Establishing the European Geological Surveys Research Area to deliver a Geological Service for Europe

Deliverable 4.4

GeoERA final report on Keyword Thesaurus and Project Vocabularies – experiences and status of the work

Authors and affiliation:
M. Schiegl [GBA]
Ch. Hörfarter [GBA]
M. Novak [GEO-ZS]
L. Kondrová [CGS]
R. Hernández, M. Sanabria Pabón [IGME]
R. van Ede [TNO]

E-mail of lead author:
martin.schiegl@geologie.ac.at

Version: 23-09-2021

This report is part of a project that has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement number 731166.

Deliverable Data		
Deliverable number	D4.4	
Dissemination level	Public	
Deliverable name	GeoERA final report on Keyword Thesaurus and Project Vocabularies – experiences and status of the work	
Work package	WP4, Semantic Harmonisation Issues	
Lead WP/Deliverable beneficiary	GBA	
Deliverable status		
Submitted (Author(s))	23/09/2021	M. Schiegl, C. Hörfarter
Reviewed	23/09/2021	M. Schiegl, C. Hörfarter
Approved (Coordinator)	27/09/2021	Jorgen Tulstrup

Abstract

The present report presents the results of work package 4 (WP4) “Semantic Harmonisation Issues” as part of the GeoERA GIP project (GIP-P)
The main goals of the framework were (1) creating a geoscientific keyword thesaurus, (2) the processing of terminological project knowledge as ‘knowledge representation’ and publishing both as RDF Linked Data. These goals were essentially achieved. The currently available geoscientific GeoERA Keyword Thesaurus 2.1 already provides more than 2500 keywords that can be used by the GeoERA projects to index their datasets and documents. The present report also describes the status quo of the project vocabulary task. There were 15 vocabularies created for 6 different GeoERA projects. Some project vocabularies are still being processed and web applications for the presentation of the use cases have not yet been developed or are in the test phase. Anyway, the existing data from the different GeoERA projects on a project vocabulary can now be used or reused for subsequent projects, e.g. the fault database. Taxonomies, classifications or categorizations that have already been started can be proposed in whole or in part for editing code lists or for extending future standards such as INSPIRE or GeoSciML. Overall, the results of this WP should provide a good basis for the entry of the geosciences into Linked Data and Semantic Web over the next few years.

Please note:

The present report aims more at a technical level for readers familiar with technologies and principles of Semantic Web and Linked Data. This refers especially to the content about “URI design” and the “documentation of vocabularies” actually created during the GeoERA project. The information compiled in this report is to be understood as a supplement to the previous reports D4.2 and D4.3 on GIP-P work package 4.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

The GeoERA IP WP4 “Semantic harmonisation issues” is intended to support two main use cases for GeoERA projects:

- T4.1 – A multilingual semantic text search through the establishment of a GeoERA Keyword Thesaurus
- T4.2 – Establishment of GeoERA project vocabularies

Both aim to ensure the interoperability of GeoERA project results and to make the resulting datasets and documents searchable and accessible. Thus, the implementation of both use cases takes into account the fundamental recommendations of the **FAIR data initiative**, which are aimed at the findability, accessibility, interoperability and re-usability of data.

The GeoERA project will provide many data sets and documents based on different data structures and different contents. In order to achieve the goals of this work package, to enable research and querying of the project results in the form of data sets and documents, we strive for the approach of "semantic harmonisation". Semantic harmonisation stands for "making databases and their attribute data consistent and compatible with regard to their meaning in language and logic". Through the use of semantic modeling and semantic web technology, we provide a platform that enables
1.) to structure the relevant multilingual processed search terms as the basis for a search system (keyword thesaurus) and
2.) each project to describe its data content (knowledge) in the form of controlled vocabularies[4] (the project vocabularies).

As part of these two areas of activity, it is now possible to clarify the meaning of a scientific concept within a project, avoid ambiguous interpretations and to show how the term should be linked and used within a project. By implementing the SKOS[1]/RDF[2] web standard principles and assigning each concept to a URI[3], we create the conditions for linked data. Consequently, the harmonisation process is supported by the now possible linking of information between different content interpretations, which is a prerequisite for making the data and data products sustainably searchable and queryable.

In the context of the T4.1 Keyword Thesaurus task, the main work consisted in the search and evaluation of significant geoscientific keywords, their compilation and modeling as a basis for an optimized semantic search system.

As for Task T4.2 “Project Vocabularies”, the main work consisted in initiating, designing and testing the development of project vocabulary data. Additionally, the suggestions for a technical infrastructure for sustainable data storage, the organization of governance and maintenance of the vocabularies have been part of this task.

While the previous mid-term reports D4.2 and D4.3 of GIP-P WP4 focused more on the individual work steps for the tasks T4.1 and T4.2, this report is intended here to provide an overview of the topics covered and the workflow within WP 4, as well as to reflect the project status and results at the time of report preparation.

[1] SKOS - Simple Knowledge Organization System - http://www.w3.org/TR/skos-reference
[2] RDF - Resource Description Framework - https://www.w3.org/TR/rdf11-primer/
[3] URI - Uniform Resource Identifier

2 GEOERA KEYWORD THESAURUS

(Participants: GBA, IGME, ISPRA, SGU, TNO, CGS, GIU, GeoZS, MBFSZ, LfU, BGRM, GTK, GEUS, BGR, HGI-CGS, LNEG)

2.1 Description

Search for data is a basic task for all data infrastructures. Here WP4 strives for a support of an optimized search and query of GeoERA data results through a keyword thesaurus. Thus, the GeoERA Keyword Thesaurus shall enable the metadata tagging of project dataset results and it builds the basis for an optimized multilingual semantic search system. To achieve this, it is needed to put all project-significant geoscientific keywords which are used to tag datasets into a hierarchy like a thesaurus.

2.1.1 General information on 'keywords'

In order to better understand what the GeoERA Keyword Thesaurus is about and which different steps were necessary within the GeoERA Keyword Thesaurus workflow, here is some general information:

- **What is a keyword and how do we define 'keyword' in this project?**

In a simple way the term 'keyword' most commonly refers to a word, concept or phrase of great significance that people use when they search for something.
Therefore, a keyword is used to tag documents or is used to find information in the web via a search system. However, the term "keyword" can also be seen as broader term for narrower concepts such as index term, subject header, descriptor, code word, and further more.
It depends on the point of view whether to use "keyword" in a linguistic context (a word that occurs in a text more often than expected), a rhetorical context (a word to reveal the main topic of something, e.g. of a presentation), in programming (a code word that is reserved by a program because the word has a special meaning) or within general information retrieval in the web (e.g. Google search engine).

In this project we use the keywords concept as "index terms" which shall be used for tagging and searching datasets. The keywords are focused on and grouped according to geoscientific topics within a thesaurus management system. This system, in which keywords are structured and modeled in relation to a particular topic is also referred to as **"subject heading system"**.

- **Why are keywords important?**

Keywords are very important for search engine optimization (SEO). Search engines aim to provide the user with the most relevant search results for his search query. For simple search engines, keywords provide the basis for e.g. stored data records to be found and retrieved. A simple search engine works much like a library catalog. There, users can enter the names of creators, editors or authors on the one hand and enter subject areas or epochs on the other. Such search categories are created as keywords so that these entries can be found.

Within GeoERA, the keywords are very important, as these words and phrases are used by the projects to identify the data sets and documents they have delivered using the metadata. Hence, keyword tagging is used to provide significant record subject information when users search for matching terms through a search engine.

- **The benefits of a keyword thesaurus**

"A thesaurus serves to guide both an indexer and a searcher in selecting the same preferred term or combination of preferred terms to represent a given subject" (https://en.wikipedia.org/wiki/Thesaurus_(information_retrieval)).

In the context of information retrieval, a thesaurus can be used to manage keywords in a form of "controlled vocabulary"[4] (as opposed to a simple keyword list) that helps index appropriate metadata for information-bearing entities (records, documents, ...). The keyword thesaurus helps to express a concept in a prescribed way to improve accuracy and retrievability. The terms for the keyword thesaurus can usually be arranged hierarchically and grouped by topics (e.g. GeoERA search categories). The hierarchy of terms within the subject thesaurus helps the indexer to narrow down the terms and thus limit semantic ambiguities. The semantically structured terms within the keyword thesaurus are easier to find due to the uniformity of presentation (e.g. spelling) and the assignment of Uniform Resource Identifiers (URIs). Furthermore, the keyword thesaurus makes it possible to manage multilingualism and to revise synonyms, which - if desired - can also be searched.

2.1.2 What is the GeoERA Keyword Thesaurus about

- The GeoERA Keyword Thesaurus is a collection of more than 2500 geoscientific index terms,
  - grouped into 16 geoscientific search categories
  - predominantly labeled in English
  - with translations into different languages (30% of the index terms are available in 22 non-English languages)
  - with unique web addresses (URIs)
  - with links to standardized codelists from INSPIRE and GeoSciML
  - including links to sources used such as GEMET thesaurus and GBA thesaurus
  - including relevant index terms from KINDRA[5] and VOGERA[6] and the OneGeology keyword lists
  - thesaurus-like modeled (SKOS/RDF) in accordance with search use cases and metadata tagging in MICKA.
  - supplied in RDF format
- The two main use cases of the GeoERA Keyword Thesaurus are:
  - The assignment of keywords to each digital product produced within GeoERA
  - The search for GeoERA project results within the GeoERA (EGDI) metadata catalogue via a semantic search system. A test system to get a bit of a feeling, what a search system could look like is available under https://schmar00.github.io/semantic-search/.
- The GeoERA Keyword Thesaurus is provided in versioned updates after improvement steps of the basic version have been carried out. The current version is 2.1. For more details on the various release notes, see also the corresponding GitHub project at: https://github.com/GeoEra-GIP/WP4-Semantics/tree/master/Keyword%20Thesaurus.

[4] Controlled Vocabulary - https://en.wikipedia.org/wiki/Controlled_vocabulary

[5] KINDRA - Knowledge Inventory for hydrogeology research - www.kindraproject.eu

[6] VOGERA - Vulnerability of Shallow Groundwater Resources to Deep Subsurface Energy-Related Activities - https://geoera.eu/projects/vogera1/

- GeoERA Keyword Thesaurus is finalized and online available within the European Geoscience Registry (currently hosted by BRGM), which shows the GeoERA Keyword Thesaurus v2.1 at: https://data.geoscience.earth/ncl/geoera/keyword. By clicking on this interface it is possible to navigate to each individual keyword. But there is also a link to download a list of all available keywords in different formats. At the top right of the page you will find download links for TTL, RDF/XML and JSON-LD. If you open the SPARQL query form at: https://data.geoscience.earth/ncl/ui/sparql-form it is possible to run a query script to display the results as a table, TXT-, JSON- or XML-file.

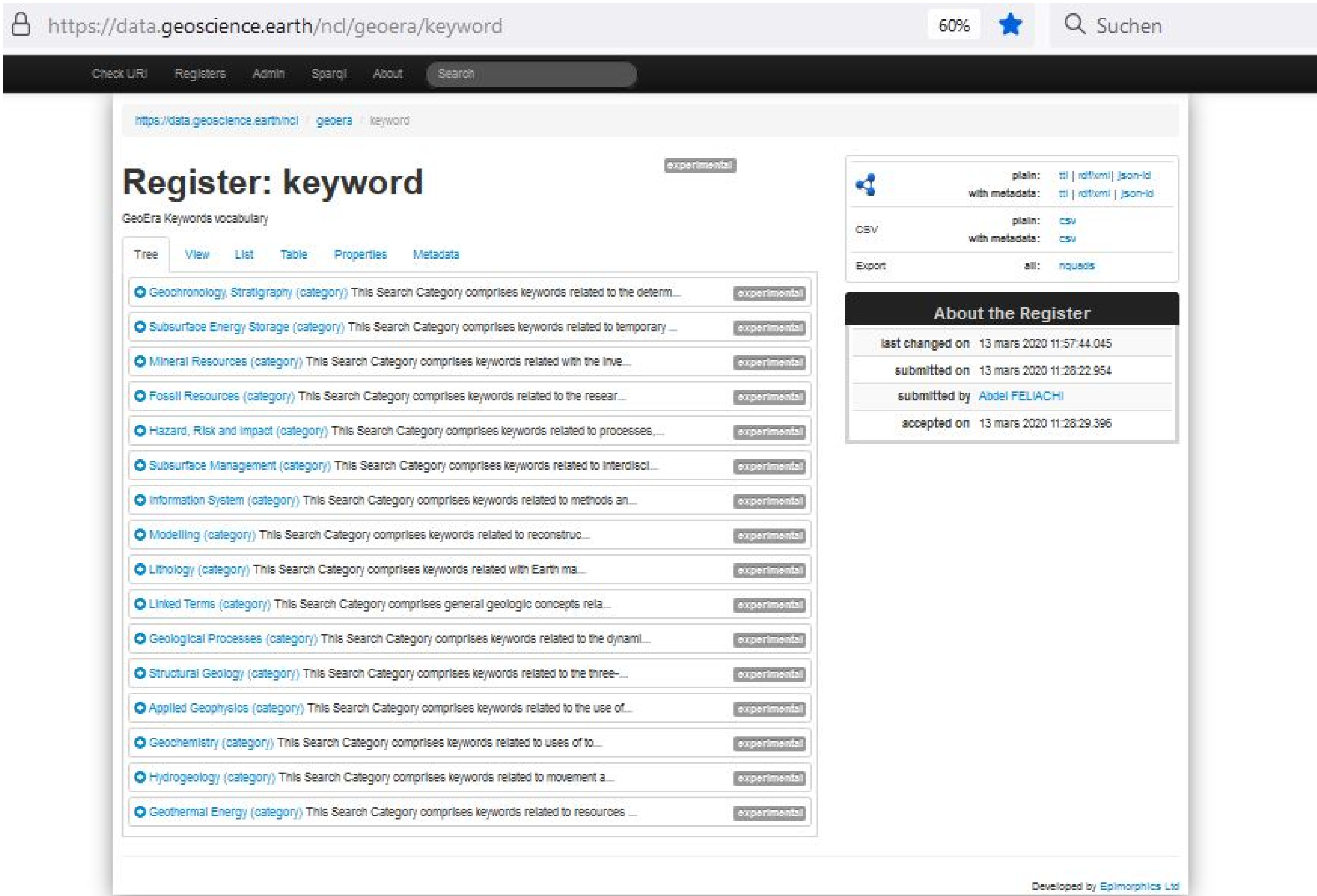

Figure 2-1 *Screenshot of the GeoERA Keyword Thesaurus application on the European Geoscience Registry web page https://data.geoscience.earth/ncl/geoera/keyword*

2.2 Workflows and results

In order to achieve the benefits described above, the task T4.1 has been subdivided into the following described 3 subtasks:

T4.1.1 – Evaluation of existing vocabularies applicable for a GeoERA Keyword Thesaurus (led by IGME)

- The first step was to identify the generic geoscientific areas or topics that form the content framework for the products of the various GeoERA projects. To define these generic topics, the abstracts and proceedings of the 14 GeoERA projects were analysed. It was agreed that these evaluated generic topics/domains would be called "search categories" Their definitions were drafted and refined in collaboration with the project partners. These search categories were and are used to group the different keywords analysed.

SEARCH CATEGORIES	DESCRIPTION
GEOCHRONOLOGY/STRATIGRAPHY	Related to the determination of relative and absolute ages of rocks, fossils, sediments and time sequences of events in the earth's history. The following disciplines are considered: stratigraphy, geological history, geological time scale, Chronostratigraphy...
GEOTHERMAL ENERGY	Principal topics related to resources, conflicts and management of geothermal energy.
GEOLOGICAL PROCESSES	Related to dynamic geological processes that act on land forms and surfaces and within the Earth. The following disciplines are considered: sedimentation, diagenesis, metamorphism, geomorphology ....
STRUCTURAL GEOLOGY	The study of the three-dimensional distribution of rock units, trying to reconstruct the movements and processes that have originated its structure, the history of movements and deformations on a global and regional scale. The following disciplines are considered: Tectonics, geologic structures.
APPLIED GEOPHYSICS	The use of geophysical techniques mainly for the exploration and exploitation of natural resources; but also applied to geological risks, environment and hydrogeology.
MODELLING	Reconstructing the geological environment and geological processes, using numerical, geostatistical and simulation techniques. The following disciplines are considered: 3D modelling, flow modelling, geochemical modelling, etc.
MINERAL RESOURCES	It focuses on the investigation and exploitation of the type of mineral resource that has economic value as a raw material.
FOSSIL RESOURCES	Research and exploitation of fossil fuels (Coal and Hydrocarbons)
INFORMATION SYSTEM	Methods and uses for managing of geoscience information systems.
LITHOLOGY	Part of the geology that deals with the study of rocks and the physical and chemical characteristics of the rocks that appear constituting a certain geological unit.
HAZARD, RISK AND IMPACT	Processes, events of natural or induced origin, including surface and subsurface activities, that can cause damage or loss of property and life in the surface and subsurface. The hazards of the following types are considered: seismicity, ground movements (e.g. surface deformation, Landslide, etc.), leakage and migration and facility hazards, climate change, pollution,....
HYDROGEOLOGY	Part of geology that studies the movement and distributions of surface and groundwater, as well as its research, prospecting, catchment and protection.
SUBSURFACE MANAGEMENT	Geoscientific contributions related to an interdisciplinary approach, which analyses, develops and manages the processes of planning and development of the subsurface, according to their environmental, economic and societal situation.
GEOCHEMISTRY	Uses of tools and principles of chemistry to study the composition and dynamics of the chemical elements in the earth. The following disciplines are considered: Geochemistry, Hydrogeochemistry, Lithogeochemistry, Organic geochemistry, etc.
SUBSURFACE ENERGY STORAGE	Temporary subsurface storage of energy (mechanical and thermal energy) for the purpose of a later reuse. This topic includes research fields dealing with exploration, testing, managing and monitoring of subsurface storage. The term subsurface storage includes geological storage (e.g. aquifer, hydrocarbon reservoir) as well as engineered subsurface storage (e.g. cavern storage, borehole thermal energy storage).

Figure 2-2 *Search categories developed as a result of the keyword thesaurus vocabulary evaluation process (T4.1)*

- The next step was to analyze existing geoscience vocabularies and code lists that could be integrated into a GeoERA Keyword Thesaurus. It would not have been very effective to first collect all the different terms individually and then model them from scratch. Therefore, an evaluation of existing vocabularies and codelists that would fit into the framework of a GeoERA Keyword Thesaurus was performed. A questionnaire has been sent to all the participants of the subtask T4.1.1, including 25 questions to obtain complete information on the existence of usable vocabularies, their domains, scope, formats, availability, granularity, etc. This questionnaire was used to select which existing terminology would be suitable for a new GeoERA/EGDI keyword thesaurus.

- The whole table with the final selection of the existing and used vocabularies and codelists for the GeoERA Keyword Thesaurus is attached in APPENDIX A.

A more detailed description of the whole workflow of T4.1.1 is available in chapter 2 in the GIP-P WP 4 report deliverable D4.2 GeoERA Keyword Thesaurus.

T4.1.2 – Compilation of the keyword thesaurus (led by GeoZS)

- The compilation of the GeoERA Keyword Thesaurus builds on the previously completed evaluation of existing vocabularies suitable for a GeoERA subject heading system. This subtask includes the integration, merging and modeling of the chosen index terms, regarding the selected and tested vocabularies resulting from subtask T4.1.1. This resulted in a selection of content-relevant terms out of these vocabualries and were assigned to the defined search categories.
- In some cases, an entire hierarchy tree was extracted from a vocabulary, which is technically a great advantage of using already existing thesaurus systems, such as the GEMET thesaurus. This made the modeling more efficient, but sometimes also more complicated. Especially when the semantic modeling of a sub-area did not quite match the defined content in terms of GeoERA search categories. In this case, the hierarchy tree was split and the terms remaining in "sub-hierarchies" were subordinated to the corresponding GeoERA search category (see Figure 2-4).
- The compilation was first made in an excel table (see Figure 2-3). This table served as the basis for integration into the GBA's thesaurus management system (see Figure 2-5) and hence for validation according to the SKOS standard. Before this implementation, however, some improvements were still necessary, such as deleting multiplications, merging terms with the same meaning, changing synonyms, handling adjectives, checking keywords with regard to their use for tagging and searching, and further more.
- After importing the keywords into the GBA Thesaurus Management System, the final product of this phase, the Keyword Thesaurus RDF file version 1.0, was sent to the project leaders for review on 14. August 2019.
- Based on the initial RDF file version from August 2019, there have been some more improvements regarding the expansion of index terms, modeling, typos, additional links and more. Therefore, the GeoERA Keyword Thesaurus is provided in **versioned updates** (for more information on this please see also the GeoERA thesaurus GitHub page). The current version available at https://data.geoscience.earth/ncl/georera/keyword is the GeoERA Keyword Thesaurus version 2.1. The version planned to be delivered by the end of the GeoERA project is the GeoERA Keyword Thesaurus version 2.1.

A more detailed description of the whole workflow of T4.1.2 is available in chapter 3 in the GIP-P WP 4 report deliverable D4.2 GeoERA Keyword Thesaurus.

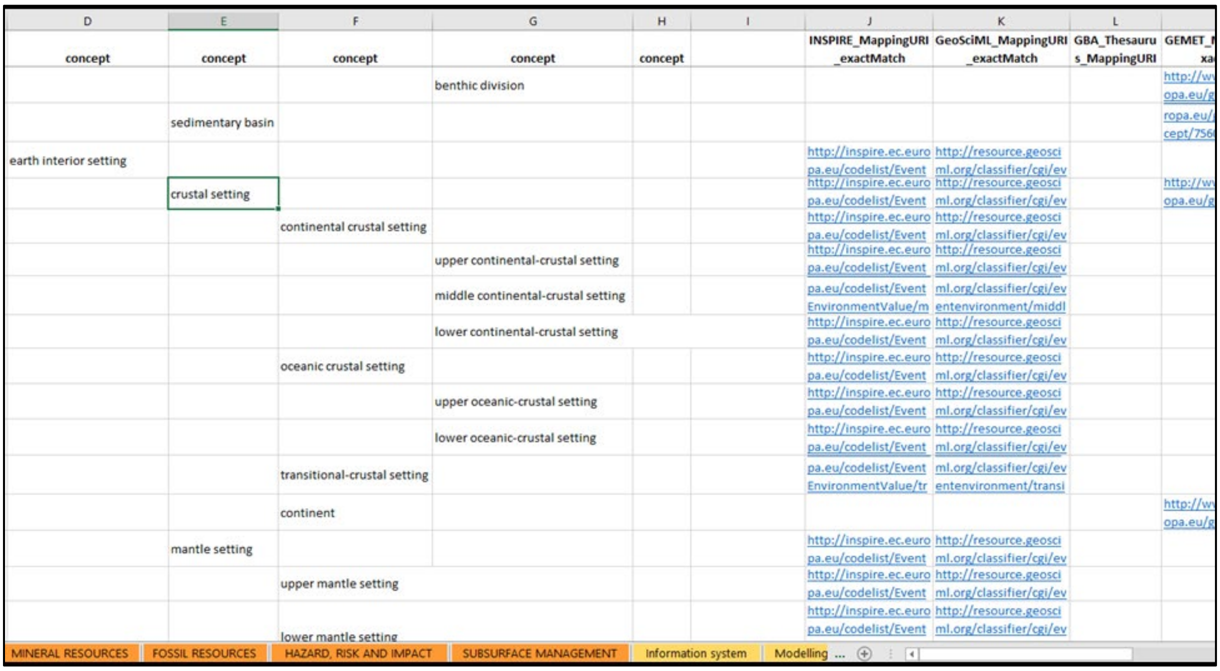

Figure 2-3 *Example of a first compilation of index terms for the GeoERA Keyword Thesaurus in an Excel sheet*

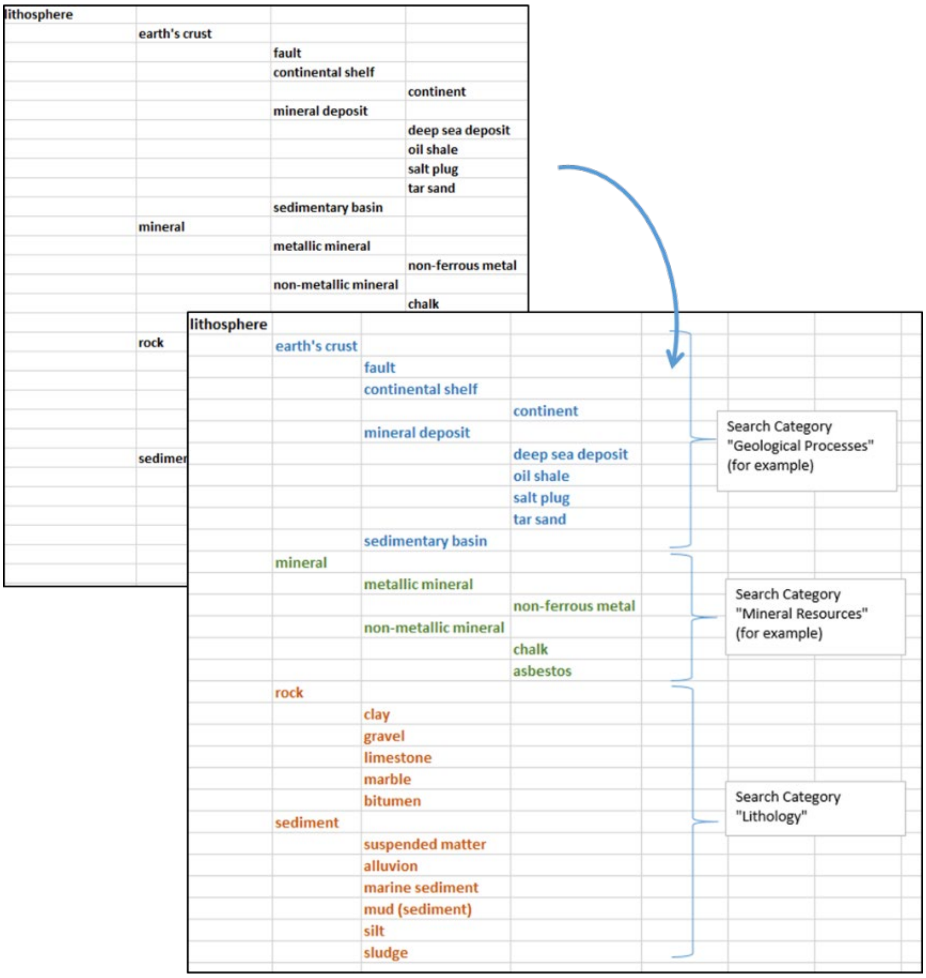

Figure 2-4 *Example of mapping GEMET terms (top left table) to the corresponding GeoERA search categories (shown in the bottom right table).*

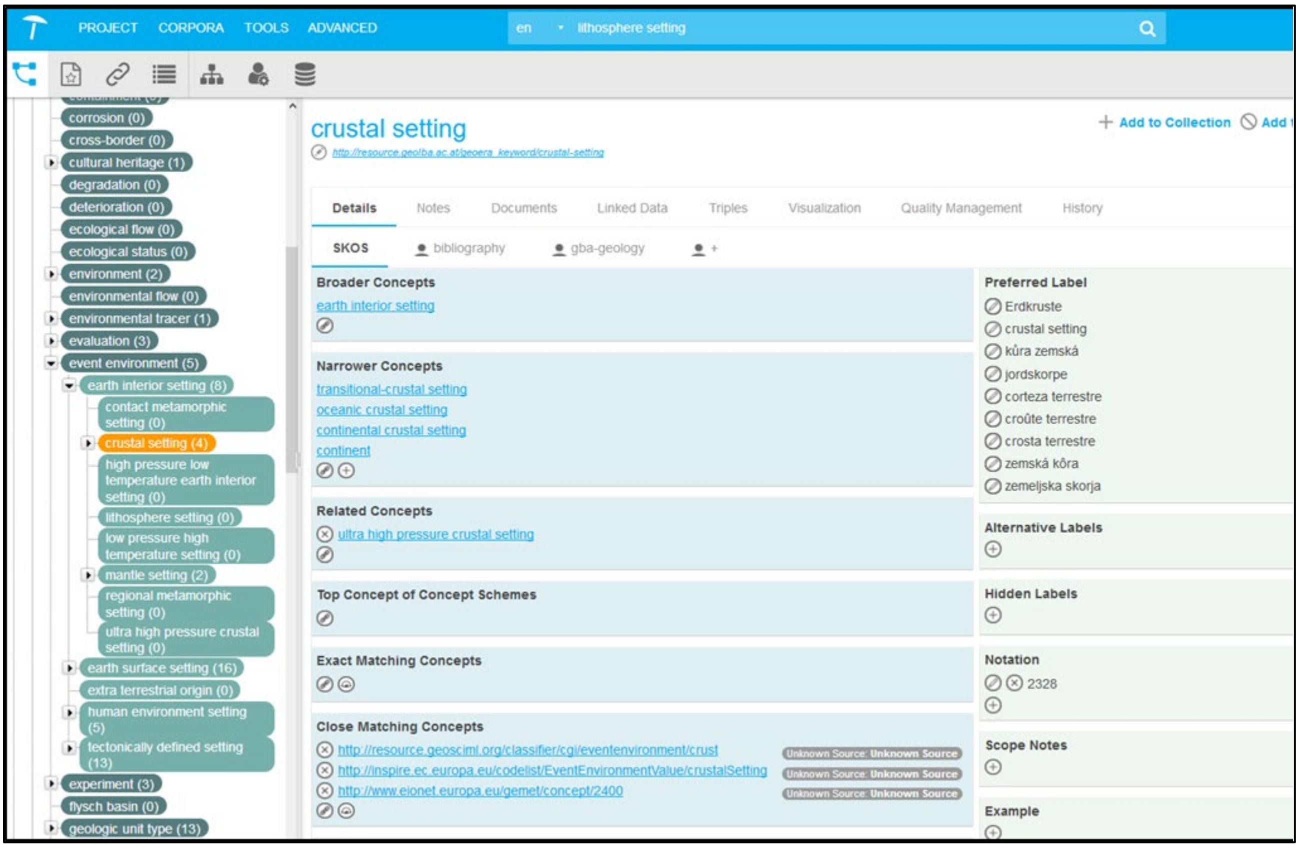

Figure 2-5 *Screenshot of the backend of the Semantic Web Company's PoolParty - the GBA thesaurus management system. This interface provides a hierarchy visualization (on the left) as well as editing of the various SKOS concept properties (multilingual preferred labels, associative properties,...).*

T4.1.3 – Governance plan, future workflows around keyword thesaurus (led by CGS)

- The aim of this subtask was to design a plan for the management of the keyword thesaurus, including workflows for governance and maintenance, in order to create a sustainable, multilingual semantic keyword system for the EGDI/GeoERA platform.

- This subtask includes the development of a workflow for change management, keyword expansion, responsibilities for content revisions such as updates, enhancements and additionally the technical infrastructure and communication during the project and after the project has ended.

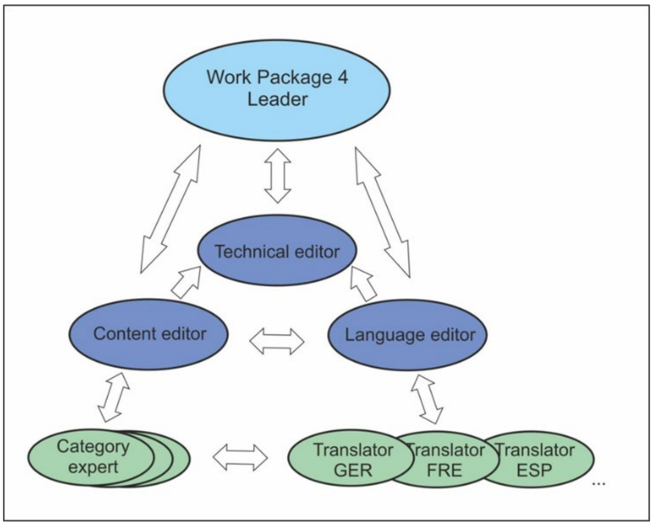

Figure 2-6 *Recommended responsibilities and communication flow for the creation and maintenance of the keyword thesaurus during the project.*

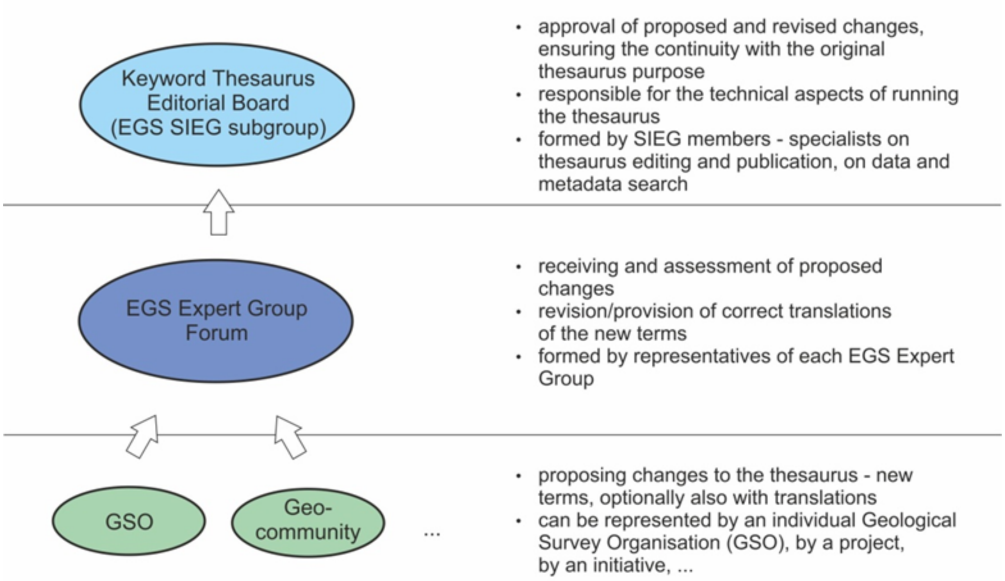

Figure 2-7 *Recommended responsibilities and communication flow for the maintenance and updates of the keyword thesaurus after the project end.*

- FUTURE MANAGEMENT – After the end of the GeoERA project, EGDI should be responsible for the management of the thesaurus. It is therefore recommended that an expert group be established at EuroGeoSurveys (EGS) level to be responsible for the technical, content and linguistic aspects of managing the thesaurus for future use in geoscience projects and research. A proposal for a future management concept is described in Figure 2-7. Without a structure like this, the thesaurus will gradually become obsolete.
- BACK UP – All files required for the operation of the GeoERA Thesaurus should be stored in a central EGDI data repository that is regularly backed up and sustainably managed. Specifically with regard to future use after the end of the GeoERA project as well. It should be discussed whether all information to be updated should be downloaded to a local system (e.g. as now for validation in the GBA PoolParty system) and then uploaded to the system again when the update is complete.
- MANAGEMENT OF THE DOMAIN OF TERMS - The domain that defines the namespace for the terms from the thesaurus should be owned and managed by EGDI. This guarantees sustainable operation in the future, so that the domain does not expire and the URIs remain unchanged. The current status is that the URI is generated via the registry system currently hosted by BRGM. The domain belongs to the EGS and is https://data.geoscience.earth/ncl/geoera/keyword.

  More detailed information on URIs and URI design can be found in chapter 3.2 URI design in GeoERA WP4 Deliverable D4.3 "GeoERA Project Vocabularies"
- MAINTENANCE OF THE SERVICE – Until the end of this project, the BRGM is responsible for this task. After the end of the project, it might be necessary to form a "Keyword Thesaurus Editorial Board" in the spirit of EGDI and to elect one or two persons who have access rights to the BRGM registry system.
- LICENSING – The thesaurus is published as Linked Open Data under the Creative Commons Licence (CC-BY 4.0) for free reuse.

A more detailed description of the whole workflow of T4.1.3 within the frame of the GeoERA project is available in chapter 4 in the GIP-P WP 4 report deliverable D4.2 “GeoERA Keyword Thesaurus”.

2.2.1 Issues and lessons learned

The following are some issues and lessons learned distilled from the activities in this T4.1 that should be actively considered in future actions:

<u>Issues:</u>

- Low level of knowledge of most of the project partners on the subject of "Linked Data" and knowledge representation
  - Misunderstandings regarding the composition of a hierarchy of terms (empty hierarchy-lines, missing hierarchy levels, mixing of viewpoints)
  - No unique labeling and problems grouping concepts into concept schemes (modeling approach, mess of categories).
  - Missing understanding in SKOS (e.g. when to use a relation and when mapping properties).
  - Reuse of existing concepts.
  - Draft of the same concept in different hierarchies.

- Clarifying responsibilities
    - Who is responsible for the maintenance and governance of the GeoERA Keyword Thesaurus when the project ends? Until now, we could just make suggestions.
    - URI design, URI persistence and web appearance of the WP4 results – These are topics still under discussion and concern the future URI design, technical management and strategy within the European Geoscience Registry System and within the EGDI portal especially after GeoERA has ended.

Lessons learned:

- Personnel resources - It happened that important communication partners left during the project. This meant that the previously jointly developed knowledge had to be communicated anew. Here it was particularly important that the most important decisions were recorded in documentation and meeting minutes. Nevertheless, this involved a time delay and sometimes a minor strategic change.
- Underestimation of the time needed for the planned activities in these large projects - Inter-European communication of complex issues is not an easy thing anyway, but the SARS-CoV-2 situation made it even more difficult. Not that the colleagues were not available, but the additional psychological burden should not go unmentioned here. It is therefore all the more important to always plan a time buffer so that the project manager does not have to invest a lot more time himself to ensure that the project achieves its goals.
- The bilateral meetings - These cost a lot of time, relatively speaking, but are definitely worth it. This made up for the online support, which in our case did not work quite as well, and it was possible to respond specifically to the questions and suggestions of the project partners.
- Creating sustainable knowledge - There has been a lot of communication on Linked Data, knowledge representation and information retrieval through the many meetings and discussions with project partners. This may had positively influenced the knowledge and the handling of this topics in a sustainable way. Especially for further programs and projects such as CSA.

2.3 Modelling

The creation of a keyword thesaurus is a complex subject. This topic represents a separate scientific area and is a part of a subject area called Search Engine Optimization (SEO). When collecting and compiling keywords for information retrieval, the questions such as "Are these terms and phrases relevant? Should they be included? Are these things that people would search for? Are they topically relevant?" are essential.
The modelling approach focusses on the use case "If somebody searches for XY, what could he/she be interested in too...? And this is one of the key differences from the project vocabularies - the modelling approach. Here, the goal is not to use the keyword thesaurus to define and link terms by their scientific classification. Rather, the goal is to link terms with the aim of providing the user with the best search result. The approach is to learn about the topic by reading about it, talking to clients (project partners), or sharing about the topic on social networks. It is important to put yourself in the shoes of the user who is looking for something. With this approach and a relatively intuitive linking of index terms focused on the use case, it is more effective to get valuable results from a search system. Of course, it is important to have the ability to periodically test the implemented modelling logic. For this reason, a

provisional application was created to test the keyword modelling approach and semantic search in the context of this task. This application can be found at https://schmar00.github.io/semantic-search/.

From a more technical point of view the modelling within the GBA thesaurus managements system was performed considering the W3C **standardised SKOS ontology** (W3C - A. Miles & S. Bechhofer, 2009) with basically modelling narrower/broader and related concepts.
SKOS mapping properties were used to link to published online resources used for keyword compilation such as GEMET, INSPIRE, and others.

At the beginning it was planned to treat each search category as a separate concept schema. Now all concepts are modelled under a single concept schema called "keyword" and the search categories are available as a custom defined property (dbPedia/Category Attributes) for almost every concept.

In order to support the HIKE project partners and their knowledge sharepoint, the modelling of their keywords took their needs into account. Therefore, a lightweight ontology was created especially for HIKE. A detailed description of specifically generated RDF properties of this light weighted ontology is available at https://gist.github.com/schmar00/ee728afd38969097d80c918a3a436dff

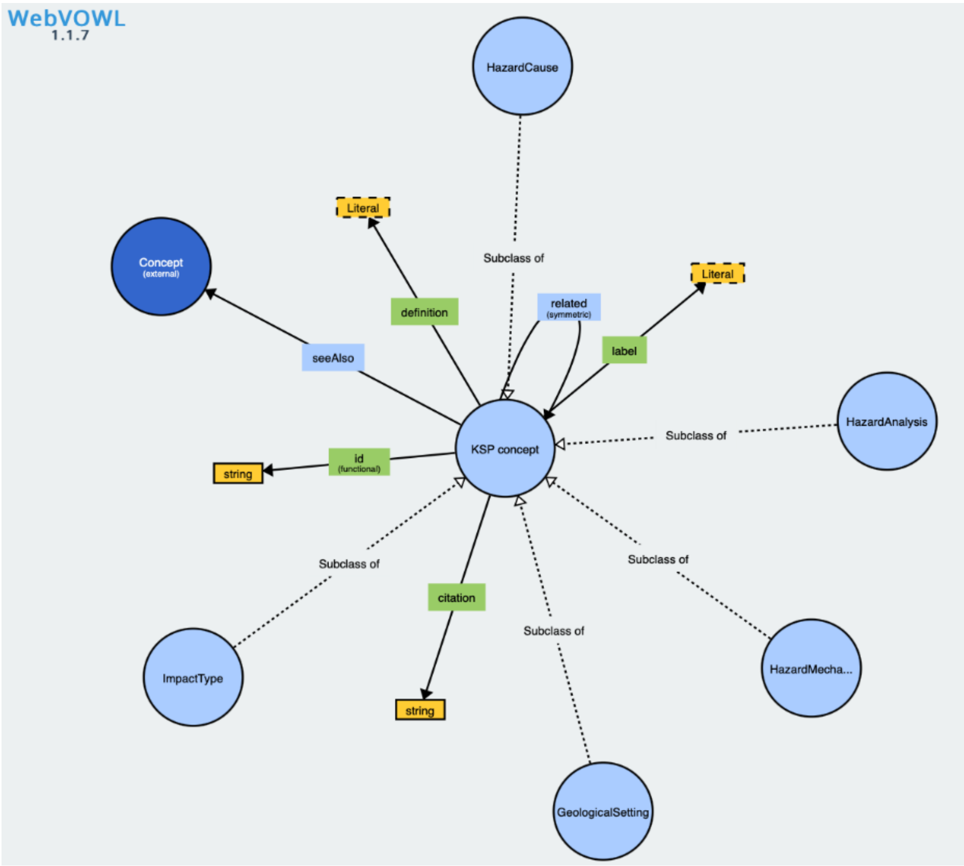

Figure 2-8 *Visualization of properties and associations of the light weight ontology for the HIKE knowledge sharepoint*

Important to know is that this ontology for HIKE has a parallel application and does not influence the modelling and application of the GeoERA keyword thesaurus for the other GeoERA projects.

2.4 Brief technical infrastructure

The input information, in this case the keywords, is delivered hierarchically structured in the form of an excel file. The actual modelling then takes place in the thesaurus management system of the GBA. The result of this modelling is then an RDF file validated according to the SKOS standard, which is transferred to the European Geoscience Registry System, into which the keyword thesaurus is then integrated and through which the keyword thesaurus is published online.

All keyword concepts have got identifiers (URIs) based on the domain name **"data.geoscience.earth"** registered by EGS.
The GeoERA Keyword Thesaurus is available online also via a web API (SPARQL - https://data.geoscience.earth/ncl/ui/sparql-form) to drive future search systems.

The GeoERA keyword thesaurus is already implemented in the EGDI metadata catalog MIcKA. There, the keywords are already usable for indexing GeoERA project datasets. Furthermore, the GeoERA keywords are also provided within the GeoERA document repository and can therefore already be used for indexing GeoERA project documents.

For more information on the development of the technical infrastructure in the GIP project as well as the European Geoscience Registry, see GeoERA GIP-P WP5 "Architecture" (led by BRGM).

2.5 Applications and use cases

- One relevant use case for the GeoERA Keyword Thesaurus is its use for metadata tagging by the project partners. This is to be done via the EGDI metadata catalogue MIcKA for EGDI datasets and obligatory for GeoERA datasets.
  See the EGDI metadata catalogue: https://egdi.geology.cz/

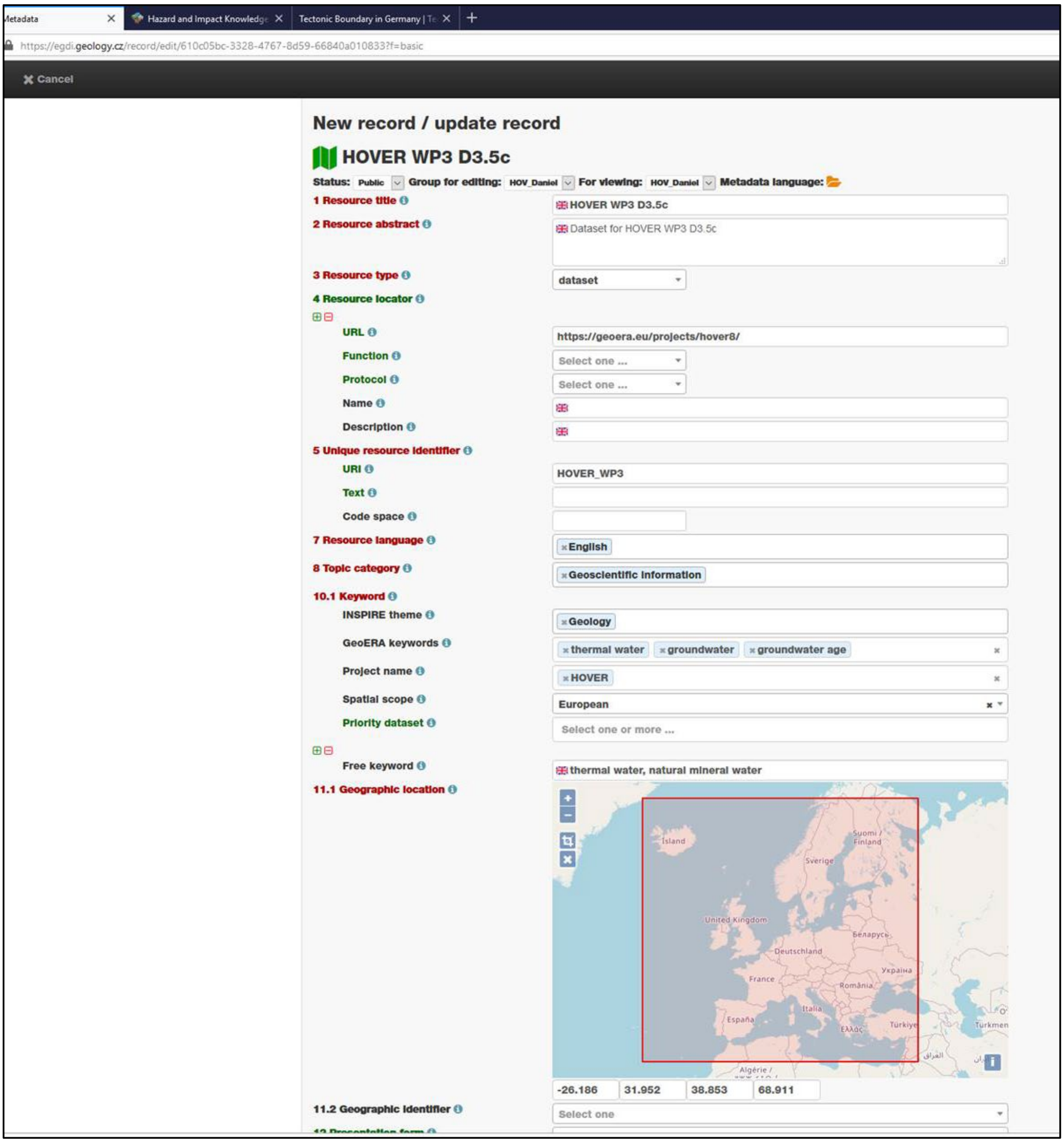

Figure 2-9 *Example form (here from the GeoERA project "HOVER") for the creation of a metadata record in the EGDI metadata catalogue - under "10.1 Keyword" the index terms for the respective data record from the GeoERA keyword thesaurus shall be assigned .*

- Another relevant use case is that the GeoERA Keyword Thesaurus shall be used as basic subject heading system for the GeoERA search engine established by IGME.
  See the system in test mode at EGDI: https://egditest01.geus.dk/searchsystem/en/GeoERA

- One application is the GBA semantic search test application created to check the modelling approach and the concerning search results:
  See and have a try on this application: https://schmar00.github.io/semantic-search/ (Figure 2-10 and Figure 2-11)

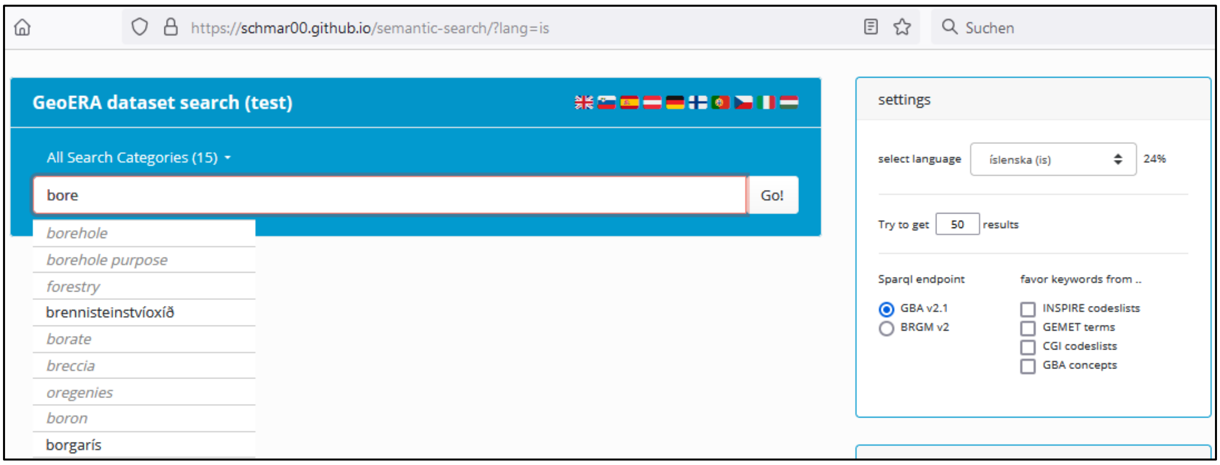

Figure 2-10 *Screenshot of the semantic search test application – the selected language here is Icelandic. When searching e.g. for the term "borehole" a few letters are enough to pop up a drop-down list with matches in English written in italics and exact matches in Icelandic written in standard font.*

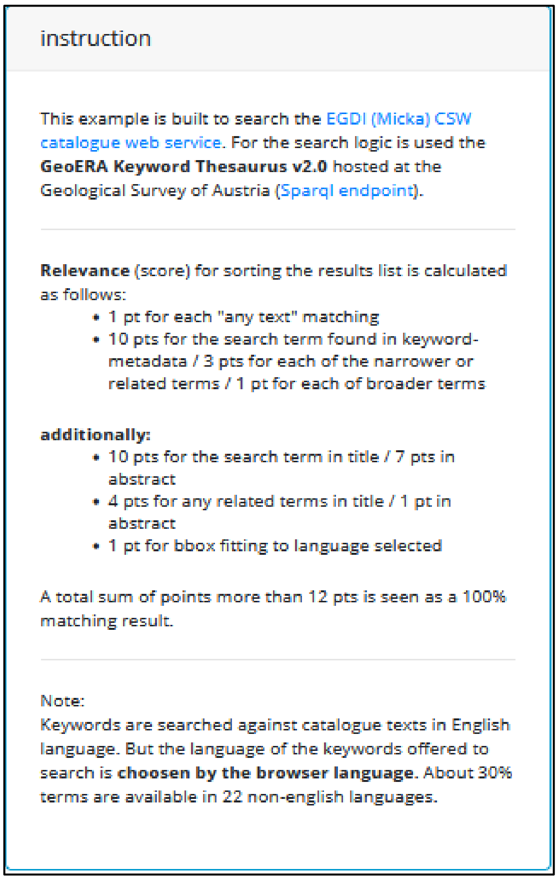

Figure 2-11 *Screenshot of the semantic search test application - additional basic instructions are given on the website how to use this semantic search application.*

- A relevant use case is the usage of the keyword thesaurus by the HIKE project. A subset of 103 (project specific) keywords were integrated into the GeoERA Keyword Thesaurus in order to support the “HIKE Knowledge Sharepoint”
  See/have a try on this application in test mode at TNO: https://www.gdngeoservices.nl/hike/
  A detailed description of the specifically generated RDF properties for this knowledge sharepoint (a light weighted ontology) is available at https://gist.github.com/schmar00/ee728afd38969097d80c918a3a436dff.

  For detailed information on the HIKE knowledge sharepoint please see the **HIKE GeoERA report deliverable D4.2b.**

- Another usage of the GeoERA Keyword Thesaurus is as basis for the search within the GeoERA document repository search implemented by GEO-ZS.
  See and have a look at: https://www.geo-zs.si/db/egdi-search/

3 GEOERA PROJECT VOCABULARIES

3.1 Description

Project vocabularies provide the opportunity to clarify expert knowledge and terminology in the form of project specific vocabulary concepts on a scientific level and to use them in datasets to code data. At the same time, parts of this vocabulary might be later included in international standards (e.g. INSPIRE or GeoSciML), if desired. By comparison, Project vocabularies are open collections of knowledge that, for example, may also contain deprecated, historical or only regionally relevant terms. In an ideal overall view, the sum of all Project vocabularies results in a knowledge database of bibliographically referenced terms that have been developed through scientific projects. Due to the consistent application of the data standards of Semantic Web and Linked Data nothing stands in the way of further use by modern technologies such as AI. This report explains what is meant by Project vocabularies in the context of GeoERA and examples of what problems, in semantics of data, can be solved by using them. In addition, project related methods and workflows around Linked Data, and SKOS in particular, are described.

An important step that could not be fully discussed during the work on the project is the distinction between the project vocabulary and the code lists or selection lists. As already described in the previous GeoERA WP4 report D4.3 “GeoERA Project Vocabularies” (Schiegl et al., 2019), a project vocabulary is a knowledge representation of the knowledge that was processed within the project period. Codelists, on the other hand, represent an editorially processed selection of terms which should or may be used specifically in relation to a specific object class in the data specification. The selection lists that are available as part of the GeoERA project therefore refer to the existing INSPIRE or GeoSciML standards. The GeoERA project vocabulary offers to the participants a technical infrastructure to describe scientific terms more precisely and sometimes even to design and formulate them for the first time. Only when the meaning of these terms has been precisely defined, is provided with a literature reference, and synonymous terms have been clarified, are these drafted terms available to be included in a standard. This step in the workflow of standardizing and creating code lists should then be supervised by a dedicated editorial team - preferably by EGS expert groups.

Experience has shown that terms, concepts and vocabularies cannot be developed as a so-called standard from the beginning. Rather, it is necessary in the first step to try out the newly designed concepts, to modify them and to check whether they are suitable as annotations for other data sets. Only in the second step essential concepts can be extracted and processed for new standardized lists or nomenclatures, or used for extending existing standards like INSPIRE codelists. It therefore also seems unrealistic for the scientific community to edit and work on existing standards without editorial support, directly by adopting from project results - in the way the EU directive INSPIRE has foreseen.

Further limitations when creating project vocabularies can mainly be found in the area of data modelling. During the project, it is difficult to explain the basics of semantic modelling of vocabularies to people who have not yet dealt with this topic. This is one of many reasons to start knowledge modelling better with SKOS and to build custom properties or ontologies only afterwards. In many cases, the preferred labels are not discussed sufficiently at the beginning or descriptive texts or literature citations are missing. In principle, ontology modelling can only begin after problems or misunderstandings in taxonomy or linguistics have been resolved.

3.2 Workflows and results

The workflows developed in the report Deliverable 4.3 GeoERA Project Vocabularies[7] couldn't be implemented due to a time shift between the scientific projects creating the content and WP4 timeline. Even at the date when this report is finalized some project teams didn't transfer all concepts into the test environment. Also due to changes in staff and redesign of URI path names the project vocabularies couldn't be integrated at the "productive system" at EGR (European Geoscience Registry) yet. Additionally, massive time delays caused by the COVID19 pandemic haven't been facilitating WP4 workflows.

In the report Deliverable 4.3 GeoERA Project Vocabularies at "3.4 Template and instruction" WP4 prepared a "quick start" manual for editing an Excel sheet in order to provide basic information for scientific concepts. This initial information is in general necessary to start a new GeoERA Project Vocabulary. This basic information was supplemented later in the modelling phase, towards the end of the project, with further relations like mapping links, other links or properties. A detailed draft description is available under https://github.com/GeoEra-GIP/WP4-Semantics.

Following issues were recognized as "Lessons learned" related to data preparation of SKOS concepts for import into SKOS/RDF triple stores:

Preparation in Excel template:
- blank spaces and line breaks where they shouldn't be (double, beginning, end, urls)
- line breaks within cell values
- wrong column headings (e.g. prefLabel@xy)
- information in the wrong columns (e.g. citation texts, links)
- use of improper characters (quotation marks, angle brackets)
- missing mandatory elements (prefLabel, description, link without citation)
- hidden information without explanation (color coding, font formatting)
- unintentionally shifted cell content (wrong columns, notation doesn't fit to label)
- hierarchically misunderstandings (empty rows, missing hierarchy levels)
- preferred label not unique (within PV), wrong linguistics

In general:
- misunderstandings in design of proper concept schemes (modelling approach, mess of categories)
- misunderstanding in SKOS relation and mapping properties (exactMatch, etc)
- incomplete information with many correction cycles (reuse of existing URIs)
- setting of SKOS labels (plural vs singular, max length, term vs phrases
- setting of SKOS descriptions (definition vs scopeNote)
- reuse of existing concepts (inScheme vs exactMatch, vocabularies vs codelists)
- use of tentative properties derived from application schemes (e.g. GeoSciML)
- merging/linking Project Vocabularies in the future (building the EGDI knowledge base)
- draft of the same concept in different templates or different hierarchies (not only citations!)

Improvements suggested:
- clarify how to prepare polyhierarchical concepts in Excel
- use a separate citations list and numbering to connect with concepts
- use a common format for citations with colon after year

[7] Schiegl M., Sőrés L., Pantaloni M., Johansson O., van Ede R. 2019: GeoERA, GIP-P Deliverable 4.3 GeoERA Project Vocabularies, available at https://geoera.eu/projects/gip-p/

- facilitate online editing (updating only) for projects
- quality checks in linguistics and ontology modelling
- label concepts for different application types (legend items vs codes for standards, categories vs individuals)

Regarding the reuse of GeoERA Project Vocabularies for extending INSPIRE codelists – governed by EGS expert groups in an official way – there wasn't a progress in establishing new workflows. In addition, this issue has been recognized and adopted by the SIEG (Spatial Information Expert Group) of EGS.

	Vocabulary draft	GeoERA Project Vocabularies	European Geoscience Register	INSPIRE register federation	INSPIRE codelists
Governance	project team	GeoERA project team	EGS expert groups	JRC INSPIRE registry team	JRC INSPIRE registry team
Publisher	project team	GIP-P WP4	BRGM	Relevant organization (BRGM)	JRC INSPIRE registry team
basic ontology	none	SKOS (https://www.w3.org/TR/skos-reference/)	REG (purl.org/linked-data/registry)	none	none
Owner	project team	GeoERA project team	EGS, EPOS?	Relevant organization	European Union
Entities	Scientific terms, codes, texts, links, citations..	concepts, concept schemes, links, codes, metadata, .	codes, ontology concepts, complete ontologies, coordinate reference systems, units of measure, spatial objects, organizations, licenses, metadata etc.	Codelists, enumerations	Codelists, enumerations
Data transfer	Excel	trig, rdf/xml, ttl, Sparql (test API)	ttl, rdf/xml, json-ld, csv, Sparql	Links to registers	rdf/xml, json, atom, csv
Data standards	none	Semantic web standards	Semantic web standards	INSPIRE standards	INSPIRE standards
Linked Data (https://5stardata.info/en/)	☆☆	☆☆☆☆☆	☆☆☆☆(☆)	☆☆☆☆	☆☆☆☆
Semantic relations	draft	pivotal	partially	partially	only parent/narrower
cross-linked vocabularies	draft	pivotal	partially	uncommon	no
scope	elaborate scientific terms, or assemble codelists	Clarification of the meaning of scientific terms, providing context, knowledge base, bibliographic references	Register codes, concepts, data and other entities	Extending INSPIRE codelists	Providing standardized codes
explanations	https://github.com/GeoEra-GIP/WP4-Semantics/tree/master/Project%20Vocabularies/templates	https://geoera.eu/wp-content/uploads/2019/11/D4.3-GeoERA-Project-Vocabularies.pdf	https://github.com/UKGovLD/registry-core/wiki/Principles-and-concepts	https://inspire.ec.europa.eu/id/document/tg/registers-and-register-federation	http://inspire.ec.europa.eu/codelist
Project Vocabularies	create concepts >> Project Vocabulary				
		Project Vocabulary >> publish concepts			
codelists	create codes	>> skip PV >>	publish codes		
INSPIRE codelists			extend INSPIRE codelists		
				use INSPIRE codelists plus official extensions	

Figure 3-1 *The table shows the differences between the creation of codelists only (blue) versus GeoERA project vocabularies (red) in relation to official INSPIRE codelists or extensions*

For the date when this report was delivered, the editing and publishing workflow processed by WP4 is located somewhere in between draft and EGR integration.

3.3 The differences in modelling

The 15 vocabularies created for 6 different projects took quite different approaches in modelling semantics. But all of them are based on the SKOS standard for RDF Linked Data. This groundbreaking decision is very important for future developments on the way to a common knowledge base. Although GeoConnect3D took the opportunity to start with a custom ontology (additional RDF properties) their vocabulary concepts for “Limits and Units” are modeled in relation to other concepts e.g. HIKE or HotLime Fault Systems. This project specific extension of SKOS is not yet described due to the late integration of GeoConnect3D vocabularies.
Almost all Project Vocabularies designed for elaboration of new categories – in the kind of codelists or extension of existing codelists – are modeled on a generic approach. All other Project Vocabularies describing named features e.g. Fault Systems/Units by HIKE, HotLime or GeoConnect3D have taken the partitive approach. Which means e.g. a named “Large Fault System” has parts of “Fault Systems” which in turn have parts of “Faults” and so on. Both modelling approaches are covered by standardized SKOS modelling with narrower/broader and related concepts. HotLime has additionally adopted GeoSparql (OGC standard) simple feature properties like sf:touches which was applied to annotate Faults delimiting (in the meaning of Geology) other Faults.

The following table describes the project vocabularies with “measured” differences in data and modelling. Critical properties regarding semantics are e.g. how many relations or attributes are given in average per concept. This kind of metrics describes how much knowledge is behind the terms and concepts.

Metrics of
GeoERA Project Vocabularies and
GeoERA Keyword Thesaurus

Mapping Relations: skos:closeMatch, skos:exactMatch, skos:broadMatch, skos:narrowMatch
Labels: skos:prefLabel, skos:altLabel, skos:hiddenLabel, hike:label

Title	URI (data.geoscience.earth)	No. of triples	Top concepts	Concepts total	Concepts created	English terms	Average per concept: Semantic Relations	Mapping Relations	Attribute Relations	Labels	hierarchical depth
Dimension stone – Commodity Type	eurolithos:CommodityCodeValue	364	1	62	53	104	1.97	0.06	1	1.06	1.87
Stone Colors	eurolithos:colorCodeValue	176	16	16	16	16	0	0	0	7	0
ornamental (unique) stone	**eurolithos:uniqueStone**	6845	**14**	**1143**	**1143**	**1147**	**1.98**	**0**	**0**	**1**	**1**
Categories of units and limits	*geoconnect3d:category*	*453*	*4*	*59*	*47*	*92*	*2.19*	*0.32*	*1.56*	*1*	
Geomanifestations	*geoconnect3d:geomanifestation*	*984*	*4*	*63*	*63*	*145*	*1.87*	*0*	*5.6*	*1*	
GC3D limits	*geoconnect3d:limits*	*209*	*1*	*20*	*20*	*35*	*1.9*	*1.05*	*2.45*	*1*	
GC3D units	*geoconnect3d:units*	*21*	*1*	*1*	*1*	*3*	*0*	*3*	*9*	*4*	
Fault offset determination	hike:FaultOffsetDetermination	93	10	10	10	20	1	0.2	1.1	1	0
Fault classification terms and parameters	hike:category	463	2	95	38	85	2.53	0.5	0.47	1.24	2.13
Tectonic Boundary Objects in Europe	**hike:faults**	19276	**18**	**2585**	**1670**	**3211**	**2.28**	**0.06**	**2.83**	**1.96**	**2.74**
Tectonic Boundary Objects in carbonate geothermal reservoirs of Europe	**hotLime:faults**	15262	**10**	**1315**	**1132**	**1894**	**2.25**	**0.01**	**2.68**	**1.75**	**2.59**
Lithological Units in carbonate geothermal reservoir settings of Europe	hotLime:units	1872	10	168	168	333	1.89	0.39	2.48	1.67	2.18
hydrochemical compounds	hover:hydrochem	560	3	49	49	96	1.88	0	2.57	2	1
geothermal energy	muse:BoreholePurposeValue	97	1	8	7	19	1.88	0	4.75	1.75	3
Concepts in the fields of geothermal energy, geology and hydogeology	muse:category	750	3	97	89	187	1.94	0.29	0.86	1.1	2.4
References	**:ref**	7425	**5**	**897**	**897**	**-**	**2.04**	**0.08**	**1.32**	**1.01**	**1**
GeoERA Keyword Thesaurus v2.1	:keyword	83319	32	2596	2596	3503	3.65	1.45	1.04	13.21	3.38

Semantic Relations: skos:broader/narrower/related, geosparql:sfTouches, geoconnect:limitTo, dcterms:replaces, Hike:related
Attribute Relations: foaf:isPrimaryTopicOf/page/depiction, dcterms:source/title/references/bibliographicCitat:, skos:notation/description/scopeNote, hike:label/id/definition, schema:hasMap

Figure 3-2 *The different Project Vocabularies plus the GeoERA Keyword Thesaurus compared by quality in semantics*

According to Alexopoulos[8] ***Semantic accuracy*** *is defined as the degree to which the semantic assertions of a model are accepted to be true. To some extent, these types of problems have been found to be related to misunderstandings of the modelling elements’ semantics and intended usage* (Alexopoulos 2020). This issue mainly affects the vocabulary “MUSE categories”, because of a missing model. A lack of domain knowledge or a lack of expertise can be ruled out when assessing the semantic accuracy.

[8]Alexopoulos P. 2020: Semantic Modeling for Data - Avoiding Pitfalls and Breaking Dilemmas, O’Reilly Media Inc. Sebastopol CA

- ***Completeness*** *of the semantic model can be defined as the degree to which elements that should be contained in the model are indeed there. The reasons why a semantic model could be incomplete are e.g.: the size and complexity, missing data sources from which the model could be extracted, or the volatility and dynamics of the domain* (Alexopoulos 2020). This was a problem for HIKE and HotLime vocabularies because the “Fault and Units information” don’t cover whole Europe. Also, a “references list” of bibliographic resources can never be complete, same as the enumeration of “Ornamental stones”.
- ***Consistency*** *means that the semantic model is free of logical or semantic contradictions* (Alexopoulos 2020). Inconsistencies mainly concern the categorization and classification of the fault systems between the two projects GeoConnect3D and HIKE, in relation to INSPIRE codelists. Much effort was also put into correcting inconsistencies between the fault systems described in project HIKE and project HotLime. MUSE categories, references list, and of course the GeoERA Keyword Thesaurus is work in progress without finalized consistency.
- ***Conciseness*** *in a semantic model is the degree to which the model does not contain redundant elements which are elements or combinations of them that already exist in the model in a different but semantically equivalent form, or that are no longer required to be in the model* (Alexopoulos 2020). Reasons for this could be the uncoordinated modelling from different parties like “Ornamental stone” vocabulary from EuroLithos, or the optimization for different applications at the same time, or even temporary workarounds that were not removed later. This problem occurred when compiling the GeoERA Keyword Thesaurus.
- ***Timeliness*** *in a semantic model can be defined as the degree to which the model contains elements that reflect the current version of the world* (Alexopoulos 2020). This was not a problem at all.
- ***Relevancy****: A semantic model is relevant when its structure and content are useful and important for a given task or application* (Alexopoulos 2020). Since all vocabularies are less likely to have been tailor-made for specific projects, this argument applies entirely.
- ***Understandability*** *or comprehensibility of a semantic model is the ease with which human consumers can understand and utilize the model's elements without misunderstanding or doubting their meaning* (Alexopoulos 2020). All vocabularies can largely only be described as understandable within their own community. Further attempts to make the vocabularies understandable for a general audience were only undertaken by GeoConnect3D, where a separate project ontology was developed. For “HotLime units”, “HOVER hydrochemical components” and “MUSE categories” it was not clear how to utilize the model elements. Please note that all quality criteria of the models are roughly estimated biased by the opinion of the authors.
- ***Trustworthiness*** *of a semantic model refers to the perception and confidence in the quality of the model by its users. This (inevitably or subjective) perception is definitely related to other quality dimensions like correctness, completeness or relevancy* (Alexopoulos 2020). The trustworthiness can generally be estimated as very high for all vocabularies due to the fact they were all created treasured manually without automatic text extraction or machine learning results.

3.4 Technical infrastructure

The input information, in this case the vocabulary concepts, are delivered by the project partners hierarchically structured in the form of an excel file. The actual modelling then takes place in the thesaurus management system of the GBA, called PoolParty. The result of this modelling is then an RDF file validated according to the SKOS standard, which is transferred to the European Geoscience Registry System, into which the project vocabulary is then integrated and published online.

Technically speaking, two different systems were used for the drafting and publication of the project vocabularies. The software PoolParty, with SKOS quality check and validation, was used in processing and in the daily updates. A linked data platform (Epimorphics, Jena Triplestore) is used for the productive operation and publication of the project vocabularies, which uses REG as the basic ontology and SKOS as an extension. For example, then a SKOS concept scheme is converted into a REG registry list - with restrictions in the URI design. Or the assignment of SKOS inScheme, for example when concepts are used multiple times in different schemes, becomes less important.

However, not all challenges relating to the technical infrastructure have been completely resolved at the time of writing this report. In particular, the management and governance of the technical infrastructure after the GeoERA project will still need to be discussed.

3.5 Applications

Various web applications were also created in connection with the project. The central application to view all project vocabularies and to search for different concepts of the different projects, called PV Viewer, was designed on GitHub and is available for further use in the central portal of EGDI. From a technical point of view, the viewer only uses HTML, JavaScript and CSS and can be integrated “standalone” in various project pages. The developers’ version on GitHub uses the test database incl. Web API (Sparql endpoint) of the GBA.

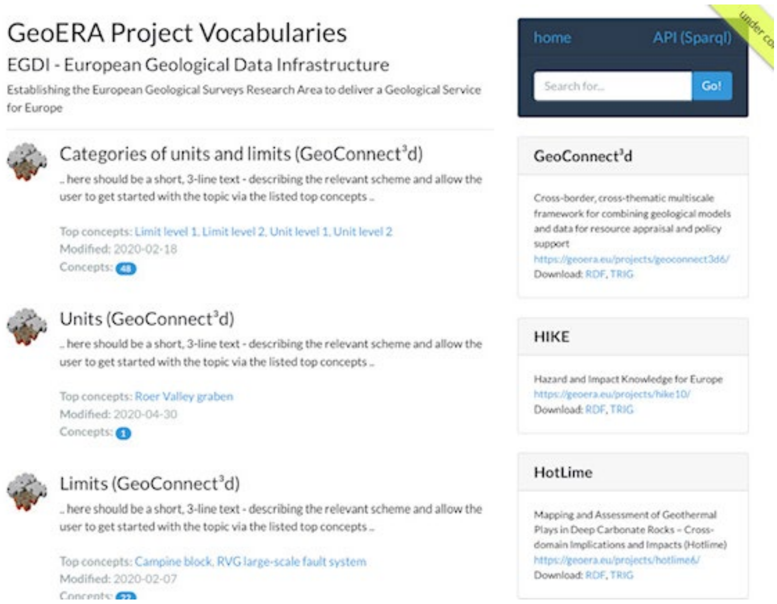

Figure 3-3 *PV Viewer at https://schmar00.github.io/project-vocabularies/*

Another version is available on GitHub, which uses RDF data directly from the "productive system" of the European Geoscience Registry "on Data.geoscience.earth.

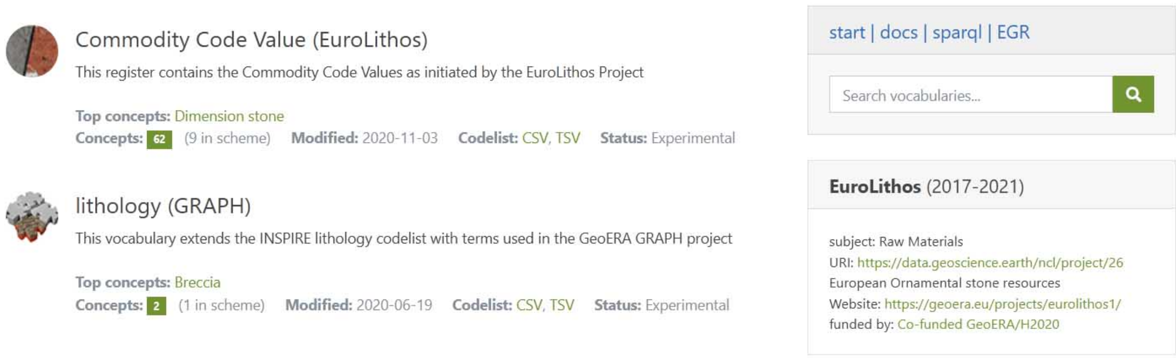

Figure 3-4 *EGDI Viewer, https://geolba.github.io/project-vocabularies/*

All development work relating to the keyword thesaurus and the project vocabularies are also documented on the GIP-P project website.

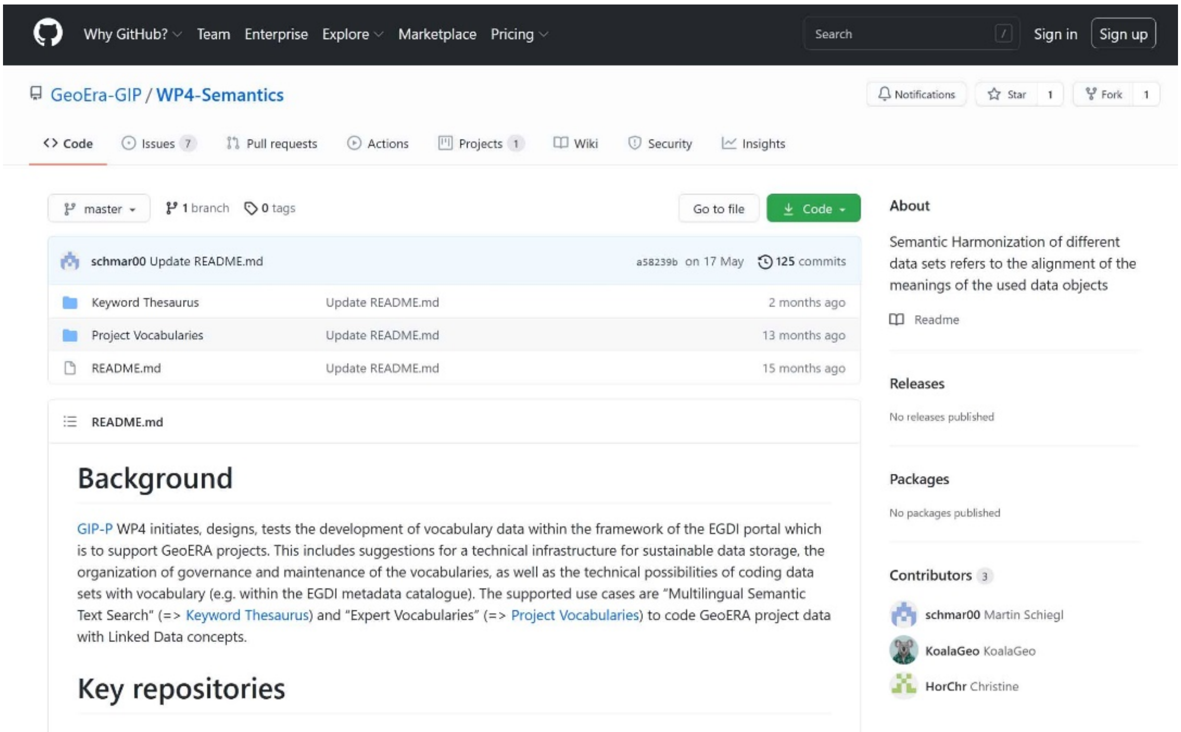

Figure 3-5 *GitHub https://github.com/GeoEra-GIP/WP4-Semantics*

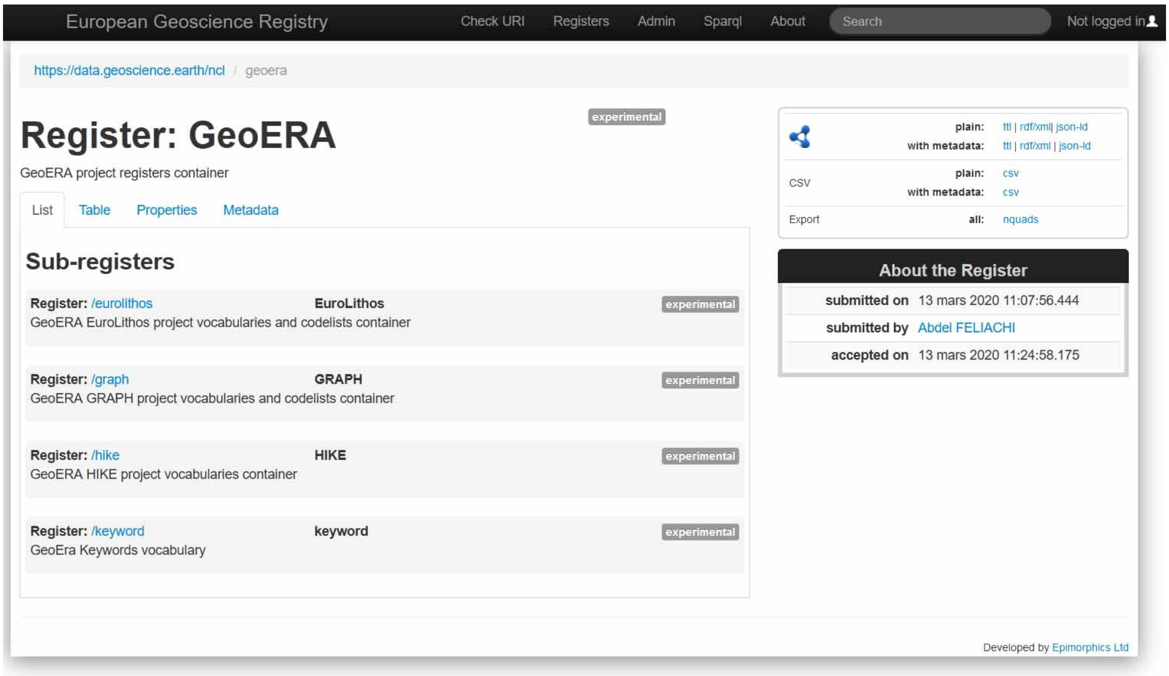

***Figure 3-6** EGR https://data.geoscience.earth/ncl/geoera*

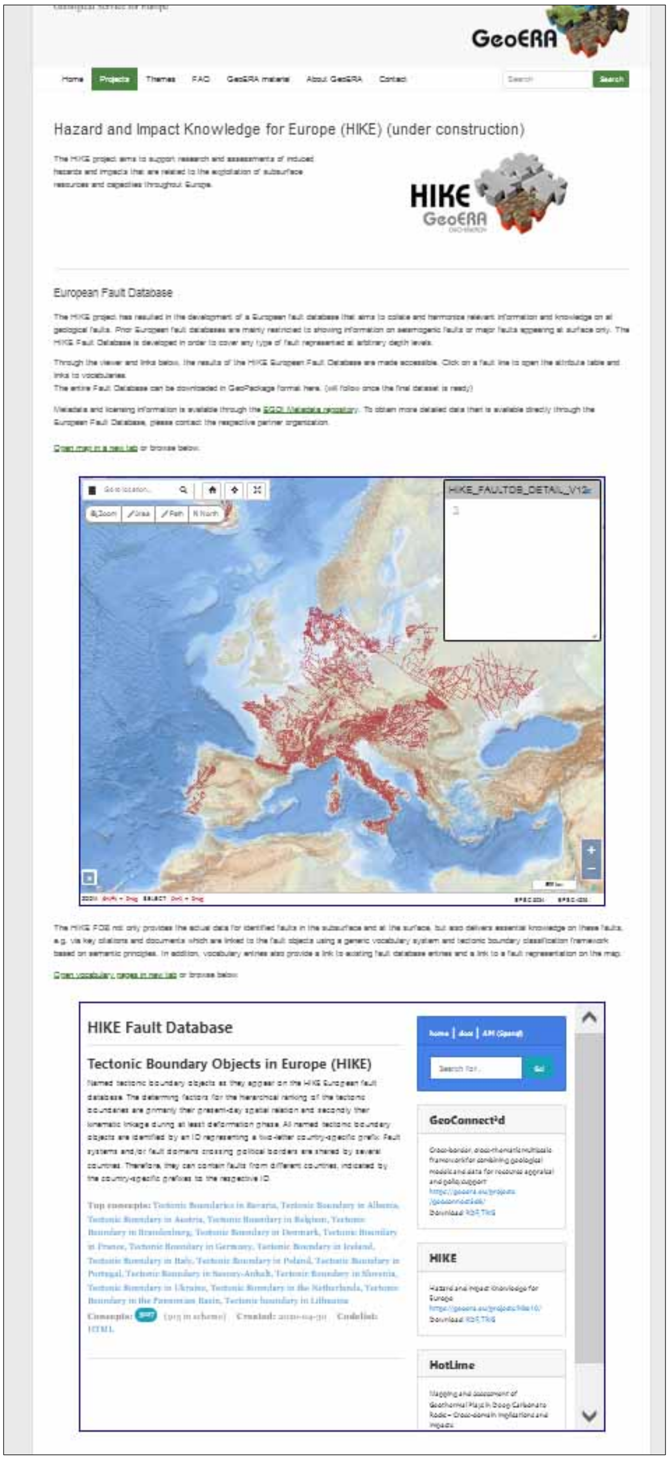

Figure 3-7 *HIKE project page*

HIKE project page for „Fault Database“
https://geoera.eu/projects/hike10/

The HIKE project page is an example of the multiple use of the developed website, where parts of the PV viewer, in this case the HIKE Fault Database, has been embedded in another page.

3.6 Use Cases

The problems in the semantic harmonisation of different data sets arise from the fact that data content uses different and non-comparable annotations and therefore cannot later be merged into a common data set. This problem can be solved by using common code lists or if these common code lists do not exist, a project vocabulary can be created as part of a joint project. The Project Vocabulary project contains the terms that have been worked out together, linked with one another in the sense of a small knowledge representation.

Semantic harmonisation

This approach was chosen by the HIKE, HotLime and GeoConnect3D projects for the naming and description of geological fault systems throughout Europe. These concepts for named fault systems (e.g. Danube fault system) were processed in so-called SKOS concepts, each with short descriptions and source references, and published as RDF Linked Data. On the other hand, the various spatial data sets of the fault systems meaning the geometry data (in our case line geometries) were merged in a central GIS database and published as WFS (REST API for GIS systems). The line geometries in the spatial data set can now be linked with the corresponding URIs of the published SKOS concepts for the geological fault systems and ultimately be used jointly via a web application. A suitable example is the Periadriatic Seam, which geologically separates two continents and can be represented on a common map performing a semantic database query over three different spatial data sets of Slovenia, Austria and Italy. In this example none of the Fault lines are indexed with the overarching "**Periadriatic-Mid-Hungarian Large-scale Fault System**". The compilation of the different lines into such a super ordinated spatial feature (without dedicated geometry data) is just done by linking concepts in a semantic way by Project Vocabularies. If concepts are linked in a wrong way it easily could be changed by improving the Project Vocabularies.

Figure 3-8 *Example of a "Large Fault System" cross-border composed of individual Fault elements in Slovenia, Austria, Italy*

Crosslinking different resources

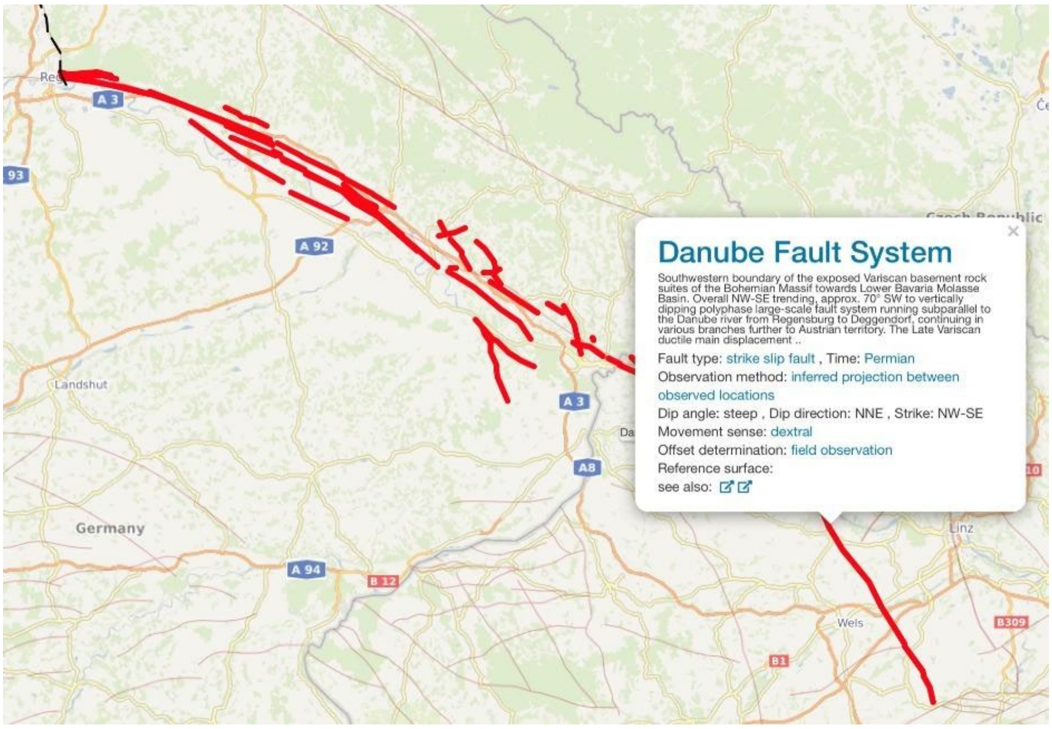

Figure 3-9 *Map depiction of Danube Fault System hyper-linked started from a PDF*

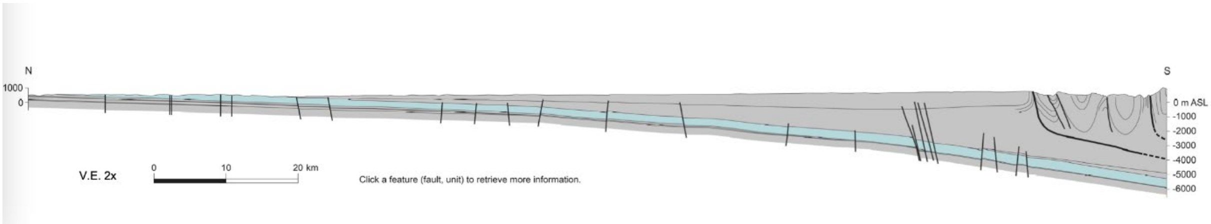

Figure 3-10 *A PDF with hyperlinks to URIs of HotLime Project Vocabulary*

The next use case is an example of the multiple reuses once a concept has been published. Here a PDF has been provided with hyperlinks to the Concept URIs, which leads directly to the PV Viewer. From there, in turn, all contextualized content - in this case - a map representation of the selected fault system can be accessed. The Project Vocabularies, implemented with Linked Data according to the principles of Semantic Web, are therefore suitable as a central storage location for linking various resources published on the web. This can be any database, for example text sections in websites, web applications including database queries or 3D models, images, PDFs, videos, etc
See also the reference links to the Italian Fault Database:

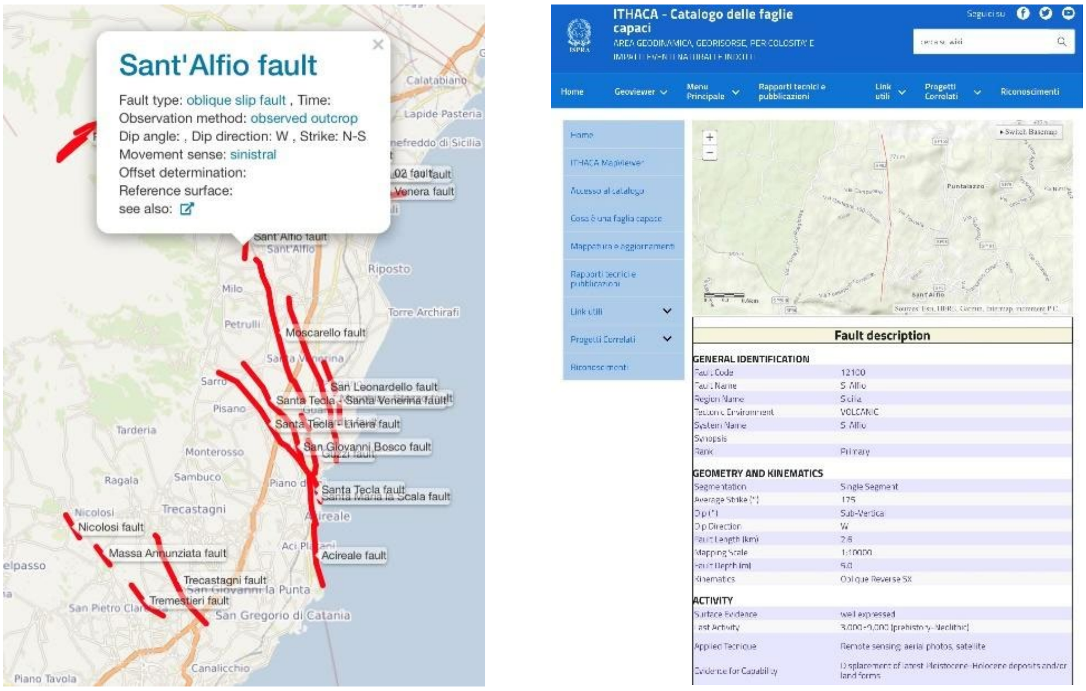

Figure 3-11 *Web application (Italian Fault Database) connected via Linked Data*

Multilingualism

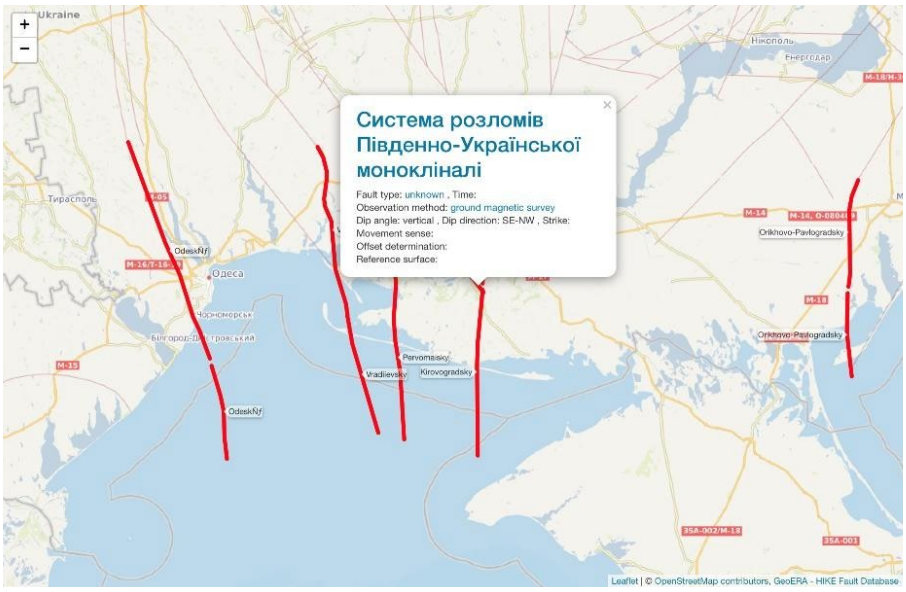

Figure 3-12 *PV Map Viewer with language setting in Ukrainian*

A great advantage, especially for European projects, when processing knowledge using RDF and Linked Data, is that the knowledge can be processed and stored in texts in any language. In the present example, the map display of the fault systems is shown in Ukrainian (if available).

Codelist extensions

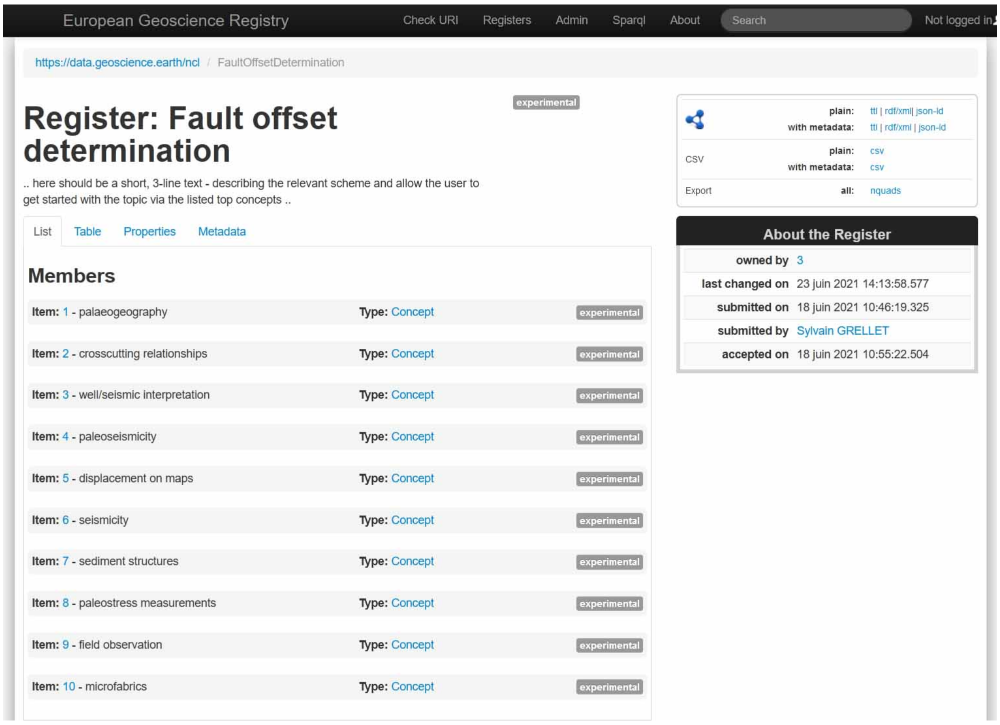

Figure 3-13 *Example of newly registered codelist*

Codelists and extensions of officially and standardized codelists like INSPIRE or GeoSciML – couldn't be realized yet, due to workflow problems described in 3.2. First experiments are online and already used by projects (e.g. borehole purposes by MUSE, Fault offset determination or Fault categories, or evaluation method by HIKE, Hydrochem. parameters by HOVER, dimension stones by EuroLithos etc.) to harmonize their annotations. In general, there is a need to clarify how EGS/EGDI will handle codelists in the future.

4 REFERENCES

4.1 Bibliographic references

- Alexopoulos P. 2020: Semantic Modeling for Data - Avoiding Pitfalls and Breaking Dilemmas, O'Reilly Media Inc. Sebastopol CA
- Auer S., Bryl V., Tramp S. (Eds.) (2014): Linked Open Data – Creating Knowledge Out of Interlinked Data, Results of the LOD2 Project, Springer
- W3C – Miles A. & Bechhofer S. (Eds.) (2009): SKOS Simple Knowledge Organization System Reference, W3C Recommendation 18 August 2009
- W3C – Schreiber G. & Raimond Y. (Eds.) (2014): RDF 1.1 Primer, W3C Working Group Note 24 June 2014

4.2 Web links

[viewed December 11, 2018]

- Archer P., Goedertier S., Loutas N. (2012): D7.1.3 Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC, ISA Interoperability Solutions for European Public Administrations, https://joinup.ec.europa.eu/document/10-rules-persistent-uris

[viewed August 30, 2021]

- W3C - SKOS Simple Knowledge Organization System, http://www.w3.org/TR/skos-reference
- W3C – RDF 1.1 Primer https://www.w3.org/TR/rdf11-primer/
- Hörfarter C., Sanabria Pabon M., Hernandez R., Novak M., Kondrova L. (2019): GeoERA Keyword Thesaurus Deliverable D.2, https://geoera.eu/wp-content/uploads/2019/11/D4.2-GeoERA-Keyword-Thesaurus.pdf
- Schiegl S., Söres L., Pantaloni M., Johansson O., van Ede R. (2019): GeoERA Project Vocabularies Deliverable D4.3, https://geoera.eu/wp-content/uploads/2019/11/D4.3-GeoERA-Project-Vocabularies.pdf
- Wikipedia - Thesaurus (information retrieval), https://en.wikipedia.org/wiki/Thesaurus_(information_retrieval)
- Wikipedia - Controlled Vocabulary - https://en.wikipedia.org/wiki/Controlled_vocabulary
- EGDI metadata catalogue: https://egdi.geology.cz/

5 APPENDIX A

5.1 Final selection of codelists and vocabularies for the GeoERA Keyword Thesaurus - at the time of report preparation

Search Category	Group	CodeList	Web	Organization	Person Of contact
APPLIED GEOPHYSIC	INSPIRE_APPLIED_GEOPHYSIC	INSPIRE Codelist BoreholePurpose	http://inspire.ec.europa.eu/codelist/BoreholePurposeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist CurveModelTypeValue	http://inspire.ec.europa.eu/codelist/CurveModelTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist ProfileTypeValue	http://inspire.ec.europa.eu/codelist/ProfileTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist StationTypeValue	http://inspire.ec.europa.eu/codelist/StationTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist SurveyTypeValue	http://inspire.ec.europa.eu/codelist/SurveyTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist SwathTypeValue	http://inspire.ec.europa.eu/codelist/SwathTypeValue	JRC INSPIRE Registry Team	
	ONEGE_APPLIED_GEOPHYSIC	OneGeology-Europe keywords database	https://gemet.bnhelp.cz/thesaurus/getTopmostConcepts?thesaurus_uri=http://www.onegeology-europe.eu/concept/&language=en	Czech Geological Survey	lucie.kondrova@geology.cz
FOSSIL RESOURCES	EarthReML_FOSSIL_RESSOURCE	EarthReML_CommodityCode	http://resource.geosciml.org/classifierscheme/cgi/2016.01/compositioncategory or https://vocabs.ands.org.au/viewById/55	CGI Geoscience Terminology Working group	Tim-McCormick or Oliver Raymond: oliver.raymond@ga.gov.au
	INSPIRE_FOSSIL_RESSOURCE_1	INSPIRE Codelist BoreholePurpose	http://inspire.ec.europa.eu/codelist/BoreholePurposeValue	JRC INSPIRE Registry Team	
	INSPIRE_FOSSIL_RESSOURCE_2	INSPIRE Codelist ClassificationAndQuantificationFrameworkValue	http://inspire.ec.europa.eu/codelist/ClassificationAndQuantificationFrameworkValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist FossilFuelClassValue	http://inspire.ec.europa.eu/codelist/FossilFuelClassValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist FossilFuelValue	http://inspire.ec.europa.eu/codelist/FossilFuelValue	JRC INSPIRE Registry Team	
GEOCHEMISTRY	INSPIRE_GEOCHEMISTRY_1	INSPIRE Codelist EventProcessValue	http://inspire.ec.europa.eu/codelist/EventProcessValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist HydroGeochemicalRockTypeValue	http://inspire.ec.europa.eu/codelist/HydroGeochemicalRockTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist NaturalGeomorphologicFeatureTypeValue	http://inspire.ec.europa.eu/codelist/NaturalGeomorphologicFeatureTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist WaterSalinityValue	http://inspire.ec.europa.eu/codelist/WaterSalinityValue	JRC INSPIRE Registry Team	
	INSPIRE_GEOCHEMISTRY_2	INSPIRE Codelist ProcessingActivityTypeValue	http://inspire.ec.europa.eu/codelist/ProcessingActivityTypeValue	JRC INSPIRE Registry Team	
	INSPIRE_GEOCHEMISTRY_3	INSPIRE Codelist ProfileElementParameterNameValue	http://inspire.ec.europa.eu/codelist/ProfileElementParameterNameValue	JRC INSPIRE Registry Team	
GEOCHRONOLOGY/STRATIGRAPHY	GeoSciML_GEOCHRONO_STRATIGRAPHY_1	CGI - Geologic Time Vocabulary - International Chronostratigraphic Chart - 2017	http://vocabs.ands.org.au/repository/api/lda/csiro/international-chronostratigraphic-chart-2017/2017/collection	Research Vocabularies Australia - Linked Data API	

GEOLOGICAL PROCESSES	GeoSciML_GEOLOGICAL_PROCESSES	GeoSciML_DeformationStyle	https://vocabs.ands.org.au/viewById/46 or http://resource.geosciml.org/classifierscheme/cgi/2016.01/deformationstyle	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_EventEnviroment	https://vocabs.ands.org.au/viewById/59	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_EventProcess	http://resource.geosciml.org/classifierscheme/cgi/2016.01/eventprocess or https://vocabs.ands.org.au/viewById/58	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_GeneticCategory	http://resource.geosciml.org/classifierscheme/cgi/2016.01/geneticcategory	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_GeologicUnitType	https://vocabs.ands.org.au/viewById/50	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_MetamorphicGrade	https://vocabs.ands.org.au/viewById/91	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
	INSPIRE_GEOLOGICAL_PROCESSES	INSPIRE Codelist AnthropogenicGeomorphologicFeatureTypeValue	http://inspire.ec.europa.eu/codelist/AnthropogenicGeomorphologicFeatureTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist EventEnvironmentValue	http://inspire.ec.europa.eu/codelist/EventEnvironmentValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist EventProcessValue	http://inspire.ec.europa.eu/codelist/EventProcessValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist GeologicUnitTypeValue	http://inspire.ec.europa.eu/codelist/GeologicUnitTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist NaturalGeomorphologicFeatureTypeValue	http://inspire.ec.europa.eu/codelist/NaturalGeomorphologicFeatureTypeValue	JRC INSPIRE Registry Team	
HAZARD, RISK AND IMPACT	EarthReML_HAZARD_RI	EarthReML_EnvironmentalImpact	http://resource.geosciml.org/classifier/cgi/environmental-impact	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
	INSPIRE_HAZARD_RI_1	INSPIRE Codelist AnthropogenicGeomorphologicFeatureTypeValue	http://inspire.ec.europa.eu/codelist/AnthropogenicGeomorphologicFeatureTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist NaturalGeomorphologicFeatureTypeValue	http://inspire.ec.europa.eu/codelist/NaturalGeomorphologicFeatureTypeValue	JRC INSPIRE Registry Team	

	INSPIRE_HAZARD_RI_2	INSPIRE Codelist ExposedElementCategoryValue	http://inspire.ec.europa.eu/codelist/ExposedElementCategoryValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist NaturalHazardCategoryValue	http://inspire.ec.europa.eu/codelist/NaturalHazardCategoryValue	JRC INSPIRE Registry Team	
GEOTHERMAL ENERGY	INSPIRE_GEOTHERMAL_ENERGY	INSPIRE Codelist ActiveWellTypeValue	http://inspire.ec.europa.eu/codelist/ActiveWellTypeValue	JRC INSPIRE Registry Team	
HYDROGEOLOGY	INSPIRE_HYDROGEOLOGY_1	INSPIRE Codelist ActiveWellTypeValue	http://inspire.ec.europa.eu/codelist/ActiveWellTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist AquiferMediaTypeValue	http://inspire.ec.europa.eu/codelist/AquiferMediaTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist AquiferTypeValue	http://inspire.ec.europa.eu/codelist/AquiferTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist ConditionOfGroundwaterValue	http://inspire.ec.europa.eu/codelist/ConditionOfGroundwaterValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist HydroGeochemicalRockTypeValue	http://inspire.ec.europa.eu/codelist/HydroGeochemicalRockTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist NaturalObjectTypeValue	http://inspire.ec.europa.eu/codelist/NaturalObjectTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist StatusCodeTypeValue	http://inspire.ec.europa.eu/codelist/StatusCodeTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist WaterPersistenceValue	http://inspire.ec.europa.eu/codelist/WaterPersistenceValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist WaterSalinityValue	http://inspire.ec.europa.eu/codelist/WaterSalinityValue	JRC INSPIRE Registry Team	
	INSPIRE_HYDROGEOLOGY_2	INSPIRE Codelist BoreholePurpose	http://inspire.ec.europa.eu/codelist/BoreholePurposeValue	JRC INSPIRE Registry Team	
	INSPIRE_HYDROGEOLOGY_3	INSPIRE Codelist OtherContaminatingActivityValue	http://inspire.ec.europa.eu/codelist/OtherContaminatingActivityValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist SoilContaminationSpecialisedZoneTypeCode	http://inspire.ec.europa.eu/codelist/SoilContaminationSpecialisedZoneTypeCode	JRC INSPIRE Registry Team	
LITHOLOGY	EARTH_LITHOLOGY	EARTh.	http://linkeddata.ge.imati.cnr.it/resource/EARTh/		
	GBA_LITHOLOGY	GBA Thesaurus Lithology	http://resource.geolba.ac.at/lithology	Geological Survey of Austria	
	GeoSciML_LITHOLOGY_1	GeoSciML_SimpleLithology	https://vocabs.ands.org.au/viewById/56	CGI Geoscience Terminology Working group	
	GeoSciML_LITHOLOGY_2	GeoSciML_CompositionCategory	http://resource.geosciml.org/classifierscheme/cgi/2016.01/compositioncategory	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_GeologicUnitType	https://vocabs.ands.org.au/viewById/50	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_MetamorphicFacies	https://vocabs.ands.org.au/viewById/90	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
	INSPIRE_LITHOLOGY_1	INSPIRE Codelist LithologyValue	http://inspire.ec.europa.eu/codelist/LithologyValue	JRC INSPIRE Registry Team	

MINERAL RESOURCES	EarthReML_MINERAL_RESOURCE	EarthReML_CommodityCode	http://resource.geosciml.org/classifierscheme/cgi/2016.01/compositioncategory or https://vocabs.ands.org.au/viewById/55	CGI Geoscience Terminology Working group	Tim-McCormick or Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_EarthResourceMaterialRole	https://vocabs.ands.org.au/viewById/78	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_EndUsePotential	http://resource.geosciml.org/classifier/cgi/end-use-potential	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_ExplorationResult	https://vocabs.ands.org.au/viewById/77	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_MineralOccurrenceType	https://vocabs.ands.org.au/viewById/76	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_MineStatus	https://vocabs.ands.org.au/viewById/126	CGI Geoscience Terminology Working group opertaing status	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_MiningActivity	http://resource.geosciml.org/classifier/cgi/mining-activity	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_ProcessingActivity	https://vocabs.ands.org.au/viewById/74	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_RawMaterialRole	https://vocabs.ands.org.au/viewById/73	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_ReportingClassificationMethod	https://vocabs.ands.org.au/viewById/125	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_ReserveAssessmentCategory	https://vocabs.ands.org.au/viewById/72	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
	GBA_MINERAL_RESOURCE	GBA Thesaurus Rohstoffgeologie (Raw Material)	http://resource.geolba.ac.at/minres.html	Geological Survey of Austria	thesaurus@geologie.ac.at
	INSPIRE_MINERAL_RESOURCE_1	INSPIRE Codelist EndusePotentialValue	http://inspire.ec.europa.eu/codelist/EndusePotentialValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist ExplorationActivityTypeValue	http://inspire.ec.europa.eu/codelist/ExplorationActivityTypeValue	JRC INSPIRE Registry Team	

		INSPIRE Codelist ExplorationResultValue	http://inspire.ec.europa.eu/codelist/ExplorationResultValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist MineralDepositGroupValue	http://inspire.ec.europa.eu/codelist/MineralDepositGroupValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist MineralDepositTypeValue	http://inspire.ec.europa.eu/codelist/MineralDepositTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist MineralOccurrenceTypeValue	http://inspire.ec.europa.eu/codelist/MineralOccurrenceTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist MineStatusValue	http://inspire.ec.europa.eu/codelist/MineStatusValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist MiningActivityTypeValue	http://inspire.ec.europa.eu/codelist/MiningActivityTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist ProcessingActivityTypeValue	http://inspire.ec.europa.eu/codelist/ProcessingActivityTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist ReserveCategoryValue	http://inspire.ec.europa.eu/codelist/ReserveCategoryValue	JRC INSPIRE Registry Team	
		INSPIRE CodelistCommodityValue	http://inspire.ec.europa.eu/codelist/CommodityCodeValue	JRC INSPIRE Registry Team	
	Minerals4EU_MINERAL_RESOURCE	Minerals4EU Metadata Keywords	http://m4eu.geology.cz/codelist	Czech Geological Survey	egdi.metadata@geology.cz
STRUCTURAL GEOLOGY	EARTH_STRUCTURAL	EARTh.	http://linkeddata.ge.imati.cnr.it/resource/EARTh/		
	GeoSciML_STRUCTURAL	GeoSciML_ContactType	https://vocabs.ands.org.au/viewById/52 or http://resource.geosciml.org/classifierscheme/cgi/2016.01/contacttype	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_DeformationStyle	https://vocabs.ands.org.au/viewById/46 or http://resource.geosciml.org/classifierscheme/cgi/2016.01/deformationstyle	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_FaultMovementSense	https://vocabs.ands.org.au/viewById/63	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_FaultType	https://vocabs.ands.org.au/viewById/68 or http://resource.geosciml.org/classifier/cgi/faulttype	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		GeoSciML_GeologicUnitType	https://vocabs.ands.org.au/viewById/50	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
	INSPIRE_STRUCTURAL	INSPIRE Codelist EventEnvironmentValue	http://inspire.ec.europa.eu/codelist/EventEnvironmentValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist EventProcessValue	http://inspire.ec.europa.eu/codelist/EventProcessValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist FaultTypeValue	http://inspire.ec.europa.eu/codelist/FaultTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist FoldProfileTypeValue	http://inspire.ec.europa.eu/codelist/FoldProfileTypeValue	JRC INSPIRE Registry Team	

SUBSURFACE MANAGEMENT	EarthReML_SUBSURFACE_MANAGEMENT	EarthReML_ExplorationActivityType	https://vocabs.ands.org.au/viewById/79	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
		EarthReML_ReserveAssessmentCategory	https://vocabs.ands.org.au/viewById/72	CGI Geoscience Terminology Working group	Oliver Raymond: oliver.raymond@ga.gov.au
	INSPIRE_SUBSURFACE_MANAGEMENT_2	INSPIRE Codelist HILUCSValue	http://inspire.ec.europa.eu/codelist/HILUCSValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist LevelOfSpatialPlanValue	http://inspire.ec.europa.eu/codelist/LevelOfSpatialPlanValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist SupplementaryRegulationValue	http://inspire.ec.europa.eu/codelist/SupplementaryRegulationValue	JRC INSPIRE Registry Team	
	INSPIRE_SUBSURFACE_MANAGEMENT_5	INSPIRE Codelist LayerTypeValue	http://inspire.ec.europa.eu/codelist/LayerTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist OtherContaminatingActivityValue	http://inspire.ec.europa.eu/codelist/OtherContaminatingActivityValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist RiskAssessmentStageValue	http://inspire.ec.europa.eu/codelist/RiskAssessmentStageValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist RiskReceptorValue	http://inspire.ec.europa.eu/codelist/RiskReceptorValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist RiskTypeValue	http://inspire.ec.europa.eu/codelist/RiskTypeValue	JRC INSPIRE Registry Team	
		INSPIRE Codelist SoilContaminationSpecialisedZoneTypeCode	http://inspire.ec.europa.eu/codelist/SoilContaminationSpecialisedZoneTypeCode	JRC INSPIRE Registry Team	
	ONEGE_SUBSURFACE_MANAGEMENT	OneGeology-Europe keywords database	https://gemet.bnhelp.cz/thesaurus/getTopmostConcepts?thesaurus_uri=http://www.onegeology-europe.eu/concept/&language=en	Czech Geological Survey	lucie.kondrova@geology.cz

Additionally, terms from the KINDRA and VOGERA keyword lists were integrated as well as HIKE keywords needed for their project.

Related to the HIKE project a subset of 103 (project specific) keywords were considered regarding the GeoERA Keyword Thesaurus in order to support the “HIKE Knowledge Sharepoint”:

Atmospheric Impacts, Biosphere impacts, Buildings and Infrastrucure damage impacts, Buildings collapse, Critical facilities out of use or malfunction, Death, Disruption of transportation, Economic Impacts, Employment rate, Environmental Impacts, Human Health and life Impacts, Infrastructure failure, Physical Injuries, Psychological impacts, Water Impacts, Instaneous deformation, emissions, explosion, facility failure, fluid spills, generated seismicity, gradual deformation, leakage and migration along constrained path, leakage and migration along unconstrained path, leakage and migration, natural seismicity, triggered seismicity, induced seismicity, seismicity, surface deformation, extraction, Anthropogenic causes, Atmospheric causes, CO2 sequestration, Compressed air energy storage, Conventional gas production, Conventional oil production, Cyclic injection and extraction, Drilling, Drought, EOR, Earthquakes, Engineering activities, Extraction, Extreme temperature, Flood, Fracking, Geothermal doublet production, Geothermal production, Hydrogen storage, Hydrogeological causes, Injection, Lightening, Mass mouvements, Natural causes, Natural gas storage, Nitrogen storage, Nuclear waste storage, Rock falls, Salt solition mining, Solid earth, Storm, Subsurface mining, Surface mining, Tornado, Tsunami, Tunnel building, Unconventional gas production, Unconventional oil production, Underground thermal storage, Volcanic eruption, Waste water injection, landslides, subsurface mining, incineration, Water production, Analytical approaches, Case studies, Datasets, Geophysical acquisition, Instruments, Interpreted datasets, Measured parameters, Methods, Modeled parameters, Modelling, Monitoring, Protocols, Satellite acquisition, experimental approaches, Basins, Lithology, Lithostratigraphy, Oregenies, Platforms, Regional geological definitions, Rock types, Shields, Stratigraphy, Structural Elements, Tectonic setting, chronostratigraphy, geomorphology.