop éloignés pour être considérés comme étant pertinents. Pour cela, nous avons contacté trois experts spécialisés dans les domaines de l'eau et de l'environnement.

Finalement, la dernière expérimentation que nous avons menée a eu pour but d'évaluer la capacité du KG à identifier des records pertinents en tenant compte des retours que nous ont faits les experts dans les deux phases précédentes.

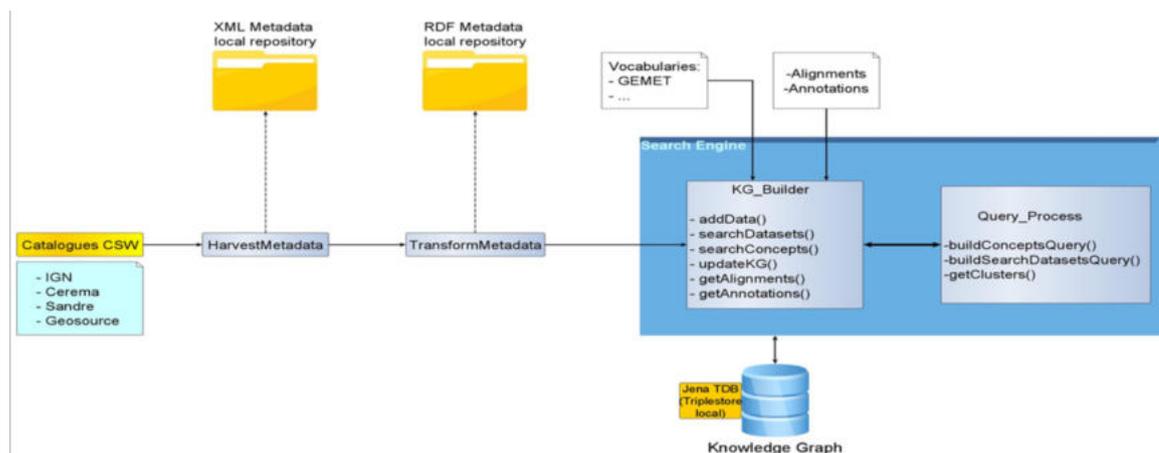


FIGURE 4.1 – Processus général de la création et de l’exploitation du KG

### 4.3 Implémentation et production du KG

Cette Section décrit une implémentation du KG que nous avons présenté dans le Chapitre 3. Cette première implémentation est une preuve de concept de la faisabilité du KG. Nous nous intéressons donc à la possibilité de moissonner et enrichir les métadonnées (c.-à-d. Processus hors-session) ainsi que la capacité du KG à permettre à l'utilisateur de formuler sa requête sous la forme d'un concept d'intérêt provenant de Wikidata, d'identifier des records pertinents pour l'utilisateur et finalement la possibilité de regrouper ces records sous la forme de clusters (c.-à-d. Session de recherche). La Figure 4.1 présente le processus général de la création et de l'exploitation du KG.

#### 4.3.1 Moissonner les records

Ici, nous décrivons tout d'abord la création du KG en nous basant sur l'approche développée dans la Section 3.2 et des ressources identifiées dans la Section 3.4. Le développement a été effectué en Java 11.

En se basant sur la Figure 3.6, les données du KG ont été réparties en plusieurs graphes nommés. Cela a pour but une meilleure lisibilité et répartition des données contenues dans le KG. Par exemple, les records bruts ont été intégrés à un graphe nommé "Records", les vocabulaires utilisés, et notamment GEMET, dans le KG ont été intégrés au graphe nommé "Vocabs", etc.

La première étape pour construire le KG est d'intégrer les records de métadonnées géographiques. Pour cela, nous avons utilisé le standard CSW pour collecter les records disponibles sur les catalogues Sandre et Cerema. Au total, 128 records Cerema et 188 records Sandre, ont été récupérés sous format XML et représentent donc l'ensemble de

records bruts.

Une fois les records bruts collectés, nous les transformons sous la forme de records RDF en utilisant un fichier<sup>1</sup> XSLT dédié, qui a été développé par SEMICEu<sup>2</sup> qui est un organe de la Commission européenne chargé d'assurer l'interopérabilité des données gouvernementales européennes sur le web sémantique, et qui permet de transformer des métadonnées conformes au standard ISO 19115 en des métadonnées conformes au standard GeoDCAT-AP. Les fichiers obtenus sont sous la forme RDF/XML.

Il faut aussi souligner que ces deux étapes peuvent être réunies en une seule (en omettant la phase de collecte des records bruts). En effet, il est possible de fournir le lien du service CSW d'un catalogue et de récupérer les records RDF sous format RDF/XML en utilisant le même fichier XSLT. Nous avons choisi d'exécuter cette tâche en deux temps pour pouvoir comparer les informations dans les records bruts et les records RDF.

#### 4.3.2 Intégration des records et de GEMET au KG

Ensuite, les records RDF ont été intégrés dans une base de données graphe Jena TDB (v3.16.0). Le choix de ce type de base de données s'explique par le fait Jena TDB est une solution proposée par Apache qui permet de stocker et de requêter des données RDF et qui est compatible avec toutes les API Apache Jena. Une base de données Jena TDB peut être décrite comme un Triple Store local sur une seule Machine et qui prend en charge des requêtes SPARQL pour accéder aux données. Il est aussi possible d'utiliser l'API Jena Fuseki pour facilement rendre disponible une base de données TDB en ligne ou sur plusieurs machines.

Les records RDF sont ajoutés dans un graphe nommé "Records" qui rassemblera tous les triplets permettant de décrire les jeux de données présents dans les catalogues choisis. L'utilisation des graphes nommés permet de mieux compartimenter les triplets du KG et d'optimiser le temps de calcul des requêtes SPARQL dessus.

Ensuite, le vocabulaire GEMET a été ajouté au KG. Pour cela, nous avons récupéré les versions française et anglaise de GEMET sous la forme de fichiers RDF et nous les avons ajoutés au KG dans un graphe nommé appelé "Vocabs".

---

1. <https://github.com/SEMICEu/iso-19139-to-dcat-ap>

2. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic>

```

SELECT DISTINCT ?gemet_concept ?wikidata_concept WHERE {

  Graph <http://www.example.com/Vocabs> {
    ?gemet_concept (skos:relatedMatch) ?dbpedia_concept.
    Filter(regex(str(?dbpedia_concept), "http://dbpedia.org" ))

    {
      SERVICE <DBPEDIA> {
        ?dbpedia_concept owl:sameAs ?wikidata_concept.
        FILTER(regex(str(?wikidata_concept), "www.wikidata.org"))
      }
    }
  }
}

SELECT DISTINCT ?gemet_concept ?wikidata_concept WHERE {

  Graph <http://www.example.com/Vocabs> {
    ?gemet_concept (skos:closeMatch) ?dbpedia_concept.
    Filter(regex(str(?dbpedia_concept), "http://dbpedia.org" ))

    {
      SERVICE <DBPEDIA> {
        ?dbpedia_concept owl:sameAs ?wikidata_concept.
        FILTER(regex(str(?wikidata_concept), "www.wikidata.org"))
      }
    }
  }
}

```

FIGURE 4.2 – Requêtes pour calculer l’alignement entre GEMET et Wikidata

### 4.3.3 Enrichissement des records et alignement des vocabulaires

Comme expliqué dans la Section 3.3, dans la mesure où l’utilisateur pourra exprimer sa requête sous la forme de concepts provenant de Wikidata, il est indispensable d’aligner ce dernier avec le vocabulaire GEMET. Pour cela, nous nous sommes basés sur deux alignements préexistants. Tout d’abord un alignement fourni dans GEMET et qui relie ce dernier à DBpedia et ensuite un deuxième alignement présent dans DBpedia qui le relie à Wikidata. L’idée est d’utiliser la transitivité pour associer les éléments GEMET et Wikidata à travers DBpedia. Ainsi, pour chaque triplet dans GEMET du type (conceptGEMET, skos:closeMatch, conceptDBpedia) et pour chaque triplet dans DBpedia du type (conceptDBpedia, owl:sameAs, conceptWikidata) nous pouvons déduire que (conceptGEMET, skos:closeMatch, conceptWikidata). Un raisonnement similaire permet de créer des triplets du type (conceptGEMET, skos:relatedMatch, conceptWikidata).

TABLE 4.1 – Liste des concepts GEMET bannis

Concept GEMET	Label GEMET	Propriété concernée
concept/1933	cyanure	skos :closeMatch
concept/11713	atrazine	skos :closeMatch
concept/6183	phénol	skos :closeMatch
concept/6346	plutonium	skos :closeMatch
concept/379	ammonium	skos :closeMatch
concept/375	ammoniac	skos :closeMatch
concept/6106	pentachlorophénol	skos :closeMatch
concept/5601	nitrite	skos :closeMatch
concept/8791	uranium	skos :closeMatch
concept/1168	dioxyde de carbone	skos :closeMatch
concept/776	benzène	skos :closeMatch
concept/4107	sulfure d'hydrogène	skos :relatedMatch
concept/11734	monoxyde de soufre	skos :relatedMatch
concept/8202	dioxyde de soufre	skos :relatedMatch

La Figure 4.2 présente la requête qui a été exécutée pour calculer les triplets composant l'alignement entre GEMET et Wikidata. Ensuite, ces triplets ont été ajoutés au graphe nommé "Alignement".

Il est à noter que certains concepts GEMET posent des problèmes pendant le calcul des triplets de l'alignement. Ils ont tous en commun le fait qu'ils sont des composants chimiques tels que le cyanure, l'atrazine, le sulfure d'hydrogène, etc. Cela est dû au fait que dans DBpedia, ces concepts alignés à un grand nombre de concepts Wikidata qui dépasse parfois 10.000 liens. Le problème est dû à l'alignement entre DBpedia et Wikidata.

Nous avons donc banni les 14 concepts concernés et nous avons calculé les triplets pour tous les autres concepts. La Table 4.1 présente la liste des concepts GEMET bannis. Il est possible d'aligner manuellement ces concepts avec Wikidata et de les ajouter au KG. Dans la mesure où les records RDF ne contiennent aucun de ces concepts, nous avons choisi de ne pas faire cette étape.

En se basant sur la Figure 3.6, et comme discuté dans la Section 3.3, la dernière étape pour constituer le KG est d'associer les différents mots-clés et thématiques présents dans les records à des concepts provenant de vocabulaires contrôlés. Nous avons développé deux approches différentes pour annoter les records. La première approche basée sur une identification automatique des thèmes et des mots-clés et des concepts GEMET à travers une requête SPARQL, et une deuxième basée sur une identification manuelle des thèmes et des mots-clés avec des concepts Wikidata. Pour cela, nous avons

```

SELECT DISTINCT ?themeLabel ?conceptGemetLabel ?conceptGEMET ?record WHERE{
  {
    GRAPH <http://www.example.com/Records> {
      ?record          a                               dcat:Dataset      ;
                      dcat:theme/skos:prefLabel     ?themeLabel      ;
    }
    GRAPH <http://www.example.com/Vocabs> {
      ?conceptGEMET    skos:prefLabel                ?conceptGemetLabel.
    }
    Filter( LANG(?themeLabel) = "fr" )
    Filter( Lang(?conceptGemetLabel) = "fr")
    Filter( (regex(" "+ str(?themeLabel), "[ ']+ str(?conceptGemetLabel), "i")))
  }
UNION
  {
    GRAPH <http://www.example.com/Records> {
      ?record          a                               dcat:Dataset      ;
                      dcat:theme/skos:prefLabel     ?theme            ;
                      dct:title                      ?title;
    }
    GRAPH <http://www.example.com/Vocabs> {
      ?conceptGEMET    skos:prefLabel                ?conceptGemetLabel.
    }
    Filter( LANG(?themeLabel) = "fr" )
    Filter( Lang(?conceptGemetLabel) = "fr")
    Filter(regex(" "+ str(?themeLabel), str(?conceptGemetLabel) + "[ ']", "i"))
  }
}

```

FIGURE 4.3 – Exemple de requête pour annoter les mots-clés et thèmes présents dans les records RDF

créé deux graphes nommés "AnnotationGEMET" et "AnnotationWikidata".

La première méthode se base sur une requête SPARQL qui permet d'associer les labels des thèmes et les labels de chaque concept GEMET. La Figure 4.3 présente la requête utilisée pour cette méthode. Pour chaque record, on sélectionne les labels en français de chaque thème (`?themeLabel`) qui sera par la suite comparé au label français de chaque concept GEMET présent dans le KG (`?conceptGemetLabel`). Si le label du thème et le label du concept GEMET ont un mot en commun, le triplet (`?record, dcat :theme, ?conceptGEMET`) est créé et ajouté au graphe nommé "AnnotationGEMET". Une méthode similaire a été appliquée aux mots-clés en sélectionnant les mots-clés à la place des thèmes dans la requête de la Figure 4.3.

La deuxième méthode pour annoter et enrichir les métadonnées se base sur une identification manuelle de chaque mot-clé et thème présents dans les records avec le concept Wikidata qui lui correspond. Ainsi, nous avons extrait tous les mots-clés et les mots-clés distincts présents dans les métadonnées. Pour chacun nous avons effectué une recherche sur Wikidata pour trouver l'associer à un concept Wikidata. Ensuite,

Concept Utilisateur (Wikidata) :

rivière Rechercher Concept

Selectionner les concepts d'intérêt :

Select	Qnode	Label Wikidata	Description Wikidata
<input type="checkbox"/>	Q1434279	rivière	figure héraldique@fr
<input type="checkbox"/>	Q4022	rivière	cours d'eau naturel@fr
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			

Comparer avec les records

FIGURE 4.4 – Exemple de résultats de recherche d'un concept à travers une interface utilisateur

ont été ajoutés au graphe nommé "AnnotationWikidata" des triplets du type (`?record, dcat :theme, ?conceptWikidata`) et (`?record, dcat :keyword, ?conceptWikidata`)

## 4.4 Exploitation du KG lors d'une session de recherche

Cette partie s'intéresse à l'exploitation du KG une fois qu'il a été créé. En se basant sur la Figure 3.1, dans cette partie nous allons nous concentrer sur la session de recherche et à décrire les différentes étapes qui la composent. Les étapes sont spécifiées dans la Figure 3.3.

### 4.4.1 Identification du concept utilisateur

Pour permettre à l'utilisateur de choisir son concept d'intérêt parmi des concepts Wikidata, il est nécessaire d'envoyer une requête SPARQL à travers l'Endpoint pour trouver les concepts qui correspondent au besoin de l'utilisateur. Il est nécessaire d'accompagner l'utilisateur dans la désignation d'un concept non ambigu, à ces fins nous avons choisi la combinaison Wikipedia-Wikidata. Wikipedia permet à l'utilisateur d'entrer dans l'ontologie Wikidata sans avoir d'expertise spécifique en ontologie. La Figure 4.4 en donne un exemple. On présente à l'utilisateur l'identifiant Wikidata du concept (i.e. *Qnode*), le label du concept ainsi que la description qui lui est associée.

La Figure 4.5 présente la requête utilisée pour retrouver un concept à partir d'une chaîne de caractères fournie par l'utilisateur. La variable `chaîne_de_caractère` est remplacée par la chaîne de caractère fournie par l'utilisateur. Il est important de ne pas extraire uniquement le label (i.e. *rdfs :label*) du concept, mais aussi les labels alternatifs

```
SELECT DISTINCT ?concept ?conceptLabel ?description WHERE{
  ?concept schema:description ?description;
  rdfs:label|rdfs:altLabel chaîne_de_caractere
SERVICE wikibase:label {bd:serviceParam wikibase:language "fr,en"}
}
```

FIGURE 4.5 – Requête pour trouver un concept Wikidata à travers son label

(i.e. *rdfs:altLabel*). En effet, il n'est pas rare que les labels alternatifs contiennent des abréviations ou d'autres appellations pour un même concept. Ainsi, pour le concept Wikidata "Q3391269" dont le label est "Plan de prévention du risque inondation", mais qui contient quatre labels alternatifs dont "PPRI" qui est une abréviation très répandue pour décrire le concept. Cela permet d'améliorer les résultats de recherche lors de la recherche d'un concept Wikidata et ainsi mieux assister l'utilisateur à faire son choix.

#### 4.4.2 Sélection et classement des records potentiellement pertinents

Une fois que l'utilisateur a choisi un concept d'intérêt, il faut choisir parmi les records présent dans le KG un ensemble de records candidats. Pour cela, une requête est effectuée sur le KG où sont extraites toutes les informations pertinentes pour présenter chaque jeu de données à l'utilisateur.

La Figure 4.6 présente la requête utilisée pour sélectionner les records depuis le KG. Cette dernière permet de récupérer l'identifiant (*?record*), le titre (*?title*), la description (*?description*) et la couverture spatiale d'un jeu de données (*?bbox*).

A cela s'ajoutent les mots-clés et le thème qui ont été associés à chaque record. Si le thème ou le mot-clé appartient aux annotations GEMET (*?theme\_gemet, ?keyword\_gemet*), ce dernier est récupéré ainsi que le concept Wikidata avec lequel il est aligné (*?themeWikidataURI, ?keywordWikidataURI*). La requête récupère aussi les concepts Wikidata qui ont été identifiés manuellement (*?wikidataTheme, ?wikidataKeyword*).

On sélectionne ensuite uniquement les records qui contiennent au moins un thème ou un mot-clé qui est assez proche du concept d'intérêt choisi par l'utilisateur pour constituer une liste de records candidats. En effet, si un record ne contient aucun mot-clé ni thème assez proche de celui choisi par l'utilisateur, on peut supposer que ce dernier n'est pas un record candidat qui pourrait l'intéresser. Pour cela, chaque mot-clé et thème des records est comparé au concept utilisateur grâce à la mesure de similarité sémantique "text" en utilisant l'outil KGTK présenté dans la Section 3.4. Pour fixer

```

SELECT DISTINCT ?title ?theme_gemet ?keyword_gemet ?themewikidataURI
?themewikidataLabel ?keywordWikidataURI ?keywordWikidataLabel
?record ?bbox ?description ?wikidataKeyword ?wikidataTheme WHERE{

    GRAPH <http://www.example.com/Records> {
        ?record          a                               dcat:Dataset      ;
                        dct:description                 ?description     ;
                        dct:spatial/dcat:bbox           ?bbox            ;
                        dct:title                       ?title           ;
    }
    FILTER(DATATYPE(?bbox)= <http://www.opengis.net/ont/geosparql#wktLiteral>)
    GRAPH <http://www.example.com/AnnotationGEMET> {
        ?record dcat:theme ?theme_gemet;
        dcat:keyword ?keyword_gemet.
    }
    GRAPH <http://www.example.com/AnnotationWikidata>{
        ?record dcat:theme ?wikidataTheme.
        ?record dcat:keyword ?wikidataKeyword
    }
    GRAPH <http://www.example.com/Alignment> {
        ?theme_gemet skos:closeMatch|skos:relatedMatch ?themewikidataURI .
        ?themewikidataURI rdfs:label ?themewikidataLabel.
        ?keyword_gemet skos:closeMatch|skos:relatedMatch ?keywordWikidataURI.
        ?keywordWikidataURI rdfs:label ?keywordWikidataLabel.
    }
}

```

FIGURE 4.6 – Requête pour extraire les records depuis le KG

le seuil à partir duquel deux concepts sont trop éloignés, une expérimentation a été conduite dans le Chapitre 4.

Une fois la liste des records candidats créée, chaque record de la liste est comparé au concept d'intérêt de l'utilisateur. Pour cela, chaque keyword et thème est comparé au concept utilisateur, et une moyenne est calculé pour associer un score de pertinence entre le concept utilisateur et chaque record. Ensuite, la liste des records candidats est classée par ordre décroissant du score de similarité global et renvoyée à l'utilisateur.

#### 4.4.3 Regrouper des records de jeux de données géographiques

Regrouper des records présuppose d'avoir un moyen fiable de les comparer. Comme discuté dans la Section 3.3, pour comparer deux records nous allons tout d'abord comparer les champs qui les composent deux à deux, pour ensuite utiliser une mesure de comparaison multicritères pour obtenir un score de similarité entre records et finalement utiliser un algorithme de clustering pour en déduire des regroupements des records similaires.

Pour comparer deux records, nous nous basons sur la comparaison de trois champs

différents : les mots-clés, les thèmes et la couverture spatiale. Pour comparer les thèmes et les mots-clés, nous utiliserons une mesure de similarité sémantique, à savoir la mesure "text" alors qu'on utilisera une mesure de similarité géographique de Hausdorff pour comparer les couvertures spatiales des deux records.

Comparer les mots-clés de deux records revient à déduire une similarité globale entre deux ensembles de mots-clés à partir des similarités des concepts deux à deux. Plusieurs méthodes peuvent être utilisées pour généraliser le calcul de similarité sur un ensemble de mots-clés. La manière la plus simple d'y arriver est de calculer une moyenne sur toutes les paires de similarités. Le même raisonnement peut être appliqué au thème des deux records pour en déduire une similarité globale de toutes les thématiques de deux records.

Une fois obtenues les trois similarités, à savoir similarité entre mots-clés, entre mots-clés et géographique, il est possible d'utiliser l'algorithme de similarité multicritère TOPSIS. Comme expliqué dans la Section 2.3, il faut définir les critères de comparaison ainsi que le poids associé à chacun de ces critères et en sortie, TOPSIS fournit une similarité globale pour chaque paire de records dans le KG.

Ensuite, un graphe complet dont les sommets sont les records du KG. Chaque arête du graphe est pondérée avec la similarité entre les deux records qui la composent. Ce graphe de similarités entre records peut être utilisé ensuite pour calculer des regroupements de records. Pour cela, nous utilisons l'algorithme Girvan-Newman présenté dans la Section 2.3.1. Nous obtenons donc en sortie un ensemble de clusters de records de métadonnées présents dans le KG. La Figure 4.7 montre un exemple des clusters obtenus représentés sous la forme d'un graphe, où chaque sommet représente un record.

## 4.5 Expérimentations

Cette Section a pour but de décrire les expérimentations qui ont été menées dans le cadre de cette thèse dans le but d'évaluer la pertinence de notre approche. Chaque expérimentation correspond à une brique fonctionnelle décrite dans la Section 4.2.

### 4.5.1 première évaluation des fonctionnalités de session

Une première expérimentation, sous la forme d'un questionnaire, a été menée pour évaluer la pertinence d'une approche fondée sur un KG pour faciliter la recherche de

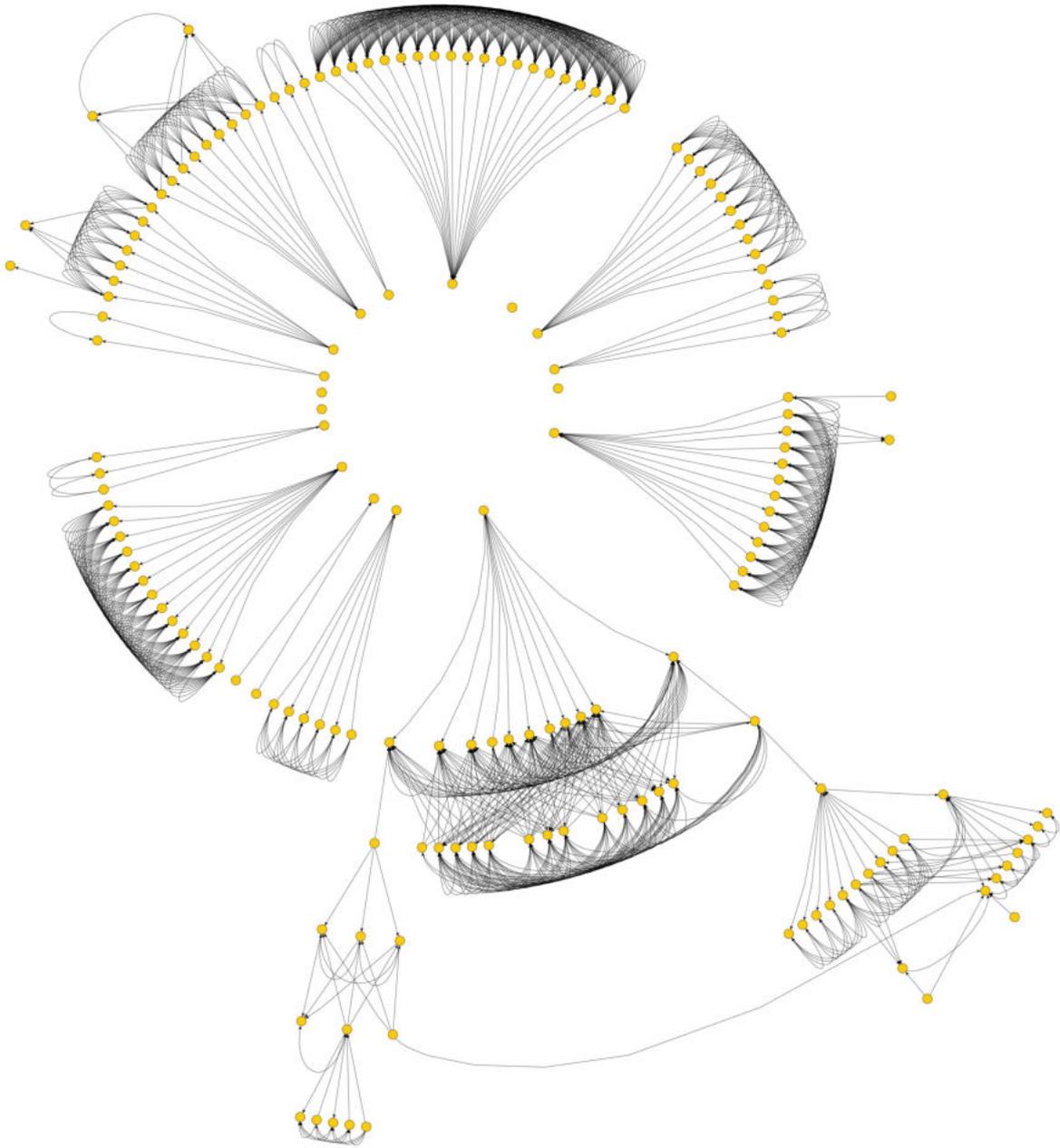


FIGURE 4.7 – Exemple de clusters de métadonnées du KG représentés sous la forme d'un graphe

jeux de données géographiques. Le questionnaire a été soumis à deux experts connaissant à la fois les métadonnées géographiques ainsi que le domaine de l'eau.

Le questionnaire se compose de trois tâches soumises aux experts. Ces tâches ont pour but de répondre à la phase 1 de la méthodologie d'évaluation présentée dans la Section 4.2. Dans cette section, nous présenterons tout d'abord les tâches qui ont été proposées aux experts, un aperçu des réponses de leurs réponses ainsi que notre analyse. Le questionnaire, tel qu'il a été présenté aux experts est disponible en Annexe A. Les réponses complètes seront disponibles en Annexe B.

#### 4.5.1.1 Tâche 1 : Pertinence de Wikipédia-Wikidata pour identifier des concepts utilisateur

Le but de cette première tâche est d'investiguer la pertinence d'utiliser Wikidata-Wikipédia comme moyen permettant à l'utilisateur d'exprimer sa requête. Notre but est de nous assurer que Wikidata contient autant de concepts relatifs aux domaines de l'eau et de l'environnement que possible. Pour cela, nous avons proposé aux experts trois méthodes pour trouver des concepts sur Wikidata, à savoir, utiliser le moteur de recherche présent sur le site de Wikidata, utiliser une requête SPARQL ou encore retrouver la page Wikipédia correspondant à un concept et par la suite trouver le concept Wikidata associé à cette page.

La question qui a été posée aux experts a été formulée comme suit :

##### **Tâche 1 :**

Nous voudrions évaluer si l'utilisation de Wikidata permet bien à l'utilisateur d'exprimer sa requête. Pouvez-vous penser à quelques concepts se rapportant au domaine de l'eau et de l'environnement et essayer de trouver les éléments Wikidata correspondants en suivant les indications ci-après. Avez-vous réussi à les trouver ? Quels concepts n'avez-vous pas réussi à trouver ?

La réponse de l'expert 1 s'est concentrée sur trois concepts qu'il a eu du mal à trouver sur Wikidata. Tout d'abord, le concept "pollution des sédiments" qui n'existe pas sur Wikidata. Il a identifié un concept proche, à savoir "pollution de l'eau", mais qui ne correspond pas exactement à sa requête. Ensuite, le concept "marais artificiel" qui n'a pas été trouvé en effectuant une recherche directement sur Wikidata malgré le

fait que ce dernier existe bien. L'expert a dû utiliser le label en anglais pour réussir à trouver le bon concept (Q1376406), car le label en français associé au concept est "phytoépuration". Un dernier concept n'a pas été trouvé par l'expert 1, à savoir, "Plan de prévention des risques inondation" à cause du libellé français qui est légèrement différent : "Plan de prévention du risque inondation".

L'expert 2 a eu du mal à trouver deux concepts. Tout d'abord "nappe phréatique" qui n'a pas pu être retrouvé directement en utilisant le moteur de recherche Wikidata alors que le concept existe bel et bien (Q583075). Ensuite le concept "station d'épuration" n'a pas été retrouvé.

En se basant sur ces retours, on peut dire que Wikidata est plus facilement explorable en passant par Wikipedia qui permet de trouver des concepts difficilement trouvables directement en utilisant l'interface Wikidata (cf. Pollution de l'eau).

Des limites ont été identifiées. Il faut parfois passer par la version anglaise. Les labels alternatifs sont utiles pour l'expression de concepts utilisateurs, mais ne sont pas présents dans Wikipédia. La plus grosse limite est que certains concepts n'existent pas dans Wikidata à la date de l'expérimentation ou ne se présentent pas sous la forme d'une page du même nom reliée à un concept Wikidata. Par exemple, le concept "pollution des sédiments" n'existe pas dans Wikidata.

D'autres concepts sont plus difficiles à trouver même en passant par Wikipedia. Par exemple, pour trouver le "plan de prévention du risque inondation", il faut prendre en compte les labels alternatifs dans Wikidata. Dans le cas de "station d'épuration", la page Wikipédia qui lui est associée renvoie directement vers la page Wikipédia "Traitement des eaux usées".

Le cas du concept "marais artificiel" est intéressant. En effet, le concept existe bien dans Wikidata avec un label français "Phytoépuration" alors que la page Wikipédia associée a comme titre "Lagunage". De plus, il n'y a aucun label alternatif associé au concept dans Wikidata. Dans ce cas, il faut avoir une bonne maîtrise du concept et de ses différents synonymes, ou effectuer une recherche en utilisant les labels en anglais.

#### 4.5.1.2 évaluer la capacité du moteur à pointer vers des records pertinents en lien avec un concept utilisateur

Pour cette deuxième tâche, nous avons procédé en deux temps. Notre premier objectif était de nous assurer que les catalogues choisis correspondaient bien au domaine d'application et d'avoir l'avis de nos experts à propos des jeux de données qu'ils contiennent. Pour cela, nous avons posé la question suivante :

**Tâche 2.1 :**

Que pensez-vous du choix des catalogues ? Pensez-vous que le KG devrait inclure des jeux de données provenant d'autres catalogues ?

À cette question, l'expert 1 a répondu que le choix des catalogues est pertinent uniquement pour de la cartographie (c.-à-d. localisation de concepts). Il a aussi ajouté que la recherche par concepts (cf. Tâche 1) échoue sur le Cerema et que le Sandre était systématiquement en Timeout. Ainsi, si un concept n'est pas dans les métadonnées, les moteurs de recherche ne trouvent rien. Ensuite, l'expert 1 pense que le KG devrait contenir plus de jeux de données pour évaluer la réponse fournie par le graphe.

L'expert 2 précise que le Sandre est le catalogue du Système d'information sur l'eau et il est donc très complet pour tout ce qui concerne les données publiques sur l'eau alors que le Cerema n'indexera que des jeux de données qu'il produit en interne. L'expert pense qu'il manque des jeux de données des collectivités et que l'idéal serait d'utiliser un catalogue national comme le Géocatalogue ou data.gouv.fr.

Ainsi, les deux experts sont d'accord quant à la pertinence des catalogues choisis, mais pensent que le KG serait plus complet si d'autres jeux de données provenant d'autres catalogues y étaient inclus et notamment des jeux de données produites ou concernant des communes. En effet, le prototype actuel contient essentiellement des jeux de données dont la couverture spatiale est nationale, régionale ou départementale.

La deuxième partie de la tâche 2 concerne les résultats obtenus en effectuant une recherche grâce au KG. Notre but est d'évaluer quels résultats obtenus en effectuant une recherche de jeux de données avec le KG sont pertinents et de quantifier la capacité de ce dernier à identifier des records que les moteurs de recherche utilisés dans le Sandre et le Cerema ne présentent pas dans leur réponse. Pour cela, nous avons exécuté trois requêtes en utilisant le KG :

- “Hydrographie” (Q182468)
- “Estuaire” (Q47053)
- “Rhône” (Q3429695)

Le choix de ces requêtes a été fait de manière à avoir un concept très général (c.-à-d. hydrographie), un concept plus précis (c.-à-d. estuaire) et une entité nommée (i.e., Rhône).

Pour chaque requête utilisateur, nous avons fourni à nos experts les 20 meilleurs résultats de recherche obtenus grâce au KG et nous leur avons demandé d’identifier lesquels sont pertinents et lesquels sont manquants. Nous nous intéressions aussi à la méthode avec laquelle les experts effectuaient leurs recherches. C’est pourquoi, nous avons posé les questions suivantes à nos experts :

**Tâche 2.2 :**

Quels jeux de données identifiés vous paraissent pertinents ?  
Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ? Y-a-t-il des jeux de données manquants ?

Les réponses détaillées de nos deux experts sont disponibles en Annexe B.

Tout d’abord, concernant les méthodes avec lesquels les experts ont effectué leurs recherches, il est intéressant de remarquer que l’expert 1 s’est aidé de la base de connaissances WikHydro qui permet de retrouver des concepts liés au domaine de l’eau ainsi que les textes qui y font référence. Les deux experts ont utilisé les moteurs de recherche mis à disposition par les portails du Sandre et du Cerema pour effectuer les requêtes que nous leur avons demandées.

## Hydrographie

Pour cette première requête, l’expert 1 a trouvé 35 résultats sur le portail du Cerema. L’expert 2 a trouvé le même nombre de résultats en entrant la chaîne de caractère

<b>Thèmes INSPIRE</b>	
<b>Mots-clés</b>	<ul style="list-style-type: none"> <li>• houle</li> <li>• état de mer</li> </ul>
<b>GEMET - INSPIRE themes, version 1.0</b>	<ul style="list-style-type: none"> <li>• Caractéristiques géographiques océanographiques</li> </ul>
<b>OPENDATA</b>	<ul style="list-style-type: none"> <li>• données ouvertes</li> </ul>
<b>theme.prodige.rdf</b>	<ul style="list-style-type: none"> <li>• Données ouvertes</li> <li>• Gestion du littoral et de la mer</li> </ul>
<b>INSPIRE Service taxonomy</b>	<ul style="list-style-type: none"> <li>• infoFeatureAccessService</li> </ul>

FIGURE 4.8 – Extrait des métadonnées associées au jeu de données :Réseau de mesures de houle du Centre d'Archivage National des Données de Houles In-Situ (CANDHIS) - France métropole et outre-mer

"Hydrographie" dans la barre de recherche. En filtrant les résultats en cochant le thème "Eaux intérieures", le nombre de résultats est 13. Concernant le Catalogue Sandre, l'expert 2 a obtenu 98 solutions en entrant la chaîne de caractères "hydrographie" dans la barre de recherche. "hydrographie" a été reconnue par le moteur de recherche. Les résultats concernent essentiellement la BD Carthage et la BD Topage, mais aussi les zonages réglementaires et de gestion, les stations de mesure, etc.

Du côté des résultats obtenus par le KG, parmi les 20 résultats proposés, l'expert 1 a jugé 6 résultats comme étant pertinents alors que l'expert 2 a jugé 3 résultats comme étant pertinents. L'expert 2 a souligné qu'en l'absence du contexte dans lequel la requête a été effectuée, il lui a été difficile de décider quels jeux de données sont pertinents. Pour cette requête, comme le concept choisi est très général, il s'est restreint à une vision très topographique et physique en excluant les jeux de données concernant la gestion de l'eau par exemple.

## Estuaire

Pour cette deuxième requête, nos deux experts ont trouvé 7 résultats sur le portail du Cerema. L'expert 2 a trouvé 3 résultats sur le portail du Sandre. Du côté des résultats obtenus grâce au KG, parmi les 20 résultats envoyés aux experts, l'expert 1 a identifié 8 résultats comme étant pertinents alors que l'expert 2 a identifié 19 résultats comme étant pertinents pour le concept d'intérêt "estuaire". Pour cette requête, l'expert 2 a considéré que comme la requête était plus précise, les résultats devaient aussi inclure les aspects relatifs à la gestion.

Il est intéressant de noter que pour cette requête, le KG a réussi à identifier un

record qui n'a été pas été identifié par les moteurs de recherche des portails du Sandre et du Cerema à savoir "réseau de la mesure de la houle". Ceci s'explique que contrairement à d'autres jeux de données, les métadonnées associées au "Réseau de mesures de houle du Centre d'Archivage National des Données de Houles In-Situ (CANDHIS) - France métropole et outre-mer"<sup>3</sup> ne contiennent à aucun moment la chaîne de caractère "Estuaire" comme on peut le voir dans la Figure 4.8 .

## Rhône

Pour le concept "Rhône", l'expert 1 trouve 36 résultats sur le Cerema. La majorité de ces résultats concernent l'occupation du sol. L'expert 2 a trouvé deux jeux de données en tapant "Rhône" dans la barre de recherche et sélectionnant le thème "Eaux intérieures" sur la barre latérale de recherche. L'expert 2 a obtenu 3 résultats en tapant "Rhône" dans la barre de recherche.

Concernant les résultats obtenus par le KG, l'expert 1 a identifié un jeu de données comme étant pertinent alors que l'expert 2 a identifié 3 résultats comme étant pertinents.

Nous remarquons que les experts ne s'accordent pas sur la notion de pertinence. En effet, cette dernière dépend du contexte autour de la recherche effectuée par l'utilisateur. L'expert 2 a choisi de considérer la pertinence dans un sens plus large pour rendre visible une variété de records qui comprennent des applications différentes englobant des records liés aux risques et à la gestion, surtout pour la requête 2. L'expert 1 a eu plus tendance à ne considérer que les records directement liés aux concepts utilisateurs.

En analysant le KG à la lumière des retours des experts, on peut se rendre compte que les concepts GEMET identifiés dans les métadonnées ne sont pas suffisants. En effet, dans le cas de plusieurs records, tous les mots-clés et thèmes n'ont pas été identifiés, et il est donc normal que le KG n'arrive pas à identifier certains records pertinents. Il serait donc intéressant de considérer une autre expérimentation où on utiliserait l'identification manuelle des mots-clés et des thèmes avec des concepts présents dans Wikidata. De plus, certains records pertinents ont été pénalisés par le grand nombre de thèmes et de mots-clés qu'ils contiennent. Dans la mesure où nous utilisons une moyenne pour calculer la pertinence globale d'un record, il n'est pas rare que des records n'ayant qu'un seul mot-clé se retrouvent bien classés.

---

3. <https://www.cdata.cerema.fr/geonetwork/srv/fre/catalog.search#/metadata/a5ded6f2-86f7-4db3-bb10-f0f621eee920>

Néanmoins, le KG a été capable d'identifier un record pertinent que les catalogues n'ont pas réussi à identifier. Cela s'explique par l'absence du terme "estuaire" dans le record en question, mais que ce dernier contienne des concepts proches du concept de la requête. Ceci est un premier indice pour identifier les situations pour lesquelles notre approche réussit à identifier des records pertinents que les approches existantes échouent à identifier.

#### 4.5.1.3 Évaluer le regroupement des records

Comme présenté dans la Section 3.2, une fonctionnalité visée par le KG est la capacité de regrouper des records sous la forme de clusters. C'est pourquoi la tâche 3 de ce questionnaire avait pour but d'évaluer la cohérence des regroupements calculés par le KG et de collecter l'avis des experts concernant la bonne méthode pour regrouper des records de jeux de données géographiques.

Pour cela, nous avons utilisé la méthode décrite dans la Section 4.4.3 pour calculer les clusters des records. Pour cela nous nous sommes basés sur trois critères et nous avons associé à chacun d'entre eux un poids représentant l'importance à chaque critère :

- similarité géographique : 0.2
- similarité sémantique entre thèmes : 0.3
- similarité sémantique entre mots-clés : 0.5

Au total, 59 clusters ont été calculés. Nous avons présenté aux experts les 11 clusters contenant tous les records qui sont apparus dans les résultats de recherche de la requête 2 de la tâche 2.1. et nous leur avons posé la question suivante :

**Tâche 3 :**

Les regroupements proposés vous semblent-ils cohérents ? Comment auriez-vous regroupé ces records ?

Les experts ont eu du mal pour comprendre la cohérence derrière les groupes qui leur ont été proposés et ont soulevé plusieurs questions. L'expert 1 n'a pas réussi à

comprendre les critères selon lesquels les groupes ont été créés et il a jugé qu'ils ne sont pas cohérents géographiquement parlant. L'expert 2 a aussi trouvé des difficultés pour saisir la manière dont ont été créés les groupes. Plusieurs d'entre eux semblent contenir des records assez mélangés sauf pour les groupes 1, 5, 10 et 11. Cela implique que pour les regroupements des records, un travail doit être mené pour améliorer la compréhension de ces derniers et de simplifier leur visualisation.

Concernant la deuxième question, l'expert 1 propose en plus des thématiques et des mots-clés de regrouper les records selon l'emprise géographique d'abord et ensuite la temporalité. L'expert 2 préconise de créer des groupes selon les thématiques, par exemple, un cluster pour les données marines, un autre pour l'hydrographie terrestre, un troisième pour les réseaux de mesures, etc. Les deux experts signalent que ce ne sont pas les seules manières de créer des clusters de records.

#### 4.5.2 Expérimentation : Fixer un seuil de similarité entre deux concepts

Le but de cette expérimentation est de fixer un seuil de similarité à partir duquel on peut affirmer que deux concepts sont suffisamment proches pour que le concept présent dans les métadonnées (c.-à-d. mot-clé ou thème) soit utilisé pour reformuler la requête de l'utilisateur, ou de manière semblable, un seuil à partir duquel, deux concepts sont trop éloignés pour être considérés comme étant proches. Pour cela, nous avons créé un questionnaire que nous avons adressé à 3 experts du domaine de l'eau et de l'environnement en France.

Nous avons repris les concepts Wikidata identifiés par les experts dans l'expérimentation 1, à savoir "nappe phréatique", "phytoépuration", "PPRI" et "station de traitement des eaux usées". Nous avons comparé les quatre concepts avec tous les concepts présents dans les métadonnées (c.-à-d. les concepts présents dans les graphes nommés "AnnotationGEMET" et "AnnotationWikidata") en utilisant la mesure de similarité sémantique "text". Nous avons classé les résultats par ordre décroissant de similarité et nous avons présenté à nos experts les 50 meilleurs résultats. Le questionnaire complet est disponible en Annexe C.

Pour chaque résultat de comparaison, nous avons fourni le label Wikidata ainsi que le label GEMET quand ce dernier est présent dans le KG. Nous avons aussi pré-calculé les chemins dans Wikidata entre le concept utilisateur et le concept des métadonnées quand ce dernier existe. Le chemin que nous avons calculé n'est donné qu'à titre indicatif dans la mesure où nous utilisons la mesure "text" pour comparer les concepts. En

effet, la mesure de similarité sémantique "text" se base sur une représentation vectorielle des concepts Wikidata.

Nous avons posé trois questions à nos experts :

**Tâche :**

1. Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté ?
2. À partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?
3. Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?

Les réponses complètes des experts seront disponibles en Annexe D

#### 4.5.2.1 Requête 1 : Nappe phréatique

Le premier concept utilisateur que nous avons soumis aux experts est le concept Wikidata (Q583075) qui correspond à "nappe phréatique".

L'expert 1 a classé et identifié 5 concepts comme étant pertinents, à savoir, "nappe phréatique", "aquifère", "eau souterraine", "niveau piézométrique" et "hydrogéologie". Les scores de similarité pour ces concepts sont compris entre 1 et 0.69. Le concept hydrogéologie, est classé 11<sup>ème</sup> dans la liste d'origine. L'expert 1 nous a aussi fourni une liste de concepts pas assez proches, mais qui englobent le concept "nappe phréatique", à savoir, eau potable, prélèvement d'eau, pluie, perméabilité, gestion de l'eau et ressource naturelle.

L'expert 2 a jugé que l'ordre global de la liste n'est pas respecté. Il a aussi identifié deux concepts comme étant proches de "nappe phréatique", à savoir nappe phréatique et aquifère (1<sup>er</sup> et 2<sup>ème</sup>), tous les autres étant trop éloignés du concept utilisateur.

L'expert 3 place la limite de pertinence aux 3 premiers résultats. Il nous a aussi fourni des indications concernant l'ordre de certains concepts qui, selon lui, devrait être classés plus haut. Par exemple, le concept "cours d'eau" qui devrait être en 6<sup>ème</sup> position au lieu de la 8<sup>ème</sup> ou encore "masse d'eau" qui devrait être mieux classée, car

sémantiquement plus proche de "nappe phréatique". L'expert 3 nous a aussi indiqué des concepts qui ne sont pas assez proches, mais qui mériteraient d'être retrouvés comme des résultats à cause de leur caractère englobant, à savoir, "qualité de l'eau" et "gestion de l'eau".

Il en ressort de cette première requête que les premiers résultats sont pertinents. Si les experts 2 et 3 ont jugé pertinents uniquement les meilleurs résultats, l'expert 1 a jugé que d'autres concepts moins bien classés étaient aussi similaires pour le concept "nappe phréatique". Pour cette requête, le score de similarité minimale pour prendre en compte tous les concepts identifiés pertinents par tous les experts est 0.69, qui correspond au score de similarité avec le concept "hydrogéologie".

Les experts 1 et 3 ont identifié un ensemble de concepts qui ne sont pas directement pertinents par rapport au concept de l'utilisateur, mais qui peuvent être intéressants par leur caractère englobant du concept "nappe phréatique". Si on veut aussi prendre en compte ces concepts, le score minimal de similarité à considérer est de 0.56 et qui correspond au concept "perméabilité".

Concernant l'ordre global des scores de similarité, les 3 experts ont jugé que le classement au niveau des 50 concepts proposés n'était pas satisfaisant. Ils ont aussi jugé que les chemins n'apportent pas une information utile.

#### 4.5.2.2 Requête 2 : Phytoépuration

Le deuxième concept utilisateur que nous avons soumis aux experts est le concept Wikidata (Q1376406) qui correspond à "phytoépuration".

L'expert 1 a jugé pertinents 6 concepts, à savoir, "pluie" (1<sup>er</sup>), "eaux usées" (8<sup>ème</sup>), "qualité de l'eau" (35<sup>ème</sup>), "gestion de l'eau" (11<sup>ème</sup>), "pollution" (14<sup>ème</sup>) et "assainissement" (15<sup>ème</sup>). Ils nous ont aussi fourni une liste de concepts pas assez proche, mais qui englobent les concepts utilisateur suivants : "ouvrage hydraulique", "prélèvement d'eau", "aménagement urbain", "gestion des déchets", "environnement" et "polluant".

L'expert 2 a jugé que seuls les 4 premiers concepts étaient pertinents, et qu'à partir du 5<sup>ème</sup> résultats, les concepts n'étaient plus proche du concept d'intérêt.

L'expert 3 a identifié 3 concepts comme étant sémantiquement proches : "eaux usées" (8<sup>ème</sup>), "assainissement" (15<sup>ème</sup>) et "qualité de l'eau" (35<sup>ème</sup>). Il souligne le fait

que les concepts pertinents sont noyés dans les résultats et qu'ils devraient apparaître plus haut dans le classement.

Pour cette requête, bien que des concepts pertinents ont été trouvés à chaque fois par les experts parmi la liste proposée ces derniers ne sont pas tous bien classés. Si les experts 1 et 2 ont jugé que le premier résultat est pertinent pour la requête utilisateur, l'expert 3 pense que le meilleur concept est trop éloigné de la requête utilisateur.

Si on veut fixer un score minimal de similarité qui prend en compte tous les concepts jugés pertinents par au moins un expert, il faut fixer le seuil à la valeur 0.56 qui correspond à la similarité avec le concept "qualité de l'eau".

Les 3 experts ont jugé que l'ordre général de classement des scores de similarité n'était pas satisfaisant. Cela se ressent le plus chez l'expert 3 pour qui le meilleur résultat ne vient qu'en 8<sup>ème</sup> position.

#### 4.5.2.3 Requête 3 : Plan de prévention du risque inondation

Le troisième concept utilisateur que nous avons soumis aux experts est le concept Wikidata (Q3391269) qui correspond à "plan de prévention du risque inondation".

L'expert 1 a jugé pertinents 8 concepts à savoir "plan de prévention du risque inondation" (1<sup>er</sup>), "crue" (28<sup>ème</sup>), "submersion marine" (2<sup>ème</sup>), "bassin versant" (11<sup>ème</sup>), "prévention", "gestion de crise" (33<sup>ème</sup>), "risque naturel" (34<sup>ème</sup>) et "catastrophe naturelle" (44<sup>ème</sup>). Il a aussi identifié 7 autres concepts englobants qui ne sont pas directement reliés à la requête, mais qui sont intéressants.

L'expert 2 n'a pas pu répondre à la question en raison de la mention "label GEMET non trouvé".

L'expert 3 a indiqué que 7 concepts étaient pertinents, à savoir, "plan de prévention du risque inondation" (1<sup>er</sup>), "submersion marine" (2<sup>ème</sup>), "protection" (3<sup>ème</sup>), "masse d'eau" (4<sup>ème</sup>), "gestion du risque" (9<sup>ème</sup>), "crue" (28<sup>ème</sup>) et "risque naturel" (34<sup>ème</sup>).

Cette requête concerne un concept très relié au domaine de l'eau en France. En effet, le plan de prévention du risque inondation est un outil permettant la mise en oeuvre de la politique de l'état Français en matière de prévention des inondations. Son but est de mettre en évidence les zones à risque, les réglementations de l'amé-

nagement et l'occupation des sols, etc. Bien que le concept existe dans Wikidata, ce dernier est très isolé dans le graphe de connaissance Wikidata. En effet, ce dernier a comme unique parent le concept "plan de prévention des risques" (Q3391265) qui n'a à son tour qu'un unique concept parent qui est "servitude d'utilité publique" (Q3480546).

Pour fixer un score minimal de similarité qui prend en compte tous les concepts jugés pertinents par au moins un expert, il faut fixer le seuil à la valeur 0.51 qui correspond au concept "catastrophe naturelle".

#### 4.5.2.4 Requête 4 : Station de traitement des eaux usées

Le dernier concept utilisateur que nous avons soumis aux experts est le concept Wikidata (Q3391269) qui correspond à "Station de traitement des eaux usées".

L'expert 1 a indiqué que 4 concepts étaient pertinents pour la requête utilisateur : "eaux usées", "ouvrage hydraulique", "assainissement" et "gestion de l'eau". De plus il nous a fourni une liste de concepts englobants le concept recherché : "gestion des déchets", "gestion des ressources naturelles", "structure construite", "polluant", "qualité de l'eau" et "pollution".

L'expert 3 a jugé que les 6 premiers concepts étaient pertinents : "eaux usées", "ouvrage hydraulique", "gestion des déchets", "gestion de l'eau", "structure construite" et "assainissement". Il nous indique aussi que deux autres concepts devraient apparaître plus haut dans la liste des similarités, à savoir, "pollution" (20<sup>ème</sup>) et "qualité de l'eau" (23<sup>ème</sup>).

Pour cette dernière requête proposée aux experts, les concepts identifiés comme étant pertinents sont classés pour la majorité d'entre eux parmi les 6 meilleurs résultats. Si on veut prendre en compte tous les concepts pertinents identifier par les experts, il faut fixer le seuil minimal de similarité à la valeur 0.68. Les experts ont émis quelques réserves concernant l'ordre de classement de certains concepts, comme par exemple "qualité de l'eau" qui n'apparaît qu'au 20<sup>ème</sup> position.

#### 4.5.2.5 Synthèse

Nous avons proposé à nos experts 4 concepts présents dans Wikidata qui ont été identifiés par les experts lors de la phase 1 de l'expérimentation et pour lesquels nous avons utilisé le KG pour calculer la similarité sémantique avec les concepts utilisés dans la description de records. Nous leur avons fourni une liste des 50 concepts présents dans

les métadonnées qui sont les plus proches du concept de l'utilisateur.

On peut aussi remarquer que les experts n'ont pas été complètement d'accord sur les concepts qu'ils considèrent comme étant pertinents pour chacune des requêtes. Cela souligne la difficulté de la tâche de décider quels concepts sont similaires. Il n'empêche que les experts ont été du même avis sur les concepts les plus pertinents.

Ainsi, d'après les retours de nos experts, il est difficile de fixer un unique seuil de similarité pour ne garder que les concepts pertinents pour le concept utilisateur. En effet, les seuils minimaux varient entre chaque requête, à savoir, 0.69 pour la requête 1, 0.56 pour la requête 2, 0.51 pour la requête 3 et 0.68 pour la requête 4.

Par exemple, si on fixe le seuil minimal comme étant 0.51, on se retrouve avec un grand nombre de résultats non pertinents pour les requêtes 1, 2 et 4. Néanmoins, ce seuil peut être utilisé comme un seuil de non-similarité à partir duquel on peut être sûrs que deux concepts ne sont pas sémantiquement similaires. Ce seuil peut être utilisé pour filtrer des records candidats dans l'étape 3 du processus de recherche. En effet, à cette étape, le but est de filtrer un maximum de records qui contiennent des mots-clés et thèmes éloignés sémantiquement du concept d'intérêt de l'utilisateur.

### 4.5.3 Expérimentation 3 : Identifier des records pertinents

Cette expérimentation s'intéresse à évaluer la capacité du KG à identifier des records qui pourraient intéresser l'utilisateur en se basant sur un concept d'intérêt.

Pour cela, nous allons comparer les résultats obtenus grâce à une recherche effectuée directement sur les catalogues du Sandre et Cerema et les résultats obtenus grâce à notre KG. Pour cette expérimentation, nous avons pris en compte les retours des experts à partir des expérimentations précédentes pour apporter plusieurs améliorations au KG.

La première amélioration par rapport à la tâche 2 de la première expérimentation est l'utilisation du graphe nommé "AnnotationWikidata" du KG. Cela veut dire qu'on se basera sur l'identification manuelle des thèmes et des mots-clés présents dans les métadonnées avec les concepts de Wikidata.

Une deuxième amélioration concerne la sélection des records candidats. En effet, nous avons fixé un seuil de similarité minimum pour filtrer des records qui n'ont aucun

concept similaire au concept d'intérêt de l'utilisateur. Pour cette expérimentation, nous avons fixé le seuil à 0.5. Ainsi, si un record ne contient aucun mot-clé ou thème qui a un score de similarité d'au moins 0.5, le record ne sera pas considéré comme un candidat.

Une troisième amélioration a été apportée à la manière dont est calculé le score global de similarité pour chaque record. Contrairement à l'expérimentation 1 où le score a été calculé grâce à une moyenne à partir des scores de similarité entre des concepts deux à deux, pour cette expérimentation, nous avons opté pour une distance Euclidienne.

Une dernière amélioration concerne les comparaisons prises en compte dans le calcul du score global de similarité pour un record. Pour cette expérimentation, les scores de similarité trop faibles (inférieurs au seuil de 0.5) sont ignorés. En effet, cette décision est motivée par la volonté de ne pas pénaliser les records qui contiennent beaucoup de mots-clés et de thèmes.

Pour cette expérimentation, nous avons choisi les requêtes suivantes :

- **Requête 1** : "plan de prévention du risque inondation"
- **Requête 2** : "nappe phréatique"

Nous avons utilisé les filtres mis à disposition dans les catalogues pour affiner à chaque fois les résultats de recherche et ne considérer que des jeux de données qui sont indexés dans le KG. Nous avons donc sélectionné uniquement les jeux de données (c.-à-d. excluant les services par exemple) dans les deux catalogues. Nous avons aussi dans Cerema sélectionné le filtre CeremaData pour ne garder que les jeux de données qui sont produits par le Cerema et que nous avons récupérés et indexés dans le KG.

#### 4.5.3.1 Requête 1 : plan de prévention du risque inondation

Pour cette requête, nous avons entré la chaîne de caractère "plan de prévention du risque naturel" sur le moteur de recherche du catalogue Cerema et celui du Sandre. Comme on peut le voir dans la Figure 4.9, le Sandre ne renvoie aucun résultat alors que le catalogue Cerema retrouve deux records comme le montre la Figure 4.10.

La Table 4.2 présente les 10 meilleurs résultats de recherche obtenus pour le concept "plan de prévention du risque inondation". Le KG a renvoyé beaucoup plus de records

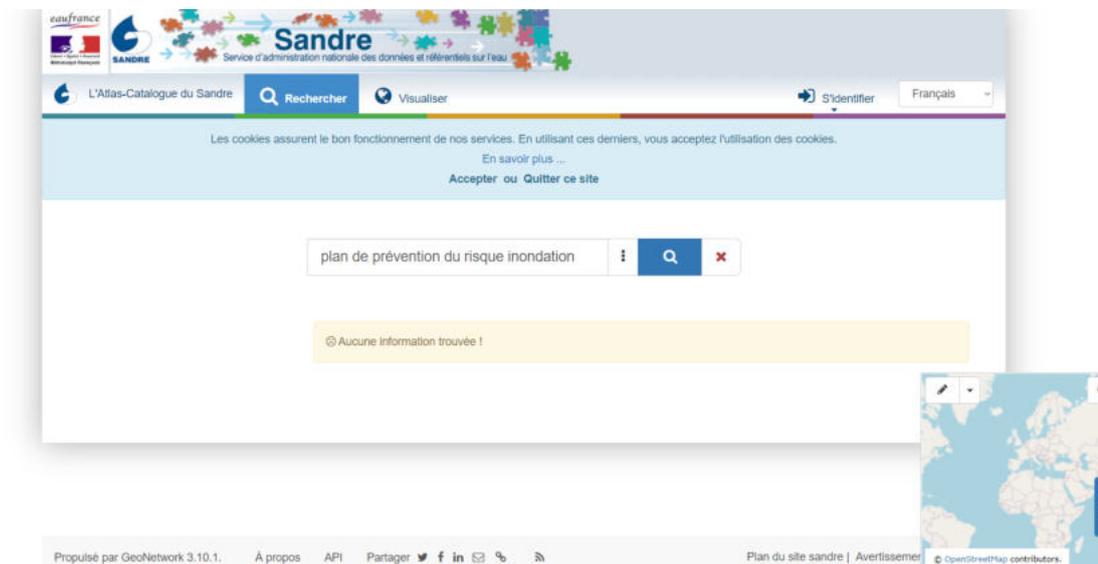


FIGURE 4.9 – Résultats de recherche pour la requête "plan de prévention du risque inondation" sur le catalogue Sandre

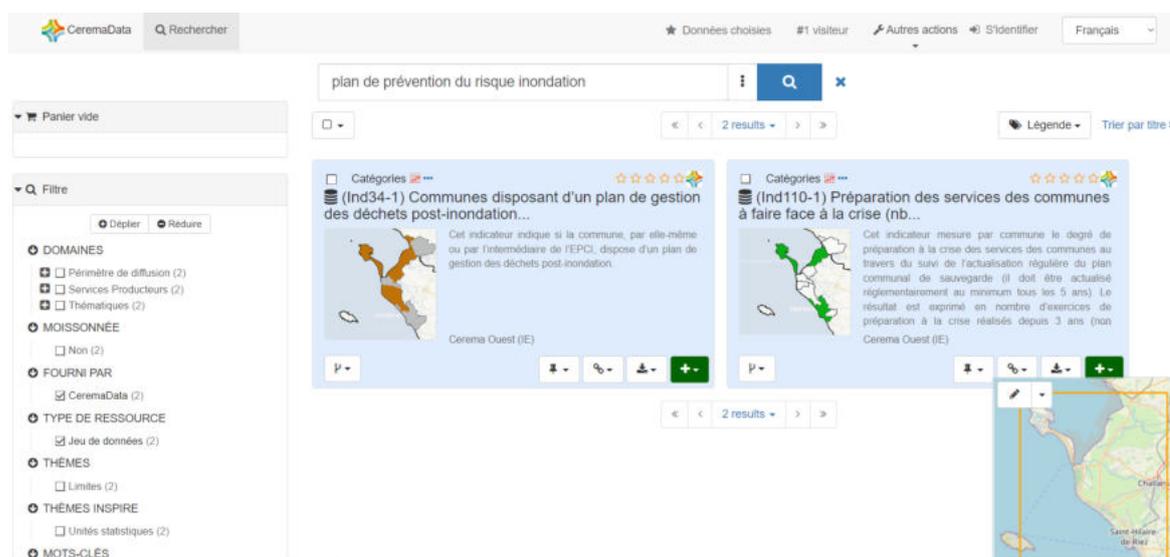


FIGURE 4.10 – Résultats de recherche pour la requête "plan de prévention du risque inondation" sur le catalogue Cerema

comme résultats de recherche. Au total, seulement 15 records ont été filtrés parmi les 231 disponibles dans le KG. On peut constater qu'un des deux résultats obtenus grâce au moteur de recherche du Cerema apparaît en première position.

Les records qui arrivent ensuite dans le classement concernent des jeux de données autour de la thématique des risques naturels, de la prévention des inondations, de la vulnérabilité aux inondations, etc. Ils sont donc tous pertinents pour le concept d'intérêt "plan de prévention du risque inondation".

Il faut aussi noter que certains records ont obtenu les mêmes scores de similarité, car ils sont décrits par les mêmes ensembles de mots-clés et de thèmes. Par exemple dans la Table 4.2, les records 4 et 5 ont obtenu le même score, les records 6 et 7 ont aussi obtenu le même score de pertinence. Cela peut s'expliquer par le fait que ces jeux de données font partie d'un unique référentiel local<sup>4</sup> établi par le Cerema sur le territoire à risques importants d'inondation.

---

4. <https://www.cdata.cerema.fr/geonetwork/srv/fre/catalog.search#/metadata/85a4a236-c420-4351-80f4-5925d6e77713>

TABLE 4.2 – Les 10 meilleurs résultats obtenus pour la requête "plan de prévention du risque inondation" grâce au KG

	<b>Titre du record</b>	<b>Score de pertinence</b>
<b>1</b>	(Ind34-1) Communes disposant d'un plan de gestion des déchets post-inondation (2019)	0.6989204123145117
<b>2</b>	(Ind32-2) Nombre de médecins et thérapeutes par commune (2019)	0.6989204123145117
<b>3</b>	(Ind37-1) Existence d'une organisation pour fournir aux administrés un premier niveau de renseignement sur les dispositifs assurantiels (2019)	0.6704153202917703
<b>4</b>	(Ind36-1) Existence de marchés publics ou d'un dispositif pour gérer les réparations d'urgence sur les ouvrages de protection (2019)	0.6702728732084887
<b>5</b>	(Ind36-2) Nombre d'interventions en période EISH (événement important pour la sécurité des ouvrages hydrauliques) de niveau « orange » par période de 3 ans (2019)	0.6702728732084887
<b>6</b>	(Ind32-1) Capacité financière de la population à faire face à la crise (écart par rapport au niveau de vie médian départemental) (2019)	0.6569339809194493
<b>7</b>	(Ind110-2) Nombre d'actions publiques organisées pour sensibiliser la population sur les risques (2019)	0.6569339809194493
<b>8</b>	(Ind35-2) Pourcentage par commune du linéaire de réseau routier de desserte locale inondable (2019)	0.6567212058568503
<b>9</b>	(Ind23-1) Nombre de bâtiments (habitat activités établissements publics) exposés à l'érosion seule (2019)	0.6567212058568503
<b>10</b>	(Ind21-1) Dommages aux logements estimés en fonction de la hauteur d'eau (2019)	0.6567212058568503

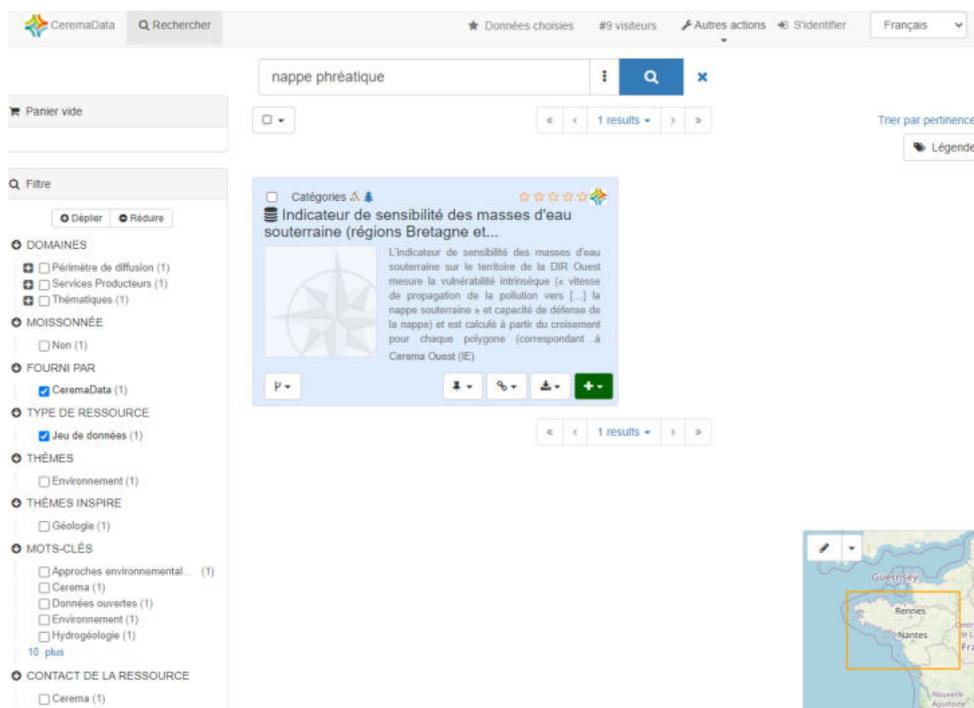


FIGURE 4.11 – Résultats de recherche pour la requête "nappe phréatique" sur le catalogue Cerema

#### 4.5.3.2 Requête 2 : nappe phréatique

Pour cette requête, nous avons saisi la chaîne de caractère "nappe phréatique" dans le moteur de recherche des catalogues Cerema et Sandre. Les Figures 4.11 et 4.12 présentent les résultats obtenus par chacun des deux catalogues. On peut constater que le catalogue du Cerema trouve un seul résultat pour cette requête alors que le Sandre trouve 16 résultats au total. Ces résultats concernent les jeux de données "*Polygones élémentaires de la pile d'entités hydrogéologiques*" et "*les entités hydrogéologiques*" en France métropolitaine et en outre-mer, "*Les zones karstiques*" en métropole et en Martinique ainsi que "*Zones altérites meubles*" en métropole.

De son côté, le KG a réussi à trouver plusieurs records candidats. Au total, seuls 13 records ont été filtrés. La Table 4.3 présente les 11 meilleurs résultats obtenus pour le concept d'intérêt "nappe phréatique". On peut remarquer les résultats présentés dans le tableau concernent "*les masses d'eau souterraines*" et "*les polygones élémentaires de masses d'eau souterraines*" en Métropole et en Outre-mer. Le dernier élément du tableau correspond au record identifié par le catalogue Cerema.

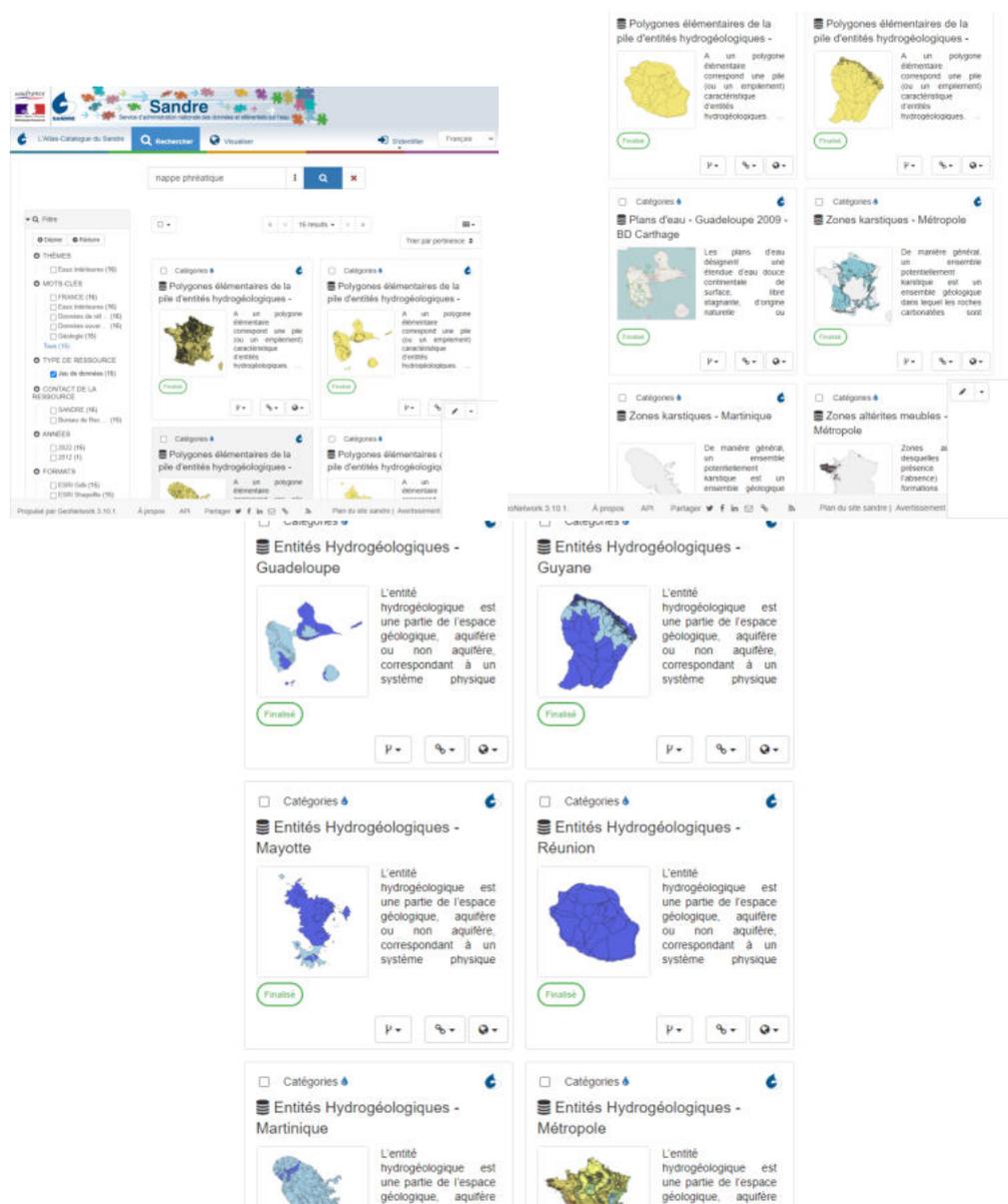


FIGURE 4.12 – Résultats de recherche pour la requête "nappe phréatique" sur le catalogue Sandre

TABLE 4.3 – Les 11 meilleurs résultats obtenus pour la requête "nappe phréatique" grâce au KG

	<b>Titre du record</b>	<b>Score de pertinence</b>
1	Polygones élémentaires de masses d'eau souterraines - Métropole - Version État des Lieux 2019	0.8253342362234256
2	Polygones élémentaires de masses d'eau souterraines - Mayotte - Version État des Lieux 2019	0.8253342362234256
3	Polygones élémentaires de masses d'eau souterraines - Guyane - Version État des Lieux 2019	0.8253342362234256
4	Polygones élémentaires de masses d'eau souterraines - Guadeloupe - Version État des Lieux 2019	0.8253342362234256
5	Polygones élémentaires de masses d'eau souterraines - France Entière - Version État des Lieux 2019	0.8253342362234256
6	Masses d'eau souterraines - Guadeloupe - Version État des Lieux 2019	0.8253342362234256
7	Masses d'eau souterraines - Martinique - Version État des Lieux 2019	0.7770700924871327
8	Polygones élémentaires de masses d'eau souterraines - Martinique - Version État des Lieux 2019	0.7770700924871327
9	Masses d'eau souterraines - Réunion - Version État des Lieux 2019	0.7757165607186189
10	Polygones élémentaires de masses d'eau souterraines - Réunion - Version État des Lieux 2019	0.7757165607186189
11	Indicateur de sensibilité des masses d'eau souterraine (régions Bretagne et Pays de la Loire)	0.7320915056646399

Les records retrouvés par le KG sont pertinents pour le concept d'intérêt "nappe phréatique" dans la mesure où une nappe phréatique fait partie des eaux souterraines. Le KG a aussi réussi à retrouver les records identifiés par le catalogue Sandre classés entre les positions 24 et 33.

#### 4.5.3.3 Synthèse

Cette dernière phase finale nous a permis d'évaluer la capacité du KG à retrouver des records pertinents tout en tenant compte des retours des experts lors des phases d'évaluation précédentes. Plusieurs améliorations ont été apportées au prototype du KG comme, par exemple l'utilisation du graphe nommé "AnnotationWikidata" comme base d'enrichissement des records des métadonnées. Une autre amélioration apportée au KG est l'utilisation du seuil minimal de similarité identifié dans la phase 3 de l'évaluation.

En nous basant sur nos échanges avec les experts, nous avons choisi deux concepts pour souligner l'apport que peut avoir notre approche basée sur le KG par rapport aux approches utilisées dans les catalogues géographiques. Pour cela, nous avons choisi deux concepts Wikidata à savoir "nappe phréatique" et "plan de prévention du risque inondation". Dans les deux cas, plusieurs records de métadonnées ne contenaient pas les chaînes de caractères associées aux deux concepts, mais contenaient à des mots-clés ou des thèmes proches sémantiquement des concepts d'intérêt.

## 4.6 Conclusion

Ce Chapitre s'est concentré sur l'évaluation de notre approche à travers l'implémentation d'un prototype du KG tel que décrit dans le Chapitre 3 et un ensemble d'expérimentations qui ont fait intervenir plusieurs experts des métadonnées géographiques et des domaines de l'eau de l'environnement.

Grâce aux expérimentations menées, nous avons démontré que notre approche est pertinente et permet d'identifier des records qui correspondent à la requête d'un utilisateur et que les moteurs de recherche des catalogues géographiques, basés sur la recherche à facettes, ne présentent pas dans leurs réponses. Notre approche est la plus efficace quand la requête utilisateur concerne un concept dont le label n'est pas présent dans les records de métadonnées, mais qui est proche sémantiquement des mots-clés ou des thématiques qui y sont présentes.

Si notre approche est effectivement pertinente, elle présente une grande charge d'ingénierie des métadonnées et ontologies existantes. En effet, certaines étapes et notamment l'enrichissement des métadonnées n'ont pas été automatisés et nécessitent une intervention humaine pour les adapter à d'autres records de métadonnées géographiques.

C'est notamment dû à la grande hétérogénéité des métadonnées géographiques quand il s'agit de décrire les thèmes et les mots-clés des jeux de données. On peut remarquer par exemple l'utilisation de plusieurs versions d'un même vocabulaire contrôlé dans plusieurs records provenant d'un même catalogue. Idéalement, pour diminuer les hétérogénéités au niveau des thèmes et des mots-clés, les producteurs de données devraient utiliser des URIs pour les décrire.

## Chapitre 5

# Conclusion et Perspectives

Ce travail de thèse s'est intéressé à la problématique de la recherche de jeux de données géographiques sur le Web. Pour cela, nous avons adopté une approche basée sur des technologies du Web sémantique et plus spécifiquement basée sur les graphes de connaissances.

Une première contribution de notre travail est l'analyse des composants nécessaires dans un tel KG pour qu'il permette la recherche de jeu de données géographiques. Une deuxième contribution est l'implémentation et la production du KG pour sur un périmètre spécifique, celui de l'eau et de l'environnement. Une troisième fonctionnalité est l'implémentation des fonctionnalités de la session de recherche exploitant le KG et qui ont été évaluées avec des experts.

Le frein essentiel à la découverte et la réutilisation des jeux de données géographiques sont les métadonnées géographiques qui les accompagnent. Nous avons constaté que cela est dû plus précisément au fait que les métadonnées ne sont pas assez homogènes et exploitables malgré l'existence de plusieurs standards internationaux et européens qui ont tenté de les harmoniser. Ces hétérogénéités sont notamment plus prononcées quand il s'agit de décrire les thèmes et les mots-clés associés à un jeu de données alors que plusieurs travaux de recherche ont montrés que ces informations sont parmi les plus importantes pour déterminer si un jeu de données est pertinent.

Les standards des métadonnées ont permis d'avoir un minimum d'homogénéité au niveau des métadonnées géographiques. Les standards ISO et les recommandations de l'OGC ont grandement amélioré la découverte et le partage des données géographiques sur le Web et ont rencontré un grand succès auprès des producteurs de données. Cela a poussé l'UE, à travers la directive INSPIRE, à rendre ces standards obligatoires dans tous les pays membres et spécifiquement pour les entités publiques productrices de

certain types de données. Malheureusement, cette obligation légale ne laisse pas le champ libre aux entités concernées d'adopter de nouveaux standards plus récents et qui pallient certaines problématiques qui ont été adressées par les nouvelles versions des standards ISO (19115-1, ISO19115-2 et ISO19115-3) notamment.

De plus, on peut constater que seules les métadonnées de description des jeux de données géographiques sont largement adoptées. En effet, plusieurs standards, notamment ceux concernant la qualité des données, sont rarement disponibles. Ce type de métadonnées est assez difficile à éditer et n'améliore pas forcément la visibilité des jeux de données. Par contre, plusieurs travaux de recherche ont montré que les métadonnées de qualité fournissent aux utilisateurs des informations précieuses sur la possibilité de réutiliser les données en question.

Les standards de métadonnées basés sur le Web sémantique deviennent de plus en plus populaires et ont prouvé leur intérêt pour diminuer les hétérogénéités dans les métadonnées associées aux jeux de données. L'adoption de standards comme DCAT, DCAT-AP ou encore schema.org par des outils de recherche de jeux de données comme Google Dataset Search nous pousse à penser qu'à l'avenir de plus en plus de producteurs de données adopteront ces standards pour permettre à leurs données de gagner en visibilité.

Si nous avons montré la pertinence de notre approche dans le cas où les records de métadonnées ne contiennent pas le label du concept d'intérêt de l'utilisateur, la création du KG a nécessité un enrichissement manuel des records de métadonnées en ce qui concerne l'identification des mots-clés et des thèmes en utilisant des concepts de Wikidata. Cette intervention manuelle a permis d'améliorer la capacité du KG à identifier des records de jeux de données géographiques pertinents pour l'utilisateur, mais nécessite d'être adaptée si les records qui ont été moissonnés viennent à changer. Cela amène notamment une problématique de maintenance du KG et une première perspective pour notre travail.

Notre méthodologie d'évaluation basée sur un échange avec des experts du domaine des métadonnées géographiques et du domaine de l'eau et de l'environnement nous a amenés à une évaluation qualitative de notre approche. En effet, après chaque entretien avec nos experts il est nécessaire d'améliorer le KG pour tenir compte de leurs retours. Cela nous a empêchés d'évaluer notre approche de manière quantitative dans la mesure où ces améliorations nécessitent parfois de grandes modifications dans le KG.

Notre approche s'est aussi basée sur la possibilité de permettre à l'utilisateur d'exprimer sa requête sous la forme d'un concept d'intérêt provenant de Wikidata. Ce choix a été motivé par le fait que Wikidata est un KG contenant des informations de sens commun sur le monde réel. De plus, Wikidata est directement relié à l'encyclopédie Wikipédia ce qui permet à l'utilisateur de choisir son concept d'intérêt sans ambiguïté. Si cette association de Wikipédia et de Wikidata permet dans la majorité des cas de retrouver le concept recherché assez facilement, et notamment pour des utilisateurs inexpérimentés dans le domaine des données liées, nos expérimentations ont montré que cela n'est toujours pas le cas. Parfois, certains concepts n'existent pas dans Wikidata et dans Wikipédia, parfois ces concepts ne sont pas disponibles en Français ou sont difficiles à trouver.

Dans le cadre de cette thèse, nous avons essentiellement exploité les champs de métadonnées concernant les mots-clés et les thèmes pour identifier des jeux de données pertinents. De futurs travaux pourraient étendre ces travaux pour prendre en compte d'autres critères importants pour un utilisateur et notamment la spatio-temporalité associée aux jeux de données.

Un autre point d'amélioration est très certainement de permettre à l'utilisateur d'exprimer sa requête grâce à plusieurs concepts d'intérêt qui pourraient potentiellement provenir de vocabulaires ou de KG différents. Avec de plus en plus d'outils comme DBpedia Spotlight permettant d'identifier des concepts présents dans des textes, il serait possible de rendre l'expérience de l'utilisateur beaucoup plus fluide.

Une approche basée sur un KG peut aussi être utilisée pour permettre une aide à la saisie des métadonnées pour des utilisateurs moins expérimentés dans la création et l'édition des métadonnées. En effet, de plus en plus de chercheurs publient leurs données de recherche et le KG pourrait leur recommander des concepts pour enrichir leurs métadonnées en se basant sur les concepts déjà présents dans le KG et qui ont été utilisés dans des contextes similaires.

Finalement, un autre aspect important des moteurs de recherche et qui pourrait être appliqué à notre approche dans des travaux futurs est la capacité des moteurs de recherche à collecter les retours des utilisateurs pour améliorer la recherche des jeux de données pour d'autres utilisateurs.



# Bibliographie

- [1] Bilal Abu-Salih. Domain-specific knowledge graphs : A survey. *Journal of Network and Computer Applications*, 185 :103076, 2021.
- [2] Ricardo Amorim, João Aguiar Castro, João Rocha, and Cristina Ribeiro. A comparison of research data management platforms : architecture, flexible metadata and interoperability. *Universal Access in the Information Society*, 16, 11 2017.
- [3] Florian Bauer and Martin Kaltenböck. Linked open data : The essentials. *Edition mono/monochrom, Vienna*, 710, 2011.
- [4] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific american*, 284(5) :34–43, 2001.
- [5] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.
- [6] Dan Brickley, Matthew Burgess, and Natasha F. Noy. Google dataset search : Building a search engine for datasets in an open web ecosystem. In *WWW*, pages 1365–1375. ACM, 2019.
- [7] Sergey Brin. The pagerank citation ranking : bringing order to the web. *Proceedings of ASIS, 1998*, 98 :161–172, 1998.
- [8] Linda Brink, Payam Barnaghi, Jeremy Tandy, Ghislain Ateazing, Robert Atkinson, Byron Cochrane, Yasmin Fathy, Raúl García Castro, Armin Haller, Andreas Harth, Krzysztof Janowicz, Şefki Kolozali, Bart Leeuwen, Maxime Lefrançois, Josh Lieberman, Andrea Perego, Danh Phuoc, Bill Roberts, Kerry Taylor, and Raphaël Troncy. Best practices for publishing, retrieving, and using spatial data on the web. *Semantic Web*, 10, 12 2017.
- [9] Jean Brodeur, Serena Coetzee, David Danko, Stéphane Garcia, and Jan Hjeltner. Geographic information metadata—an outlook from the international standardization perspective. *ISPRS International Journal of Geo-Information*, 8(6), 2019.
- [10] Bénédicte Bucher, Esa Tiainen, Thomas Ellett von Brasch, Paul Janssen, Dimitris Kotzinos, Marjan Čeh, Martijn Rijdsdijk, Erwin Folmer, Marie-Dominique Van

- Damme, and Mehdi Zhral. Conciliating perspectives from mapping agencies and web of data on successful european sdis : Toward a european geographic knowledge graph. *ISPRS International Journal of Geo-Information*, 9(2) :62, 2020.
- [11] Adriane Chapman, Elena Simperl, Laura Koesten, George Konstantinidis, Luis-Daniel Ibáñez, Emilia Kacprzak, and Paul Groth. Dataset search : a survey. *VLDB J.*, 29(1) :251–272, 2020.
- [12] Zugang Chen and Yaping Yang. Semantic relatedness algorithm for keyword sets of geographic metadata. *Cartography and Geographic Information Science*, 47(2) :125–140, 2020.
- [13] Shivani Choudhary, Tarun Luthra, Ashima Mittal, and Rajat Singh. A survey of knowledge graph embedding and their applications. *arXiv preprint arXiv :2107.07842*, 2021.
- [14] Sarah Dahir and Abderrahim El Qadi. A query expansion method based on topic modeling and dbpedia features. *International Journal of Information Management Data Insights*, 1(2) :100043, 2021.
- [15] Sarah Dahir, Jalil Elhassouni, Abderrahim El Qadi, and Hamid Bennis. Medical query expansion using semantic sources dbpedia and wikidata. In *ISIC*, pages 195–201, 2021.
- [16] Sarah Dahir, Hamid Khalifi, and Abderrahim El Qadi. Query expansion using dbpedia and wordnet. In *Proceedings of the ArabWIC 6th Annual International Conference Research Track*, pages 1–6, 2019.
- [17] Dayany Díaz-Corona, María de Luna, Javier Lacasta, and Javier Nogueras-Iso. The quality of thesauri used in geographic metadata. In *Proc. AGILE 2017 conference*, 2017.
- [18] Lisa Ehrlinger and Wolfram Wöß. Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48(1-4) :2, 2016.
- [19] European Union EU. Inspire metadata implementing rules, 2013.
- [20] European Union EU. Inspire metadata technical guidelines, 2013.
- [21] European Parliament. Directive 2007/2/ec establishing an infrastructure for spatial information in the european community (inspire) (oj l 108, 25.4.2007, pp. 1-14), 2007.
- [22] Nikolaos Fanourakis, Vasilis Efthymiou, Dimitris Kotzinos, and Vassilis Christophides. Knowledge graph embedding methods for entity alignment : An experimental review. *arXiv preprint arXiv :2203.09280*, 2022.

- [23] Jernej Flisar and Vili Podgorelec. Document enrichment using dbpedia ontology for short text classification. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*, pages 1–9, 2018.
- [24] Jernej Flisar and Vili Podgorelec. Improving short text classification using information from dbpedia ontology. *Fundamenta Informaticae*, 172(3) :261–297, 2020.
- [25] Jian-Bo Gao, Bao-Wen Zhang, and Xiao-Hua Chen. A wordnet-based semantic similarity measurement combining edge-counting and information content theory. *Engineering Applications of Artificial Intelligence*, 39 :80–88, 2015.
- [26] Florent Geniet, Valérie Gouet-Brunet, and Mathieu Brédif. ALEGORIA : Joint Multimodal Search and Spatial Navigation into the Geographic Iconographic Heritage. In *Demo at the 30th ACM International Conference on Multimedia (MM '22)*, Lisbon, Portugal, October 2022.
- [27] John Goodwin, Catherine Dolbear, and Glen Hart. Geographical linked data : The administrative geography of great britain on the semantic web. *Transactions in GIS*, 12 :19–30, 2008.
- [28] Stephen C Guptill. Metadata and data catalogues. *Geographical information systems*, 2 :677–692, 1999.
- [29] Peter Haase, Andriy Nikolov, Johannes Trame, Artem Kozlov, and Daniel M Herzig. Alexa, ask wikidata! voice interaction with knowledge graphs using amazon alexa. In *ISWC (Posters, Demos & Industry Tracks)*, 2017.
- [30] Harry Halpin, Patrick J Hayes, James P McCusker, Deborah L McGuinness, and Henry S Thompson. When owl : sameas isn't the same : An analysis of identity in linked data. In *International semantic web conference*, pages 305–320. Springer, 2010.
- [31] Fayçal Hamdi, Nathalie Abadie, Bénédicte Bucher, and Abdelfettah Feliachi. Geomrdf : A geodata converter with a fine-grained structured representation of geometry in the web. *arXiv preprint arXiv :1503.04864*, 2015.
- [32] Thomas Herve, Sara Lafia, and Werner Kuhn. Search facets and ranking in geospatial dataset search. In *11th International Conference on Geographic Information Science (GIScience 2021)-Part I*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.
- [33] Ian Horrocks, Oliver Kutz, and Ulrike Sattler. The even more irresistible sroiq. *Kr*, 6 :57–67, 2006.
- [34] Filip Ilievski, Kartik Shenoy, Nicholas Klein, Hans Chalupsky, and Pedro Szekely. Capturing concept similarity with knowledge graphs.

- [35] Filip Ilievski, Pedro Szekely, Gleb Satyukov, and Amandeep Singh. User-friendly comparison of similarity algorithms on wikidata. *arXiv preprint arXiv :2108.05410*, 2021.
- [36] Organisation internationale de standardisation (ISO). Iso 19119 :2005, 2005.
- [37] Organisation internationale de standardisation (ISO). Métadonnées - implémentation du modèle xml, 2007.
- [38] Organisation internationale de standardisation (ISO). Ontologie - partie 1 : Framework, 2012.
- [39] Organisation internationale de standardisation (ISO). Ontologie - partie 2 : Règles pour développer des ontologies owl, 2015.
- [40] Organisation internationale de standardisation (ISO). Métadonnées - extensions pour l'acquisition et le traitement, 2019.
- [41] Organisation internationale de standardisation (ISO). Métadonnées - schéma d'implémentation xml des concepts fondamentaux, 2019.
- [42] ISO 19157 :2013. Geographic information - data quality.
- [43] Paul Jaccard. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.*, 44 :223–270, 1908.
- [44] Emilia Kacprzak, Laura Koesten, Luis Daniel Ibáñez, Tom Blount, Jeni Tennison, and Elena Simperl. Characterising dataset search - an analysis of search logs and data requests. *J. Web Semant.*, 55 :37–55, 2019.
- [45] Leila Benaissa Kaddar and Farah Ben-Naoum. Novel approach for semantic similarity cross ontology. *Indonesian Journal of Electrical Engineering and Computer Science*, 26(1) :493–504, 2022.
- [46] Ali Kalantari, Amirrudin Kamsin, Halim Shukri Kamaruddin, Nader Ale Ebrahim, Abdullah Gani, Ali Ebrahimi, and Shahaboddin Shamshirband. A bibliometric approach to tracking big data research trends. *Journal of Big Data*, 4(1) :1–18, 2017.
- [47] Fabian Kirstein, Benjamin Dittwald, Simon Dutkowski, Yury Glikman, Sonja Schimmler, and Manfred Hauswirth. Linked data in the european data portal : A comprehensive platform for applying dcat-ap. In *International Conference on Electronic Government*, pages 192–204. Springer, 2019.
- [48] Alexander Kotsev, Marco Minghini, Robert Tomas, Vlado Cetl, and Michael Lutz. From spatial data infrastructures to data spaces—a technological perspective on the evolution of european sdis. *ISPRS International Journal of Geo-Information*, 9(3) :176, 2020.

- [49] Sven R. Kunze and Sören Auer. Dataset retrieval. In *2013 IEEE Seventh International Conference on Semantic Computing, Irvine, CA, USA, September 16-18, 2013*, pages 1–8. IEEE Computer Society, 2013.
- [50] Javier Lacasta, Javier Nogueras-Iso, Rubén Béjar, Pedro R Muro-Medrano, and F Javier Zarazaga-Soria. A web ontology service to facilitate interoperability within a spatial data infrastructure : Applicability to discovery. *Data & Knowledge Engineering*, 63(3) :947–971, 2007.
- [51] Sara Lafia, Andrew Turner, and Werner Kuhn. *Improving discovery of open civic data*. 2018.
- [52] Young-Jou Lai, Ting-Yun Liu, and Ching-Lai Hwang. Topsis for modm. *European journal of operational research*, 76(3) :486–500, 1994.
- [53] Yuanhua Lv and ChengXiang Zhai. When documents are very long, bm25 fails! In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1103–1104, 2011.
- [54] Joan Masó. Geospatial web services. In *Springer Handbook of Geographic Information*, pages 493–530. Springer, 2022.
- [55] Pablo N Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. Dbpedia spotlight : shedding light on the web of documents. In *Proceedings of the 7th international conference on semantic systems*, pages 1–8, 2011.
- [56] D Nebert. Developing spatial data infrastructures : The sdi cookbook v. 2.0. *Global Spatial Data Infrastructure*, 2 :39–56, 2004.
- [57] Douglas D Nebert. The sdi cookbook. <http://www.gsdi.org/pubs.html>, 2001.
- [58] Natalya Fridman Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. Industry-scale knowledge graphs : lessons and challenges. *Commun. ACM*, 62(8) :36–43, 2019.
- [59] OGC. Ogc apis - building blocks for location. [ogcapi.ogc.org](http://ogcapi.ogc.org).
- [60] OGC. Ogc api - environmental data retrieval. [github.com/opengeospatial/ogcapi-environmentaldata-retrieval](https://github.com/opengeospatial/ogcapi-environmentaldata-retrieval), 2022.
- [61] OGC. Ogc api - features. [github.com/opengeospatial/ogcapi-features](https://github.com/opengeospatial/ogcapi-features), 2022.
- [62] OGC. Ogc api features implementations. [github.com/opengeospatial/ogcapifeatures/tree/master](https://github.com/opengeospatial/ogcapifeatures/tree/master), 2022.
- [63] OGC. Ogc api records - encodings. [docs.ogc.org/drafts/20-004.html#clause-encodings](https://docs.ogc.org/drafts/20-004.html#clause-encodings), 2022.

- [64] Jonathan Oliver, Chun Cheng, and Yanggui Chen. Tlsh—a locality sensitive hash. In *2013 Fourth Cybercrime and Trustworthy Computing Workshop*, pages 7–13. IEEE, 2013.
- [65] Pierre-Henri Paris. Assessing the quality of owl : sameas links. In *European Semantic Web Conference*, pages 304–313. Springer, 2018.
- [66] Andrea Perego, Vlado Cetl, Anders Friis-Christensen, and Michael Lutz. Geodcat-ap : Representing geographic metadata by using the “dcat application profile for data portals in europe”. In *Joint UNECE/UNGGIM Europe Workshop on Integrating Geospatial and Statistical Standards, Stockholm, Sweden*, 2017.
- [67] Joe Raad, Wouter Beek, Frank van Harmelen, Nathalie Pernelle, and Fatiha Saïs. Detecting erroneous identity links on the web using network metrics. In *International semantic web conference*, pages 391–407. Springer, 2018.
- [68] Stephen Robertson, Hugo Zaragoza, et al. The probabilistic relevance framework : Bm25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4) :333–389, 2009.
- [69] Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. Learning a health knowledge graph from electronic medical records. *Scientific reports*, 7(1) :1–11, 2017.
- [70] Stefano De Sabbata and Tumasch Reichenbacher. Criteria of geographic relevance : an experimental study. *International Journal of Geographical Information Science*, 26, 2012.
- [71] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5) :513–523, 1988.
- [72] François Scharffe, Laurent Bihanic, Gabriel Képéklian, Ghislain Atemezang, Raphaël Troncy, Franck Cotton, Fabien Gandon, Serena Villata, Jérôme Euzenat, Zhengjie Fan, et al. Enabling linked data publication with the datalift platform. In *Workshops at the twenty-sixth aaii conference on artificial intelligence*, 2012.
- [73] Manjula K Shenoy, KC Shet, and U Dinesh Acharya. A new similarity measure for taxonomy based on edge counting. *International Journal of Web & Semantic Technology*, 3(4) :23, 2012.
- [74] J Simoes and A Cerciello. Serving geospatial data using modern and legacy standards : a case study from the urban health domain. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48 :4, 2022.
- [75] Thabet Slimani, B Ben Yaghlane, and Khaled Mellouli. A new similarity measure based on edge counting. *International Journal of Computer and Information Engineering*, 2(11) :3851–3855, 2008.

- [76] Kai Sun, Yunqiang Zhu, and Jia Song. Progress and challenges on entity alignment of geographic knowledge bases. *ISPRS International Journal of Geo-Information*, 8(2) :77, 2019.
- [77] Sandeep Tata and Jignesh M Patel. Estimating the selectivity of tf-idf based cosine similarity predicates. *ACM Sigmod Record*, 36(2) :7–12, 2007.
- [78] M. Filomena Teodoro, Anacleto Correia, and PAULO NUNES. Knowledge management in geospatial information context. a preliminary statistical approach -a case study. *WSEAS Transactions on Business and Economics*, 14, 02 2017.
- [79] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR, 2016.
- [80] Daniel Tunkelang. Faceted search. *Synthesis lectures on information concepts, retrieval, and services*, 1(1) :1–80, 2009.
- [81] Manuel Antonio Ureña-Cámara, Javier Nogueras-Iso, Javier Lacasta, and Francisco Javier Ariza-López. A method for checking the quality of geographic metadata based on iso 19157. *International Journal of Geographical Information Science*, 33(1) :1–27, 2019.
- [82] Manuel Antonio Ureña-Cámara, Javier Nogueras-Iso, Javier Lacasta, and Francisco Javier Ariza-López. A method for checking the quality of geographic metadata based on iso 19157. *International Journal of Geographical Information Science*, 33(1) :1–27, 2019.
- [83] Bernhard Vockner and Manfred Mittlböck. Geo-enrichment and semantic enhancement of metadata sets to augment discovery in geoportals. *ISPRS International Journal of Geo-Information*, 3(1) :345–367, 2014.
- [84] Bifan Wei, Jun Liu, Qinghua Zheng, Wei Zhang, Xiaoyu Fu, and Boqin Feng. A survey of faceted search. *Journal of Web engineering*, pages 041–064, 2013.
- [85] Zhibiao Wu and Martha Palmer. Verb semantics and lexical selection. *arXiv preprint cmp-lg/9406033*, 1994.
- [86] Li Yujian and Liu Bo. A normalized levenshtein distance metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6) :1091–1095, 2007.
- [87] Zhang Yun-tao, Gong Ling, and Wang Yong-cheng. An improved tf-idf approach for text classification. *Journal of Zhejiang University-Science A*, 6(1) :49–55, 2005.
- [88] Edmundas Kazimieras Zavadskas, Zenonas Turskis, and Simona Kildienė. State of art surveys of overviews on mcdm/madm methods. *Technological and economic development of economy*, 20(1) :165–179, 2014.

- 
- [89] Kaisheng Zeng, Chengjiang Li, Lei Hou, Juanzi Li, and Ling Feng. A comprehensive survey of entity alignment for knowledge graphs. *AI Open*, 2 :1–13, 2021.
- [90] Mehdi Zrhal, Bénédicte Bucher, Fayçal Hamdi, and Marie-Dominique Van Damme. Identifying the key resources and missing elements to build a knowledge graph dedicated to spatial dataset search. *Procedia Computer Science*, 207 :2911–2920, 2022.
- [91] Mehdi Zrhal, Bénédicte Bucher, Marie-Dominique Van Damme, and Fayçal Hamdi. Spatial dataset search : Building a dedicated knowledge graph. *AGILE : GIScience Series*, 2 :1–5, 2021.

**Annexe A**

# **Questionnaire Experts**

# Questionnaire Expert

Le but de ce questionnaire est d'avoir votre retour d'expert, expert connaissant à la fois les métadonnées géo et le domaine de l'eau, sur une approche fondée sur un Graphe de Connaissances (KG) pour faciliter la recherche de jeux de données géographiques.

Il comprend des petites tâches auxquelles nous espérons que vous pourrez consacrer 10 à 20mn chacune pour un total de 1h pour traiter l'ensemble de ce questionnaire.

Vos réponses permettront de faire évoluer le prototype actuel en vue de conduire un test avec des utilisateurs moins avertis qui ne connaissent pas les métadonnées géographiques.

L'analyse des réponses sera partagée avec les experts ayant répondu, si vous souhaitez participer à un rendu de vive voix merci de l'indiquer.

## Tâche 1 : évaluer la pertinence de Wikidata dans le domaine applicatif de l'eau.

Un premier apport du KG est de permettre à l'utilisateur d'exprimer sa requête non pas en utilisant les champs de métadonnées, mais sous la forme d'un concept d'intérêt de sens commun, qui n'est pas forcément présent dans les concepts utilisés pour décrire les jeux de données dans les métadonnées classiques.

Le KG permettra par la suite de relier ce concept utilisateur aux concepts métadonnées.

Nous voudrions évaluer si l'utilisation de Wikidata permet bien à l'utilisateur d'exprimer sa requête. Pouvez-vous penser à quelques concepts se rapportant au domaine de l'eau et de l'environnement et essayer de trouver les éléments Wikidata correspondants en suivant les indications ci-après. Avez-vous réussi à les trouver ? Quels concepts n'avez-vous pas réussi à trouver ?

Indication sur comment trouver un élément Wikidata : Une recherche peut être effectuée directement sur [Wikidata](#) en utilisant la barre de recherche, en utilisant une requête SPARQL [ici](#), ou en trouvant la page [Wikipedia](#) correspondante et en cliquant sur "Élément Wikidata" dans la barre latérale :



## Tâche 2 : évaluer la capacité du moteur à pointer vers des records pertinents en lien avec un concept utilisateur

Dans cette partie, notre but est de comparer les jeux de données obtenus en utilisant le KG à ceux obtenus en utilisant des catalogues du domaine applicatif : [Cerema](#) et [Sandre](#).

*Que pensez-vous du choix des catalogues ? Pensez-vous que le KG devrait inclure des jeux de données provenant d'autres catalogues ? :*

1. Pour le concept "Hydrographie" (Q182468), le KG identifie les records suivants comme étant les 20 meilleurs :

- Territoires de vigilance crues - Métropole
- Territoires de Compétence Crues (TCC) - Métropole
- Contrats de Milieu - Guadeloupe
- Contrats de Milieu - Martinique
- Masses d'eau souterraines - Martinique - Version État des Lieux 2019
- Polygones élémentaires de masses d'eau souterraines - Guyane - Version État des Lieux 2019
- (Ind34-1) Communes disposant d'un plan de gestion des déchets post-inondation (2019)
- (Ind110-2) Nombre d'actions publiques organisées pour sensibiliser la population sur les risques (2019)
- (Ind110-1) Préparation des services des communes à faire face à la crise (nb exercices PCS depuis 3 ans) (2019)
- (Ind37-1) Existence d'une organisation pour fournir aux administrés un premier niveau de renseignement sur les dispositifs assurantiels (2019)
- Indicateur de sensibilité des masses d'eau souterraine (régions Bretagne et Pays de la Loire)
- Segment de classement de continuité écologique - Guadeloupe
- Zones hydrographiques - Métropole 2017 - BD Carthage
- Zones hydrographiques - Guyane 2015 - BD Carthage
- Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en PACA - Découpage par département
- Recensement des Installations Terminales Embranchées (ITE)
- Nœuds hydrographiques - Métropole 2016 - BD Carthage
- Contrats de Milieu - France entière
- Polygones élémentaires de masses d'eau souterraines - Mayotte - Version État des Lieux 2019

- Amphibiens et dispositifs de franchissement des infrastructures de transport terrestre

*Quels jeux de données identifiés vous paraissent pertinents ? Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ? Y-a-t-il des jeux de données manquants ? :*

2. Pour le concept "Estuaire" (Q47053), le KG identifie les jeux de données suivants comme étant les 20 meilleurs :

- Masses d'eau de transition - Guyane - Version Rapportage 2010
- Réseau de mesures de houle du Centre d'Archivage National des Données de Houles In-Situ (CANDHIS) - France métropole et outre-mer
- Recensement des Installations Terminales Embranchées (ITE)
- Masses d'eau côtières - Réunion - Version Rapportage 2010
- Masses d'eau Côtières - Guadeloupe - Version Rapportage 2016
- Masses d'eau Côtières - Guyane - Version Rapportage 2016
- Hydroecorégions de niveau 1 (HER-1)
- Tronçons élémentaires de masse d'eau cours d'eau - Guadeloupe - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Martinique - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Réunion - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2010
- Lieux de surveillance des eaux littorales - SEL
- Masses d'eau cours d'eau - Martinique - Version Rapportage 2010
- Masses d'eau cours d'eau - Guyane - Version Rapportage 2010
- Masses d'eau de transition - Guyane - Version Rapportage 2016
- Masses d'eau de transition - France entière - Version Rapportage 2016
- Masses d'eau de transition - Martinique - Version Rapportage 2016
- Vigilance crues - VIC
- Nœuds hydrographiques - Métropole 2016 - BD Carthage

Quels jeux de données identifiés vous paraissent pertinents ? Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ? Y-a-t-il des jeux de données manquants ?

3. Pour le concept "Rhone" (Q3429695), le KG identifie les jeux de données suivants comme étant les 20 meilleurs :

- Tronçons élémentaires de masse d'eau cours d'eau - Guadeloupe - Version Rapportage 2010
- Masses d'eau cours d'eau - Martinique - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Martinique - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2010
- Masses d'eau cours d'eau - Guyane - Version Rapportage 2010
- Nœuds hydrographiques - Métropole 2016 - BD Carthage
- Lieux de surveillance des eaux littorales - SEL
- Contextes piscicoles - Métropole
- Stations de mesure de la qualité des eaux superficielles continentales (STQ) - Martinique
- Stations de mesure de la qualité des eaux superficielles continentales (STQ) - Mayotte
- Unités de méthanisation France métropole
- Communes administratives au 1er janvier 2019
- Départements administratifs au 1er janvier 2018
- Tronçons élémentaires de masse d'eau cours d'eau - Réunion - Version Rapportage 2010
- Masses d'eau de transition - Guyane - Version Rapportage 2016
- Masses d'eau cours d'eau - Réunion - Version Rapportage 2010
- Hydroecorégions de niveau 1 (HER-1)
- Points de prélèvements associés aux stations de mesure de la qualité des eaux superficielles continentales (STQ) - France entière
- Points de prélèvements associés aux stations de la mesure de la qualité des eaux superficielles continentales (STQ) - Guadeloupe

Quels jeux de données identifiés vous paraissent pertinents ? Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ? Y-a-t-il des jeux de données manquants ?

### Tâche 3 : évaluer le regroupement des réponses

Le KG est employé pour regrouper les records trouvés en réponse à un concept d'intérêt pour faciliter l'exploration de ces réponses. Il s'agit de regroupement multicritères (couverture géographique, thèmes, mot-clés). Le moteur présente alors les 10 groupes les plus pertinents dans le cas de recherche du concept "Estuaire" (Q47053).

Nous avons surligné les records obtenus par le moteur lors de la recherche du concept "Estuaire" pour vous aider à facilement les retrouver.

#### Groupe 1 :

- Lieux de surveillance des eaux littorales - Métropole
- Masses d'eau Côtières - Metropole - Version Rapportage 2016
- Masses d'eau Côtières - Guadeloupe - Version Rapportage 2016
- Masses d'eau côtières - Guyane - Version Rapportage 2010
- Masses d'eau Côtières - Reunion - Version Rapportage 2016
- Lieux de surveillance des eaux littorales - Mayotte
- Masses d'eau Côtières - Mayotte - Version Rapportage 2016
- Masses d'eau de transition - Guyane - Version Rapportage 2010
- Masses d'eau Côtières - Guyane - Version Rapportage 2016
- Lieux de surveillance des eaux littorales - SEL
- Masses d'eau côtières - Réunion - Version Rapportage 2010
- Lieux de surveillance des eaux littorales - Réunion
- Masses d'eau côtières - Guadeloupe - Version Rapportage 2010@fr

#### Groupe 2 :

- Zones d'hydrographie de texture - Guyane 2015 - BD Carthage
- Points d'eau isolés - Guyane 2015 - BD Carthage
- Sous-secteurs hydrographiques - Guyane 2015 - BD Carthage
- Zones hydrographiques - Guyane 2015 - BD Carthage
- Nœuds hydrographiques - Guyane 2015 - BD Carthage
- Stations de mesure de la qualité des eaux superficielles continentales (STQ) - Guyane

- Masses d'eau plan d'eau - Guyane - Version Rapportage 2016
- Cours d'eau - Guyane 2015 - BD Carthage
- **Masses d'eau de transition - Guyane - Version Rapportage 2016**
- Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2016
- Secteurs hydrographiques - Guyane 2015 - BD Carthage
- Masses d'eau plan d'eau - Guyane - Version Rapportage 2010
- Trait de côte - Guyane 2015 - BD Carthage
- Eléments hydrographiques de surface - Guyane 2015 - BD Carthage@fr

#### Groupe 3 :

- Départements administratifs au 1er janvier 2019
- Départements administratifs au 1er janvier 2018
- **Masses d'eau de transition - France entière - Version Rapportage 2016**
- Tronçons élémentaires de masse d'eau cours d'eau - France entière - Version Rapportage 2016
- Circonscriptions administratives de bassin 2018
- Croisement entre les plans d'eau et les communes 2008 - France entière
- Points de prélèvements associés aux stations de mesure de la qualité des eaux superficielles continentales (STQ) - France entière
- Communes administratives au 1er janvier 2019
- Masses d'eau plan d'eau - France entière - Version Rapportage 2016@fr

#### Groupe 4 :

- Hydrographie - ETH
- **Nœuds hydrographiques - Métropole 2016 - BD Carthage**
- Zones hydrographiques - Métropole 2017 - BD Carthage
- Sous-secteurs hydrographiques - Métropole 2017 - BD Carthage
- Tronçons hydrographiques - Métropole 2016 - BD Carthage
- Tronçons élémentaires de masse d'eau cours d'eau - Métropole - Version Rapportage 2016
- Cours d'eau - Métropole 2017 - BD Carthage
- Eléments hydrographiques de surface - Métropole 2017 - BD Carthage
- **Vigilance crues - VIC**
- Zones d'hydrographie de texture - Métropole 2016 - BD Carthage
- Secteurs hydrographiques - Métropole 2016 - BD Carthage
- Points d'eau isolés - Métropole 2017 - BD Carthage
- Tronçons hydrographiques - Métropole 2017 - BD Carthage@fr

#### Groupe 5 :

- **Recensement des Installations Terminales Embranchées (ITE)**

- Amphibiens et dispositifs de franchissement des infrastructures de transport terrestre@fr

Groupe 6 :

- Tronçons élémentaires de masse d'eau cours d'eau - Guadeloupe - Version Rapportage 2010@fr

Groupe 7 :

- Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2010
- Masses d'eau cours d'eau - Guyane - Version Rapportage 2010@fr

Groupe 8 :

- Masses d'eau cours d'eau - Martinique - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Martinique - Version Rapportage 2010@fr

Groupe 9 :

- Tronçons élémentaires de masse d'eau cours d'eau - Réunion - Version Rapportage 2010
- Masses d'eau cours d'eau - Réunion - Version Rapportage 2010@fr

Groupe 10 :

- Hydroecorégions de niveau 1 (HER-1)@fr

Groupe 11 :

- Réseau de mesures de houle du Centre d'Archivage National des Données de Houles In-Situ (CANDHIS) - France métropole et outre-mer@fr

*Les regroupements proposés vous semblent-ils cohérents ? Comment auriez-vous regroupé ces records ?*

**Annexe B**

**Réponses des experts au  
questionnaire**

## Annexe 2 : Réponses de l'expert 1

Tâche 1 : évaluer la pertinence de Wikidata dans le domaine applicatif de l'eau.

Nous voudrions évaluer si l'utilisation de Wikidata permet bien à l'utilisateur d'exprimer sa requête. Pouvez-vous penser à quelques concepts[1] se rapportant au domaine de l'eau et de l'environnement et essayer de trouver les éléments Wikidata correspondants en suivant les indications ci-après. Avez-vous réussi à les trouver ? Quels concepts n'avez-vous pas réussi à trouver ?

1. Pollution des sédiments :
  - recherche directe sur wikidata : 0 éléments trouvés
  - recherche SPARQL sur wikidata (via l'assistant) : 0 éléments trouvés[1]
  - recherche Wikipédia : parmi les éléments listés [https://fr.wikipedia.org/wiki/Pollution\\_de\\_l'eau](https://fr.wikipedia.org/wiki/Pollution_de_l'eau) se rapproche le plus du concepts → <https://www.wikidata.org/wiki/Q183129>
3. Marais artificiel :
  - recherche directe sur wikidata : → non trouvé en français!  
<https://www.wikidata.org/wiki/Q1376406> via constructed wetland (en)
4. Plan de prévention des risques inondation (PPRI) : non trouvé, le libellé en français sur wikidata est « Plan de prévention du risque inondation » (d'autres risques sont bien écrits avec « des » pourtant!).

Tâche 2 : évaluer la capacité du moteur à pointer vers des records pertinents en lien avec un concept utilisateur

*Que pensez-vous du choix des catalogues ? Pensez-vous que le GC devrait inclure des jeux de données provenant d'autres catalogues ? :*

Choix pertinents uniquement pour de la cartographie (i.e. localisation de concepts). Par contre, la recherche par concept échoue (Cf. Tâche 1) échoue sur Cerema et Sandre était systématiquement en timeout ! Si le concept n'est pas dans les métadonnées, les moteurs de recherche Cerema/Sandre ne trouve rien (ce n'est pas une découverte^^)

Oui, le graphe de connaissance devrait inclure des jeux de données pour pouvoir évaluer la réponse fournie par le graphe.

- 1. Pour le concept "Hydrographie" (Q182468), le GC identifie les records suivants comme étant les 20 meilleurs :**

- Territoires de vigilance crues - Métropole
- Territoires de Compétence Crues (TCC) - Métropole
- Contrats de Milieu - Guadeloupe
- Contrats de Milieu - Martinique

- Masses d'eau souterraines - Martinique - Version État des Lieux 2019
- Polygones élémentaires de masses d'eau souterraines - Guyane - Version État des Lieux 2019
- (Ind34-1) Communes disposant d'un plan de gestion des déchets post-inondation (2019)
- (Ind110-2) Nombre d'actions publiques organisées pour sensibiliser la population sur les risques (2019)
- (Ind110-1) Préparation des services des communes à faire face à la crise (nb exercices PCS depuis 3 ans) (2019)
- (Ind37-1) Existence d'une organisation pour fournir aux administrés un premier niveau de renseignement sur les dispositifs assurantiels (2019)
- Indicateur de sensibilité des masses d'eau souterraine (régions Bretagne et Pays de la Loire)
- Segment de classement de continuité écologique - Guadeloupe
- Zones hydrographiques - Métropole 2017 - BD Carthage
- Zones hydrographiques - Guyane 2015 - BD Carthage
- Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en PACA - Découpage par département
- Recensement des Installations Terminales Embranchées (ITE)
- Nœuds hydrographiques - Métropole 2016 - BD Carthage
- Contrats de Milieu - France entière
- Polygones élémentaires de masses d'eau souterraines - Mayotte - Version État des Lieux 2019
- Amphibiens et dispositifs de franchissement des infrastructures de transport terrestre

*Quels jeux de données identifiés vous paraissent pertinents ? Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ? Y-a-t-il des jeux de données manquants ? :*

Pertinent → surligné bleu

Cerema : Hydrographie, 35 jeux de données dans cette catégorie

Wikhydro (base de connaissances) : La recherche remonte les concepts liés ainsi que les textes y faisant référence dont certains sont BD Carthage ; certains résultats sont donc identiques, mais d'autres manques dus à la différence de point de vue : Cerema = jeu de données, Wikhydro = référence aux données dans les textes

**2. Pour le concept "Estuaire" (Q47053), le GC identifie les jeux de données suivants comme étant les 20 meilleurs :**

- Masses d'eau de transition - Guyane - Version Rapportage 2010
- Réseau de mesures de houle du Centre d'Archivage National des Données de Houles In-Situ (CANDHIS) - France métropole et outre-mer

- Recensement des Installations Terminales Embranchées (ITE)
- Masses d'eau côtières - Réunion - Version Rapportage 2010
- Masses d'eau Côtières - Guadeloupe - Version Rapportage 2016
- Masses d'eau Côtières - Guyane - Version Rapportage 2016
- Hydroecorégions de niveau 1 (HER-1)
- Tronçons élémentaires de masse d'eau cours d'eau - Guadeloupe - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Martinique - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Réunion - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2010
- Lieux de surveillance des eaux littorales - SEL
- Masses d'eau cours d'eau - Martinique - Version Rapportage 2010
- Masses d'eau cours d'eau - Guyane - Version Rapportage 2010
- Masses d'eau de transition - Guyane - Version Rapportage 2016
- Masses d'eau de transition - France entière - Version Rapportage 2016
- Masses d'eau de transition - Martinique - Version Rapportage 2016
- Vigilance crues - VIC
- Nœuds hydrographiques - Métropole 2016 - BD Carthage

*Quels jeux de données identifiés vous paraissent pertinents ? Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ? Y-a-t-il des jeux de données manquants ?*

Pertinent → surligné bleu

Cerema : Estuaires, 7 résultats. Les filtres à gauche limitent les jeux de données. Il est intéressant de voir que « Information géo-scientifiques » limite à 3 jeux de données. Aucun recoupement avec le GC.

Wikhydro : Aucun texte ne mentionne des jeux de données (étonnant!).

**3. Pour le concept "Rhône" (Q3429695), le GC identifie les jeux de données suivants comme étant les 20 meilleurs :**

- Tronçons élémentaires de masse d'eau cours d'eau - Guadeloupe - Version Rapportage 2010
- Masses d'eau cours d'eau - Martinique - Version Rapportage 2010
- Tronçons élémentaires de masse d'eau cours d'eau - Martinique - Version Rapportage 2010

- Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2010
- Masses d'eau cours d'eau - Guyane - Version Rapportage 2010
- **Nœuds hydrographiques - Métropole 2016 - BD Carthage**
- Lieux de surveillance des eaux littorales - SEL
- Contextes piscicoles - Métropole
- Stations de mesure de la qualité des eaux superficielles continentales (STQ) - Martinique
- Stations de mesure de la qualité des eaux superficielles continentales (STQ) - Mayotte
- Unités de méthanisation France métropole
- Communes administratives au 1er janvier 2019
- Départements administratifs au 1er janvier 2018
- Tronçons élémentaires de masse d'eau cours d'eau - Réunion - Version Rapportage 2010
- Masses d'eau de transition - Guyane - Version Rapportage 2016
- Masses d'eau cours d'eau - Réunion - Version Rapportage 2010
- Hydroecorégions de niveau 1 (HER-1)
- Points de prélèvements associés aux stations de mesure de la qualité des eaux superficielles continentales (STQ) - France entière
- Points de prélèvements associés aux stations de la mesure de la qualité des eaux superficielles continentales (STQ) - Guadeloupe

*Quels jeux de données identifiés vous paraissent pertinents ? Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ? Y-a-t-il des jeux de données manquants ?*

Pertinent → **surligné bleu** (il n'y a pas de jeux de données Métropole?)

Cerema : Rhône, 36 jeux de données très orienté Occupation du sol

Wikhydro : Beaucoup d'articles sur ce concept, mais pas de jeu de données (de nouveau, différence de point de vue).

Tâche 3 : évaluer le regroupement des réponses

Le GC est employé pour regrouper les informations trouvées en réponse à un concept d'intérêt pour faciliter l'exploration de ces réponses. Il s'agit de regroupement multicritères (couverture géographique, thèmes, mot-clés). Le moteur présente alors les 10 groupes les plus pertinents dans le cas de recherche du concept "Estuaire" (*deuxième requête de la tâche 2*).

Nous avons surligné les informations obtenues par le moteur lors de la recherche du concept "Estuaire" pour vous aider à facilement les retrouver.

- Groupe 1 :

- Lieux de surveillance des eaux littorales - Métropole
- Masses d'eau Côtieres - Metropole - Version Rapportage 2016
- Masses d'eau Côtieres - Guadeloupe - Version Rapportage 2016
- Masses d'eau côtières - Guyane - Version Rapportage 2010
- Masses d'eau Côtieres - Reunion - Version Rapportage 2016
- Lieux de surveillance des eaux littorales - Mayotte
- Masses d'eau Côtieres - Mayotte - Version Rapportage 2016
- Masses d'eau de transition - Guyane - Version Rapportage 2010
- Masses d'eau Côtieres - Guyane - Version Rapportage 2016
- Lieux de surveillance des eaux littorales - SEL
- Masses d'eau côtières - Réunion - Version Rapportage 2010
- Lieux de surveillance des eaux littorales - Réunion
- Masses d'eau côtières - Guadeloupe - Version Rapportage 2010@fr
- Groupe 2 :
  - Zones d'hydrographie de texture - Guyane 2015 - BD Carthage
  - Points d'eau isolés - Guyane 2015 - BD Carthage
  - Sous-secteurs hydrographiques - Guyane 2015 - BD Carthage
  - Zones hydrographiques - Guyane 2015 - BD Carthage
  - Nœuds hydrographiques - Guyane 2015 - BD Carthage
  - Stations de mesure de la qualité des eaux superficielles continentales (STQ) - Guyane
  - Masses d'eau plan d'eau - Guyane - Version Rapportage 2016
  - Cours d'eau - Guyane 2015 - BD Carthage
  - Masses d'eau de transition - Guyane - Version Rapportage 2016
  - Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2016
  - Secteurs hydrographiques - Guyane 2015 - BD Carthage
  - Masses d'eau plan d'eau - Guyane - Version Rapportage 2010
  - Trait de côte - Guyane 2015 - BD Carthage
  - Eléments hydrographiques de surface - Guyane 2015 - BD Carthage@fr
- Groupe 3 :

- Départements administratifs au 1er janvier 2019
- Départements administratifs au 1er janvier 2018
- Masses d'eau de transition - France entière - Version Rapportage 2016
- Tronçons élémentaires de masse d'eau cours d'eau - France entière - Version Rapportage 2016
- Circonscriptions administratives de bassin 2018
- Croisement entre les plans d'eau et les communes 2008 - France entière
- Points de prélèvements associés aux stations de mesure de la qualité des eaux superficielles continentales (STQ) - France entière
- Communes administratives au 1er janvier 2019
- Masses d'eau plan d'eau - France entière - Version Rapportage 2016@fr
- Groupe 4 :
  - Hydrographie - ETH
  - Nœuds hydrographiques - Métropole 2016 - BD Carthage
  - Zones hydrographiques - Métropole 2017 - BD Carthage
  - Sous-secteurs hydrographiques - Métropole 2017 - BD Carthage
  - Tronçons hydrographiques - Métropole 2016 - BD Carthage
  - Tronçons élémentaires de masse d'eau cours d'eau - Métropole - Version Rapportage 2016
  - Cours d'eau - Métropole 2017 - BD Carthage
  - Éléments hydrographiques de surface - Métropole 2017 - BD Carthage
  - Vigilance crues - VIC
  - Zones d'hydrographie de texture - Métropole 2016 - BD Carthage
  - Secteurs hydrographiques - Métropole 2016 - BD Carthage
  - Points d'eau isolés - Métropole 2017 - BD Carthage
  - Tronçons hydrographiques - Métropole 2017 - BD Carthage@fr
- Groupe 5 :
  - Recensement des Installations Terminales Embranchées (ITE)
  - Amphibiens et dispositifs de franchissement des infrastructures de transport terrestre@fr
- Groupe 6 :
  - Tronçons élémentaires de masse d'eau cours d'eau - Guadeloupe - Version Rapportage 2010@fr
- Groupe 7 :
  - Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2010

- Masses d'eau cours d'eau - Guyane - Version Rapportage 2010@fr
- Groupe 8 :
  - Masses d'eau cours d'eau - Martinique - Version Rapportage 2010
  - Tronçons élémentaires de masse d'eau cours d'eau - Martinique - Version Rapportage 2010@fr
- Groupe 9 :
  - Tronçons élémentaires de masse d'eau cours d'eau - Réunion - Version Rapportage 2010
  - Masses d'eau cours d'eau - Réunion - Version Rapportage 2010@fr
- Groupe 10 :
  - Hydroecorégions de niveau 1 (HER-1)@fr
- Groupe 11 :
  - Réseau de mesures de houle du Centre d'Archivage National des Données de Houles In-Situ (CANDHIS) - France métropole et outre-mer@fr

*Les regroupements proposés vous semblent-ils cohérents ? Comment auriez-vous regroupé ces records ?*

Problème de compréhension des regroupements : je vois mal les critères employés pour avoir les 11 groupes (même avec un classement par pertinence, c'est pas très clair). Ainsi sur les Masses d'eau de transition, pourquoi le groupe 1 ne regroupe pas Guyane 2010 et 2016. Pourquoi la métropole n'y est pas ?

Les regroupements proposés ne semblent pas cohérents (au moins sur le critère géographique).

Hors thématique et mot-clef, un premier regroupement serait justement l'emprise (attention : la France regroupe la Terre entière ou pas, alors que la France Métropolitaine ne recouvre qu'une partie et la France continentale pas la Corse!). Un second regroupement serait la temporalité.

Pour la thématique et les mots-clefs, la difficulté réside dans le sens des mots employés : Rhône est un bon exemple (fleuve, département, unité administrative, etc.), Estuaire montre aussi que la notion géomorphologique et géographique, voire cartographique ou d'usage donne des résultats très disparates ... Ces thématiques ne sont pas que géographiques, le regroupement impliquerait la connaissance de ces domaines attachés au concept.

Temps passé : 2h10 !!!!

---

```
[1]SELECT DISTINCT ?item ?itemLabel WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
  {
    SELECT DISTINCT ?item WHERE {
      ?item p:P31 ?statement0.
      ?statement0 (ps:P31/(wdt:P279*)) wd:Q183129.
      ?item p:P373 ?statement1.
```

```
?statement1 (ps:P373) "Sediments".
}
LIMIT 100
}}
```

---

## Annexe 3 : Réponses de l'expert 2

Tâche 1 : évaluer la pertinence de Wikidata dans le domaine applicatif de l'eau.

*Nous voudrions évaluer si l'utilisation de Wikidata permet bien à l'utilisateur d'exprimer sa requête. Pouvez-vous penser à quelques concepts se rapportant au domaine de l'eau et de l'environnement et essayer de trouver les éléments Wikidata correspondants en suivant les indications ci-après. Avez-vous réussi à les trouver ? Quels concepts n'avez-vous pas réussi à trouver ?*

- nappe phréatique (pourtant le concept anglais phreatic zone comporte bien la traduction française nappe phréatique)
- station d'épuration

Tâche 2 : évaluer la capacité du moteur à pointer vers des records pertinents en lien avec un concept utilisateur

*Que pensez-vous du choix des catalogues ? Pensez-vous que le KG devrait inclure des jeux de données provenant d'autres catalogues ? :*

Le Sandre est le catalogue du Système d'information sur l'eau, il est donc très complet pour toute ce qui concerne données publiques sur l'eau.

Le catalogue du Cerema va référencer uniquement les jeux de données qu'il produit.

Il va manquer les jeux de données des collectivités, etc.

L'idéal serait d'utiliser un catalogue national (Géocatalogue et/ou data.gouv.fr)

- 1. Pour le concept "Hydrographie" (Q182468), le KG identifie les records suivants comme étant les 20 meilleurs :**

*Quels jeux de données identifiés vous paraissent pertinents ?*

*Note : c'est assez subjectif, car on n'a pas défini le terme Hydrographie ni pourquoi on cherchait ce terme. Je me suis donc restreint à une vision très topographique et physique (en excluant la gestion de l'eau, par ex.)*

- ~~Territoires de vigilance crues – Métropole~~
- ~~Territoires de Compétence Crues (TCC) – Métropole~~
- ~~Contrats de Milieu – Guadeloupe~~
- ~~Contrats de Milieu – Martinique~~
- ~~Masses d'eau souterraines – Martinique – Version État des Lieux 2019~~
- ~~Polygones élémentaires de masses d'eau souterraines – Guyane – Version État des Lieux 2019~~

- ~~(Ind34 1) Communes disposant d'un plan de gestion des déchets post inondation (2019)~~
- ~~(Ind110 2) Nombre d'actions publiques organisées pour sensibiliser la population sur les risques (2019)~~
- ~~(Ind110 1) Préparation des services des communes à faire face à la crise (nb exercices PGS depuis 3 ans) (2019)~~
- ~~(Ind37 1) Existence d'une organisation pour fournir aux administrés un premier niveau de renseignement sur les dispositifs assurantiels (2019)~~
- ~~Indicateur de sensibilité des masses d'eau souterraine (régions Bretagne et Pays de la Loire)~~
- ~~Segment de classement de continuité écologique - Guadeloupe~~
- Zones hydrographiques - Métropole 2017 - BD Carthage
- Zones hydrographiques - Guyane 2015 - BD Carthage
- Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en PACA - Découpage par département
- ~~Recensement des Installations Terminales Embranchées (ITE)~~
- Nœuds hydrographiques - Métropole 2016 - BD Carthage
- ~~Contrats de Milieu - France entière~~
- ~~Polygones élémentaires de masses d'eau souterraines - Mayotte - Version État des Lieux 2019~~
- ~~Amphibiens et dispositifs de franchissement des infrastructures de transport terrestre~~

*Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ?*

Dans le catalogue Cerema, en tapant Hydrographie dans la barre de recherche :

1. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Région Provence Alpes Côte d'Azur (service WFS)*
2. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Corse - Découpage par département (service WFS)*
3. *Ruissellement à la Réunion (974): Résultats Exzeco (service WFS)*
4. *Indicateur de sensibilité des masses d'eau souterraine (régions Bretagne et Pays de la Loire)*
5. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Corse - Découpage par département*
6. *Ruissellement à la Réunion (974): Résultats Exzeco*
7. *Parc éolien en mer de Dunkerque - MNT bathymétriques de la zone d'étude*
8. *Position des nadirs (centre des clichés) de l'Ortho-littorale v2*
9. *Cartographie de l'artificialisation du trait de côte*
10. *Position des nadirs (centres des clichés) de l'Ortho-littorale v3*
11. *Hauteurs d'eau calculées (Ortho-littorale v3)*
12. *Ouvrages et aménagements littoraux (métropole et outre-mer)*
13. *Nappe d'eau souterraine lyonnaise – modélisation des écoulements et de la température au pas de 20 m*

14. *Nappe d'eau souterraine lyonnaise – modélisation des écoulements et de la température au pas de 200 m*
15. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de la Haute Garonne (31)*
16. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de la Haute Garonne (31) (service WFS)*
17. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de la Lozère (48) (service WFS)*
18. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département des Pyrénées Orientales (66)*
19. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de l'Aveyron (12) (service WFS)*
20. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de l'Aveyron (12)*
21. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de l'Hérault (34) (service WFS)*
22. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de l'Hérault (34)*
23. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de la Lozère (48)*
24. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de l'Aude (11) (service WFS)*
25. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de l'Aude (11)*
26. *Ennoiement pour le scénario + 60 cm, +1 m, +1.4 m et +1.8 m - PACA*
27. *Ennoiement pour le scénario + 60 cm, +1 m, +1.4 m et +1.8 m - PACA (service WFS)*
28. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en PACA - Découpage par département*
29. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en PACA - Découpage par département (service WFS)*
30. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département du Gard (30) (service WFS)*
31. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département du Gard (30)*
32. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département du Tarn (81) (service WFS)*
33. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département du Tarn (81)*
34. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département du Tarn et Garonne (82) (service WFS)*
35. *Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département du Tarn et Garonne (82)*

*Y-a-t-il des jeux de données manquants ?*

*Non, mais quelques jeux de données en plus.*

Dans le catalogue Cerema, en cochant le thème Eaux intérieures :

1. Ruissellement à la Réunion (974): Résultats Exzeco
2. Cartographie de l'artificialisation du trait de côte
3. Ouvrages et aménagements littoraux (métropole et outre-mer)
4. Nappe d'eau souterraine lyonnaise – modélisation des écoulements et de la température au pas de 20 m
5. Nappe d'eau souterraine lyonnaise – modélisation des écoulements et de la température au pas de 200 m
6. Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de la Haute Garonne (31)
7. Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Découpage par département
8. Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département des Pyrénées Orientales (66)
9. Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de l'Aveyron (12)
10. Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de l'Hérault (34)
11. Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département de la Lozère (48)
12. Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département du Gard (30)
13. Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département du Tarn (81)
14. Ruissellement sur l'Arc Méditerranéen: Résultats Exzeco en Occitanie - Département du Tarn et Garonne (82)

*Y-a-t-il des jeux de données manquants ?*

*Non, mais quelques jeux de données en plus.*

Dans le catalogue Sandre, en tapant Hydrographie (mot-clé reconnu) dans la barre de recherche :

1. Bassins DCE
2. Bassins hydrographiques- Mayotte 2021- BD Topage®
3. Bassins hydrographiques- Métropole 2019- BD Topage®
4. Bassins versant spécifiques- Métropole- Version État des Lieux 2019
5. Bassins versant topographiques- Mayotte 2021- BD Topage®
6. Bassins versant topographiques- Métropole 2019- BD Topage®
7. BD Topage®
8. Communes administratives au 1er janvier 2018
9. Communes administratives au 1er janvier 2019
10. Communes administratives au 1er janvier 2020
11. Communes administratives au 1er janvier 2021
12. Contrats de Milieu- France entière
13. Cours d'eau- Guyane 2015- BD Carthage
14. Cours d'eau- Mayotte 2013- BD Carthage
15. Cours d'eau- Mayotte 2021- BD Topage®

16. Cours d'eau- Métropole 2017- BD Carthage
17. Cours d'eau- Métropole 2019- BD Topage®
18. Cours d'eau- Réunion 2012- BD Carthage
19. Eléments hydrographiques de surface- Guyane 2015- BD Carthage
20. Hydrographie- ETH
21. Limite des 1 mille de la laisse de basse mer- Métropole
22. Masses d'eau cours d'eau- France entière- Version Rapportage 2010
23. Masses d'eau Cours d'eau- Guadeloupe- Version Etat des Lieux 2013
24. Masses d'eau Cours d'eau- Guadeloupe- Version Rapportage 2016
25. Masses d'eau Cours d'eau- Guyane- Version Rapportage 2016
26. Masses d'eau Cours d'eau- Martinique- Version Etat des Lieux 2013
27. Masses d'eau Cours d'eau- Martinique- Version Rapportage 2016
28. Masses d'eau Cours d'eau- Mayotte- Version Etat des Lieux 2013
29. Masses d'eau Cours d'eau- Mayotte- Version Rapportage 2016
30. Masses d'eau Cours d'eau- Metropole- Version Rapportage 2016
31. Masses d'eau cours d'eau- Métropole- Version Etat des Lieux 2013
32. Masses d'eau Cours d'eau- Reunion- Version Rapportage 2016
33. Masses d'eau cours d'eau- Réunion- Version Rapportage 2010
34. Masses d'eau Cours d'eau- Version Rapportage 2016- France entière
35. Masses d'eau plan d'eau- France entière- Version Rapportage 2010
36. Masses d'eau plan d'eau- France entière- Version Rapportage 2016
37. Masses d'eau plan d'eau- Guyane- Version Etat des lieux 2013
38. Masses d'eau plan d'eau- Guyane- Version Rapportage 2016
39. Masses d'eau plan d'eau- Martinique- Version Etat des lieux 2013
40. Masses d'eau plan d'eau- Metropole- Version Rapportage 2016
41. Masses d'eau plan d'eau- Métropole- Version Rapportage 2010
42. Masses d'eau plan d'eau- Réunion- Version Rapportage 2016
43. Nœuds hydrographiques- Guyane 2015- BD Carthage
44. Nœuds hydrographiques- Mayotte 2021- BD Topage®
45. Nœuds hydrographiques- Métropole 2017- BD Carthage
46. Nœuds hydrographiques- Métropole 2019- BD Topage®
47. Plans d'eau- Métropole 2017- BD Carthage
48. Points d'eau isolés- Guyane 2015- BD Carthage
49. Points d'eau isolés- Métropole 2016- BD Carthage
50. Points d'eau isolés- Métropole 2017- BD Carthage
51. Régions hydrographiques- Guyane 2015- BD Carthage
52. Régions hydrographiques- Métropole 2016- BD Carthage
53. Régions hydrographiques- Métropole 2017- BD Carthage
54. Schémas d'Aménagement et de Gestion des Eaux (SAGE)- France entière
55. Schémas d'Aménagement et de Gestion des Eaux (SAGE)- Métropole
56. Schémas d'Aménagement et de Gestion des Eaux (SAGE)- Réunion
57. Secteurs hydrographiques- Guyane 2015- BD Carthage
58. Segment de classement de continuité écologique- Guadeloupe
59. Segment de classement de continuité écologique- Liste 1 et 2- France entière
60. Segment de classement de continuité écologique- Métropole
61. Segments du Domaine Public Fluvial (DPF)- Métropole
62. Service de visualisation cartographique (WMS) du référentiel hydrographique BD CarTHAgE Guyane

63. Service de visualisation cartographique (WMS) du référentiel hydrographique BD CarTHAgE Métropole
64. Sous-secteurs hydrographiques- Guyane 2015- BD Carthage
65. Sous-Secteurs hydrographiques- Mayotte 2013- BD Carthage
66. Sous-secteurs hydrographiques- Métropole 2016- BD Carthage
67. Stations de mesure de la qualité des eaux superficielles continentales (STQ)- France entière
68. Stations de mesure de la qualité des eaux superficielles continentales (STQ)- Guadeloupe
69. Stations de mesure de la qualité des eaux superficielles continentales (STQ)- Métropole
70. Stations de mesure de la qualité des eaux superficielles continentales- STQ
71. Territoires de vigilance crues- Métropole
72. Tronçons de vigilance crues- Métropole
73. Tronçons hydrographiques- Guyane 2015- BD Carthage
74. Tronçons hydrographiques- Mayotte 2021- BD Topage®
75. Tronçons hydrographiques- Métropole 2019- BD Topage®
76. Tronçons élémentaires de masse d'eau cours d'eau- France entière- Version Rapportage 2010
77. Tronçons élémentaires de masse d'eau cours d'eau- France entière- Version Rapportage 2016
78. Tronçons élémentaires de masse d'eau cours d'eau- France Entière- Version État des Lieux 2019
79. Tronçons élémentaires de masse d'eau cours d'eau- Guadeloupe- Version Rapportage 2016
80. Tronçons élémentaires de masse d'eau cours d'eau- Guadeloupe- Version État des Lieux 2019
81. Tronçons élémentaires de masse d'eau cours d'eau- Guyane- Version Rapportage 2016
82. Tronçons élémentaires de masse d'eau cours d'eau- Guyane- Version État des Lieux 2019
83. Tronçons élémentaires de masse d'eau cours d'eau- Martinique- Version Etat des Lieux 2013
84. Tronçons élémentaires de masse d'eau cours d'eau- Martinique- Version Rapportage 2016
85. Tronçons élémentaires de masse d'eau cours d'eau- Martinique- Version État des Lieux 2019
86. Tronçons élémentaires de masse d'eau cours d'eau- Mayotte- Version Rapportage 2016
87. Tronçons élémentaires de masse d'eau cours d'eau- Mayotte- Version État des Lieux 2019
88. Tronçons élémentaires de masse d'eau cours d'eau- Métropole- Version Rapportage 2016
89. Tronçons élémentaires de masse d'eau cours d'eau- Métropole- Version État des Lieux 2019
90. Tronçons élémentaires de masse d'eau cours d'eau- Reunion- Version Rapportage 2016

91. Tronçons élémentaires de masse d'eau cours d'eau- Réunion- Version État des Lieux 2019
92. Vigilance crues- VIC
93. Zonages de gestion de la qualité des ressources (ZGR)
94. Zonages Planification (ZPL)
95. Zones d'hydrographie de texture- Guyane 2015- BD Carthage
96. Zones de régulations pour l'environnement (ZRPE)
97. Zones hydrographiques- Guyane 2015- BD Carthage
98. Zones hydrographiques- Martinique 2004- BD Carthage

On trouve donc tout ce qui est BD Carthage et BD Topage (ce qui est correct), mais aussi des zonages de gestion et réglementaires, des stations de mesure

**2. Pour le concept “Estuaire” (Q47053), le KG identifie les jeux de données suivants comme étant les 20 meilleurs :**

*Quels jeux de données identifiés vous paraissent pertinents ?*

*Ici j'ai considéré qu'en cherchant estuaire, la recherche étant plus précise on s'intéressait également aux aspects gestion (mais ça reste subjectif)*

- *Masses d'eau de transition - Guyane - Version Rapportage 2010*
- *Réseau de mesures de houle du Centre d'Archivage National des Données de Houles In-Situ (CANDHIS) - France métropole et outre-mer*
- ~~*Recensement des Installations Terminales Embranchées (ITE)*~~
- *Masses d'eau côtières - Réunion - Version Rapportage 2010*
- *Masses d'eau Côtières - Guadeloupe - Version Rapportage 2016*
- *Masses d'eau Côtières - Guyane - Version Rapportage 2016*
- *Hydroecorégions de niveau 1 (HER-1)*
- *Tronçons élémentaires de masse d'eau cours d'eau - Guadeloupe - Version Rapportage 2010*
- *Tronçons élémentaires de masse d'eau cours d'eau - Martinique - Version Rapportage 2010*
- *Tronçons élémentaires de masse d'eau cours d'eau - Réunion - Version Rapportage 2010*
- *Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2010*
- *Lieux de surveillance des eaux littorales - SEL*
- *Masses d'eau cours d'eau - Martinique - Version Rapportage 2010*
- *Masses d'eau cours d'eau - Guyane - Version Rapportage 2010*
- *Masses d'eau de transition - Guyane - Version Rapportage 2016*
- *Masses d'eau de transition - France entière - Version Rapportage 2016*
- *Masses d'eau de transition - Martinique - Version Rapportage 2016*
- *Vigilance crues - VIC*
- *Nœuds hydrographiques - Métropole 2016 - BD Carthage*

*Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ?*

Dans le catalogue Cerema,, en cherchant Estuaire dans la barre de recherche :

1. Indice de sensibilité morpho-sédimentaire du trait de côte du littoral français (indice ESI - Environnement Sensitivity Index)
2. Ortho littorale v3
3. Ortho littorale v2 (2011-2014)
4. Indice de sensibilité morpho-sédimentaire de l'estran du littoral français (indice ESI - Environnement Sensitivity Index)
5. Nombre de catastrophes naturelles liées à la mer par commune
6. Evolution du trait de côte (histogrammes)
7. Indicateur national de l'érosion côtière

*Y-a-t-il des jeux de données manquants ?*

Oui, par ex. le réseau de mesure de la houle.

Dans le catalogue Sandre, en tapant Estuaire dans la barre de recherche :

1. Trait de côte- Guyane 2015- BD Carthage
2. Zones vulnérables- Historique
3. Zones vulnérables- Métropole

*Y-a-t-il des jeux de données manquants ?*

Oui, beaucoup.

**3. Pour le concept "Rhone" (Q3429695), le KG identifie les jeux de données suivants comme étant les 20 meilleurs :**

*Quels jeux de données identifiés vous paraissent pertinents ?*

- ~~○ Tronçons élémentaires de masse d'eau cours d'eau - Guadeloupe - Version Rapportage 2010~~
- ~~○ Masses d'eau cours d'eau - Martinique - Version Rapportage 2010~~
- ~~○ Tronçons élémentaires de masse d'eau cours d'eau - Martinique - Version Rapportage 2010~~
- ~~○ Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2010~~
- ~~○ Masses d'eau cours d'eau - Guyane - Version Rapportage 2010~~
- Nœuds hydrographiques - Métropole 2016 - BD Carthage
- Lieux de surveillance des eaux littorales - SEL
- ~~○ Contextes piscicoles - Métropole~~
- ~~○ Stations de mesure de la qualité des eaux superficielles continentales (STQ) - Martinique~~

- ~~○ Stations de mesure de la qualité des eaux superficielles continentales (STQ) – Mayotte~~
- ~~○ Unités de méthanisation France métropole~~
- ~~○ Communes administratives au 1er janvier 2019~~
- ~~○ Départements administratifs au 1er janvier 2018~~
- ~~○ Tronçons élémentaires de masse d'eau cours d'eau – Réunion – Version Rapportage 2010~~
- ~~○ Masses d'eau de transition – Guyane – Version Rapportage 2016~~
- ~~○ Masses d'eau cours d'eau – Réunion – Version Rapportage 2010~~
- ~~○ Hydroecorégions de niveau 1 (HER-1)~~
- Points de prélèvements associés aux stations de mesure de la qualité des eaux superficielles continentales (STQ) - France entière
- ~~○ Points de prélèvements associés aux stations de la mesure de la qualité des eaux superficielles continentales (STQ) – Guadeloupe~~

*Rechercher de votre côté directement dans les catalogues, comment recherchez-vous, avez-vous trouvé les mêmes résultats ?*

Dans le catalogue Cerema, en tapant Rhône dans la barre de recherche et thème = 'Eaux intérieures' :

1. Nappe d'eau souterraine lyonnaise – modélisation des écoulements et de la température au pas de 200 m
2. Nappe d'eau souterraine lyonnaise – modélisation des écoulements et de la température au pas de 20 m

*Y-a-t-il des jeux de données manquants ?*

*Je ne pense pas que dans liste de départ il y avait des jeux de données Cerema.*

Dans le catalogue Sandre, en tapant Rhône dans la barre de recherche :

1. Bassins hydrographiques- Mayotte 2021- BD Topage®
2. Bassins hydrographiques- Métropole 2019- BD Topage®
3. Zones hydrographiques- Guyane 2015- BD Carthage

*Y-a-t-il des jeux de données manquants ?*

*Oui beaucoup*

Tâche 3 : évaluer le regroupement des réponses

Le KG est employé pour regrouper les records trouvés en réponse à un concept d'intérêt pour faciliter l'exploration de ces réponses. Il s'agit de regroupement multicritères (couverture géographique, thèmes, mot-clés). Le moteur présente alors les 10 groupes les plus pertinents dans le cas de recherche du concept "Estuaire" (*deuxième requête de la tâche 2*).

Nous avons surligné les records obtenus par le moteur lors de la recherche du concept "Estuaire" pour vous aider à facilement les retrouver.

- Groupe 1 :
  - Lieux de surveillance des eaux littorales - Métropole
  - Masses d'eau Côtières - Metropole - Version Rapportage 2016
  - Masses d'eau Côtières - Guadeloupe - Version Rapportage 2016
  - Masses d'eau côtières - Guyane - Version Rapportage 2010
  - Masses d'eau Côtières - Reunion - Version Rapportage 2016
  - Lieux de surveillance des eaux littorales - Mayotte
  - Masses d'eau Côtières - Mayotte - Version Rapportage 2016
  - Masses d'eau de transition - Guyane - Version Rapportage 2010
  - Masses d'eau Côtières - Guyane - Version Rapportage 2016
  - Lieux de surveillance des eaux littorales - SEL
  - Masses d'eau côtières - Réunion - Version Rapportage 2010
  - Lieux de surveillance des eaux littorales - Réunion
  - Masses d'eau côtières - Guadeloupe - Version Rapportage 2010@fr
- Groupe 2 :
  - Zones d'hydrographie de texture - Guyane 2015 - BD Carthage
  - Points d'eau isolés - Guyane 2015 - BD Carthage
  - Sous-secteurs hydrographiques - Guyane 2015 - BD Carthage
  - Zones hydrographiques - Guyane 2015 - BD Carthage
  - Nœuds hydrographiques - Guyane 2015 - BD Carthage
  - Stations de mesure de la qualité des eaux superficielles continentales (STQ) - Guyane
  - Masses d'eau plan d'eau - Guyane - Version Rapportage 2016
  - Cours d'eau - Guyane 2015 - BD Carthage
  - Masses d'eau de transition - Guyane - Version Rapportage 2016
  - Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2016
  - Secteurs hydrographiques - Guyane 2015 - BD Carthage
  - Masses d'eau plan d'eau - Guyane - Version Rapportage 2010

- Trait de côte - Guyane 2015 - BD Carthage
- Eléments hydrographiques de surface - Guyane 2015 - BD Carthage@fr
- Groupe 3 :
  - Départements administratifs au 1er janvier 2019
  - Départements administratifs au 1er janvier 2018
  - Masses d'eau de transition - France entière - Version Rapportage 2016
  - Tronçons élémentaires de masse d'eau cours d'eau - France entière - Version Rapportage 2016
  - Circonscriptions administratives de bassin 2018
  - Croisement entre les plans d'eau et les communes 2008 - France entière
  - Points de prélèvements associés aux stations de mesure de la qualité des eaux superficielles continentales (STQ) - France entière
  - Communes administratives au 1er janvier 2019
  - Masses d'eau plan d'eau - France entière - Version Rapportage 2016@fr
- Groupe 4 :
  - Hydrographie - ETH
  - Nœuds hydrographiques - Métropole 2016 - BD Carthage
  - Zones hydrographiques - Métropole 2017 - BD Carthage
  - Sous-secteurs hydrographiques - Métropole 2017 - BD Carthage
  - Tronçons hydrographiques - Métropole 2016 - BD Carthage
  - Tronçons élémentaires de masse d'eau cours d'eau - Métropole - Version Rapportage 2016
  - Cours d'eau - Métropole 2017 - BD Carthage
  - Eléments hydrographiques de surface - Métropole 2017 - BD Carthage
  - Vigilance crues - VIC
  - Zones d'hydrographie de texture - Métropole 2016 - BD Carthage
  - Secteurs hydrographiques - Métropole 2016 - BD Carthage
  - Points d'eau isolés - Métropole 2017 - BD Carthage
  - Tronçons hydrographiques - Métropole 2017 - BD Carthage@fr
- Groupe 5 :
  - Recensement des Installations Terminales Embranchées (ITE)
  - Amphibiens et dispositifs de franchissement des infrastructures de transport terrestre@fr
- Groupe 6 :

- Tronçons élémentaires de masse d'eau cours d'eau - Guadeloupe - Version Rapportage 2010@fr
- Groupe 7 :
  - Tronçons élémentaires de masse d'eau cours d'eau - Guyane - Version Rapportage 2010
  - Masses d'eau cours d'eau - Guyane - Version Rapportage 2010@fr
- Groupe 8 :
  - Masses d'eau cours d'eau - Martinique - Version Rapportage 2010
  - Tronçons élémentaires de masse d'eau cours d'eau - Martinique - Version Rapportage 2010@fr
- Groupe 9 :
  - Tronçons élémentaires de masse d'eau cours d'eau - Réunion - Version Rapportage 2010
  - Masses d'eau cours d'eau - Réunion - Version Rapportage 2010@fr
- Groupe 10 :
  - Hydroecorégions de niveau 1 (HER-1)@fr
- Groupe 11 :
  - Réseau de mesures de houle du Centre d'Archivage National des Données de Houles In-Situ (CANDHIS) - France métropole et outre-mer@fr

*Les regroupements proposés vous semblent-ils cohérents ? Comment auriez-vous regroupé ces records ?*

Pas facile. Le groupe 1 semble contenir des jeux de données de thématique marine, mais pour beaucoup des groupes suivants ça semble assez mélangé (sauf 5, 10, 11)

Pour moi il y aurait comme groupes : données marines, hydrographie terrestre, réseaux de mesure, zones de gestion, infrastructures de transport, risques., outremer (intersection possible entre les groupes)

Mais d'autres choix sont possibles. Ça mérite qu'on en discute de vive voix !

## Annexe C

# Questionnaire pour la similarité sémantique entre deux concepts du domaine de l'eau

# Expérimentation :

## Fixer un seuil de similarité sémantique minimal entre deux concepts du domaine de l'eau

### Introduction

Dans le cadre de ma thèse, j'étudie comment interroger des catalogues de données géographiques en utilisant des concepts, dits "concepts utilisateur", relatifs à l'application - pour laquelle l'utilisateur a besoin de données géographiques-. Ces concepts utilisateurs sont potentiellement différents de ceux utilisés dans les métadonnées qui décrivent les jeux, et donc dans les catalogues actuels. Lorsque le concept est un nom de lieu, des services permettent de passer de la requête de l'utilisateur, e.g. "France", à une requête sur l'extension spatiale du jeu. Je m'intéresse aux concepts de la requête relatifs à une thématique. Dans les métadonnées, les concepts relatifs à la thématique correspondent aux valeurs des propriétés "keyword" et "topic category", je les appelle "concepts des métadonnées". Pour cela, je cherche à associer à un "concept utilisateur" les "concepts des métadonnées", c'est-à-dire les concepts utilisés pour renseigner ces propriétés, qui sont suffisamment proches pour qu'on considère qu'ils sont pertinents pour interroger des catalogues en vue de construire l'application. Le but de ce questionnaire est d'avoir votre retour d'experts sur la façon dont je mesure la proximité entre un concept utilisateur et les concepts métadonnées, dans le cas particulier du domaine de l'eau.

Une première expérience avec des experts a permis d'évaluer la pertinence de recourir à l'ontologie de sens commun Wikidata pour exprimer des concepts utilisateurs, correspondant à des concepts d'intérêt applicatifs, dans le domaine de l'eau. Pour cette expérience, nous avons choisi 4 concepts Wikidata, parmi ceux suggérés par les experts, à savoir : "nappe phréatique", "station d'épuration", "marais artificiel" et "PPRI".

Les concepts des métadonnées ont été récupérés de la façon suivante. Les métadonnées de catalogues du domaine, conformes au standard INSPIRE, ont été moissonnées. J'ai ensuite extrait les valeurs des champs "Topics Category" et "Keyword" de tous les enregistrements (un enregistrement = une description d'un jeu de données à l'aide de divers champs). Ces valeurs sont des chaînes de caractères. Elles ont été utilisées pour extraire des concepts du thésaurus GEMET (standard utilisé pour documenter Topic Category dans INSPIRE) ainsi qu'avec l'ontologie de sens commun Wikidata.. Au total, 205 concepts différents ont été identifiés. Pour les concepts GEMET, un alignement a été réalisé avec Wikidata, de sorte que tous les concepts, utilisateurs ou métadonnées, sont intégrés dans un graphe dans lequel nous pouvons utiliser une mesure de similarité sémantique.

Notre but est de fixer un seuil à partir duquel on peut affirmer que deux concepts sont suffisamment proches pour que le concept métadonnées soit utilisé pour reformuler la requête de l'utilisateur. Pour cela, nous avons comparé chaque concept utilisateur à tous les concepts métadonnées en utilisant la mesure de similarité sémantique et gardé les 50 meilleurs résultats. Par ailleurs, dans une majorité des cas, nous avons affiché également le chemin entre les concepts.

## Recherche des concepts suffisamment proches de : “Nappe phréatique” (Q583075)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté?**

**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**

**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

- **i. Label GEMET <-> Label Wikidata : score de similarité [chemin entre le concept utilisateur et le concept métadonnées]**

1. label Gemet non trouvé <-> nappe phréatique@fr : 1.0  
[nappe phréatique@fr]
2. aquifère@fr <-> aquifère@fr : 0.9400091171264648  
[[nappe phréatique@fr : aquifère@fr]]
3. eaux souterraines@fr <-> eau souterraine@fr : 0.8649603128433228  
[[eau souterraine@fr : eaux intérieures@fr], (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
4. label Gemet non trouvé <-> bassin versant@fr : 0.7749752998352051  
[[bassin versant@fr : zone géographique@fr], (étendue d'eau@fr : zone géographique@fr), (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
5. label Gemet non trouvé <-> bassin hydrographique@fr : 0.7529347538948059  
Chemin non trouvé
6. fleuve@fr <-> courant d'eau@fr : 0.7424725294113159  
[[courant d'eau@fr : cours d'eau naturel@fr], (cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
7. label Gemet non trouvé <-> masse d'eau@fr : 0.7240350246429443  
[[masse d'eau@fr : étendue d'eau@fr], (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
8. fleuve@fr <-> cours d'eau@fr : 0.7219454646110535  
[[cours d'eau@fr : eaux intérieures@fr], (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
9. nappe phréatique@fr <-> niveau piézométrique@fr : 0.7217990159988403  
[[niveau piézométrique@fr : surface@fr], (surface@fr : espace à deux dimensions@fr), (espace à deux dimensions@fr : espace mathématique@fr), (espace mathématique@fr : ensemble@fr), (ensemble@fr : classe@fr), (classe@fr : entité composée@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]

10. estuaire@fr <-> estuaire@fr : 0.7091968059539795  
 [(estuaire@fr : étendue d'eau@fr), (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
11. hydrogéologie@fr <-> hydrogéologie@fr : 0.6929916143417358  
 [(hydrogéologie@fr : hydrologie@fr), (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
12. label Gemet non trouvé <-> submersion marine@fr : 0.6904162764549255  
 [(submersion marine@fr : inondation@fr), (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (objet artificiel physique@fr : entité artificielle@fr), (réceptif@fr : objet artificiel physique@fr), (réservoir@fr : réceptif@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
13. fleuve@fr <-> cours d'eau naturel@fr : 0.6870043277740479  
 [(cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
14. label Gemet non trouvé <-> conchyliculture@fr : 0.6838634610176086  
 [(conchyliculture@fr : pêche@fr), (pêche@fr : chasse@fr), (chasse@fr : action@fr), (action@fr : convention@fr), (convention@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
15. entité hydrologique@fr <-> étendue d'eau@fr : 0.6789747476577759  
 [(zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
16. littoral@fr <-> littoral@fr : 0.6748140454292297  
 [(littoral@fr : rivage@fr), (rivage@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
17. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.6704185009002686  
 [(qualité de l'eau@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
18. label Gemet non trouvé <-> gestion de l'eau@fr : 0.6675404906272888  
 [(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : relation asymétrique@fr), (relation asymétrique@fr : relation binaire@fr), (relation binaire@fr : relation@fr), (représentation@fr : relation@fr), (représentation@fr : entité artificielle@fr), (objet artificiel physique@fr : entité artificielle@fr), (réceptif@fr : objet artificiel physique@fr), (réservoir@fr : réceptif@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
19. pluie@fr <-> pluie@fr : 0.6633356809616089  
 [(pluie@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr :

*structure@fr*), (*stockage d'eau@fr : stockage@fr*), (*réservoir d'eau@fr : stockage d'eau@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]

20. *ressource naturelle@fr* <-> *ressource naturelle@fr* : 0.6476647853851318  
[[*ressource naturelle@fr : matériau brut@fr*], (*matériau brut@fr : bien@fr*), (*produit@fr : bien@fr*), (*œuvre@fr : produit@fr*), (*stockage d'eau@fr : œuvre@fr*), (*réservoir d'eau@fr : stockage d'eau@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]
21. *hydrologie@fr* <-> *hydrologie@fr* : 0.643667459487915  
[[*hydrologie@fr : science@fr*], (*science@fr : système de connaissances@fr*), (*système de connaissances@fr : ensemble@fr*), (*ensemble@fr : entité composée@fr*), (*structure@fr : entité composée@fr*), (*structure architecturale@fr : structure@fr*), (*réservoir@fr : structure architecturale@fr*), (*réservoir d'eau@fr : réservoir@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]
22. label Gemet non trouvé <-> label Wikidata introuvable : 0.6301834583282471  
*Chemin non trouvé*
23. *lac@fr* <-> *lac@fr* : 0.6288014650344849  
[[*lac@fr : eau calme@fr*], (*réservoir d'eau@fr : eau calme@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]
24. *fleuve@fr* <-> *rivière@fr* : 0.6218131184577942  
[[*rivière@fr : cours d'eau naturel@fr*], (*cours d'eau naturel@fr : cours d'eau@fr*), (*cours d'eau@fr : eaux intérieures@fr*), (*eau calme@fr : eaux intérieures@fr*), (*réservoir d'eau@fr : eau calme@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]
25. label Gemet non trouvé <-> *source d'énergie@fr* : 0.6207959651947021  
[[*source d'énergie@fr : source@fr*], (*source@fr : objet@fr*), (*entité artificielle@fr : objet@fr*), (*produit@fr : entité artificielle@fr*), (*œuvre@fr : produit@fr*), (*stockage d'eau@fr : œuvre@fr*), (*réservoir d'eau@fr : stockage d'eau@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]
26. *ferme (exploitation)@fr* <-> *ferme@fr* : 0.6167993545532227  
[[*ferme@fr : propriété agricole@fr*], (*propriété agricole@fr : complexe immobilier@fr*), (*complexe immobilier@fr : structure architecturale@fr*), (*réservoir@fr : structure architecturale@fr*), (*réservoir d'eau@fr : réservoir@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]
27. *eau usée@fr* <-> *eaux usées@fr* : 0.6070104837417603  
[[*eaux usées@fr : eau non potable@fr*], (*eau non potable@fr : eau liquide@fr*), (*eau liquide@fr : boisson@fr*), (*boisson@fr : nourriture@fr*), (*nourriture@fr : objet tangible@fr*), (*structure architecturale@fr : objet tangible@fr*), (*réservoir@fr : structure architecturale@fr*), (*réservoir d'eau@fr : réservoir@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]
28. *port@fr* <-> *havre@fr* : 0.6061320304870605  
[[*havre@fr : baie@fr*], (*baie@fr : zone aquatique@fr*), (*eaux intérieures@fr : zone aquatique@fr*), (*eau calme@fr : eaux intérieures@fr*), (*réservoir d'eau@fr : eau calme@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]
29. label Gemet non trouvé <-> *crue@fr* : 0.6060789823532104  
[[*crue@fr : catastrophe naturelle@fr*], (*catastrophe naturelle@fr : phénomène naturel@fr*), (*phénomène naturel@fr : phénomène@fr*), (*phénomène@fr : entité@fr*), (*entité composée@fr : entité@fr*), (*structure@fr : entité composée@fr*), (*structure architecturale@fr : structure@fr*), (*réservoir@fr : structure architecturale@fr*), (*réservoir d'eau@fr : réservoir@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]
30. label Gemet non trouvé <-> *prélèvement d'eau@fr* : 0.600567102432251  
[[*prélèvement d'eau@fr : infrastructure@fr*], (*infrastructure@fr : œuvre@fr*), (*stockage d'eau@fr : œuvre@fr*), (*réservoir d'eau@fr : stockage d'eau@fr*), (*aquifère@fr : réservoir d'eau@fr*), (*nappe phréatique@fr : aquifère@fr*)]

31. label Gemet non trouvé <-> saliculture@fr : 0.59828782081604  
 [(saliculture@fr : production@fr), (production@fr : activité économique@fr), (activité économique@fr : action@fr), (action@fr : convention@fr), (convention@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
32. label Gemet non trouvé <-> unité administrative locale@fr : 0.5975565910339355  
 [(unité administrative locale@fr : division territoriale d'un niveau d'imbrication donné@fr), (division territoriale d'un niveau d'imbrication donné@fr : entité territoriale administrative@fr), (entité territoriale administrative@fr : entité territoriale géographique humaine@fr), (entité territoriale géographique humaine@fr : entité géographique artificielle@fr), (entité géographique artificielle@fr : entité géographique humaine@fr), (objet artificiel physique@fr : entité artificielle@fr), (récepteur@fr : objet artificiel physique@fr), (réservoir@fr : récepteur@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
33. érosion@fr <-> érosion@fr : 0.5942178964614868  
 [(érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (convertisseur@fr : entité@fr), (outil@fr : convertisseur@fr), (récepteur@fr : outil@fr), (réservoir@fr : récepteur@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
34. fleuve@fr <-> ruisseau@fr : 0.5931576490402222  
 [(ruisseau@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
35. santé publique et assainissement@fr <-> assainissement@fr : 0.5848271250724792  
 [(assainissement@fr : hygiène@fr), (hygiène@fr : soins de santé@fr), (soins de santé@fr : service@fr), (service@fr : produit@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
36. gaz@fr <-> gaz@fr : 0.5823714137077332  
 [(gaz@fr : fluide@fr), (fluide@fr : substance@fr), (substance@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
37. risque naturel@fr <-> risque naturel@fr : 0.5795302391052246  
 [(risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (produit@fr : entité artificielle@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
38. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5750667452812195  
 [(catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
39. label Gemet non trouvé <-> BD Carthage@fr : 0.5737375020980835  
 Chemin non trouvé
40. biogaz@fr <-> biogaz@fr : 0.5729211568832397  
 [(biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : combustible@fr), (combustible@fr : matière énergétique@fr), (matière énergétique@fr : matériau@fr), (matériau@fr : produit@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
41. eau potable@fr <-> eau potable@fr : 0.5725058913230896

*[(eau potable@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

42. label Gemet non trouvé <-> Gestion des ressources naturelles@fr :

0.5694080591201782

*[(Gestion des ressources naturelles@fr : maltraitance@fr), (maltraitance@fr : problème sociétal@fr), (problème sociétal@fr : problème@fr), (problème@fr : condition@fr), (condition@fr : état@fr), (état@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

43. label Gemet non trouvé <-> perméabilité@fr : 0.5676491856575012

*[(perméabilité@fr : propriété physique@fr), (propriété physique@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

44. hydrographie@fr <-> hydrographie@fr : 0.5633646845817566

*[(hydrographie@fr : sciences de la Terre@fr), (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

45. lac@fr <-> ensemble de lacs@fr : 0.563302755355835

*[(ensemble de lacs@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

46. label Gemet non trouvé <-> cordon littoral@fr : 0.5623621344566345

*Chemin non trouvé*

47. biotope@fr <-> biotope@fr : 0.5620291829109192

*[(biotope@fr : entité géographique naturelle@fr), (entité géographique naturelle@fr : entité géographique@fr), (entité géographique@fr : entité@fr), (convertisseur@fr : entité@fr), (outil@fr : convertisseur@fr), (récipient@fr : outil@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

48. label Gemet non trouvé <-> plan d'eau@fr : 0.5611335635185242

*[(plan d'eau@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

49. gestion des déchets@fr <-> gestion des déchets@fr : 0.5611010193824768

*[(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

50. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.5600736737251282

*[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure architecturale@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

## Recherche des concepts suffisamment proches de : "Phytoépuration" (Q1376406)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté?**

**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**

**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

- **i. Label GEMET <-> Label Wikidata : score de similarité**  
**[chemin entre le concept utilisateur et le concept métadonnées ]**
1. pluie@fr <-> pluie@fr : 0.732921838760376  
[[pluie@fr : phénomène naturel@fr], (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
  2. label Gemet non trouvé <-> bassin versant@fr : 0.6982258558273315  
[[bassin versant@fr : zone géographique@fr], (étendue d'eau@fr : zone géographique@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
  3. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.6933472156524658  
[[ouvrage hydraulique@fr : ouvrage d'art@fr], (ouvrage d'art@fr : structure construite@fr), (structure construite@fr : objet géographique@fr), (objet géographique naturel@fr : objet géographique@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
  4. label Gemet non trouvé <-> submersion marine@fr : 0.6883678436279297  
[[submersion marine@fr : inondation@fr], (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
  5. label Gemet non trouvé <-> masse d'eau@fr : 0.6847524642944336  
[[masse d'eau@fr : étendue d'eau@fr], (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
  6. label Gemet non trouvé <-> prélèvement d'eau@fr : 0.66660475730896  
[[prélèvement d'eau@fr : infrastructure@fr], (infrastructure@fr : œuvre@fr), (œuvre@fr : produit@fr), (produit@fr : entité artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
  7. estuaire@fr <-> estuaire@fr : 0.6564208269119263  
[[estuaire@fr : écosystème@fr], (zone humide@fr : écosystème@fr), (Phytoépuration@fr : zone humide@fr)]
  8. eau usée@fr <-> eaux usées@fr : 0.6487299203872681  
[[eaux usées@fr : eau non potable@fr], (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

9. label Gemet non trouvé <-> nappe phréatique@fr : 0.6477580070495605  
 [(nappe phréatique@fr : aquifère@fr), (aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : eau calme@fr), (eau calme@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
10. hydrologie@fr <-> hydrologie@fr : 0.6444834470748901  
 [(hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : système@fr), (Q106865236 : système@fr), (Q106865236 : objet naturel@fr), (objet physique naturel@fr : objet naturel@fr), (objet géographique naturel@fr : objet physique naturel@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
11. label Gemet non trouvé <-> gestion de l'eau@fr : 0.6390826106071472  
 [(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : aptitude@fr), (aptitude@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
12. aquifère@fr <-> aquifère@fr : 0.6373130679130554  
 [(aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : eau calme@fr), (eau calme@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
13. protection contre les crues@fr <-> inondation@fr : 0.6344674229621887  
 [(inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
14. pollution industrielle@fr <-> pollution@fr : 0.6167508959770203  
 [(pollution@fr : effet pervers@fr), (effet pervers@fr : résultat@fr), (résultat@fr : entité observable@fr), (entité observable@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
15. santé publique et assainissement@fr <-> assainissement@fr : 0.6126349568367004  
 [(assainissement@fr : facteur écologique@fr), (facteur écologique@fr : corrélation@fr), (corrélation@fr : relation@fr), (représentation@fr : relation@fr), (représentation@fr : entité artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
16. nappe phréatique@fr <-> niveau piézométrique@fr : 0.6110435724258423  
 [(niveau piézométrique@fr : surface@fr), (surface@fr : espace à deux dimensions@fr), (espace à deux dimensions@fr : espace mathématique@fr), (espace mathématique@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
17. label Gemet non trouvé <-> saliculture@fr : 0.6026057004928589  
 [(saliculture@fr : production@fr), (production@fr : processus@fr), (processus@fr : changement@fr), (changement@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité physique@fr : entité spatio-temporelle@fr), (objet tangible@fr : entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
18. biogaz@fr <-> biogaz@fr : 0.5992053747177124  
 [(biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : combustible@fr), (combustible@fr : matière énergétique@fr), (matière énergétique@fr : matériau@fr), (matériau@fr : produit@fr), (produit@fr :

*entité artificielle@fr*, (*entité géographique artificielle@fr : entité artificielle@fr*), (*forme de relief artificielle@fr : entité géographique artificielle@fr*), (*Phytoépuration@fr : forme de relief artificielle@fr*)

19. eaux souterraines@fr <-> eau souterraine@fr : 0.5980356931686401  
[[*eau souterraine@fr : eaux intérieures@fr*], (*eaux intérieures@fr : zone aquatique@fr*), (*zone aquatique@fr : étendue d'eau@fr*), (*zone humide@fr : étendue d'eau@fr*), (*Phytoépuration@fr : zone humide@fr*)]
20. label Gemet non trouvé <-> crue@fr : 0.5977764129638672  
[[*crue@fr : catastrophe naturelle@fr*], (*catastrophe naturelle@fr : phénomène naturel@fr*), (*phénomène naturel@fr : phénomène@fr*), (*phénomène@fr : entité@fr*), (*entité géographique@fr : entité@fr*), (*entité géographique artificielle@fr : entité géographique@fr*), (*forme de relief artificielle@fr : entité géographique artificielle@fr*), (*Phytoépuration@fr : forme de relief artificielle@fr*)]
21. urbanisme@fr <-> aménagement urbain@fr : 0.5961806774139404  
[[*aménagement urbain@fr : urbanisme@fr*], (*urbanisme@fr : géographie humaine@fr*), (*géographie humaine@fr : géographie@fr*), (*géographie@fr : objet de la recherche@fr*), (*objet de la recherche@fr : objet@fr*), (*entité artificielle@fr : objet@fr*), (*entité géographique artificielle@fr : entité artificielle@fr*), (*forme de relief artificielle@fr : entité géographique artificielle@fr*), (*Phytoépuration@fr : forme de relief artificielle@fr*)]
22. ferme (exploitation)@fr <-> ferme@fr : 0.5928126573562622  
[[*ferme@fr : établissement rural@fr*], (*établissement rural@fr : localité@fr*), (*localité@fr : entité géographique artificielle@fr*), (*forme de relief artificielle@fr : entité géographique artificielle@fr*), (*Phytoépuration@fr : forme de relief artificielle@fr*)]
23. label Gemet non trouvé <-> conchyliculture@fr : 0.5910194516181946  
[[*conchyliculture@fr : pêche@fr*], (*pêche@fr : chasse@fr*), (*chasse@fr : action@fr*), (*action@fr : occurrence@fr*), (*occurrence@fr : entité temporelle@fr*), (*entité temporelle@fr : entité spatio-temporelle@fr*), (*entité physique@fr : entité spatio-temporelle@fr*), (*objet tangible@fr : entité physique@fr*), (*objet physico-géographique@fr : objet tangible@fr*), (*forme de relief artificielle@fr : objet physico-géographique@fr*), (*Phytoépuration@fr : forme de relief artificielle@fr*)]
24. hydrogéologie@fr <-> hydrogéologie@fr : 0.5909260511398315  
[[*hydrogéologie@fr : hydrologie@fr*], (*hydrologie@fr : science@fr*), (*science@fr : système de connaissances@fr*), (*système de connaissances@fr : ensemble@fr*), (*ensemble@fr : entité composée@fr*), (*entité composée@fr : entité@fr*), (*entité géographique@fr : entité@fr*), (*entité géographique artificielle@fr : entité géographique@fr*), (*forme de relief artificielle@fr : entité géographique artificielle@fr*), (*Phytoépuration@fr : forme de relief artificielle@fr*)]
25. label Gemet non trouvé <-> Plan de prévention du risque inondation@fr : 0.5880736708641052  
*Chemin non trouvé*
26. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5866144895553589  
[[*catastrophe naturelle@fr : phénomène naturel@fr*], (*phénomène naturel@fr : phénomène@fr*), (*phénomène@fr : entité@fr*), (*entité géographique@fr : entité@fr*), (*entité géographique artificielle@fr : entité géographique@fr*), (*forme de relief artificielle@fr : entité géographique artificielle@fr*), (*Phytoépuration@fr : forme de relief artificielle@fr*)]
27. développement urbain@fr <-> urbanisme@fr : 0.5849825143814087  
[[*urbanisme@fr : géographie humaine@fr*], (*géographie humaine@fr : géographie@fr*), (*géographie@fr : objet de la recherche@fr*), (*objet de la recherche@fr : objet@fr*), (*entité artificielle@fr : objet@fr*), (*entité géographique artificielle@fr : entité artificielle@fr*), (*forme de relief artificielle@fr : entité géographique artificielle@fr*), (*Phytoépuration@fr : forme de relief artificielle@fr*)]
28. littoral@fr <-> littoral@fr : 0.5812649726867676  
[[*littoral@fr : rivage@fr*], (*rivage@fr : objet physico-géographique@fr*), (*forme de relief artificielle@fr : objet physico-géographique@fr*), (*Phytoépuration@fr : forme de relief artificielle@fr*)]

29. hydrographie@fr <-> hydrographie@fr : 0.579881489276886  
 [(hydrographie@fr : sciences de la Terre@fr), (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
30. risque naturel@fr <-> risque naturel@fr : 0.5705837607383728  
 [(risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
31. fleuve@fr <-> cours d'eau naturel@fr : 0.5690486431121826  
 [(cours d'eau naturel@fr : objet géographique naturel@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
32. label Gemet non trouvé <-> label Wikidata introuvable : 0.5682553648948669  
 Chemin non trouvé
33. gestion des déchets@fr <-> gestion des déchets@fr : 0.567966639995575  
 [(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
34. ressource naturelle@fr <-> ressource naturelle@fr : 0.5668643712997437  
 [(ressource naturelle@fr : ressource@fr), (ressource@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
35. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.5630149841308594  
 [(qualité de l'eau@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
36. label Gemet non trouvé <-> masse d'eau de surface@fr : 0.5613843202590942  
 [(masse d'eau de surface@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
37. fleuve@fr <-> cours d'eau@fr : 0.5561116933822632  
 [(cours d'eau@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
38. fleuve@fr <-> courant d'eau@fr : 0.5535402894020081  
 [(courant d'eau@fr : cours d'eau naturel@fr), (cours d'eau naturel@fr : objet géographique naturel@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
39. port@fr <-> havre@fr : 0.5492718815803528  
 [(havre@fr : baie@fr), (baie@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
40. érosion@fr <-> érosion@fr : 0.5444740056991577  
 [(érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité@fr)]

*géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*

41. **agriculture@fr <-> agriculture@fr : 0.5412713289260864**  
*[(agriculture@fr : agriculture et sylviculture@fr), (agriculture et sylviculture@fr : économie@fr), (économie@fr : système social@fr), (système social@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
42. **environnement@fr <-> environnement@fr : 0.5334920883178711**  
*Chemin non trouvé*
43. **polluant@fr <-> polluant@fr : 0.531983494758606**  
*[(polluant@fr : substance physique@fr), (substance physique@fr : entité physique@fr), (objet tangible@fr : entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
44. **lac@fr <-> lac@fr : 0.5294979214668274**  
*[(lac@fr : modelé@fr), (forme de relief artificielle@fr : modelé@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
45. **entité hydrologique@fr <-> étendue d'eau@fr : 0.5264604687690735**  
*[(zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]*
46. **gestion@fr <-> label Wikidata introuvable : 0.5256054997444153**  
*Chemin non trouvé*
47. **diversité biologique@fr <-> biodiversité@fr : 0.5250712633132935**  
*[(biodiversité@fr : diversité@fr), (diversité@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
48. **label Gemet non trouvé <-> La Réunion@fr : 0.5243727564811707**  
*Chemin non trouvé*
49. **label Gemet non trouvé <-> bassin hydrographique@fr : 0.5230716466903687**  
*Chemin non trouvé*
50. **label Gemet non trouvé <-> réseau de transport@fr : 0.522017776966095**  
*[(réseau de transport@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*

## Recherche des concepts suffisamment proches de : “Plan de prévention du risque inondation” (Q3391269)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté?  
A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**

*(Note Mehdi : il n'y a pas de chemins entre PPRI et les autres concepts. PPRI est isolé dans le graphe de Wikidata)*

### Liste des comparaisons : (meilleurs 50 résultats)

1. label Gemet non trouvé <-> Plan de prévention du risque inondation@fr : 1.0
2. label Gemet non trouvé <-> submersion marine@fr : 0.699097752571106
3. label Gemet non trouvé <-> protection@fr : 0.65545654296875
4. label Gemet non trouvé <-> masse d'eau@fr : 0.6312357187271118
5. estuaire@fr <-> estuaire@fr : 0.6220836639404297
6. aquifère@fr <-> aquifère@fr : 0.6191027164459229
7. label Gemet non trouvé <-> directive-cadre sur l'eau@fr : 0.6143583655357361
8. label Gemet non trouvé <-> conchyliculture@fr : 0.6103661060333252
9. communication en matière de risques@fr <-> gestion du risque@fr : 0.6078090071678162
10. port@fr <-> havre@fr : 0.6073088049888611
11. label Gemet non trouvé <-> bassin versant@fr : 0.606191873550415
12. label Gemet non trouvé <-> label Wikidata introuvable : 0.5888326168060303
13. plan@fr <-> plan@fr : 0.5848706364631653
14. fleuve@fr <-> cours d'eau@fr : 0.5825957655906677
15. hydrographie@fr <-> hydrographie@fr : 0.5800302624702454
16. politique@fr <-> politique@fr : 0.5799053907394409
17. hydrogéologie@fr <-> hydrogéologie@fr : 0.5739492177963257
18. label Gemet non trouvé <-> bassin hydrographique@fr : 0.5727195143699646
19. label Gemet non trouvé <-> prévention@fr : 0.567156195640564
20. pluie@fr <-> pluie@fr : 0.5633394718170166
21. label Gemet non trouvé <-> Auvergne-Rhône-Alpes@fr : 0.5619726181030273
22. ferme (exploitation)@fr <-> ferme@fr : 0.5597040057182312
23. label Gemet non trouvé <-> nappe phréatique@fr : 0.5578835010528564
24. label Gemet non trouvé <-> Guadeloupe@fr : 0.5573101043701172
25. label Gemet non trouvé <-> cordon littoral@fr : 0.5549724102020264
26. label Gemet non trouvé <-> Guyane@fr : 0.554639995098114
27. label Gemet non trouvé <-> Politique environnementale de l'Union européenne@fr : 0.5531107187271118
28. label Gemet non trouvé <-> crue@fr : 0.5518550872802734
29. planification@fr <-> planification@fr : 0.5512016415596008
30. label Gemet non trouvé <-> saliculture@fr : 0.5446555614471436
31. politique de l'environnement@fr <-> politique environnementale@fr : 0.5440297722816467
32. fleuve@fr <-> courant d'eau@fr : 0.5439935922622681
33. gestion de crise@fr <-> gestion de crise@fr : 0.537813663482666

34. risque naturel@fr <-> risque naturel@fr : 0.5369033813476562
35. ressource naturelle@fr <-> ressource naturelle@fr : 0.5360186100006104
36. littoral@fr <-> littoral@fr : 0.5313910245895386
37. réserve naturelle nationale@fr <-> aire protégée@fr : 0.530930757522583
38. label Gemet non trouvé <-> source d'énergie@fr : 0.5305038690567017
39. surveillance@fr <-> surveillance@fr : 0.5277665853500366
40. label Gemet non trouvé <-> Bretagne@fr : 0.5256977081298828
41. label Gemet non trouvé <-> gestion de l'eau@fr : 0.5250158309936523
42. environnement@fr <-> environnement@fr : 0.5216457843780518
43. fleuve@fr <-> rivière@fr : 0.5166248083114624
44. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5143988132476807
45. label Gemet non trouvé <-> Mayotte@fr : 0.5137329697608948
46. source d'énergie renouvelable@fr <-> énergie renouvelable@fr : 0.5136480331420898
47. risque@fr <-> risque@fr : 0.5124579071998596
48. eau usée@fr <-> eaux usées@fr : 0.5105905532836914
49. fleuve@fr <-> cours d'eau naturel@fr : 0.5103414058685303
50. label Gemet non trouvé <-> lycée@fr : 0.5079976320266724

## Recherche des concepts suffisamment proches de : “Station de traitement des eaux usées”

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté?**

**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**

**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

**Label GEMET <-> Label Wikidata : score de similarité**

**[chemin entre le concept utilisateur et le concept métadonnées]**

1. eau usée@fr <-> eaux usées@fr : 0.7805310487747192  
[(eaux usées@fr : eau non potable@fr), (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
2. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.7642497420310974  
[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
3. gestion des déchets@fr <-> gestion des déchets@fr : 0.761856198310852  
[(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
4. label Gemet non trouvé <-> gestion de l'eau@fr : 0.7167595624923706  
[(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : aptitude@fr), (aptitude@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
5. bâtiment public@fr <-> structure construite@fr : 0.6913622617721558  
[(équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
6. santé publique et assainissement@fr <-> assainissement@fr : 0.6888487935066223  
[(assainissement@fr : activité@fr), (activité@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (ensemble architectural@fr : groupe de constructions et de bâtiments@fr), (complexe immobilier@fr : ensemble architectural@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
7. label Gemet non trouvé <-> saliculture@fr : 0.6875752210617065  
[(saliculture@fr : production@fr), (production@fr : processus@fr), (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de

*bâtiments@fr : groupe d'entités géographiques@fr), (groupe de constructions et de bâtiments@fr : structure construite@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*

8. label Gemet non trouvé <-> station d'observation@fr : 0.6836385726928711  
*[(station d'observation@fr : équipement@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
9. label Gemet non trouvé <-> source d'énergie@fr : 0.6789122819900513  
*[(source d'énergie@fr : source@fr), (source@fr : objet@fr), (objet@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
10. urbanisme@fr <-> aménagement urbain@fr : 0.6710588335990906  
*[(aménagement urbain@fr : urbanisme@fr), (urbanisme@fr : géographie humaine@fr), (géographie humaine@fr : société@fr), (société@fr : groupe d'humains@fr), (organisation@fr : groupe d'humains@fr), (firme@fr : organisation@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
11. polluant@fr <-> polluant@fr : 0.6678998470306396  
*[(polluant@fr : substance physique@fr), (substance physique@fr : entité physique@fr), (objet tangible@fr : entité physique@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*
12. captage@fr <-> zone de chalandise@fr : 0.664354681968689  
*[(zone de chalandise@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
13. label Gemet non trouvé <-> masse d'eau@fr : 0.6608922481536865  
*[(masse d'eau@fr : étendue d'eau@fr), (étendue d'eau@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
14. hydrogéologie@fr <-> hydrogéologie@fr : 0.6581923961639404  
*[(hydrogéologie@fr : hydrologie@fr), (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*
15. hydrologie@fr <-> hydrologie@fr : 0.655835747718811  
*[(hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : classe@fr), (forme juridique@fr : classe@fr), (personne morale@fr : forme juridique@fr), (firme@fr : personne morale@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
16. label Gemet non trouvé <-> Gestion des ressources naturelles@fr : 0.653746485710144  
*[(Gestion des ressources naturelles@fr : maltraitance@fr), (maltraitance@fr : problème sociétal@fr), (problème sociétal@fr : problème@fr), (problème@fr : condition@fr), (condition@fr : état@fr), (état@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
17. chaleur@fr <-> chaleur@fr : 0.6528462171554565  
*[(chaleur@fr : grandeur physique@fr), (grandeur physique@fr : quantité@fr), (quantité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*

18. développement urbain@fr <-> urbanisme@fr : 0.6490142345428467  
 [(urbanisme@fr : géographie humaine@fr), (géographie humaine@fr : société@fr), (société@fr : groupe d'humains@fr), (organisation@fr : groupe d'humains@fr), (firme@fr : organisation@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
19. géologie@fr <-> géologie@fr : 0.6452451944351196  
 [(géologie@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : classe@fr), (forme juridique@fr : classe@fr), (personne morale@fr : forme juridique@fr), (firme@fr : personne morale@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
20. pollution industrielle@fr <-> pollution@fr : 0.6422240734100342  
 [(pollution@fr : effet pervers@fr), (effet pervers@fr : résultat@fr), (résultat@fr : entité observable@fr), (entité observable@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
21. label Gemet non trouvé <-> infrastructure de transport@fr : 0.64194655418396  
 [(infrastructure de transport@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
22. label Gemet non trouvé <-> bassin hydrographique@fr : 0.6412675380706787  
 Chemin non trouvé
23. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.6363604664802551  
 [(qualité de l'eau@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
24. label Gemet non trouvé <-> bien culturel@fr : 0.6321249604225159  
 [(bien culturel@fr : objet concret@fr), (objet tangible@fr : objet concret@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
25. label Gemet non trouvé <-> label Wikidata introuvable : 0.6305616497993469  
 Chemin non trouvé
26. label Gemet non trouvé <-> nappe phréatique@fr : 0.6300790905952454  
 [(nappe phréatique@fr : aquifère@fr), (aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : réservoir@fr), (réservoir@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
27. hydrographie@fr <-> hydrographie@fr : 0.6294116973876953  
 [(hydrographie@fr : sciences de la Terre@fr), (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
28. infrastructure@fr <-> infrastructure@fr : 0.628700852394104  
 [(infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
29. label Gemet non trouvé <-> submersion marine@fr : 0.6275973320007324  
 [(submersion marine@fr : inondation@fr), (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr :

entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]

30. biogaz@fr <-> biogaz@fr : 0.6261656284332275  
[[biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : gaz@fr), (gaz@fr : fluide@fr), (fluide@fr : substance@fr), (substance@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
31. logement@fr <-> gestion de logement@fr : 0.6251828670501709  
Chemin non trouvé
32. aquifère@fr <-> aquifère@fr : 0.6246013641357422  
[[aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : réservoir@fr), (réservoir@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
33. communication en matière de risques@fr <-> gestion du risque@fr :  
0.6228855848312378  
[[gestion du risque@fr : fonction@fr), (fonction@fr : usage@fr), (usage@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
34. monument@fr <-> monument de Hongrie@fr : 0.6182126998901367  
[[monument de Hongrie@fr : bien patrimonial national@fr), (bien patrimonial national@fr : site patrimonial@fr), (site patrimonial@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
35. label Gemet non trouvé <-> label Wikidata introuvable : 0.6179236173629761  
Chemin non trouvé
36. label Gemet non trouvé <-> bassin versant@fr : 0.6176934242248535  
[[bassin versant@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
37. habitat@fr <-> habitat@fr : 0.6176526546478271  
[[habitat@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
38. ressource naturelle@fr <-> ressource naturelle@fr : 0.6150418519973755  
[[ressource naturelle@fr : matériau brut@fr), (matériau brut@fr : bien@fr), (produit@fr : bien@fr), (œuvre@fr : produit@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
39. risque naturel@fr <-> risque naturel@fr : 0.6141432523727417  
[[risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (objet@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
40. érosion@fr <-> érosion@fr : 0.6131541728973389  
[[érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
41. polluant@fr <-> label Wikidata introuvable : 0.6121931076049805  
Chemin non trouvé

42. label Gemet non trouvé <-> réseau de télécommunications@fr :  
0.6115759015083313  
[[réseau de télécommunications@fr : infrastructure@fr], (infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
43. surveillance@fr <-> surveillance@fr : 0.6091445684432983  
[[surveillance@fr : activité@fr], (Q18390984 : activité@fr), (Q18390984 : outil de gestion du risque@fr), (outil de gestion du risque@fr : outil de gestion@fr), (outil de gestion@fr : outil@fr), (outil@fr : équipement@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
44. estuaire@fr <-> estuaire@fr : 0.6066519021987915  
[[estuaire@fr : étendue d'eau@fr], (étendue d'eau@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
45. dommage@fr <-> dégât@fr : 0.6065858602523804  
[[dégât@fr : processus@fr], (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (ensemble architectural@fr : groupe de constructions et de bâtiments@fr), (complexe immobilier@fr : ensemble architectural@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
46. logement@fr <-> appartement@fr : 0.6065312623977661  
[[appartement@fr : logement@fr], (logement@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
47. label Gemet non trouvé <-> transition écologique@fr : 0.6048821210861206  
[[transition écologique@fr : changement@fr], (changement@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité spatio-temporelle@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
48. chauffage à distance@fr <-> réseau de chaleur@fr : 0.603854775428772  
[[réseau de chaleur@fr : infrastructure@fr], (infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
49. zonage@fr <-> zonage@fr : 0.6028785705566406  
[[zonage@fr : planification@fr], (planification@fr : processus@fr), (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (groupe de constructions et de bâtiments@fr : structure construite@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
50. label Gemet non trouvé <-> unité statistique@fr : 0.6022796630859375  
[[unité statistique@fr : unité@fr], (unité@fr : objet abstrait@fr), (objet abstrait@fr : entité non-physique@fr), (entité non-physique@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]

## Annexe D

# Réponses au questionnaire pour la similarité sémantique entre deux concepts du domaine de l'eau

# Expérimentation :

## Fixer un seuil de similarité sémantique minimal entre deux concepts du domaine de l'eau

### Introduction

Dans le cadre de ma thèse, j'étudie comment interroger des catalogues de données géographiques en utilisant des concepts, dits "concepts utilisateur", relatifs à l'application - pour laquelle l'utilisateur a besoin de données géographiques-. Ces concepts utilisateurs sont potentiellement différents de ceux utilisés dans les métadonnées qui décrivent les jeux, et donc dans les catalogues actuels. Lorsque le concept est un nom de lieu, des services permettent de passer de la requête de l'utilisateur, e.g. "France", à une requête sur l'extension spatiale du jeu. Je m'intéresse aux concepts de la requête relatifs à une thématique. Dans les métadonnées, les concepts relatifs à la thématique correspondent aux valeurs des propriétés "keyword" et "topic category", je les appelle "concepts des métadonnées". Pour cela, je cherche à associer à un "concept utilisateur" les "concepts des métadonnées", c'est-à-dire les concepts utilisés pour renseigner ces propriétés, qui sont suffisamment proches pour qu'on considère qu'ils sont pertinents pour interroger des catalogues en vue de construire l'application. Le but de ce questionnaire est d'avoir votre retour d'experts sur la façon dont je mesure la proximité entre un concept utilisateur et les concepts métadonnées, dans le cas particulier du domaine de l'eau.

Une première expérience avec des experts a permis d'évaluer la pertinence de recourir à l'ontologie de sens commun Wikidata pour exprimer des concepts utilisateurs, correspondant à des concepts d'intérêt applicatifs, dans le domaine de l'eau. Pour cette expérience, nous avons choisi 4 concepts Wikidata, parmi ceux suggérés par les experts, à savoir : "nappe phréatique", "station d'épuration", "marais artificiel" et "PPRI".

Les concepts des métadonnées ont été récupérés de la façon suivante. Les métadonnées de catalogues du domaine, conformes au standard INSPIRE, ont été moissonnées. J'ai ensuite extrait les valeurs des champs "Topics Category" et "Keyword" de tous les enregistrements (un enregistrement = une description d'un jeu de données à l'aide de divers champs). Ces valeurs sont des chaînes de caractères. Elles ont été utilisées pour extraire des concepts du thésaurus GEMET (standard utilisé pour documenter Topic Category dans INSPIRE) ainsi qu'avec l'ontologie de sens commun Wikidata.. Au total, 205 concepts différents ont été identifiés. Pour les concepts GEMET, un alignement a été réalisé avec Wikidata, de sorte que tous les concepts, utilisateurs ou métadonnées, sont intégrés dans un graphe dans lequel nous pouvons utiliser une mesure de similarité sémantique.

Notre but est de fixer un seuil à partir duquel on peut affirmer que deux concepts sont suffisamment proches pour que le concept métadonnées soit utilisé pour reformuler la requête de l'utilisateur. Pour cela, nous avons comparé chaque concept utilisateur à tous les concepts métadonnées en utilisant la mesure de similarité sémantique et gardé les 50 meilleurs résultats. Par ailleurs, dans une majorité des cas, nous avons affiché également le chemin entre les concepts.

## Méthode employée par Timothée :

En vert : concepts proches/pertinents. Ils sont classés par ordre de pertinence.

En orange : concepts pas assez proches mais plutôt englobants. Non classés.

En rouge : concepts jugés comme trop éloignés. Non classés.

## Recherche des concepts suffisamment proches de : “Nappe phréatique” (Q583075)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté ?**  
**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**  
**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

- **i. Label GEMET <-> Label Wikidata : score de similarité [chemin entre le concept utilisateur et le concept métadonnées]**
  1. label Gemet non trouvé <-> nappe phréatique@fr : 1.0  
[nappe phréatique@fr]
  2. aquifère@fr <-> aquifère@fr : 0.9400091171264648  
[(nappe phréatique@fr : aquifère@fr)]
  3. eaux souterraines@fr <-> eau souterraine@fr : 0.8649603128433228  
[(eau souterraine@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
  4. nappe phréatique@fr <-> niveau piézométrique@fr : 0.7217990159988403  
[(niveau piézométrique@fr : surface@fr), (surface@fr : espace à deux dimensions@fr), (espace à deux dimensions@fr : espace mathématique@fr), (espace mathématique@fr : ensemble@fr), (ensemble@fr : classe@fr), (classe@fr : entité composée@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
  5. hydrogéologie@fr <-> hydrogéologie@fr : 0.6929916143417358  
[(hydrogéologie@fr : hydrologie@fr), (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
  6. eau potable@fr <-> eau potable@fr : 0.5725058913230896  
[(eau potable@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
  7. label Gemet non trouvé <-> prélèvement d'eau@fr : 0.600567102432251

*[(prélèvement d'eau@fr : infrastructure@fr), (infrastructure@fr : œuvre@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

8. pluie@fr <-> pluie@fr : 0.6633356809616089  
*[(pluie@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
9. label Gemet non trouvé <-> perméabilité@fr : 0.5676491856575012  
*[(perméabilité@fr : propriété physique@fr), (propriété physique@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
10. label Gemet non trouvé <-> gestion de l'eau@fr : 0.6675404906272888  
*[(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : relation asymétrique@fr), (relation asymétrique@fr : relation binaire@fr), (relation binaire@fr : relation@fr), (représentation@fr : relation@fr), (représentation@fr : entité artificielle@fr), (objet artificiel physique@fr : entité artificielle@fr), (récipient@fr : objet artificiel physique@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
11. ressource naturelle@fr <-> ressource naturelle@fr : 0.6476647853851318  
*[(ressource naturelle@fr : matériau brut@fr), (matériau brut@fr : bien@fr), (produit@fr : bien@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
12. label Gemet non trouvé <-> bassin versant@fr : 0.7749752998352051  
*[(bassin versant@fr : zone géographique@fr), (étendue d'eau@fr : zone géographique@fr), (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
13. label Gemet non trouvé <-> bassin hydrographique@fr : 0.7529347538948059  
*Chemin non trouvé*
14. fleuve@fr <-> courant d'eau@fr : 0.7424725294113159  
*[(courant d'eau@fr : cours d'eau naturel@fr), (cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
15. label Gemet non trouvé <-> masse d'eau@fr : 0.7240350246429443  
*[(masse d'eau@fr : étendue d'eau@fr), (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
16. fleuve@fr <-> cours d'eau@fr : 0.7219454646110535  
*[(cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
17. estuaire@fr <-> estuaire@fr : 0.7091968059539795  
*[(estuaire@fr : étendue d'eau@fr), (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
18. label Gemet non trouvé <-> submersion marine@fr : 0.6904162764549255

*[(submersion marine@fr : inondation@fr), (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (objet artificiel physique@fr : entité artificielle@fr), (récipient@fr : objet artificiel physique@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**19. fleuve@fr <-> cours d'eau naturel@fr : 0.6870043277740479**

*[(cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**20. label Gemet non trouvé <-> conchyliculture@fr : 0.6838634610176086**

*[(conchyliculture@fr : pêche@fr), (pêche@fr : chasse@fr), (chasse@fr : action@fr), (action@fr : convention@fr), (convention@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**21. entité hydrologique@fr <-> étendue d'eau@fr : 0.6789747476577759**

*[(zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**22. littoral@fr <-> littoral@fr : 0.6748140454292297**

*[(littoral@fr : rivage@fr), (rivage@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**23. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.6704185009002686**

*[(qualité de l'eau@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**24. hydrologie@fr <-> hydrologie@fr : 0.643667459487915**

*[(hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**25. label Gemet non trouvé <-> label Wikidata introuvable : 0.6301834583282471**

*Chemin non trouvé*

**26. lac@fr <-> lac@fr : 0.6288014650344849**

*[(lac@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**27. fleuve@fr <-> rivière@fr : 0.6218131184577942**

*[(rivière@fr : cours d'eau naturel@fr), (cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**28. label Gemet non trouvé <-> source d'énergie@fr : 0.6207959651947021**

*[(source d'énergie@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (produit@fr : entité artificielle@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

29. ferme (exploitation)@fr <-> ferme@fr : 0.6167993545532227  
[[ferme@fr : propriété agricole@fr], (propriété agricole@fr : complexe immobilier@fr), (complexe immobilier@fr : structure architecturale@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
30. eau usée@fr <-> eaux usées@fr : 0.6070104837417603  
[[eaux usées@fr : eau non potable@fr], (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
31. port@fr <-> havre@fr : 0.6061320304870605  
[[havre@fr : baie@fr], (baie@fr : zone aquatique@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
32. label Gemet non trouvé <-> crue@fr : 0.6060789823532104  
[[crue@fr : catastrophe naturelle@fr], (catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
33. label Gemet non trouvé <-> saliculture@fr : 0.59828782081604  
[[saliculture@fr : production@fr], (production@fr : activité économique@fr), (activité économique@fr : action@fr), (action@fr : convention@fr), (convention@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
34. label Gemet non trouvé <-> unité administrative locale@fr : 0.5975565910339355  
[[unité administrative locale@fr : division territoriale d'un niveau d'imbrication donné@fr], (division territoriale d'un niveau d'imbrication donné@fr : entité territoriale administrative@fr), (entité territoriale administrative@fr : entité territoriale géographique humaine@fr), (entité territoriale géographique humaine@fr : entité géographique artificielle@fr), (entité géographique artificielle@fr : entité géographique humaine@fr), (objet artificiel physique@fr : entité artificielle@fr), (récipient@fr : objet artificiel physique@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
35. érosion@fr <-> érosion@fr : 0.5942178964614868  
[[érosion@fr : phénomène naturel@fr], (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (convertisseur@fr : entité@fr), (outil@fr : convertisseur@fr), (récipient@fr : outil@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
36. fleuve@fr <-> ruisseau@fr : 0.5931576490402222  
[[ruisseau@fr : cours d'eau@fr], (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
37. santé publique et assainissement@fr <-> assainissement@fr : 0.5848271250724792  
[[assainissement@fr : hygiène@fr], (hygiène@fr : soins de santé@fr), (soins de santé@fr : service@fr), (service@fr : produit@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
38. gaz@fr <-> gaz@fr : 0.5823714137077332  
[[gaz@fr : fluide@fr], (fluide@fr : substance@fr), (substance@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]

39. **risque naturel@fr <-> risque naturel@fr : 0.5795302391052246**  
*[(risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (produit@fr : entité artificielle@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
40. **catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5750667452812195**  
*[(catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
41. **label Gemet non trouvé <-> BD Carthage@fr : 0.5737375020980835**  
*Chemin non trouvé*
42. **biogaz@fr <-> biogaz@fr : 0.5729211568832397**  
*[(biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : combustible@fr), (combustible@fr : matière énergétique@fr), (matière énergétique@fr : matériau@fr), (matériau@fr : produit@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
43. **label Gemet non trouvé <-> Gestion des ressources naturelles@fr : 0.5694080591201782**  
*[(Gestion des ressources naturelles@fr : maltraitance@fr), (maltraitance@fr : problème sociétal@fr), (problème sociétal@fr : problème@fr), (problème@fr : condition@fr), (condition@fr : état@fr), (état@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
44. **hydrographie@fr <-> hydrographie@fr : 0.5633646845817566**  
*[(hydrographie@fr : sciences de la Terre@fr), (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
45. **lac@fr <-> ensemble de lacs@fr : 0.563302755355835**  
*[(ensemble de lacs@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
46. **label Gemet non trouvé <-> cordon littoral@fr : 0.5623621344566345**  
*Chemin non trouvé*
47. **biotope@fr <-> biotope@fr : 0.5620291829109192**  
*[(biotope@fr : entité géographique naturelle@fr), (entité géographique naturelle@fr : entité géographique@fr), (entité géographique@fr : entité@fr), (convertisseur@fr : entité@fr), (outil@fr : convertisseur@fr), (récipient@fr : outil@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
48. **label Gemet non trouvé <-> plan d'eau@fr : 0.5611335635185242**  
*[(plan d'eau@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
49. **gestion des déchets@fr <-> gestion des déchets@fr : 0.5611010193824768**

[(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]

50. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.5600736737251282

[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure architecturale@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]

## Recherche des concepts suffisamment proches de : "Phytoépuration" (Q1376406)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté ?**  
**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**  
**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

- **i. Label GEMET <-> Label Wikidata : score de similarité [chemin entre le concept utilisateur et le concept métadonnées]**

1. pluie@fr <-> pluie@fr : 0.732921838760376

[(pluie@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

2. eau usée@fr <-> eaux usées@fr : 0.6487299203872681

[(eaux usées@fr : eau non potable@fr), (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

3. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.5630149841308594

[(qualité de l'eau@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

4. label Gemet non trouvé <-> gestion de l'eau@fr : 0.6390826106071472

[(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : aptitude@fr), (aptitude@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

5. **pollution industrielle@fr <-> pollution@fr : 0.6167508959770203**  
*[(pollution@fr : effet pervers@fr), (effet pervers@fr : résultat@fr), (résultat@fr : entité observable@fr), (entité observable@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
6. **santé publique et assainissement@fr <-> assainissement@fr : 0.6126349568367004**  
*[(assainissement@fr : facteur écologique@fr), (facteur écologique@fr : corrélation@fr), (corrélation@fr : relation@fr), (représentation@fr : relation@fr), (représentation@fr : entité artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
7. **label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.6933472156524658**  
*[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure construite@fr), (structure construite@fr : objet géographique@fr), (objet géographique naturel@fr : objet géographique@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]*
  
8. **label Gemet non trouvé <-> prélèvement d'eau@fr : 0.66660475730896**  
*[(prélèvement d'eau@fr : infrastructure@fr), (infrastructure@fr : œuvre@fr), (œuvre@fr : produit@fr), (produit@fr : entité artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
9. **urbanisme@fr <-> aménagement urbain@fr : 0.5961806774139404**  
*[(aménagement urbain@fr : urbanisme@fr), (urbanisme@fr : géographie humaine@fr), (géographie humaine@fr : géographie@fr), (géographie@fr : objet de la recherche@fr), (objet de la recherche@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
10. **gestion des déchets@fr <-> gestion des déchets@fr : 0.567966639995575**  
*[(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
11. **environnement@fr <-> environnement@fr : 0.5334920883178711**  
*Chemin non trouvé*
  
12. **polluant@fr <-> polluant@fr : 0.531983494758606**  
*[(polluant@fr : substance physique@fr), (substance physique@fr : entité physique@fr), (objet tangible@fr : entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
13. **label Gemet non trouvé <-> bassin versant@fr : 0.6982258558273315**  
*[(bassin versant@fr : zone géographique@fr), (étendue d'eau@fr : zone géographique@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]*
  
14. **label Gemet non trouvé <-> submersion marine@fr : 0.6883678436279297**  
*[(submersion marine@fr : inondation@fr), (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
15. **label Gemet non trouvé <-> masse d'eau@fr : 0.6847524642944336**
  
16. **estuaire@fr <-> estuaire@fr : 0.6564208269119263**

*[[estuaire@fr : écosystème@fr], (zone humide@fr : écosystème@fr), (Phytoépuration@fr : zone humide@fr)]*

17. label Gemet non trouvé <-> nappe phréatique@fr : 0.6477580070495605  
*[[nappe phréatique@fr : aquifère@fr], (aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : eau calme@fr), (eau calme@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]*
  
18. hydrologie@fr <-> hydrologie@fr : 0.6444834470748901  
*[[hydrologie@fr : science@fr], (science@fr : système de connaissances@fr), (système de connaissances@fr : système@fr), (Q106865236 : système@fr), (Q106865236 : objet naturel@fr), (objet physique naturel@fr : objet naturel@fr), (objet géographique naturel@fr : objet physique naturel@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]*
  
19. aquifère@fr <-> aquifère@fr : 0.6373130679130554  
*[[aquifère@fr : réservoir d'eau@fr], (réservoir d'eau@fr : eau calme@fr), (eau calme@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]*
  
20. protection contre les crues@fr <-> inondation@fr : 0.6344674229621887  
*[[inondation@fr : catastrophe naturelle@fr], (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
21. nappe phréatique@fr <-> niveau piézométrique@fr : 0.6110435724258423  
*[[niveau piézométrique@fr : surface@fr], (surface@fr : espace à deux dimensions@fr), (espace à deux dimensions@fr : espace mathématique@fr), (espace mathématique@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
22. label Gemet non trouvé <-> saliculture@fr : 0.6026057004928589  
*[[saliculture@fr : production@fr], (production@fr : processus@fr), (processus@fr : changement@fr), (changement@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité physique@fr : entité spatio-temporelle@fr), (objet tangible@fr : entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
23. biogaz@fr <-> biogaz@fr : 0.5992053747177124  
*[[biogaz@fr : combustible gazeux@fr], (combustible gazeux@fr : combustible@fr), (combustible@fr : matière énergétique@fr), (matière énergétique@fr : matériau@fr), (matériau@fr : produit@fr), (produit@fr : entité artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
24. eaux souterraines@fr <-> eau souterraine@fr : 0.5980356931686401  
*[[eau souterraine@fr : eaux intérieures@fr], (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]*
  
25. label Gemet non trouvé <-> crue@fr : 0.5977764129638672  
*[[crue@fr : catastrophe naturelle@fr], (catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*

26. ferme (exploitation)@fr <-> ferme@fr : 0.5928126573562622  
[[ferme@fr : établissement rural@fr], (établissement rural@fr : localité@fr), (localité@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
27. label Gemet non trouvé <-> conchyliculture@fr : 0.5910194516181946  
[[conchyliculture@fr : pêche@fr], (pêche@fr : chasse@fr), (chasse@fr : action@fr), (action@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité physique@fr : entité spatio-temporelle@fr), (objet tangible@fr : entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
28. hydrogéologie@fr <-> hydrogéologie@fr : 0.5909260511398315  
[[hydrogéologie@fr : hydrologie@fr], (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
29. label Gemet non trouvé <-> Plan de prévention du risque inondation@fr :  
0.5880736708641052  
Chemin non trouvé
30. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5866144895553589  
[[catastrophe naturelle@fr : phénomène naturel@fr], (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
31. développement urbain@fr <-> urbanisme@fr : 0.5849825143814087  
[[urbanisme@fr : géographie humaine@fr], (géographie humaine@fr : géographie@fr), (géographie@fr : objet de la recherche@fr), (objet de la recherche@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
32. littoral@fr <-> littoral@fr : 0.5812649726867676  
[[littoral@fr : rivage@fr], (rivage@fr : objet physico-géographique@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
33. hydrographie@fr <-> hydrographie@fr : 0.579881489276886  
[[hydrographie@fr : sciences de la Terre@fr], (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
34. risque naturel@fr <-> risque naturel@fr : 0.5705837607383728  
[[risque naturel@fr : danger@fr], (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
35. fleuve@fr <-> cours d'eau naturel@fr : 0.5690486431121826  
[[cours d'eau naturel@fr : objet géographique naturel@fr], (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
36. label Gemet non trouvé <-> label Wikidata introuvable : 0.5682553648948669  
Chemin non trouvé

37. **ressource naturelle@fr** <-> **ressource naturelle@fr** : 0.5668643712997437  
[[ressource naturelle@fr : ressource@fr], (ressource@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
38. **label Gemet non trouvé** <-> **masse d'eau de surface@fr** : 0.5613843202590942  
[[masse d'eau de surface@fr : étendue d'eau@fr], (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
39. **fleuve@fr** <-> **cours d'eau@fr** : 0.5561116933822632  
[[cours d'eau@fr : eaux intérieures@fr], (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
40. **fleuve@fr** <-> **courant d'eau@fr** : 0.5535402894020081  
[[courant d'eau@fr : cours d'eau naturel@fr], (cours d'eau naturel@fr : objet géographique naturel@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
41. **port@fr** <-> **havre@fr** : 0.5492718815803528  
[[havre@fr : baie@fr], (baie@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
42. **érosion@fr** <-> **érosion@fr** : 0.5444740056991577  
[[érosion@fr : phénomène naturel@fr], (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
43. **agriculture@fr** <-> **agriculture@fr** : 0.5412713289260864  
[[agriculture@fr : agriculture et sylviculture@fr], (agriculture et sylviculture@fr : économie@fr), (économie@fr : système social@fr), (système social@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
44. **lac@fr** <-> **lac@fr** : 0.5294979214668274  
[[lac@fr : modelé@fr], (forme de relief artificielle@fr : modelé@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
45. **entité hydrologique@fr** <-> **étendue d'eau@fr** : 0.5264604687690735  
[[zone humide@fr : étendue d'eau@fr], (Phytoépuration@fr : zone humide@fr)]
46. **gestion@fr** <-> **label Wikidata introuvable** : 0.5256054997444153  
Chemin non trouvé
47. **diversité biologique@fr** <-> **biodiversité@fr** : 0.5250712633132935  
[[biodiversité@fr : diversité@fr], (diversité@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
48. **label Gemet non trouvé** <-> **La Réunion@fr** : 0.5243727564811707  
Chemin non trouvé
49. **label Gemet non trouvé** <-> **bassin hydrographique@fr** : 0.5230716466903687  
Chemin non trouvé
50. **label Gemet non trouvé** <-> **réseau de transport@fr** : 0.522017776966095

*[(réseau de transport@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*

## Recherche des concepts suffisamment proches de : “Plan de prévention du risque inondation” (Q3391269)

***Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté?  
A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?***

*(Note Mehdi : il n'y a pas de chemins entre PPRI et les autres concepts. PPRI est isolé dans le graphe de Wikidata)*

### Liste des comparaisons : (meilleurs 50 résultats)

1. label Gemet non trouvé <-> Plan de prévention du risque inondation@fr : 1.0
2. label Gemet non trouvé <-> crue@fr : 0.5518550872802734
3. label Gemet non trouvé <-> submersion marine@fr : 0.699097752571106
4. label Gemet non trouvé <-> bassin versant@fr : 0.606191873550415
5. label Gemet non trouvé <-> prévention@fr : 0.567156195640564
6. gestion de crise@fr <-> gestion de crise@fr : 0.537813663482666
7. risque naturel@fr <-> risque naturel@fr : 0.5369033813476562
8. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5143988132476807
  
9. pluie@fr <-> pluie@fr : 0.5633394718170166
10. label Gemet non trouvé <-> protection@fr : 0.65545654296875
11. communication en matière de risques@fr <-> gestion du risque@fr : 0.6078090071678162
12. fleuve@fr <-> cours d'eau@fr : 0.5825957655906677
13. hydrographie@fr <-> hydrographie@fr : 0.5800302624702454
14. politique@fr <-> politique@fr : 0.5799053907394409
15. politique de l'environnement@fr <-> politique environnementale@fr : 0.5440297722816467
  
16. label Gemet non trouvé <-> masse d'eau@fr : 0.6312357187271118
17. estuaire@fr <-> estuaire@fr : 0.6220836639404297
18. aquifère@fr <-> aquifère@fr : 0.6191027164459229
19. label Gemet non trouvé <-> directive-cadre sur l'eau@fr : 0.6143583655357361
20. label Gemet non trouvé <-> conchyliculture@fr : 0.6103661060333252
21. port@fr <-> havre@fr : 0.6073088049888611
22. label Gemet non trouvé <-> label Wikidata introuvable : 0.5888326168060303
23. plan@fr <-> plan@fr : 0.5848706364631653
24. hydrogéologie@fr <-> hydrogéologie@fr : 0.5739492177963257
25. label Gemet non trouvé <-> bassin hydrographique@fr : 0.5727195143699646

26. label Gemet non trouvé <-> Auvergne-Rhône-Alpes@fr : 0.5619726181030273
27. ferme (exploitation)@fr <-> ferme@fr : 0.5597040057182312
28. label Gemet non trouvé <-> nappe phréatique@fr : 0.5578835010528564
29. label Gemet non trouvé <-> Guadeloupe@fr : 0.5573101043701172
30. label Gemet non trouvé <-> cordon littoral@fr : 0.5549724102020264
31. label Gemet non trouvé <-> Guyane@fr : 0.554639995098114
32. label Gemet non trouvé <-> Politique environnementale de l'Union européenne@fr : 0.5531107187271118
33. planification@fr <-> planification@fr : 0.5512016415596008
34. label Gemet non trouvé <-> saliculture@fr : 0.5446555614471436
35. fleuve@fr <-> courant d'eau@fr : 0.5439935922622681
36. ressource naturelle@fr <-> ressource naturelle@fr : 0.5360186100006104
37. littoral@fr <-> littoral@fr : 0.5313910245895386
38. réserve naturelle nationale@fr <-> aire protégée@fr : 0.530930757522583
39. label Gemet non trouvé <-> source d'énergie@fr : 0.5305038690567017
40. surveillance@fr <-> surveillance@fr : 0.5277665853500366
41. label Gemet non trouvé <-> Bretagne@fr : 0.5256977081298828
42. label Gemet non trouvé <-> gestion de l'eau@fr : 0.5250158309936523
43. environnement@fr <-> environnement@fr : 0.5216457843780518
44. fleuve@fr <-> rivière@fr : 0.5166248083114624
45. label Gemet non trouvé <-> Mayotte@fr : 0.5137329697608948
46. source d'énergie renouvelable@fr <-> énergie renouvelable@fr : 0.5136480331420898
47. risque@fr <-> risque@fr : 0.5124579071998596
48. eau usée@fr <-> eaux usées@fr : 0.5105905532836914
49. fleuve@fr <-> cours d'eau naturel@fr : 0.5103414058685303
50. label Gemet non trouvé <-> lycée@fr : 0.5079976320266724

## Recherche des concepts suffisamment proches de : “Station de traitement des eaux usées”

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté?**

**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**

**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

**Label GEMET <-> Label Wikidata : score de similarité**

**[chemin entre le concept utilisateur et le concept métadonnées]**

1. eau usée@fr <-> eaux usées@fr : 0.7805310487747192  
[(eaux usées@fr : eau non potable@fr), (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
2. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.7642497420310974  
[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
3. santé publique et assainissement@fr <-> assainissement@fr : 0.6888487935066223  
[(assainissement@fr : activité@fr), (activité@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (ensemble architectural@fr : groupe de constructions et de bâtiments@fr), (complexe immobilier@fr : ensemble architectural@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
4. label Gemet non trouvé <-> gestion de l'eau@fr : 0.7167595624923706  
[(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : aptitude@fr), (aptitude@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
5. gestion des déchets@fr <-> gestion des déchets@fr : 0.761856198310852  
[(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
6. label Gemet non trouvé <-> Gestion des ressources naturelles@fr : 0.653746485710144  
[(Gestion des ressources naturelles@fr : maltraitance@fr), (maltraitance@fr : problème sociétal@fr), (problème sociétal@fr : problème@fr), (problème@fr : condition@fr), (condition@fr : état@fr), (état@fr :

*propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*

7. bâtiment public@fr <-> structure construite@fr : 0.6913622617721558  
*[(équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
8. polluant@fr <-> polluant@fr : 0.6678998470306396  
*[(polluant@fr : substance physique@fr), (substance physique@fr : entité physique@fr), (objet tangible@fr : entité physique@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*
9. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.6363604664802551  
*[(qualité de l'eau@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
10. pollution industrielle@fr <-> pollution@fr : 0.6422240734100342  
*[(pollution@fr : effet pervers@fr), (effet pervers@fr : résultat@fr), (résultat@fr : entité observable@fr), (entité observable@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
11. label Gemet non trouvé <-> saliculture@fr : 0.6875752210617065  
*[(saliculture@fr : production@fr), (production@fr : processus@fr), (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (groupe de constructions et de bâtiments@fr : structure construite@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
12. label Gemet non trouvé <-> station d'observation@fr : 0.6836385726928711  
*[(station d'observation@fr : équipement@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
13. label Gemet non trouvé <-> source d'énergie@fr : 0.6789122819900513  
*[(source d'énergie@fr : source@fr), (source@fr : objet@fr), (objet@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
14. urbanisme@fr <-> aménagement urbain@fr : 0.6710588335990906  
*[(aménagement urbain@fr : urbanisme@fr), (urbanisme@fr : géographie humaine@fr), (géographie humaine@fr : société@fr), (société@fr : groupe d'humains@fr), (organisation@fr : groupe d'humains@fr), (firme@fr : organisation@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
15. captage@fr <-> zone de chalandise@fr : 0.664354681968689  
*[(zone de chalandise@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
16. label Gemet non trouvé <-> masse d'eau@fr : 0.6608922481536865  
*[(masse d'eau@fr : étendue d'eau@fr), (étendue d'eau@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*

17. hydrogéologie@fr <-> hydrogéologie@fr : 0.6581923961639404  
[[hydrogéologie@fr : hydrologie@fr], (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
18. hydrologie@fr <-> hydrologie@fr : 0.655835747718811  
[[hydrologie@fr : science@fr], (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : classe@fr), (forme juridique@fr : classe@fr), (personne morale@fr : forme juridique@fr), (firme@fr : personne morale@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
19. chaleur@fr <-> chaleur@fr : 0.6528462171554565  
[[chaleur@fr : grandeur physique@fr], (grandeur physique@fr : quantité@fr), (quantité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
20. développement urbain@fr <-> urbanisme@fr : 0.6490142345428467  
[[urbanisme@fr : géographie humaine@fr], (géographie humaine@fr : société@fr), (société@fr : groupe d'humains@fr), (organisation@fr : groupe d'humains@fr), (firme@fr : organisation@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
21. géologie@fr <-> géologie@fr : 0.6452451944351196  
[[géologie@fr : science de la nature@fr], (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : classe@fr), (forme juridique@fr : classe@fr), (personne morale@fr : forme juridique@fr), (firme@fr : personne morale@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
22. label Gemet non trouvé <-> infrastructure de transport@fr : 0.64194655418396  
[[infrastructure de transport@fr : objet géographique@fr], (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
23. label Gemet non trouvé <-> bassin hydrographique@fr : 0.6412675380706787  
Chemin non trouvé
24. label Gemet non trouvé <-> bien culturel@fr : 0.6321249604225159  
[[bien culturel@fr : objet concret@fr], (objet tangible@fr : objet concret@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
25. label Gemet non trouvé <-> label Wikidata introuvable : 0.6305616497993469  
Chemin non trouvé
26. label Gemet non trouvé <-> nappe phréatique@fr : 0.6300790905952454  
[[nappe phréatique@fr : aquifère@fr], (aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : réservoir@fr), (réservoir@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
27. hydrographie@fr <-> hydrographie@fr : 0.6294116973876953  
[[hydrographie@fr : sciences de la Terre@fr], (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (complexe immobilier@fr : structure

*architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*

28. **infrastructure@fr <-> infrastructure@fr : 0.628700852394104**  
*[(infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*
29. **label Gemet non trouvé <-> submersion marine@fr : 0.6275973320007324**  
*[(submersion marine@fr : inondation@fr), (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
30. **biogaz@fr <-> biogaz@fr : 0.6261656284332275**  
*[(biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : gaz@fr), (gaz@fr : fluide@fr), (fluide@fr : substance@fr), (substance@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
31. **logement@fr <-> gestion de logement@fr : 0.6251828670501709**  
Chemin non trouvé
32. **aquifère@fr <-> aquifère@fr : 0.6246013641357422**  
*[(aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : réservoir@fr), (réservoir@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*
33. **communication en matière de risques@fr <-> gestion du risque@fr : 0.6228855848312378**  
*[(gestion du risque@fr : fonction@fr), (fonction@fr : usage@fr), (usage@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*
34. **monument@fr <-> monument de Hongrie@fr : 0.6182126998901367**  
*[(monument de Hongrie@fr : bien patrimonial national@fr), (bien patrimonial national@fr : site patrimonial@fr), (site patrimonial@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
35. **label Gemet non trouvé <-> label Wikidata introuvable : 0.6179236173629761**  
Chemin non trouvé
36. **label Gemet non trouvé <-> bassin versant@fr : 0.6176934242248535**  
*[(bassin versant@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
37. **habitat@fr <-> habitat@fr : 0.6176526546478271**  
*[(habitat@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*
38. **ressource naturelle@fr <-> ressource naturelle@fr : 0.6150418519973755**  
*[(ressource naturelle@fr : matériau brut@fr), (matériau brut@fr : bien@fr), (produit@fr : bien@fr), (œuvre@fr : produit@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*
39. **risque naturel@fr <-> risque naturel@fr : 0.6141432523727417**

*[(risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (objet@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*

40. **érosion@fr <-> érosion@fr : 0.6131541728973389**

*[(érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*

41. **polluant@fr <-> label Wikidata introuvable : 0.6121931076049805**

*Chemin non trouvé*

42. **label Gemet non trouvé <-> réseau de télécommunications@fr : 0.6115759015083313**

*[(réseau de télécommunications@fr : infrastructure@fr), (infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*

43. **surveillance@fr <-> surveillance@fr : 0.6091445684432983**

*[(surveillance@fr : activité@fr), (Q18390984 : activité@fr), (Q18390984 : outil de gestion du risque@fr), (outil de gestion du risque@fr : outil de gestion@fr), (outil de gestion@fr : outil@fr), (outil@fr : équipement@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*

44. **estuaire@fr <-> estuaire@fr : 0.6066519021987915**

*[(estuaire@fr : étendue d'eau@fr), (étendue d'eau@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*

45. **dommage@fr <-> dégât@fr : 0.6065858602523804**

*[(dégât@fr : processus@fr), (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (ensemble architectural@fr : groupe de constructions et de bâtiments@fr), (complexe immobilier@fr : ensemble architectural@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*

46. **logement@fr <-> appartement@fr : 0.6065312623977661**

*[(appartement@fr : logement@fr), (logement@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*

47. **label Gemet non trouvé <-> transition écologique@fr : 0.6048821210861206**

*[(transition écologique@fr : changement@fr), (changement@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité spatio-temporelle@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*

48. **chauffage à distance@fr <-> réseau de chaleur@fr : 0.603854775428772**

*[(réseau de chaleur@fr : infrastructure@fr), (infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]*

49. **zonage@fr <-> zonage@fr : 0.6028785705566406**

*[(zonage@fr : planification@fr), (planification@fr : processus@fr), (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (groupe de constructions et de bâtiments@fr : structure construite@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]*

50. **label Gemet non trouvé <-> unité statistique@fr : 0.6022796630859375**

*[(unité statistique@fr : unité@fr), (unité@fr : objet abstrait@fr), (objet abstrait@fr : entité non-physique@fr), (entité non-physique@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]*

# Expérimentation :

## Fixer un seuil de similarité sémantique minimal entre deux concepts du domaine de l'eau

### Introduction

Dans le cadre de ma thèse, j'étudie comment interroger des catalogues de données géographiques en utilisant des concepts, dits "concepts utilisateur", relatifs à l'application - pour laquelle l'utilisateur a besoin de données géographiques-. Ces concepts utilisateurs sont potentiellement différents de ceux utilisés dans les métadonnées qui décrivent les jeux, et donc dans les catalogues actuels. Lorsque le concept est un nom de lieu, des services permettent de passer de la requête de l'utilisateur, e.g. "France", à une requête sur l'extension spatiale du jeu. Je m'intéresse aux concepts de la requête relatifs à une thématique. Dans les métadonnées, les concepts relatifs à la thématique correspondent aux valeurs des propriétés "keyword" et "topic category", je les appelle "concepts des métadonnées". Pour cela, je cherche à associer à un "concept utilisateur" les "concepts de métadonnées", c'est-à-dire les concepts utilisés pour renseigner ces propriétés, qui sont suffisamment proches pour qu'on considère qu'ils sont pertinents pour interroger des catalogues en vue de construire l'application. Le but de ce questionnaire est d'avoir votre retour d'experts sur la façon dont je mesure la proximité entre un concept utilisateur et les concepts métadonnées, dans le cas particulier du domaine de l'eau.

Une première expérience avec des experts a permis d'évaluer la pertinence de recourir à l'ontologie de sens commun Wikidata pour exprimer des concepts utilisateurs, correspondant à des concepts d'intérêt applicatifs, dans le domaine de l'eau. Pour cette expérience, nous avons choisi 4 concepts Wikidata, parmi ceux suggérés par les experts, à savoir : "nappe phréatique", "station d'épuration", "marais artificiel" et "PPRI".

Les concepts des métadonnées ont été récupérés de la façon suivante. Les métadonnées de catalogues du domaine, conformes au standard INSPIRE, ont été moissonnées. J'ai ensuite extrait les valeurs des champs "Topics Category" et "Keyword" de tous les enregistrements (un enregistrement = une description d'un jeu de données à l'aide de divers champs). Ces valeurs sont des chaînes de caractères. Elles ont été utilisées pour extraire des concepts du thésaurus GEMET (standard utilisé pour documenter Topic Category dans INSPIRE) ainsi qu'avec l'ontologie de sens commun Wikidata.. Au total, 205 concepts différents ont été identifiés. Pour les concepts GEMET, un alignement a été réalisé avec Wikidata, de sorte que tous les concepts, utilisateurs ou métadonnées, sont intégrés dans un graphe dans lequel nous pouvons utiliser une mesure de similarité sémantique.

Notre but est de fixer un seuil à partir duquel on peut affirmer que deux concepts sont suffisamment proches pour que le concept métadonnées soit utilisé pour reformuler la requête de l'utilisateur. Pour cela, nous avons comparé chaque concept utilisateur à tous les concepts métadonnées en utilisant la mesure de similarité sémantique et gardé les 50 meilleurs résultats. Par ailleurs, dans une majorité des cas, nous avons affiché également le chemin entre les concepts.

## Recherche des concepts suffisamment proches de : "Nappe phréatique" (Q583075)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté?**

**Plutôt pas**

**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**

**3 : tout est faux sauf. (nappe phréatique@fr : aquifère@fr). Globalement, la cohérence est erronée.**

**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

**Non**

**NB : label Gemet non trouvé : étonnant de ne pas avoir « bassin versant » dans GEMET, ou bassin hydro, masse d'eau... qui sont typiquement des sémantiques issues des textes européens :**

**<https://www.eionet.europa.eu/gemet/fr/concept/1215>**

Mis en forme : Couleur de police : Rouge foncé

Mis en forme : Couleur de police : Rouge foncé

Mis en forme : Couleur de police : Rouge foncé

Mis en forme : Couleur de police : Rouge foncé

Mis en forme : Couleur de police : Rouge foncé

### Liste des comparaisons : (meilleurs 50 résultats)

- i. Label GEMET <-> Label Wikidata : score de similarité**  
*[chemin entre le concept utilisateur et le concept métadonnées]*

- label Gemet non trouvé <-> nappe phréatique@fr : 1.0  
*[nappe phréatique@fr]*
- aquifère@fr <-> aquifère@fr : 0.9400091171264648  
*[[nappe phréatique@fr : aquifère@fr]]*
- eaux souterraines@fr <-> eau souterraine@fr : 0.8649603128433228  
*[[eau souterraine@fr : eaux intérieures@fr], (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
- label Gemet non trouvé <-> bassin versant@fr : 0.7749752998352051  
*[[bassin versant@fr : zone géographique@fr], (étendue d'eau@fr : zone géographique@fr), (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
- label Gemet non trouvé <-> bassin hydrographique@fr : 0.7529347538948059  
*Chemin non trouvé*
- fleuve@fr <-> courant d'eau@fr : 0.7424725294113159  
*[[courant d'eau@fr : cours d'eau naturel@fr], (cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
- label Gemet non trouvé <-> masse d'eau@fr : 0.7240350246429443

- [[masse d'eau@fr : étendue d'eau@fr], (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
8. fleuve@fr <-> cours d'eau@fr : 0.7219454646110535  
[[cours d'eau@fr : eaux intérieures@fr], (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
9. nappe phréatique@fr <-> niveau piézométrique@fr : 0.7217990159988403  
[[niveau piézométrique@fr : surface@fr], (surface@fr : espace à deux dimensions@fr), (espace à deux dimensions@fr : espace mathématique@fr), (espace mathématique@fr : ensemble@fr), (ensemble@fr : classe@fr), (classe@fr : entité composée@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
10. estuaire@fr <-> estuaire@fr : 0.7091968059539795  
[[estuaire@fr : étendue d'eau@fr], (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
11. hydrogéologie@fr <-> hydrogéologie@fr : 0.6929916143417358  
[[hydrogéologie@fr : hydrologie@fr], (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
12. label Gemet non trouvé <-> submersion marine@fr : 0.6904162764549255  
[[submersion marine@fr : inondation@fr], (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (objet artificiel physique@fr : entité artificielle@fr), (réceptif@fr : objet artificiel physique@fr), (réservoir@fr : réceptif@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
13. fleuve@fr <-> cours d'eau naturel@fr : 0.6870043277740479  
[[cours d'eau naturel@fr : cours d'eau@fr], (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
14. label Gemet non trouvé <-> conchyliculture@fr : 0.6838634610176086  
[[conchyliculture@fr : pêche@fr], (pêche@fr : chasse@fr), (chasse@fr : action@fr), (action@fr : convention@fr), (convention@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
15. entité hydrologique@fr <-> étendue d'eau@fr : 0.6789747476577759  
[[zone aquatique@fr : étendue d'eau@fr], (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
16. littoral@fr <-> littoral@fr : 0.6748140454292297  
[[littoral@fr : rivage@fr], (rivage@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
17. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.6704185009002686  
[[qualité de l'eau@fr : qualité@fr], (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage

- d'eau@fr : stockage@fr, (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
18. label Gemet non trouvé <-> gestion de l'eau@fr : 0.6675404906272888  
*[(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : relation asymétrique@fr), (relation asymétrique@fr : relation binaire@fr), (relation binaire@fr : relation@fr), (représentation@fr : relation@fr), (représentation@fr : entité artificielle@fr), (objet artificiel physique@fr : entité artificielle@fr), (récipient@fr : objet artificiel physique@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
19. pluie@fr <-> pluie@fr : 0.6633356809616089  
*[(pluie@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
20. ressource naturelle@fr <-> ressource naturelle@fr : 0.6476647853851318  
*[(ressource naturelle@fr : matériau brut@fr), (matériau brut@fr : bien@fr), (produit@fr : bien@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
21. hydrologie@fr <-> hydrologie@fr : 0.643667459487915  
*[(hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
22. label Gemet non trouvé <-> label Wikidata introuvable : 0.6301834583282471  
*Chemin non trouvé*
23. lac@fr <-> lac@fr : 0.6288014650344849  
*[(lac@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
24. fleuve@fr <-> rivière@fr : 0.6218131184577942  
*[(rivière@fr : cours d'eau naturel@fr), (cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
25. label Gemet non trouvé <-> source d'énergie@fr : 0.6207959651947021  
*[(source d'énergie@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (produit@fr : entité artificielle@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
26. ferme (exploitation)@fr <-> ferme@fr : 0.6167993545532227  
*[(ferme@fr : propriété agricole@fr), (propriété agricole@fr : complexe immobilier@fr), (complexe immobilier@fr : structure architecturale@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
27. eau usée@fr <-> eaux usées@fr : 0.6070104837417603  
*[(eaux usées@fr : eau non potable@fr), (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
28. port@fr <-> havre@fr : 0.6061320304870605

- [(havre@fr : baie@fr), (baie@fr : zone aquatique@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
29. label Gemet non trouvé <-> crue@fr : 0.6060789823532104  
*[(crue@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
30. label Gemet non trouvé <-> prélèvement d'eau@fr : 0.600567102432251  
*[(prélèvement d'eau@fr : infrastructure@fr), (infrastructure@fr : œuvre@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
31. label Gemet non trouvé <-> saliculture@fr : 0.59828782081604  
*[(saliculture@fr : production@fr), (production@fr : activité économique@fr), (activité économique@fr : action@fr), (action@fr : convention@fr), (convention@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
32. label Gemet non trouvé <-> unité administrative locale@fr : 0.5975565910339355  
*[(unité administrative locale@fr : division territoriale d'un niveau d'imbrication donné@fr), (division territoriale d'un niveau d'imbrication donné@fr : entité territoriale administrative@fr), (entité territoriale administrative@fr : entité territoriale géographique humaine@fr), (entité territoriale géographique humaine@fr : entité géographique artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (objet artificiel physique@fr : entité artificielle@fr), (récipient@fr : objet artificiel physique@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
33. érosion@fr <-> érosion@fr : 0.5942178964614868  
*[(érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (convertisseur@fr : entité@fr), (outil@fr : convertisseur@fr), (récipient@fr : outil@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
34. fleuve@fr <-> ruisseau@fr : 0.5931576490402222  
*[(ruisseau@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
35. santé publique et assainissement@fr <-> assainissement@fr : 0.5848271250724792  
*[(assainissement@fr : hygiène@fr), (hygiène@fr : soins de santé@fr), (soins de santé@fr : service@fr), (service@fr : produit@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
36. gaz@fr <-> gaz@fr : 0.5823714137077332  
*[(gaz@fr : fluide@fr), (fluide@fr : substance@fr), (substance@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
37. risque naturel@fr <-> risque naturel@fr : 0.5795302391052246  
*[(risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (produit@fr : entité artificielle@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
38. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5750667452812195

*[(catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

39. label Gemet non trouvé <-> BD Carthage@fr : 0.5737375020980835

*Chemin non trouvé*

40. biogaz@fr <-> biogaz@fr : 0.5729211568832397

*[(biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : combustible@fr), (combustible@fr : matière énergétique@fr), (matière énergétique@fr : matériau@fr), (matériau@fr : produit@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

41. eau potable@fr <-> eau potable@fr : 0.5725058913230896

*[(eau potable@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

42. label Gemet non trouvé <-> Gestion des ressources naturelles@fr :

0.5694080591201782

*[(Gestion des ressources naturelles@fr : maltraitance@fr), (maltraitance@fr : problème sociétal@fr), (problème sociétal@fr : problème@fr), (problème@fr : condition@fr), (condition@fr : état@fr), (état@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

43. label Gemet non trouvé <-> perméabilité@fr : 0.5676491856575012

*[(perméabilité@fr : propriété physique@fr), (propriété physique@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

44. hydrographie@fr <-> hydrographie@fr : 0.5633646845817566

*[(hydrographie@fr : sciences de la Terre@fr), (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

45. lac@fr <-> ensemble de lacs@fr : 0.563302755355835

*[(ensemble de lacs@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

46. label Gemet non trouvé <-> cordon littoral@fr : 0.5623621344566345

*Chemin non trouvé*

47. biotope@fr <-> biotope@fr : 0.5620291829109192

*[(biotope@fr : entité géographique naturelle@fr), (entité géographique naturelle@fr : entité géographique@fr), (entité géographique@fr : entité@fr), (convertisseur@fr : entité@fr), (outil@fr : convertisseur@fr), (récipient@fr : outil@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

48. label Gemet non trouvé <-> plan d'eau@fr : 0.5611335635185242

*[(plan d'eau@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

49. gestion des déchets@fr <-> gestion des déchets@fr : 0.5611010193824768

[(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]

50. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.5600736737251282  
[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure architecturale@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]

## Recherche des concepts suffisamment proches de : "Phytoépuration" (Q1376406)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté ?  
A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**

**5 : il n'y a pas de lien entre masse d'eau et étendue d'eau.**

**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

**Non**

Mis en forme : Couleur de police : Rouge foncé

Mis en forme : Couleur de police : Rouge foncé

Mis en forme : Couleur de police : Rouge foncé

### Liste des comparaisons : (meilleurs 50 résultats)

- i. Label GEMET <-> Label Wikidata : score de similarité [chemin entre le concept utilisateur et le concept métadonnées ]**

- pluie@fr <-> pluie@fr : 0.732921838760376  
[(pluie@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
- label Gemet non trouvé <-> bassin versant@fr : 0.6982258558273315  
[(bassin versant@fr : zone géographique@fr), (étendue d'eau@fr : zone géographique@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
- label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.6933472156524658  
[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure construite@fr), (structure construite@fr : objet géographique@fr), (objet géographique naturel@fr : objet géographique@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
- label Gemet non trouvé <-> submersion marine@fr : 0.6883678436279297  
[(submersion marine@fr : inondation@fr), (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

5. label Gemet non trouvé <-> masse d'eau@fr : 0.6847524642944336  
*[(masse d'eau@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]*
  
6. label Gemet non trouvé <-> prélèvement d'eau@fr : 0.66660475730896  
*[(prélèvement d'eau@fr : infrastructure@fr), (infrastructure@fr : œuvre@fr), (œuvre@fr : produit@fr), (produit@fr : entité artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
7. estuaire@fr <-> estuaire@fr : 0.6564208269119263  
*[(estuaire@fr : écosystème@fr), (zone humide@fr : écosystème@fr), (Phytoépuration@fr : zone humide@fr)]*
  
8. eau usée@fr <-> eaux usées@fr : 0.6487299203872681  
*[(eaux usées@fr : eau non potable@fr), (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
9. label Gemet non trouvé <-> nappe phréatique@fr : 0.6477580070495605  
*[(nappe phréatique@fr : aquifère@fr), (aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : eau calme@fr), (eau calme@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]*
  
10. hydrologie@fr <-> hydrologie@fr : 0.6444834470748901  
*[(hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : système@fr), (Q106865236 : système@fr), (Q106865236 : objet naturel@fr), (objet physique naturel@fr : objet naturel@fr), (objet géographique naturel@fr : objet physique naturel@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]*
  
11. label Gemet non trouvé <-> gestion de l'eau@fr : 0.6390826106071472  
*[(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : aptitude@fr), (aptitude@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
12. aquifère@fr <-> aquifère@fr : 0.6373130679130554  
*[(aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : eau calme@fr), (eau calme@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]*
  
13. protection contre les crues@fr <-> inondation@fr : 0.6344674229621887  
*[(inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
14. pollution industrielle@fr <-> pollution@fr : 0.6167508959770203  
*[(pollution@fr : effet pervers@fr), (effet pervers@fr : résultat@fr), (résultat@fr : entité observable@fr), (entité observable@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
  
15. santé publique et assainissement@fr <-> assainissement@fr : 0.6126349568367004  
*[(assainissement@fr : facteur écologique@fr), (facteur écologique@fr : corrélation@fr), (corrélation@fr : relation@fr), (représentation@fr : relation@fr), (représentation@fr : entité artificielle@fr), (entité*

*géographique artificielle@fr : entité artificielle@fr, (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*

16. **nappe phréatique@fr <-> niveau piézométrique@fr : 0.6110435724258423**  
[[niveau piézométrique@fr : surface@fr], (surface@fr : espace à deux dimensions@fr), (espace à deux dimensions@fr : espace mathématique@fr), (espace mathématique@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
17. **label Gemet non trouvé <-> saliculture@fr : 0.6026057004928589**  
[[saliculture@fr : production@fr], (production@fr : processus@fr), (processus@fr : changement@fr), (changement@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité physique@fr : entité spatio-temporelle@fr), (objet tangible@fr : entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
18. **biogaz@fr <-> biogaz@fr : 0.5992053747177124**  
[[biogaz@fr : combustible gazeux@fr], (combustible gazeux@fr : combustible@fr), (combustible@fr : matière énergétique@fr), (matière énergétique@fr : matériau@fr), (matériau@fr : produit@fr), (produit@fr : entité artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
19. **eaux souterraines@fr <-> eau souterraine@fr : 0.5980356931686401**  
[[eau souterraine@fr : eaux intérieures@fr], (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
20. **label Gemet non trouvé <-> crue@fr : 0.5977764129638672**  
[[crue@fr : catastrophe naturelle@fr], (catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
21. **urbanisme@fr <-> aménagement urbain@fr : 0.5961806774139404**  
[[aménagement urbain@fr : urbanisme@fr], (urbanisme@fr : géographie humaine@fr), (géographie humaine@fr : géographie@fr), (géographie@fr : objet de la recherche@fr), (objet de la recherche@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
22. **ferme (exploitation)@fr <-> ferme@fr : 0.5928126573562622**  
[[ferme@fr : établissement rural@fr], (établissement rural@fr : localité@fr), (localité@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
23. **label Gemet non trouvé <-> conchyliculture@fr : 0.5910194516181946**  
[[conchyliculture@fr : pêche@fr], (pêche@fr : chasse@fr), (chasse@fr : action@fr), (action@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité physique@fr : entité spatio-temporelle@fr), (objet tangible@fr : entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
24. **hydrogéologie@fr <-> hydrogéologie@fr : 0.5909260511398315**  
[[hydrogéologie@fr : hydrologie@fr], (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

25. label Gemet non trouvé <-> Plan de prévention du risque inondation@fr :  
0.5880736708641052  
*Chemin non trouvé*
26. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5866144895553589  
[[catastrophe naturelle@fr : phénomène naturel@fr], (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
27. développement urbain@fr <-> urbanisme@fr : 0.5849825143814087  
[[urbanisme@fr : géographie humaine@fr], (géographie humaine@fr : géographie@fr), (géographie@fr : objet de la recherche@fr), (objet de la recherche@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
28. littoral@fr <-> littoral@fr : 0.5812649726867676  
[[littoral@fr : rivage@fr], (rivage@fr : objet physico-géographique@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
29. hydrographie@fr <-> hydrographie@fr : 0.579881489276886  
[[hydrographie@fr : sciences de la Terre@fr], (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
30. risque naturel@fr <-> risque naturel@fr : 0.5705837607383728  
[[risque naturel@fr : danger@fr], (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
31. fleuve@fr <-> cours d'eau naturel@fr : 0.5690486431121826  
[[cours d'eau naturel@fr : objet géographique naturel@fr], (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
32. label Gemet non trouvé <-> label Wikidata introuvable : 0.5682553648948669  
*Chemin non trouvé*
33. gestion des déchets@fr <-> gestion des déchets@fr : 0.567966639995575  
[[gestion des déchets@fr : Durabilité et gestion de l'environnement@fr], (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
34. ressource naturelle@fr <-> ressource naturelle@fr : 0.5668643712997437  
[[ressource naturelle@fr : ressource@fr], (ressource@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
35. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.5630149841308594  
[[qualité de l'eau@fr : qualité@fr], (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

36. label Gemet non trouvé <-> masse d'eau de surface@fr : 0.5613843202590942  
 [(masse d'eau de surface@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr),  
 (Phytoépuration@fr : zone humide@fr)]
37. fleuve@fr <-> cours d'eau@fr : 0.5561116933822632  
 [(cours d'eau@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr :  
 étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
38. fleuve@fr <-> courant d'eau@fr : 0.5535402894020081  
 [(courant d'eau@fr : cours d'eau naturel@fr), (cours d'eau naturel@fr : objet géographique naturel@fr),  
 (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
39. port@fr <-> havre@fr : 0.5492718815803528  
 [(havre@fr : baie@fr), (baie@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone  
 humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
40. érosion@fr <-> érosion@fr : 0.5444740056991577  
 [(érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr :  
 entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité  
 géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr :  
 forme de relief artificielle@fr)]
41. agriculture@fr <-> agriculture@fr : 0.5412713289260864  
 [(agriculture@fr : agriculture et sylviculture@fr), (agriculture et sylviculture@fr : économie@fr),  
 (économie@fr : système social@fr), (système social@fr : système@fr), (système@fr : entité  
 composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité  
 géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique  
 artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
42. environnement@fr <-> environnement@fr : 0.5334920883178711  
 Chemin non trouvé
43. polluant@fr <-> polluant@fr : 0.531983494758606  
 [(polluant@fr : substance physique@fr), (substance physique@fr : entité physique@fr), (objet tangible@fr :  
 entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr :  
 objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
44. lac@fr <-> lac@fr : 0.5294979214668274  
 [(lac@fr : modelé@fr), (forme de relief artificielle@fr : modelé@fr), (Phytoépuration@fr : forme de relief  
 artificielle@fr)]
45. entité hydrologique@fr <-> étendue d'eau@fr : 0.5264604687690735  
 [(zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
46. gestion@fr <-> label Wikidata introuvable : 0.5256054997444153  
 Chemin non trouvé
47. diversité biologique@fr <-> biodiversité@fr : 0.5250712633132935  
 [(biodiversité@fr : diversité@fr), (diversité@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr :  
 entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité  
 géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr :  
 forme de relief artificielle@fr)]
48. label Gemet non trouvé <-> La Réunion@fr : 0.5243727564811707  
 Chemin non trouvé
49. label Gemet non trouvé <-> bassin hydrographique@fr : 0.5230716466903687  
 Chemin non trouvé

50. label Gemet non trouvé <-> réseau de transport@fr : 0.522017776966095  
[(réseau de transport@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

## Recherche des concepts suffisamment proches de : "Plan de prévention du risque inondation" (Q3391269)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté ?  
A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**

**Impossible de répondre en raison de la mention « label Gemet non trouvé ». Les PPRI étant un concept purement français, il n'y a évidemment pas de relation avec GEMET**

Mis en forme : Couleur de police : Rouge foncé

*(Note Mehdi : il n'y a pas de chemins entre PPRI et les autres concepts. PPRI est isolé dans le graphe de Wikidata)*

### Liste des comparaisons : (meilleurs 50 résultats)

1. label Gemet non trouvé <-> Plan de prévention du risque inondation@fr : 1.0
2. label Gemet non trouvé <-> submersion marine@fr : 0.699097752571106
3. label Gemet non trouvé <-> protection@fr : 0.65545654296875
4. label Gemet non trouvé <-> masse d'eau@fr : 0.6312357187271118
5. estuaire@fr <-> estuaire@fr : 0.6220836639404297
6. aquifère@fr <-> aquifère@fr : 0.6191027164459229
7. label Gemet non trouvé <-> directive-cadre sur l'eau@fr : 0.6143583655357361
8. label Gemet non trouvé <-> conchyliculture@fr : 0.6103661060333252
9. communication en matière de risques@fr <-> gestion du risque@fr : 0.6078090071678162
10. port@fr <-> havre@fr : 0.6073088049888611
11. label Gemet non trouvé <-> bassin versant@fr : 0.606191873550415
12. label Gemet non trouvé <-> label Wikidata introuvable : 0.5888326168060303
13. plan@fr <-> plan@fr : 0.5848706364631653
14. fleuve@fr <-> cours d'eau@fr : 0.5825957655906677
15. hydrographie@fr <-> hydrographie@fr : 0.5800302624702454
16. politique@fr <-> politique@fr : 0.5799053907394409
17. hydrogéologie@fr <-> hydrogéologie@fr : 0.5739492177963257
18. label Gemet non trouvé <-> bassin hydrographique@fr : 0.5727195143699646
19. label Gemet non trouvé <-> prévention@fr : 0.567156195640564
20. pluie@fr <-> pluie@fr : 0.5633394718170166
21. label Gemet non trouvé <-> Auvergne-Rhône-Alpes@fr : 0.5619726181030273
22. ferme (exploitation)@fr <-> ferme@fr : 0.5597040057182312
23. label Gemet non trouvé <-> nappe phréatique@fr : 0.5578835010528564

24. label Gemet non trouvé <-> Guadeloupe@fr : 0.5573101043701172
25. label Gemet non trouvé <-> cordon littoral@fr : 0.5549724102020264
26. label Gemet non trouvé <-> Guyane@fr : 0.554639995098114
27. label Gemet non trouvé <-> Politique environnementale de l'Union européenne@fr :  
0.5531107187271118
28. label Gemet non trouvé <-> crue@fr : 0.5518550872802734
29. planification@fr <-> planification@fr : 0.5512016415596008
30. label Gemet non trouvé <-> saliculture@fr : 0.5446555614471436
31. politique de l'environnement@fr <-> politique environnementale@fr : 0.5440297722816467
32. fleuve@fr <-> courant d'eau@fr : 0.5439935922622681
33. gestion de crise@fr <-> gestion de crise@fr : 0.537813663482666
34. risque naturel@fr <-> risque naturel@fr : 0.5369033813476562
35. ressource naturelle@fr <-> ressource naturelle@fr : 0.5360186100006104
36. littoral@fr <-> littoral@fr : 0.5313910245895386
37. réserve naturelle nationale@fr <-> aire protégée@fr : 0.530930757522583
38. label Gemet non trouvé <-> source d'énergie@fr : 0.5305038690567017
39. surveillance@fr <-> surveillance@fr : 0.5277665853500366
40. label Gemet non trouvé <-> Bretagne@fr : 0.5256977081298828
41. label Gemet non trouvé <-> gestion de l'eau@fr : 0.5250158309936523
42. environnement@fr <-> environnement@fr : 0.5216457843780518
43. fleuve@fr <-> rivière@fr : 0.5166248083114624
44. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5143988132476807
45. label Gemet non trouvé <-> Mayotte@fr : 0.5137329697608948
46. source d'énergie renouvelable@fr <-> énergie renouvelable@fr : 0.5136480331420898
47. risque@fr <-> risque@fr : 0.5124579071998596
48. eau usée@fr <-> eaux usées@fr : 0.5105905532836914
49. fleuve@fr <-> cours d'eau naturel@fr : 0.5103414058685303
50. label Gemet non trouvé <-> lycée@fr : 0.5079976320266724

## Recherche des concepts suffisamment proches de : "Station de traitement des eaux usées"

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté?**  
**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**  
**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

**Label GEMET <-> Label Wikidata : score de similarité**

**[chemin entre le concept utilisateur et le concept métadonnées]**

1. eau usée@fr <-> eaux usées@fr : 0.7805310487747192  
[(eaux usées@fr : eau non potable@fr), (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
2. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.7642497420310974  
[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
3. gestion des déchets@fr <-> gestion des déchets@fr : 0.761856198310852  
[(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
4. label Gemet non trouvé <-> gestion de l'eau@fr : 0.7167595624923706  
[(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : aptitude@fr), (aptitude@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
5. bâtiment public@fr <-> structure construite@fr : 0.6913622617721558  
[(équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
6. santé publique et assainissement@fr <-> assainissement@fr : 0.6888487935066223  
[(assainissement@fr : activité@fr), (activité@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (ensemble architectural@fr : groupe de constructions et de bâtiments@fr), (complexe immobilier@fr : ensemble architectural@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
7. label Gemet non trouvé <-> saliculture@fr : 0.6875752210617065  
[(saliculture@fr : production@fr), (production@fr : processus@fr), (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de

bâtiments@fr : groupe d'entités géographiques@fr, (groupe de constructions et de bâtiments@fr : structure construite@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]

8. label Gemet non trouvé <-> station d'observation@fr : 0.6836385726928711  
[(station d'observation@fr : équipement@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
9. label Gemet non trouvé <-> source d'énergie@fr : 0.6789122819900513  
[(source d'énergie@fr : source@fr), (source@fr : objet@fr), (objet@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
10. urbanisme@fr <-> aménagement urbain@fr : 0.6710588335990906  
[(aménagement urbain@fr : urbanisme@fr), (urbanisme@fr : géographie humaine@fr), (géographie humaine@fr : société@fr), (société@fr : groupe d'humains@fr), (organisation@fr : groupe d'humains@fr), (firme@fr : organisation@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
11. polluant@fr <-> polluant@fr : 0.6678998470306396  
[(polluant@fr : substance physique@fr), (substance physique@fr : entité physique@fr), (objet tangible@fr : entité physique@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
12. captage@fr <-> zone de chalandise@fr : 0.664354681968689  
[(zone de chalandise@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
13. label Gemet non trouvé <-> masse d'eau@fr : 0.6608922481536865  
[(masse d'eau@fr : étendue d'eau@fr), (étendue d'eau@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
14. hydrogéologie@fr <-> hydrogéologie@fr : 0.6581923961639404  
[(hydrogéologie@fr : hydrologie@fr), (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
15. hydrologie@fr <-> hydrologie@fr : 0.655835747718811  
[(hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : classe@fr), (forme juridique@fr : classe@fr), (personne morale@fr : forme juridique@fr), (firme@fr : personne morale@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
16. label Gemet non trouvé <-> Gestion des ressources naturelles@fr : 0.653746485710144  
[(Gestion des ressources naturelles@fr : maltraitance@fr), (maltraitance@fr : problème sociétal@fr), (problème sociétal@fr : problème@fr), (problème@fr : condition@fr), (condition@fr : état@fr), (état@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
17. chaleur@fr <-> chaleur@fr : 0.6528462171554565  
[(chaleur@fr : grandeur physique@fr), (grandeur physique@fr : quantité@fr), (quantité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]

18. développement urbain@fr <-> urbanisme@fr : 0.6490142345428467  
[[urbanisme@fr : géographie humaine@fr], (géographie humaine@fr : société@fr), (société@fr : groupe d'humains@fr), (organisation@fr : groupe d'humains@fr), (firme@fr : organisation@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
19. géologie@fr <-> géologie@fr : 0.6452451944351196  
[[géologie@fr : science de la nature@fr], (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : classe@fr), (forme juridique@fr : classe@fr), (personne morale@fr : forme juridique@fr), (firme@fr : personne morale@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
20. pollution industrielle@fr <-> pollution@fr : 0.6422240734100342  
[[pollution@fr : effet pervers@fr], (effet pervers@fr : résultat@fr), (résultat@fr : entité observable@fr), (entité observable@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
21. label Gemet non trouvé <-> infrastructure de transport@fr : 0.64194655418396  
[[infrastructure de transport@fr : objet géographique@fr], (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
22. label Gemet non trouvé <-> bassin hydrographique@fr : 0.6412675380706787  
Chemin non trouvé
23. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.6363604664802551  
[[qualité de l'eau@fr : qualité@fr], (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
24. label Gemet non trouvé <-> bien culturel@fr : 0.6321249604225159  
[[bien culturel@fr : objet concret@fr], (objet tangible@fr : objet concret@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
25. label Gemet non trouvé <-> label Wikidata introuvable : 0.6305616497993469  
Chemin non trouvé
26. label Gemet non trouvé <-> nappe phréatique@fr : 0.6300790905952454  
[[nappe phréatique@fr : aquifère@fr], (aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : réservoir@fr), (réservoir@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
27. hydrographie@fr <-> hydrographie@fr : 0.6294116973876953  
[[hydrographie@fr : sciences de la Terre@fr], (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
28. infrastructure@fr <-> infrastructure@fr : 0.628700852394104  
[[infrastructure@fr : œuvre@fr], (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
29. label Gemet non trouvé <-> submersion marine@fr : 0.6275973320007324  
[[submersion marine@fr : inondation@fr], (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr :

- entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
30. biogaz@fr <-> biogaz@fr : 0.6261656284332275  
[[biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : gaz@fr), (gaz@fr : fluide@fr), (fluide@fr : substance@fr), (substance@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
31. logement@fr <-> gestion de logement@fr : 0.6251828670501709  
Chemin non trouvé
32. aquifère@fr <-> aquifère@fr : 0.6246013641357422  
[[aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : réservoir@fr), (réservoir@fr : structure architecturale@fr), (structure architecturale@fr : complexe immobilier@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
33. communication en matière de risques@fr <-> gestion du risque@fr :  
0.6228855848312378  
[[gestion du risque@fr : fonction@fr), (fonction@fr : usage@fr), (usage@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
34. monument@fr <-> monument de Hongrie@fr : 0.6182126998901367  
[[monument de Hongrie@fr : bien patrimonial national@fr), (bien patrimonial national@fr : site patrimonial@fr), (site patrimonial@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (objet géographique@fr : équipement@fr : structure construite@fr), (équipement@fr : structure construite@fr), (station de traitement des eaux usées@fr : usine@fr)]
35. label Gemet non trouvé <-> label Wikidata introuvable : 0.6179236173629761  
Chemin non trouvé
36. label Gemet non trouvé <-> bassin versant@fr : 0.6176934242248535  
[[bassin versant@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
37. habitat@fr <-> habitat@fr : 0.6176526546478271  
[[habitat@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
38. ressource naturelle@fr <-> ressource naturelle@fr : 0.6150418519973755  
[[ressource naturelle@fr : matériau brut@fr), (matériau brut@fr : bien@fr), (produit@fr : bien@fr), (œuvre@fr : produit@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
39. risque naturel@fr <-> risque naturel@fr : 0.6141432523727417  
[[risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (objet@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
40. érosion@fr <-> érosion@fr : 0.6131541728973389  
[[érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
41. polluant@fr <-> label Wikidata introuvable : 0.6121931076049805  
Chemin non trouvé

42. label Gemet non trouvé <-> réseau de télécommunications@fr :  
0.6115759015083313  
[[réseau de télécommunications@fr : infrastructure@fr], (infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
43. surveillance@fr <-> surveillance@fr : 0.6091445684432983  
[[surveillance@fr : activité@fr], (Q18390984 : activité@fr), (Q18390984 : outil de gestion du risque@fr), (outil de gestion du risque@fr : outil de gestion@fr), (outil de gestion@fr : outil@fr), (outil@fr : équipement@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
44. estuaire@fr <-> estuaire@fr : 0.6066519021987915  
[[estuaire@fr : étendue d'eau@fr], (étendue d'eau@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
45. dommage@fr <-> dégât@fr : 0.6065858602523804  
[[dégât@fr : processus@fr], (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (ensemble architectural@fr : groupe de constructions et de bâtiments@fr), (complexe immobilier@fr : ensemble architectural@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
46. logement@fr <-> appartement@fr : 0.6065312623977661  
[[appartement@fr : logement@fr], (logement@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
47. label Gemet non trouvé <-> transition écologique@fr : 0.6048821210861206  
[[transition écologique@fr : changement@fr], (changement@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité spatio-temporelle@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
48. chauffage à distance@fr <-> réseau de chaleur@fr : 0.603854775428772  
[[réseau de chaleur@fr : infrastructure@fr], (infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
49. zonage@fr <-> zonage@fr : 0.6028785705566406  
[[zonage@fr : planification@fr], (planification@fr : processus@fr), (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (groupe de constructions et de bâtiments@fr : structure construite@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
50. label Gemet non trouvé <-> unité statistique@fr : 0.6022796630859375  
[[unité statistique@fr : unité@fr], (unité@fr : objet abstrait@fr), (objet abstrait@fr : entité non-physique@fr), (entité non-physique@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]

# Expérimentation :

## Fixer un seuil de similarité sémantique minimal entre deux concepts du domaine de l'eau

### Introduction

Dans le cadre de ma thèse, j'étudie comment interroger des catalogues de données géographiques en utilisant des concepts, dits "concepts utilisateur", relatifs à l'application - pour laquelle l'utilisateur a besoin de données géographiques-. Ces concepts utilisateurs sont potentiellement différents de ceux utilisés dans les métadonnées qui décrivent les jeux, et donc dans les catalogues actuels. Lorsque le concept est un nom de lieu, des services permettent de passer de la requête de l'utilisateur, e.g. "France", à une requête sur l'extension spatiale du jeu. Je m'intéresse aux concepts de la requête relatifs à une thématique. Dans les métadonnées, les concepts relatifs à la thématique correspondent aux valeurs des propriétés "keyword" et "topic category", je les appelle "concepts des métadonnées". Pour cela, je cherche à associer à un "concept utilisateur" les "concepts de métadonnées", c'est-à-dire les concepts utilisés pour renseigner ces propriétés, qui sont suffisamment proches pour qu'on considère qu'ils sont pertinents pour interroger des catalogues en vue de construire l'application. Le but de ce questionnaire est d'avoir votre retour d'experts sur la façon dont je mesure la proximité entre un concept utilisateur et les concepts métadonnées, dans le cas particulier du domaine de l'eau.

Une première expérience avec des experts a permis d'évaluer la pertinence de recourir à l'ontologie de sens commun Wikidata pour exprimer des concepts utilisateurs, correspondant à des concepts d'intérêt applicatifs, dans le domaine de l'eau. Pour cette expérience, nous avons choisi 4 concepts Wikidata, parmi ceux suggérés par les experts, à savoir : "nappe phréatique", "station d'épuration", "marais artificiel" et "PPRI".

Les concepts des métadonnées ont été récupérés de la façon suivante. Les métadonnées de catalogues du domaine, conformes au standard INSPIRE, ont été moissonnées. J'ai ensuite extrait les valeurs des champs "Topics Category" et "Keyword" de tous les enregistrements (un enregistrement = une description d'un jeu de données à l'aide de divers champs). Ces valeurs sont des chaînes de caractères. Elles ont été utilisées pour extraire des concepts du thésaurus GEMET (standard utilisé pour documenter Topic Category dans INSPIRE) ainsi qu'avec l'ontologie de sens commun Wikidata.. Au total, 205 concepts différents ont été identifiés. Pour les concepts GEMET, un alignement a été réalisé avec Wikidata, de sorte que tous les concepts, utilisateurs ou métadonnées, sont intégrés dans un graphe dans lequel nous pouvons utiliser une mesure de similarité sémantique.

Notre but est de fixer un seuil à partir duquel on peut affirmer que deux concepts sont suffisamment proches pour que le concept métadonnées soit utilisé pour reformuler la requête de l'utilisateur. Pour cela, nous avons comparé chaque concept utilisateur à tous les concepts métadonnées en utilisant la mesure de similarité sémantique et gardé les 50 meilleurs résultats. Par ailleurs, dans une majorité des cas, nous avons affiché également le chemin entre les concepts.

**Commentaire [NP1]:** LE chemin ; son unicité est pour moi un problème. Peut être que considérer simultanément plusieurs chemins renforcerait le poids thématique de certains items, les faisant grimper dans la hiérarchie des propositions

## Recherche des concepts suffisamment proches de : “Nappe phréatique” (Q583075)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté ?**  
**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**  
**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

- i. Label GEMET <-> Label Wikidata : score de similarité**  
*[chemin entre le concept utilisateur et le concept métadonnées]*

- label Gemet non trouvé <-> nappe phréatique@fr : 1.0  
*[nappe phréatique@fr]*
- aquifère@fr <-> aquifère@fr : 0.9400091171264648  
*[[nappe phréatique@fr : aquifère@fr]]*
- eaux souterraines@fr <-> eau souterraine@fr : 0.8649603128433228  
*[[eau souterraine@fr : eaux intérieures@fr], (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
- label Gemet non trouvé <-> bassin versant@fr : 0.7749752998352051  
*[[bassin versant@fr : zone géographique@fr], (étendue d'eau@fr : zone géographique@fr), (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
- label Gemet non trouvé <-> bassin hydrographique@fr : 0.7529347538948059  
*Chemin non trouvé*
- fleuve@fr <-> courant d'eau@fr : 0.7424725294113159  
*[[courant d'eau@fr : cours d'eau naturel@fr], (cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
- label Gemet non trouvé <-> masse d'eau@fr : 0.7240350246429443  
*[[masse d'eau@fr : étendue d'eau@fr], (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
- fleuve@fr <-> cours d'eau@fr : 0.7219454646110535  
*[[cours d'eau@fr : eaux intérieures@fr], (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*
- nappe phréatique@fr <-> niveau piézométrique@fr : 0.7217990159988403  
*[[niveau piézométrique@fr : surface@fr], (surface@fr : espace à deux dimensions@fr), (espace à deux dimensions@fr : espace mathématique@fr), (espace mathématique@fr : ensemble@fr), (ensemble@fr : classe@fr), (classe@fr : entité composée@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

**Commentaire [NP2]:** Pour moi :  
•nappe phréatique@fr  
•aquifère@fr  
•masse d'eau@fr (parent)  
ensuite trop éloigné MAIS certains peuvent être pertinent par leur caractère englobant (sortir l'utilisateur d'un 'minimum local')

**Commentaire [NP3]:** Le label gemet existe :  
<https://www.eionet.europa.eu/gemet/fr/concept/12104>

**Commentaire [NP4]:** Dans gemet, la définition indique Phreatic water is synonymous with groundwater, le score de similarité me semblerait devoir être plus élevé

**Commentaire [NP5]:** Le label gemet existe  
<https://www.eionet.europa.eu/gemet/fr/concept/1215>

**Commentaire [NP6]:** Je préférerais voir cours d'eau à cette place n°6

**Commentaire [NP7]:** Sémantiquement plus proche de nappe phréatique, cf [https://fr.wikipedia.org/wiki/Masse\\_d'eau](https://fr.wikipedia.org/wiki/Masse_d'eau)

10. estuaire@fr <-> estuaire@fr : 0.7091968059539795  
 [(estuaire@fr : étendue d'eau@fr), (zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
11. hydrogéologie@fr <-> hydrogéologie@fr : 0.6929916143417358  
 [(hydrogéologie@fr : hydrologie@fr), (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
12. label Gemet non trouvé <-> submersion marine@fr : 0.6904162764549255  
 [(submersion marine@fr : inondation@fr), (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (objet artificiel physique@fr : entité artificielle@fr), (réceptif@fr : objet artificiel physique@fr), (réservoir@fr : réceptif@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
13. fleuve@fr <-> cours d'eau naturel@fr : 0.6870043277740479  
 [(cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
14. label Gemet non trouvé <-> conchyliculture@fr : 0.6838634610176086  
 [(conchyliculture@fr : pêche@fr), (pêche@fr : chasse@fr), (chasse@fr : action@fr), (action@fr : convention@fr), (convention@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
15. entité hydrologique@fr <-> étendue d'eau@fr : 0.6789747476577759  
 [(zone aquatique@fr : étendue d'eau@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
16. littoral@fr <-> littoral@fr : 0.6748140454292297  
 [(littoral@fr : rivage@fr), (rivage@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
17. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.6704185009002686  
 [(qualité de l'eau@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
18. label Gemet non trouvé <-> gestion de l'eau@fr : 0.6675404906272888  
 [(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : relation asymétrique@fr), (relation asymétrique@fr : relation binaire@fr), (relation binaire@fr : relation@fr), (représentation@fr : relation@fr), (représentation@fr : entité artificielle@fr), (objet artificiel physique@fr : entité artificielle@fr), (réceptif@fr : objet artificiel physique@fr), (réservoir@fr : réceptif@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
19. pluie@fr <-> pluie@fr : 0.6633356809616089  
 [(pluie@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr :

**Commentaire [NP8]:** Bizarrement loin du score de cours d'eau

**Commentaire [NP9]:** Mériterait de popper, comme keyword englobant

**Commentaire [NP10]:** Label gemet existe, <https://www.eionet.europa.eu/gemet/fr/concept/9195>

Mériterait de popper, comme keyword englobant

*structure@fr*, (*stockage d'eau@fr* : *stockage@fr*), (*réservoir d'eau@fr* : *stockage d'eau@fr*), (*aquifère@fr* : *réservoir d'eau@fr*), (*nappe phréatique@fr* : *aquifère@fr*)

20. *ressource naturelle@fr* <-> *ressource naturelle@fr* : 0.6476647853851318

*[[ressource naturelle@fr : matériau brut@fr], (matériau brut@fr : bien@fr), (produit@fr : bien@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

21. *hydrologie@fr* <-> *hydrologie@fr* : 0.643667459487915

*[[hydrologie@fr : science@fr], (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

22. label Gemet non trouvé <-> label Wikidata introuvable : 0.6301834583282471

*Chemin non trouvé*

Commentaire [NP11]: ???

23. *lac@fr* <-> *lac@fr* : 0.6288014650344849

*[[lac@fr : eau calme@fr], (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

24. *fleuve@fr* <-> *rivière@fr* : 0.6218131184577942

*[[rivière@fr : cours d'eau naturel@fr], (cours d'eau naturel@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

Commentaire [NP12]: Equivalence de thesaurus très surprenante

25. label Gemet non trouvé <-> *source d'énergie@fr* : 0.6207959651947021

*[[source d'énergie@fr : source@fr], (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (produit@fr : entité artificielle@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

Commentaire [NP13]: Totalement hors de propos

26. *ferme (exploitation)@fr* <-> *ferme@fr* : 0.6167993545532227

*[[ferme@fr : propriété agricole@fr], (propriété agricole@fr : complexe immobilier@fr), (complexe immobilier@fr : structure architecturale@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

Commentaire [NP14]: Totalement hors de propos

27. *eau usée@fr* <-> *eaux usées@fr* : 0.6070104837417603

*[[eaux usées@fr : eau non potable@fr], (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

28. *port@fr* <-> *havre@fr* : 0.6061320304870605

*[[havre@fr : baie@fr], (baie@fr : zone aquatique@fr), (eaux intérieures@fr : zone aquatique@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

29. label Gemet non trouvé <-> *crue@fr* : 0.6060789823532104

*[[crue@fr : catastrophe naturelle@fr], (catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

Commentaire [NP15]: Label gemet <https://www.eionet.europa.eu/gemet/fr/concept/3298>

30. label Gemet non trouvé <-> *prélèvement d'eau@fr* : 0.600567102432251

*[[prélèvement d'eau@fr : infrastructure@fr], (infrastructure@fr : œuvre@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

31. label Gemet non trouvé <-> saliculture@fr : 0.59828782081604  
 [(saliculture@fr : production@fr), (production@fr : activité économique@fr), (activité économique@fr : action@fr), (action@fr : convention@fr), (convention@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
32. label Gemet non trouvé <-> unité administrative locale@fr : 0.5975565910339355  
 [(unité administrative locale@fr : division territoriale d'un niveau d'imbrication donné@fr), (division territoriale d'un niveau d'imbrication donné@fr : entité territoriale administrative@fr), (entité territoriale administrative@fr : entité territoriale géographique humaine@fr), (entité territoriale géographique humaine@fr : entité géographique artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (objet artificiel physique@fr : entité artificielle@fr), (récipient@fr : objet artificiel physique@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
33. érosion@fr <-> érosion@fr : 0.5942178964614868  
 [(érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (convertisseur@fr : entité@fr), (outil@fr : convertisseur@fr), (récipient@fr : outil@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
34. fleuve@fr <-> ruisseau@fr : 0.5931576490402222  
 [(ruisseau@fr : cours d'eau@fr), (cours d'eau@fr : eaux intérieures@fr), (eau calme@fr : eaux intérieures@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
35. santé publique et assainissement@fr <-> assainissement@fr : 0.5848271250724792  
 [(assainissement@fr : hygiène@fr), (hygiène@fr : soins de santé@fr), (soins de santé@fr : service@fr), (service@fr : produit@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
36. gaz@fr <-> gaz@fr : 0.5823714137077332  
 [(gaz@fr : fluide@fr), (fluide@fr : substance@fr), (substance@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
37. risque naturel@fr <-> risque naturel@fr : 0.5795302391052246  
 [(risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (produit@fr : entité artificielle@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
38. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5750667452812195  
 [(catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
39. label Gemet non trouvé <-> BD Carthage@fr : 0.5737375020980835  
 Chemin non trouvé
40. biogaz@fr <-> biogaz@fr : 0.5729211568832397  
 [(biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : combustible@fr), (combustible@fr : matière énergétique@fr), (matière énergétique@fr : matériau@fr), (matériau@fr : produit@fr), (œuvre@fr : produit@fr), (stockage d'eau@fr : œuvre@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]
41. eau potable@fr <-> eau potable@fr : 0.5725058913230896

**Commentaire [NP16]:** Comment le gaz peut arriver devant l'eau ?

*[(eau potable@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

42. label Gemet non trouvé <-> Gestion des ressources naturelles@fr :

0.5694080591201782

*[(Gestion des ressources naturelles@fr : maltraitance@fr), (maltraitance@fr : problème sociétal@fr), (problème sociétal@fr : problème@fr), (problème@fr : condition@fr), (condition@fr : état@fr), (état@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

43. label Gemet non trouvé <-> perméabilité@fr : 0.5676491856575012

*[(perméabilité@fr : propriété physique@fr), (propriété physique@fr : propriété@fr), (propriété@fr : entité@fr), (entité composée@fr : entité@fr), (structure@fr : entité composée@fr), (stockage@fr : structure@fr), (stockage d'eau@fr : stockage@fr), (réservoir d'eau@fr : stockage d'eau@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

44. hydrographie@fr <-> hydrographie@fr : 0.5633646845817566

*[(hydrographie@fr : sciences de la Terre@fr), (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

45. lac@fr <-> ensemble de lacs@fr : 0.563302755355835

*[(ensemble de lacs@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

46. label Gemet non trouvé <-> cordon littoral@fr : 0.5623621344566345

*Chemin non trouvé*

47. biotope@fr <-> biotope@fr : 0.5620291829109192

*[(biotope@fr : entité géographique naturelle@fr), (entité géographique naturelle@fr : entité géographique@fr), (entité géographique@fr : entité@fr), (convertisseur@fr : entité@fr), (outil@fr : convertisseur@fr), (récipient@fr : outil@fr), (réservoir@fr : récipient@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

48. label Gemet non trouvé <-> plan d'eau@fr : 0.5611335635185242

*[(plan d'eau@fr : eau calme@fr), (réservoir d'eau@fr : eau calme@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

49. gestion des déchets@fr <-> gestion des déchets@fr : 0.5611010193824768

*[(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

50. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.5600736737251282

*[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure architecturale@fr), (réservoir@fr : structure architecturale@fr), (réservoir d'eau@fr : réservoir@fr), (aquifère@fr : réservoir d'eau@fr), (nappe phréatique@fr : aquifère@fr)]*

## Recherche des concepts suffisamment proches de : "Phytoépuration" (Q1376406)

**Commentaire [NP17]:** Résultats très décevants ! seuls 3 items mériteraient pour moi de sortir, ils sont noyés

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté ?**  
**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**  
**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

• **i. Label GEMET <-> Label Wikidata : score de similarité**  
**[chemin entre le concept utilisateur et le concept métadonnées ]**

1. pluie@fr <-> pluie@fr : 0.732921838760376  
[[pluie@fr : phénomène naturel@fr], (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
2. label Gemet non trouvé <-> bassin versant@fr : 0.6982258558273315  
[[bassin versant@fr : zone géographique@fr], (étendue d'eau@fr : zone géographique@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
3. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.6933472156524658  
[[ouvrage hydraulique@fr : ouvrage d'art@fr], (ouvrage d'art@fr : structure construite@fr), (structure construite@fr : objet géographique@fr), (objet géographique naturel@fr : objet géographique@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
4. label Gemet non trouvé <-> submersion marine@fr : 0.6883678436279297  
[[submersion marine@fr : inondation@fr], (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
5. label Gemet non trouvé <-> masse d'eau@fr : 0.6847524642944336  
[[masse d'eau@fr : étendue d'eau@fr], (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
6. label Gemet non trouvé <-> prélèvement d'eau@fr : 0.66660475730896  
[[prélèvement d'eau@fr : infrastructure@fr], (infrastructure@fr : œuvre@fr), (œuvre@fr : produit@fr), (produit@fr : entité artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
7. estuaire@fr <-> estuaire@fr : 0.6564208269119263  
[[estuaire@fr : écosystème@fr], (zone humide@fr : écosystème@fr), (Phytoépuration@fr : zone humide@fr)]
8. eau usée@fr <-> eaux usées@fr : 0.6487299203872681  
[[eaux usées@fr : eau non potable@fr], (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]

**Commentaire [NP18]:** Premier item sémantiquement proche, mais bien trop éloigné !

9. label Gemet non trouvé <-> nappe phréatique@fr : 0.6477580070495605  
 [(nappe phréatique@fr : aquifère@fr), (aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : eau calme@fr), (eau calme@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
10. hydrologie@fr <-> hydrologie@fr : 0.6444834470748901  
 [(hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : système@fr), (Q106865236 : système@fr), (Q106865236 : objet naturel@fr), (objet physique naturel@fr : objet naturel@fr), (objet géographique naturel@fr : objet physique naturel@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
11. label Gemet non trouvé <-> gestion de l'eau@fr : 0.6390826106071472  
 [(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : aptitude@fr), (aptitude@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
12. aquifère@fr <-> aquifère@fr : 0.6373130679130554  
 [(aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : eau calme@fr), (eau calme@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
13. protection contre les crues@fr <-> inondation@fr : 0.6344674229621887  
 [(inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : source de risque@fr), (source de risque@fr : source@fr : objet@fr), (entité artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
14. pollution industrielle@fr <-> pollution@fr : 0.6167508959770203  
 [(pollution@fr : effet pervers@fr), (effet pervers@fr : résultat@fr), (résultat@fr : entité observable@fr), (entité observable@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
15. santé publique et assainissement@fr <-> assainissement@fr : 0.6126349568367004  
 [(assainissement@fr : facteur écologique@fr), (facteur écologique@fr : corrélation@fr), (corrélation@fr : relation@fr), (représentation@fr : relation@fr), (représentation@fr : entité artificielle@fr), (entité géographique artificielle@fr : entité artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
16. nappe phréatique@fr <-> niveau piézométrique@fr : 0.6110435724258423  
 [(niveau piézométrique@fr : surface@fr), (surface@fr : espace à deux dimensions@fr), (espace à deux dimensions@fr : espace mathématique@fr), (espace mathématique@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
17. label Gemet non trouvé <-> saliculture@fr : 0.6026057004928589  
 [(saliculture@fr : production@fr), (production@fr : processus@fr), (processus@fr : changement@fr), (changement@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité physique@fr : entité spatio-temporelle@fr), (objet tangible@fr : entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
18. biogaz@fr <-> biogaz@fr : 0.5992053747177124  
 [(biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : combustible@fr), (combustible@fr : matière énergétique@fr), (matière énergétique@fr : matériau@fr), (matériau@fr : produit@fr), (produit@fr :

**Commentaire [NP19]:** La phytoépuration est une technique d'assainissement, cet item devrait popper en premier !

- entité artificielle@fr*, (*entité géographique artificielle@fr* : *entité artificielle@fr*), (*forme de relief artificielle@fr* : *entité géographique artificielle@fr*), (*Phytoépuration@fr* : *forme de relief artificielle@fr*)
19. **eaux souterraines@fr <-> eau souterraine@fr : 0.5980356931686401**  
[[*eau souterraine@fr* : *eaux intérieures@fr*], (*eaux intérieures@fr* : *zone aquatique@fr*), (*zone aquatique@fr* : *étendue d'eau@fr*), (*zone humide@fr* : *étendue d'eau@fr*), (*Phytoépuration@fr* : *zone humide@fr*)]
20. **label Gemet non trouvé <-> crue@fr : 0.5977764129638672**  
[[*crue@fr* : *catastrophe naturelle@fr*], (*catastrophe naturelle@fr* : *phénomène naturel@fr*), (*phénomène naturel@fr* : *phénomène@fr*), (*phénomène@fr* : *entité@fr*), (*entité géographique@fr* : *entité@fr*), (*entité géographique artificielle@fr* : *entité géographique@fr*), (*forme de relief artificielle@fr* : *entité géographique artificielle@fr*), (*Phytoépuration@fr* : *forme de relief artificielle@fr*)]
21. **urbanisme@fr <-> aménagement urbain@fr : 0.5961806774139404**  
[[*aménagement urbain@fr* : *urbanisme@fr*], (*urbanisme@fr* : *géographie humaine@fr*), (*géographie humaine@fr* : *géographie@fr*), (*géographie@fr* : *objet de la recherche@fr*), (*objet de la recherche@fr* : *objet@fr*), (*entité géographique artificielle@fr* : *entité géographique@fr*), (*entité géographique artificielle@fr* : *entité artificielle@fr*), (*forme de relief artificielle@fr* : *entité géographique artificielle@fr*), (*Phytoépuration@fr* : *forme de relief artificielle@fr*)]
22. **ferme (exploitation)@fr <-> ferme@fr : 0.5928126573562622**  
[[*ferme@fr* : *établissement rural@fr*], (*établissement rural@fr* : *localité@fr*), (*localité@fr* : *entité géographique artificielle@fr*), (*forme de relief artificielle@fr* : *entité géographique artificielle@fr*), (*Phytoépuration@fr* : *forme de relief artificielle@fr*)]
23. **label Gemet non trouvé <-> conchyliculture@fr : 0.5910194516181946**  
[[*conchyliculture@fr* : *pêche@fr*], (*pêche@fr* : *chasse@fr*), (*chasse@fr* : *action@fr*), (*action@fr* : *occurrence@fr*), (*occurrence@fr* : *entité temporelle@fr*), (*entité temporelle@fr* : *entité spatio-temporelle@fr*), (*entité physique@fr* : *entité spatio-temporelle@fr*), (*objet tangible@fr* : *entité physique@fr*), (*objet physico-géographique@fr* : *objet tangible@fr*), (*forme de relief artificielle@fr* : *objet physico-géographique@fr*), (*Phytoépuration@fr* : *forme de relief artificielle@fr*)]
24. **hydrogéologie@fr <-> hydrogéologie@fr : 0.5909260511398315**  
[[*hydrogéologie@fr* : *hydrologie@fr*], (*hydrologie@fr* : *science@fr*), (*science@fr* : *système de connaissances@fr*), (*système de connaissances@fr* : *ensemble@fr*), (*ensemble@fr* : *entité composée@fr*), (*entité composée@fr* : *entité@fr*), (*entité géographique@fr* : *entité@fr*), (*entité géographique artificielle@fr* : *entité géographique@fr*), (*forme de relief artificielle@fr* : *entité géographique artificielle@fr*), (*Phytoépuration@fr* : *forme de relief artificielle@fr*)]
25. **label Gemet non trouvé <-> Plan de prévention du risque inondation@fr : 0.5880736708641052**  
*Chemin non trouvé*
26. **catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5866144895553589**  
[[*catastrophe naturelle@fr* : *phénomène naturel@fr*], (*phénomène naturel@fr* : *phénomène@fr*), (*phénomène@fr* : *entité@fr*), (*entité géographique@fr* : *entité@fr*), (*entité géographique artificielle@fr* : *entité géographique@fr*), (*forme de relief artificielle@fr* : *entité géographique artificielle@fr*), (*Phytoépuration@fr* : *forme de relief artificielle@fr*)]
27. **développement urbain@fr <-> urbanisme@fr : 0.5849825143814087**  
[[*urbanisme@fr* : *géographie humaine@fr*], (*géographie humaine@fr* : *géographie@fr*), (*géographie@fr* : *objet de la recherche@fr*), (*objet de la recherche@fr* : *objet@fr*), (*entité géographique artificielle@fr* : *entité géographique@fr*), (*entité géographique artificielle@fr* : *entité artificielle@fr*), (*forme de relief artificielle@fr* : *entité géographique artificielle@fr*), (*Phytoépuration@fr* : *forme de relief artificielle@fr*)]
28. **littoral@fr <-> littoral@fr : 0.5812649726867676**  
[[*littoral@fr* : *rivage@fr*], (*rivage@fr* : *objet physico-géographique@fr*), (*forme de relief artificielle@fr* : *objet physico-géographique@fr*), (*Phytoépuration@fr* : *forme de relief artificielle@fr*)]

29. hydrographie@fr <-> hydrographie@fr : 0.579881489276886  
 [(hydrographie@fr : sciences de la Terre@fr), (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
30. risque naturel@fr <-> risque naturel@fr : 0.5705837607383728  
 [(risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (entité géographique artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
31. fleuve@fr <-> cours d'eau naturel@fr : 0.5690486431121826  
 [(cours d'eau naturel@fr : objet géographique naturel@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
32. label Gemet non trouvé <-> label Wikidata introuvable : 0.5682553648948669  
 Chemin non trouvé
33. gestion des déchets@fr <-> gestion des déchets@fr : 0.567966639995575  
 [(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
34. ressource naturelle@fr <-> ressource naturelle@fr : 0.5668643712997437  
 [(ressource naturelle@fr : ressource@fr), (ressource@fr : source@fr), (source@fr : objet@fr), (entité géographique artificielle@fr : objet@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
35. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.5630149841308594  
 [(qualité de l'eau@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]
36. label Gemet non trouvé <-> masse d'eau de surface@fr : 0.5613843202590942  
 [(masse d'eau de surface@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
37. fleuve@fr <-> cours d'eau@fr : 0.5561116933822632  
 [(cours d'eau@fr : eaux intérieures@fr), (eaux intérieures@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
38. fleuve@fr <-> courant d'eau@fr : 0.5535402894020081  
 [(courant d'eau@fr : cours d'eau naturel@fr), (cours d'eau naturel@fr : objet géographique naturel@fr), (zone humide@fr : objet géographique naturel@fr), (Phytoépuration@fr : zone humide@fr)]
39. port@fr <-> havre@fr : 0.5492718815803528  
 [(havre@fr : baie@fr), (baie@fr : zone aquatique@fr), (zone aquatique@fr : étendue d'eau@fr), (zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]
40. érosion@fr <-> érosion@fr : 0.5444740056991577  
 [(érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr)]

Commentaire [NP20]: Autre item, un peu plus englobant

*géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*

41. agriculture@fr <-> agriculture@fr : 0.5412713289260864  
*[(agriculture@fr : agriculture et sylviculture@fr), (agriculture et sylviculture@fr : économie@fr), (économie@fr : système social@fr), (système social@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
42. environnement@fr <-> environnement@fr : 0.5334920883178711  
*Chemin non trouvé*
43. polluant@fr <-> polluant@fr : 0.531983494758606  
*[(polluant@fr : substance physique@fr), (substance physique@fr : entité physique@fr), (objet tangible@fr : entité physique@fr), (objet physico-géographique@fr : objet tangible@fr), (forme de relief artificielle@fr : objet physico-géographique@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
44. lac@fr <-> lac@fr : 0.5294979214668274  
*[(lac@fr : modelé@fr), (forme de relief artificielle@fr : modelé@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
45. entité hydrologique@fr <-> étendue d'eau@fr : 0.5264604687690735  
*[(zone humide@fr : étendue d'eau@fr), (Phytoépuration@fr : zone humide@fr)]*
46. gestion@fr <-> label Wikidata introuvable : 0.5256054997444153  
*Chemin non trouvé*
47. diversité biologique@fr <-> biodiversité@fr : 0.5250712633132935  
*[(biodiversité@fr : diversité@fr), (diversité@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité géographique@fr : entité@fr), (entité géographique artificielle@fr : entité géographique@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*
48. label Gemet non trouvé <-> La Réunion@fr : 0.5243727564811707  
*Chemin non trouvé*
49. label Gemet non trouvé <-> bassin hydrographique@fr : 0.5230716466903687  
*Chemin non trouvé*
50. label Gemet non trouvé <-> réseau de transport@fr : 0.522017776966095  
*[(réseau de transport@fr : entité géographique artificielle@fr), (forme de relief artificielle@fr : entité géographique artificielle@fr), (Phytoépuration@fr : forme de relief artificielle@fr)]*

## Recherche des concepts suffisamment proches de : "Plan de prévention du risque inondation" (Q3391269)

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté?**  
**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**

*(Note Mehdi : il n'y a pas de chemins entre PPRI et les autres concepts. PPRI est isolé dans le graphe de Wikidata)*

### Liste des comparaisons : (meilleurs 50 résultats)

1. label Gemet non trouvé <-> Plan de prévention du risque inondation@fr : 1.0
2. label Gemet non trouvé <-> submersion marine@fr : 0.699097752571106
3. label Gemet non trouvé <-> protection@fr : 0.65545654296875
4. label Gemet non trouvé <-> masse d'eau@fr : 0.6312357187271118
5. estuaire@fr <-> estuaire@fr : 0.6220836639404297
6. aquifère@fr <-> aquifère@fr : 0.6191027164459229
7. label Gemet non trouvé <-> directive-cadre sur l'eau@fr : 0.6143583655357361
8. label Gemet non trouvé <-> conchyliculture@fr : 0.6103661060333252
9. communication en matière de risques@fr <-> gestion du risque@fr : 0.6078090071678162
10. port@fr <-> havre@fr : 0.6073088049888611
11. label Gemet non trouvé <-> bassin versant@fr : 0.606191873550415
12. label Gemet non trouvé <-> label Wikidata introuvable : 0.5888326168060303
13. plan@fr <-> plan@fr : 0.5848706364631653
14. fleuve@fr <-> cours d'eau@fr : 0.5825957655906677
15. hydrographie@fr <-> hydrographie@fr : 0.5800302624702454
16. politique@fr <-> politique@fr : 0.5799053907394409
17. hydrogéologie@fr <-> hydrogéologie@fr : 0.5739492177963257
18. label Gemet non trouvé <-> bassin hydrographique@fr : 0.5727195143699646
19. label Gemet non trouvé <-> prévention@fr : 0.567156195640564
20. pluie@fr <-> pluie@fr : 0.5633394718170166
21. label Gemet non trouvé <-> Auvergne-Rhône-Alpes@fr : 0.5619726181030273
22. ferme (exploitation)@fr <-> ferme@fr : 0.5597040057182312
23. label Gemet non trouvé <-> nappe phréatique@fr : 0.5578835010528564
24. label Gemet non trouvé <-> Guadeloupe@fr : 0.5573101043701172
25. label Gemet non trouvé <-> cordon littoral@fr : 0.5549724102020264
26. label Gemet non trouvé <-> Guyane@fr : 0.554639995098114
27. label Gemet non trouvé <-> Politique environnementale de l'Union européenne@fr : 0.5531107187271118
28. label Gemet non trouvé <-> crue@fr : 0.5518550872802734
29. planification@fr <-> planification@fr : 0.5512016415596008
30. label Gemet non trouvé <-> saliculture@fr : 0.5446555614471436
31. politique de l'environnement@fr <-> politique environnementale@fr : 0.5440297722816467
32. fleuve@fr <-> courant d'eau@fr : 0.5439935922622681
33. gestion de crise@fr <-> gestion de crise@fr : 0.537813663482666

Commentaire [NP21]: non

Commentaire [NP22]: non

Commentaire [NP23]: non

Commentaire [NP24]: non

Commentaire [NP25]: surprenante association !!! générique avec une de ses instances !

Commentaire [NP26]: pourquoi ca apparait, avec ce score ???

Commentaire [NP27]: confusion instance/générique

Commentaire [NP28]: Devrait popper en premier

34. risque naturel@fr <-> risque naturel@fr : 0.5369033813476562
35. ressource naturelle@fr <-> ressource naturelle@fr : 0.5360186100006104
36. littoral@fr <-> littoral@fr : 0.5313910245895386
37. réserve naturelle nationale@fr <-> aire protégée@fr : 0.530930757522583
38. label Gemet non trouvé <-> source d'énergie@fr : 0.5305038690567017
39. surveillance@fr <-> surveillance@fr : 0.5277665853500366
40. label Gemet non trouvé <-> Bretagne@fr : 0.5256977081298828
41. label Gemet non trouvé <-> gestion de l'eau@fr : 0.5250158309936523
42. environnement@fr <-> environnement@fr : 0.5216457843780518
43. fleuve@fr <-> rivière@fr : 0.5166248083114624
44. catastrophe naturelle@fr <-> catastrophe naturelle@fr : 0.5143988132476807
45. label Gemet non trouvé <-> Mayotte@fr : 0.5137329697608948
46. source d'énergie renouvelable@fr <-> énergie renouvelable@fr : 0.5136480331420898
47. risque@fr <-> risque@fr : 0.5124579071998596
48. eau usée@fr <-> eaux usées@fr : 0.5105905532836914
49. fleuve@fr <-> cours d'eau naturel@fr : 0.5103414058685303
50. label Gemet non trouvé <-> lycée@fr : 0.5079976320266724

**Commentaire [NP29]:** Un ppri préserve d'un risque naturel, devrait popper plus haut

## Recherche des concepts suffisamment proches de : "Station de traitement des eaux usées"

**Est-ce que dans la liste ordonnée ci-dessous, l'ordre du plus proche au plus lointain vous semble respecté ?**  
**A partir de quel résultat estimez-vous que les concepts sont trop éloignés sémantiquement ?**  
**Dans les cas où un chemin est affiché, est-ce que ce chemin vous apporte une information utile ?**

### Liste des comparaisons : (meilleurs 50 résultats)

**Label GEMET <-> Label Wikidata : score de similarité**

**[chemin entre le concept utilisateur et le concept métadonnées]**

1. eau usée@fr <-> eaux usées@fr : 0.7805310487747192  
[(eaux usées@fr : eau non potable@fr), (eau non potable@fr : eau liquide@fr), (eau liquide@fr : boisson@fr), (boisson@fr : nourriture@fr), (nourriture@fr : objet tangible@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
2. label Gemet non trouvé <-> ouvrage hydraulique@fr : 0.7642497420310974  
[(ouvrage hydraulique@fr : ouvrage d'art@fr), (ouvrage d'art@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
3. gestion des déchets@fr <-> gestion des déchets@fr : 0.761856198310852  
[(gestion des déchets@fr : Durabilité et gestion de l'environnement@fr), (Durabilité et gestion de l'environnement@fr : génie environnemental@fr), (génie environnemental@fr : ingénierie@fr), (ingénierie@fr : secteur économique@fr), (secteur économique@fr : système@fr), (système@fr : entité composée@fr), (entité composée@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
4. label Gemet non trouvé <-> gestion de l'eau@fr : 0.7167595624923706  
[(gestion de l'eau@fr : gestion des ressources naturelles@fr), (gestion des ressources naturelles@fr : maîtrise@fr), (maîtrise@fr : aptitude@fr), (aptitude@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
5. bâtiment public@fr <-> structure construite@fr : 0.6913622617721558  
[(équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
6. santé publique et assainissement@fr <-> assainissement@fr : 0.6888487935066223  
[(assainissement@fr : activité@fr), (activité@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (ensemble architectural@fr : groupe de constructions et de bâtiments@fr), (complexe immobilier@fr : ensemble architectural@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
7. label Gemet non trouvé <-> saliculture@fr : 0.6875752210617065  
[(saliculture@fr : production@fr), (production@fr : processus@fr), (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de

**Commentaire [NP30]:** Devrait popper en premier

**Commentaire [NP31]:** Limite de pertinence

bâtiments@fr : groupe d'entités géographiques@fr), (groupe de constructions et de bâtiments@fr : structure construite@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]

8. label Gemet non trouvé <-> station d'observation@fr : 0.6836385726928711  
[(station d'observation@fr : équipement@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
9. label Gemet non trouvé <-> source d'énergie@fr : 0.6789122819900513  
[(source d'énergie@fr : source@fr), (source@fr : objet@fr), (objet@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
10. urbanisme@fr <-> aménagement urbain@fr : 0.6710588335990906  
[(aménagement urbain@fr : urbanisme@fr), (urbanisme@fr : géographie humaine@fr), (géographie humaine@fr : société@fr), (société@fr : groupe d'humains@fr), (organisation@fr : groupe d'humains@fr), (firme@fr : organisation@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
11. polluant@fr <-> polluant@fr : 0.6678998470306396  
[(polluant@fr : substance physique@fr), (substance physique@fr : entité physique@fr), (objet tangible@fr : entité physique@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
12. captage@fr <-> zone de chalandise@fr : 0.664354681968689  
[(zone de chalandise@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
13. label Gemet non trouvé <-> masse d'eau@fr : 0.6608922481536865  
[(masse d'eau@fr : étendue d'eau@fr), (étendue d'eau@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
14. hydrogéologie@fr <-> hydrogéologie@fr : 0.6581923961639404  
[(hydrogéologie@fr : hydrologie@fr), (hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
15. hydrologie@fr <-> hydrologie@fr : 0.655835747718811  
[(hydrologie@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : classe@fr), (forme juridique@fr : classe@fr), (personne morale@fr : forme juridique@fr), (firme@fr : personne morale@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
16. label Gemet non trouvé <-> Gestion des ressources naturelles@fr : 0.653746485710144  
[(Gestion des ressources naturelles@fr : maltraitance@fr), (maltraitance@fr : problème sociétal@fr), (problème sociétal@fr : problème@fr), (problème@fr : condition@fr), (condition@fr : état@fr), (état@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
17. chaleur@fr <-> chaleur@fr : 0.6528462171554565  
[(chaleur@fr : grandeur physique@fr), (grandeur physique@fr : quantité@fr), (quantité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]

18. développement urbain@fr <-> urbanisme@fr : 0.6490142345428467  
 [(urbanisme@fr : géographie humaine@fr), (géographie humaine@fr : société@fr), (société@fr : groupe d'humains@fr), (organisation@fr : groupe d'humains@fr), (firme@fr : organisation@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
19. géologie@fr <-> géologie@fr : 0.6452451944351196  
 [(géologie@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : classe@fr), (forme juridique@fr : classe@fr), (personne morale@fr : forme juridique@fr), (firme@fr : personne morale@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
20. pollution industrielle@fr <-> pollution@fr : 0.6422240734100342  
 [(pollution@fr : effet pervers@fr), (effet pervers@fr : résultat@fr), (résultat@fr : entité observable@fr), (entité observable@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
21. label Gemet non trouvé <-> infrastructure de transport@fr : 0.64194655418396  
 [(infrastructure de transport@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
22. label Gemet non trouvé <-> bassin hydrographique@fr : 0.6412675380706787  
 Chemin non trouvé
23. qualité de l'eau@fr <-> qualité de l'eau@fr : 0.6363604664802551  
 [(qualité de l'eau@fr : qualité@fr), (qualité@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
24. label Gemet non trouvé <-> bien culturel@fr : 0.6321249604225159  
 [(bien culturel@fr : objet concret@fr), (objet tangible@fr : objet concret@fr), (structure architecturale@fr : objet tangible@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
25. label Gemet non trouvé <-> label Wikidata introuvable : 0.6305616497993469  
 Chemin non trouvé
26. label Gemet non trouvé <-> nappe phréatique@fr : 0.6300790905952454  
 [(nappe phréatique@fr : aquifère@fr), (aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : réservoir@fr), (réservoir@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
27. hydrographie@fr <-> hydrographie@fr : 0.6294116973876953  
 [(hydrographie@fr : sciences de la Terre@fr), (sciences de la Terre@fr : science de la nature@fr), (science de la nature@fr : science@fr), (science@fr : système de connaissances@fr), (système de connaissances@fr : ensemble@fr), (ensemble@fr : entité composée@fr), (structure@fr : entité composée@fr), (structure architecturale@fr : structure@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
28. infrastructure@fr <-> infrastructure@fr : 0.628700852394104  
 [(infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
29. label Gemet non trouvé <-> submersion marine@fr : 0.6275973320007324  
 [(submersion marine@fr : inondation@fr), (inondation@fr : catastrophe naturelle@fr), (catastrophe naturelle@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : phénomène@fr)]

**Commentaire [NP32]:** Pourrait apparaître plus haut

**Commentaire [NP33]:** Devrait évidemment être beaucoup plus eau, dans les premières

- entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
30. biogaz@fr <-> biogaz@fr : 0.6261656284332275  
 [(biogaz@fr : combustible gazeux@fr), (combustible gazeux@fr : gaz@fr), (gaz@fr : fluide@fr), (fluide@fr : substance@fr), (substance@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
31. logement@fr <-> gestion de logement@fr : 0.6251828670501709  
 Chemin non trouvé
32. aquifère@fr <-> aquifère@fr : 0.6246013641357422  
 [(aquifère@fr : réservoir d'eau@fr), (réservoir d'eau@fr : réservoir@fr), (réservoir@fr : structure architecturale@fr), (structure architecturale@fr : complexe immobilier@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
33. communication en matière de risques@fr <-> gestion du risque@fr :  
 0.6228855848312378  
 [(gestion du risque@fr : fonction@fr), (fonction@fr : usage@fr), (usage@fr : propriété@fr), (propriété@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
34. monument@fr <-> monument de Hongrie@fr : 0.6182126998901367  
 [(monument de Hongrie@fr : bien patrimonial national@fr), (bien patrimonial national@fr : site patrimonial@fr), (site patrimonial@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (objet géographique@fr : équipement@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
35. label Gemet non trouvé <-> label Wikidata introuvable : 0.6179236173629761  
 Chemin non trouvé
36. label Gemet non trouvé <-> bassin versant@fr : 0.6176934242248535  
 [(bassin versant@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
37. habitat@fr <-> habitat@fr : 0.6176526546478271  
 [(habitat@fr : zone géographique@fr), (zone géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
38. ressource naturelle@fr <-> ressource naturelle@fr : 0.6150418519973755  
 [(ressource naturelle@fr : matériau brut@fr), (matériau brut@fr : bien@fr), (produit@fr : bien@fr), (œuvre@fr : produit@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
39. risque naturel@fr <-> risque naturel@fr : 0.6141432523727417  
 [(risque naturel@fr : danger@fr), (danger@fr : source de risque@fr), (source de risque@fr : source@fr), (source@fr : objet@fr), (objet@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
40. érosion@fr <-> érosion@fr : 0.6131541728973389  
 [(érosion@fr : phénomène naturel@fr), (phénomène naturel@fr : phénomène@fr), (phénomène@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
41. polluant@fr <-> label Wikidata introuvable : 0.6121931076049805  
 Chemin non trouvé

Commentaire [NP34]: Sérieux ???

42. label Gemet non trouvé <-> réseau de télécommunications@fr :  
0.6115759015083313  
[[réseau de télécommunications@fr : infrastructure@fr], (infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
43. surveillance@fr <-> surveillance@fr : 0.6091445684432983  
[[surveillance@fr : activité@fr], (Q18390984 : activité@fr), (Q18390984 : outil de gestion du risque@fr), (outil de gestion du risque@fr : outil de gestion@fr), (outil de gestion@fr : outil@fr), (outil@fr : équipement@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
44. estuaire@fr <-> estuaire@fr : 0.6066519021987915  
[[estuaire@fr : étendue d'eau@fr], (étendue d'eau@fr : objet physico-géographique@fr), (objet physico-géographique@fr : objet géographique@fr), (structure construite@fr : objet géographique@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
45. dommage@fr <-> dégât@fr : 0.6065858602523804  
[[dégât@fr : processus@fr], (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (ensemble architectural@fr : groupe de constructions et de bâtiments@fr), (complexe immobilier@fr : ensemble architectural@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
46. logement@fr <-> appartement@fr : 0.6065312623977661  
[[appartement@fr : logement@fr], (logement@fr : structure architecturale@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
47. label Gemet non trouvé <-> transition écologique@fr : 0.6048821210861206  
[[transition écologique@fr : changement@fr], (changement@fr : occurrence@fr), (occurrence@fr : entité temporelle@fr), (entité temporelle@fr : entité spatio-temporelle@fr), (entité spatio-temporelle@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]
48. chauffage à distance@fr <-> réseau de chaleur@fr : 0.603854775428772  
[[réseau de chaleur@fr : infrastructure@fr], (infrastructure@fr : œuvre@fr), (structure architecturale@fr : œuvre@fr), (complexe immobilier@fr : structure architecturale@fr), (usine@fr : complexe immobilier@fr), (station de traitement des eaux usées@fr : usine@fr)]
49. zonage@fr <-> zonage@fr : 0.6028785705566406  
[[zonage@fr : planification@fr], (planification@fr : processus@fr), (processus@fr : série@fr), (série@fr : groupe@fr), (groupe d'entités géographiques@fr : groupe@fr), (groupe de constructions et de bâtiments@fr : groupe d'entités géographiques@fr), (groupe de constructions et de bâtiments@fr : structure construite@fr), (équipement@fr : structure construite@fr), (usine@fr : équipement@fr), (station de traitement des eaux usées@fr : usine@fr)]
50. label Gemet non trouvé <-> unité statistique@fr : 0.6022796630859375  
[[unité statistique@fr : unité@fr], (unité@fr : objet abstrait@fr), (objet abstrait@fr : entité non-physique@fr), (entité non-physique@fr : entité@fr), (entité économique@fr : entité@fr), (firme@fr : entité économique@fr), (usine@fr : firme@fr), (station de traitement des eaux usées@fr : usine@fr)]