International Cartographic Association
Association Cartographique Internationale

Making Marine and Freshwater Data FAIR: Refining Geospatial Search Requests Using Ontologies and Thesauri

Pekka Latvala [a,*], Markus Konkol [b], Juha Oksanen [a], Henning Sten Hansen [c]

[a] *Finnish Geospatial Research Institute FGI, National Land Survey of Finland, pekka.latvala@nls.fi, juha.oksanen@nls.fi*
[b] *52°North Spatial Information Research GmbH, m.konkol@52north.org*
[c] *Aalborg University, hsh@plan.aau.dk*

* Corresponding author

Keywords: Metadata Search, Thesaurus, GEMET, SPARQL, OGC API Records, Ontology

Abstract:

A search request usually returns only those results that have metadata including the search term. However, there might be relevant items that are not shown to the user although they have a semantically related term in the metadata. The metadata may be described with terms that are synonyms to the query, or they may contain other related terms. We propose a solution to this problem by presenting a list of related term suggestions that are retrieved from thesaurus data.

The AquaINFRA is a European Open Science Cloud (EOSC 2022) project that develops a research data infrastructure for supporting actions relating to freshwater and marine research. The infrastructure contains an AquaINFRA Data Space that includes data and services that are made available following the FAIR (Findable, Accessible, Interoperable, Reusable) principles. In particular, the project is developing a Data Discovery and Access Service (DDAS) that provides a federated metadata search and access via the OGC API – Records, Features, and Coverages interfaces to various background services (e.g., EMODnet). In addition, the DDAS includes the AquaINFRA Data Lake, which is used for storing certain datasets locally. The central gateway to find and access aquatic digital resources is the AquaINFRA Interaction Platform (AquaINFRA 2025) that acts as a user interface for the DDAS.

The work was started by investigating several ontologies and thesauri to find a suitable one for the purpose. The GEMET (GEneral Multilingual Environmental Thesaurus) contains environmental terminology that is written with up to 37 different languages (GEMET 2025). The GEMET data consists of concepts (Figure 1) that each can have links to other concepts that are modelled with *broader*, *narrower* or *related* relations. In our implementation, the GEMET data was downloaded in the RDF (Resource Description Framework) format and inserted to the Apache Jena Fuseki SPARQL Server application.

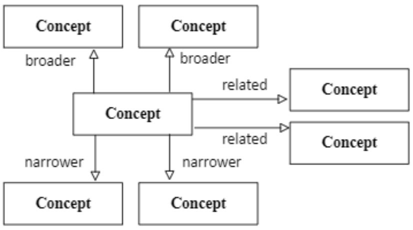
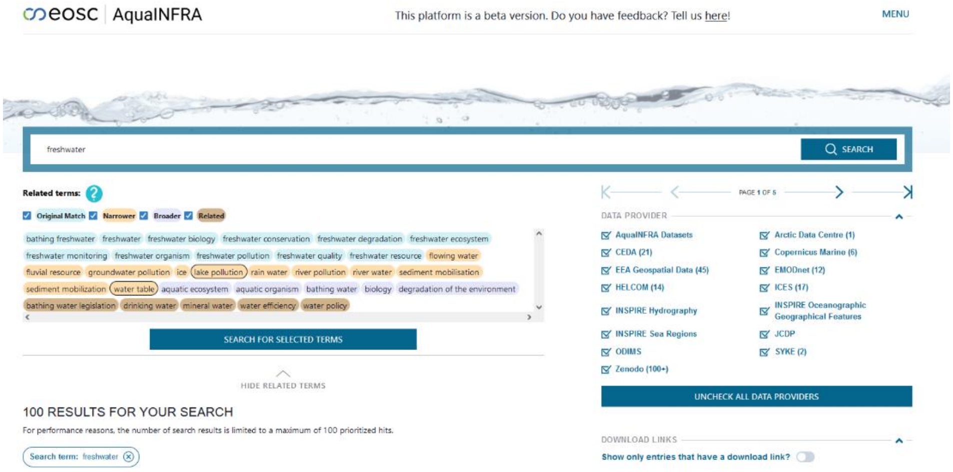


Figure 1: Structure of the GEMET data. Each concept can have links to broader, narrower and related concepts.

A Python-based web service was created for retrieving the related terms information from the thesaurus data. The web service formulates SPARQL queries based on the user's search term and sends them to the Fuseki server. The web service contains parameters for including the broader, narrower, and related terms to the results. The language support is limited to the English language. The web service's response is returned in JSON format and includes results divided into the following four categories: The o*riginalMatch* category contains terms that match directly with the query word (e.g., watercourse). The b*roader* category contains the broader terms of the o*riginalMatch* results (e.g., hydrosphere). This category is limited only to the first-level parents because the broader concepts quickly become too broad to be relevant to the original query term. The *narrower* category contains all narrower terms of the o*riginalMatch* results (e.g., river). The r*elated* category includes all related concepts of the *originalMatch* results (e.g., hydrographic basin), together with their narrower terms if the related and the narrower search parameters are used at the same time in the query.

Abstracts of the International Cartographic Association, 10, 165, 2025
32nd International Cartographic Conference (ICC 2025), 17–22 August 2025, Vancouver, Canada.
https://doi.org/10.5194/ica-abs-10-165-2025 © Author(s) 2025. CC BY 4.0 License.

The GEMET English language data consists mainly of the base forms of the terms with certain exceptions. To support the inflectional inputs, such as plurals, the web service performs morphological analysis to the input by using the NLTK (Natural Language Toolkit) Python library's WordNetLemmatizer functionality. Inputs that consist of two words are handled by splitting them into separate words and lemmatizing only the last one. After that, the lemmatized word is re-combined with the other one. In the SPARQL queries, the user's original query and the lemmatized one are used together to also match the inflectional terms in the GEMET data.



This platform is a beta version. Do you have feedback? Tell us here!

freshwater

Related terms: ?

☑ Original Match ☑ Narrower ☑ Broader ☑ Related

bathing freshwater | freshwater | freshwater biology | freshwater conservation | freshwater degradation | freshwater ecosystem | freshwater monitoring | freshwater organism | freshwater pollution | freshwater quality | freshwater resource | flowing water | fluvial resource | groundwater pollution | ice | lake pollution | rain water | river pollution | river water | sediment mobilisation | sediment mobilization | water table | aquatic ecosystem | aquatic organism | bathing water | biology | degradation of the environment | bathing water legislation | drinking water | mineral water | water efficiency | water policy

100 RESULTS FOR YOUR SEARCH

For performance reasons, the number of search results is limited to a maximum of 100 prioritized hits.

Search term: freshwater ⊗

DATA PROVIDER

☑ AquaINFRA Datasets
☑ Arctic Data Centre (1)
☑ CEDA (21)
☑ Copernicus Marine (6)
☑ EEA Geospatial Data (45)
☑ EMODnet (12)
☑ HELCOM (14)
☑ ICES (17)
☑ INSPIRE Hydrography
☑ INSPIRE Oceanographic Geographical Features
☑ INSPIRE Sea Regions
☑ JCDP
☑ ODIMS
☑ SYKE (2)
☑ Zenodo (100+)

DOWNLOAD LINKS

Show only entries that have a download link? ○

Figure 2: Related Term Suggestions in the AquaINFRA Interaction Platform.

The web service was integrated with the AquaINFRA Interaction Platform (Figure 2). The related terms search is executed at the same time as the metadata search. The results are visualized as a list of terms where the categories are presented with different colours. The terms can be selected individually from the list and added to the metadata search. In addition, users can click on the checkboxes to include or exclude related terms categories. The added terms are separated by commas in the metadata search query field. This creates an OR filter query to the DDAS that returns results that match any of the query terms. If an individual term consists of two words, the words are separated by the space character. This creates an AND filter query where the results must match both words.

The presented approach could be extended to support other languages than English by including a language parameter to the web service. This would also require performing the morphological analysis in the supported languages. The completeness of the GEMET data varies between the languages. Therefore, the extended language support could require also enhancing the GEMET contents.

Acknowledgements

The AquaINFRA project has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No 101094434.

The GEMET data is available under the Creative Commons Attribution 4.0 International (CC BY 4.0) license: https://creativecommons.org/licenses/by/4.0/

The authors wish to acknowledge CSC – IT Center for Science, Finland, and the Open Geospatial Information Infrastructure for Research (Geoportti, urn:nbn:fi:research-infras-2016072513) for computational resources.

References

AquaINFRA, 2025, AquaINFRA Interaction Platform, Available at: https://aquainfra.dev.52north.org/ (Accessed: 2025-05-14)

EOSC, 2022. Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC). https://doi.org/10.2777/935288

GEMET, 2025. About GEMET - GEneral Multilingual Environmental Thesaurus. Available at: https://www.eionet.europa.eu/gemet/en/about/ (Accessed: 2025-05-14)