HIKE

GeoERA

GEO-ENERGY

Hazard and Impact Knowledge for Europe

Deliverable D5.2b

User manual Fault Database + Knowledge base

Author and affiliation:
R. van Ede [TNO]

E-mail of lead author:
Rob.vanEde@tno.nl

Version: 2021.10.18

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. Scientific work is co-funded by the Geological Surveys and national funds allocated for science within the period 2018-2021.

GeoERA

Deliverable Data		
Deliverable number	D5.2b	
Dissemination level	Public	
Deliverable name	User manual Fault Database + Knowledge base	
Work package	WP5 / WP2 / WP4	
Lead WP/Deliverable beneficiary	GBA / TNO	
Deliverable status		
Author(s)	2021/07/07	R. van Ede [TNO]
Reviewed by (WP Lead)	2021/10/13	E. Hintersberger [GBA]
Approved by (Project Lead)	2021/10/18	Hans Doornenbal [TNO]

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Document background and scope

Energy, mineral resources and groundwater are intrinsically related to the UN Sustainable Development Goals. Exploitation of these resources is not always without risk to citizens and the environment. The HIKE project aims to stimulate the development of common information repositories and a knowledge sharing infrastructure in order to support induced hazard and risk assessments at the geological survey organizations and other research institutions. Ultimately this could lead to a more effective and uniform approach on management of subsurface risks.

This document presents an overview of the procedures for data delivery as well as data retrieval, in the form of a "quick start guide". For more details about the data model and attributes, the Fault Data Catalogue can be consulted (GeoERA_HIKE_D2.1b).

2 DATA DELIVERY

During the course of the HIKE project, the partners have delivered their fault data and relevant documents to the project team, who have taken care of standards-adherence and performed data quality checks before incorporating the data into the fault database. The compiled European Fault Database was then delivered to the EGDI portal for dissemination.

After the project, updating then-loaded data will be possible by delivering a new GeoPackage to the EGDI support team. The entire existing dataset from the partner will then be removed from the European Fault Database before the newly delivered dataset is loaded and made available.

2.1 European Fault Database

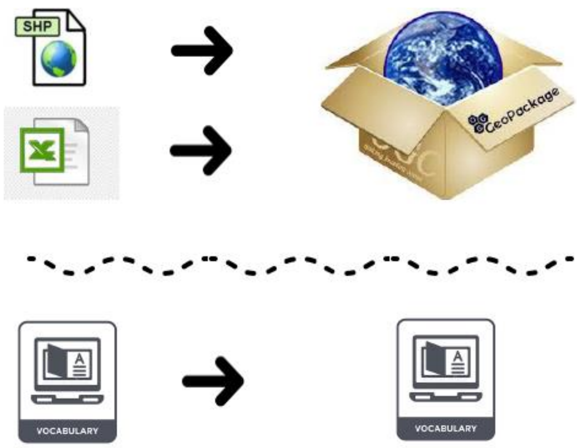
Data was delivered on a per-partner basis, in the form of a GeoPackage, containing a geometry table and an attribute table, both following strict templates, which were included in the zip-file that was shared with the partners. The templates and examples are available online here.

This chapter is based on the working document "Preparation of HIKE data for delivery" that was shared with all partners during the development of the European Fault Database.

2.1.1 Format

The HIKE fault data shall be delivered inside a GeoPackage container.
Fault data consist of the geometry with attributes relevant for the geometry, named FaultGeometries*Country,* as well as the fault attributes spreadsheet, named FaultAttributes*Country*.

During the project, the project vocabulary spreadsheet could be delivered separately. After the project, there is no possibility for changing the project vocabulary.



The contents of these three files has been defined in the technical specifications (GeoERA_HIKE_D5.1a) that were shared earlier. The updated specifications and details are listed in the Fault Data Catalogue (GeoERA_HIKE_D2.1b). The most important properties for the fault data are outlined below.

2.1.2 Data content – geometry data

Geometry type: 2D line (can be multiline/polyline).
Coordinate reference system: ETRS89 / LCC Europe (EPSG:3034 (see GeoERA_HIKE_D2.1b § 3.2.1).
Attributes: All attributes belonging to the geometry are described in GeoERA_HIKE_D2.1b § 4.9.
Id for linking to the other parts: "Id" column. Several geometries can relate to the same fault and therefore share the same Id.

2.1.3 Data content – attribute data

Excel template sheet as shared in the annex of this document (zip file) on the HIKE project website. All textual values are predefined based on international standards and shall only contain the values as defined in the template sheet. Definitions for attributes are available as tooltips. Furthermore, all fault attributes and the associated values are described in GeoERA_HIKE_D2.1b, § 4.
Id for linking to other parts: "Id" column. Several attribute rows can relate to the same fault geometry, for example where time-dependent attribute sets (e.g. varying fault type) for several geochronological eras are available.

2.1.4 Cookbook for actual creation of the GeoPackage

The most straightforward way to do this, is with the open-source GIS application QGIS (https://www.qgis.org/en/site/forusers/download.html), following this procedure (note: this will work for QGIS version 3.8 and higher):

- Create a new Project
- Add the shapefile (1): Layer > Add Layer > Add Vector Layer… . Choose Source of type ESRI Shapefile. (Note that you can add other kinds of geometry files if you don't use shapefiles) You can also just drag and drop a geometry file into the Layers Window of QGIS. Make sure that the CRS for the shapefile is EPSG:3034 - ETRS89 / LCC Europe.
- Add the fault attributes Excel file (2) by also dragging and dropping it into the Layers Window of QGIS. QGIS will ask which sheet from the Excel file to add. Choose FAULTS.
- Choose Processing > Toolbox to open the GeoProcessing Toolbox. Then choose Database > Package Layers (double-click). Click the Input layers selection button, check both your layers and then OK. Click the Destination GeoPackage selection button > Save to file... and choose a File name (suggested name: InstituteName_YYYY-MM-DD.gpkg). Click Save.

Alternatively, a geopackage can be created in ArcGIS via the Data Management tool "Create SQLite Database" (https://desktop.arcgis.com/de/arcmap/latest/tools/data-management-toolbox/create-sqlite-database.htm):

- Create a new Project
- Choose in the ArcToolbox > Data Management Tools > Workspace > Create SQLite Database (double-click). Choose a File name (suggested name: InstituteName_YYYY-MM-DD.gpkg). Choose from the Spatial Type drop-down menu the option "GEOPACKAGE_1.2". Click OK.

- Open ArcCatalog and add the shapefile (1): Right click on the generated geopackage > Import > FeatureClass (Single). Choose Source of type ESRI Shapefile. (Note that you can add other kinds of geometry files if you don't use shapefiles). Make sure that the CRS for the shapefile is EPSG:3034 - ETRS89 / LCC Europe. Click OK.
- Add the fault attributes excel file (2) by clicking right again on the geopackage, > Import > Table (Single). In "Input Rows" you will be asked which sheet from the Excel file to add. Choose FAULTS$. Name the output table "FAULTS". Click OK.

2.1.5 Project vocabulary

An excel template along with guidance on how to populate it was circulated and support was provided. Since the project vocabularies cannot be changed after the project, instructions are not included in this user manual.

2.1.6 Delivery / update of the data

After final project data delivery (later than August 2021), partners will be able to send an updated GeoPackage for their region to EGDI support (contact@europe-geology.eu). All previous data from that partner will then be replaced by the newly delivered data.
Note that in order for the update to succeed, it is extremely important that the delivered GeoPackage strictly adheres to the specifications as described earlier in this chapter. Any deviation from the specifications might break functionality of the fault database for the specific dataset. Special care needs to be taken regarding the use of existing fault Id's: The Id is used to join the vocabulary to the fault data. Therefore, if an existing fault Id is used in the updated dataset, this fault will be linked to the existing vocabulary item.

2.1.6.1 Partner data update workflow for EGDI support

First some background:
The HIKE GeoPackage consists of the following components:

- FaultAttributes table per country / partner – Holds the fault attributes in HIKE format
- FaultGeometries table per country / partner – Holds the fault geometries in HIKE format
- CodelistURIs table per code list – Holds the valid code list values and corresponding URI's
- VocabConnection table – Holds the project vocabulary URI's and HIKE fault Id's
- HIKE_FAULTDB_DETAIL view – Combines all of the above for the detailed map layer
- HIKE_FAULTDB_OVERVIEW view – Combines all of the above for the overview map layer
- HIKEPolygonGeometries table – Holds area features supporting the European Fault Database
- GeoPackage system tables

In order to incorporate newly delivered partner data into the HIKE European Fault Database, the contents of only 2 tables (FaultAttributes*Country* and FaultGeometries*Country*) in the HIKE GeoPackage need to be replaced with the newly delivered data. The database views inside the GeoPackage will then take care of the inclusion in the map layers without any extra input.
Note that the contents of the newly delivered data will need to be strictly following the HIKE specifications (as described in GeoERA_HIKE_D2.1b). If invalid values are loaded, the links to

INSPIRE code lists and HIKE project vocabularies will not be established for the offending records, and the harmonization of the European Fault Database as a whole will be compromised.

Once the GeoPackage is updated, it will need to be reloaded onto the EGDI platform and the 2 views need to be published under the exact same name as before, so that the online maps will automatically include the new dataset. Also, the GeoPackage download link in the HIKE Wordpress page (https://geoera.eu/projects/hike10/faultdatabase/) needs to be updated to point at the most recent version.

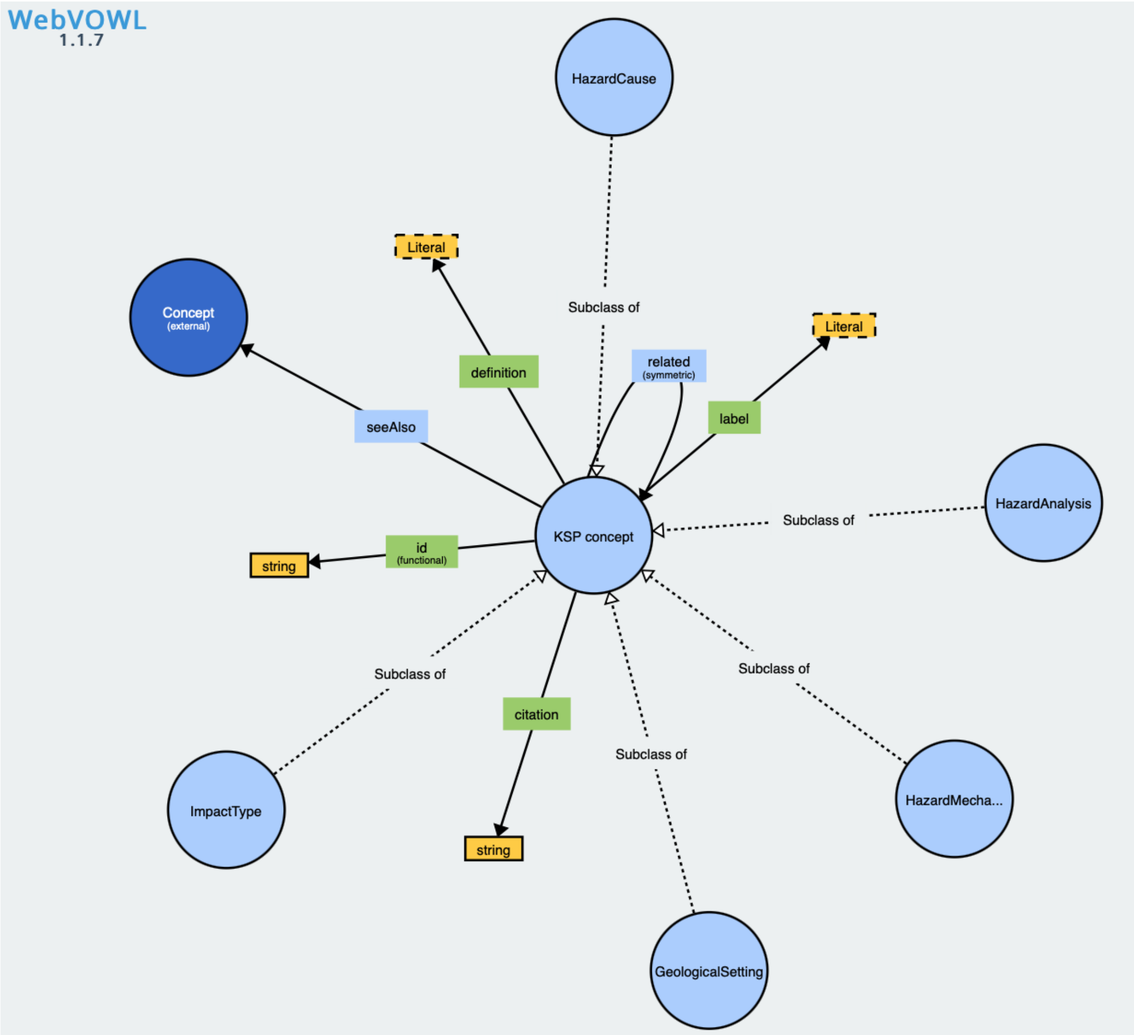
2.2 Knowledge Base

2.2.1 Uploading documents

For uploading documents to the HIKE Knowledge Base, an account to the [EGDI administration module]() is needed. This account can be requested through the project lead ([Serge van Gessel]()). Once logged in, documents and related files can be uploaded.

2.2.1.1 Keywords

Since the HIKE Knowledge Base relies on keywords for presenting related documents, it is important that the keywords that are included while uploading the documents are the ones that are included in the HIKE Knowledge Base. All applicable keywords are organized in five subcategories:



Overview of the keywords that are included in the Knowledge Base:

Category	Keyword
https://data.geoscience.earth/def/hike#HazardCause	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
https://data.geoscience.earth/def/hike#HazardMechanism	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

Category	Keyword
https://data.geoscience.earth/def/hike#HazardAnalysis	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
https://data.geoscience.earth/def/hike#ImpactType	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

The full definition of the keywords, including relationships and other languages can be retrieved by running following SPARQL query on the keyword thesaurus SPARQL endpoint (https://resource.geolba.ac.at/PoolParty/sparql/keyword):

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
PREFIX hike:<https://data.geoscience.earth/def/hike#>

SELECT
?kspClass ?uri
(MIN(COALESCE(?hLe,?sLe)) AS ?enLabel) #override missing hike:label
(GROUP_CONCAT(DISTINCT(CONCAT(STR(?hL),"(",lang(?hL),")")); separator = '|') as
?allKspLabels)
(GROUP_CONCAT(DISTINCT(CONCAT(STR(?sL),"(",lang(?sL),")")); separator = '|') as
?allSkosLabels)
(MIN(?i) AS ?id)
(GROUP_CONCAT(DISTINCT(CONCAT(STR(?d),"(",lang(?d),")")); separator = '|') as
?definitions)
(GROUP_CONCAT(DISTINCT(STR(?c)); separator = '|') as ?citations)
(GROUP_CONCAT(DISTINCT(CONCAT(STR(?nL),"[",STRAFTER(STR(?n),"keyword/"),"]")); separator
= '|') as ?sameClassNarrower)
(GROUP_CONCAT(DISTINCT(CONCAT(STR(?saL),"[",STRAFTER(STR(?sa),"keyword/"),"]"));
separator = '|') as ?kspSeeAlso)
(GROUP_CONCAT(DISTINCT(CONCAT(STR(?sdrL),"[",STRAFTER(STR(?sdr),"keyword/"),"]"));
separator = '|') as ?skosDirectlyRelated)
(GROUP_CONCAT(DISTINCT(CONCAT(STR(?hrL),"[",STRAFTER(STR(?hr),"keyword/"),"]"));
separator = '|') as ?kspRelated)

WHERE {
VALUES ?p {skos:prefLabel skos:altLabel}
VALUES ?dr {skos:narrower skos:broader skos:related}
?uri a hike:KSP; rdf:type ?kspClass; ?p ?sL; skos:prefLabel ?sLe; ?dr ?sdr #everything
what sticks on uri
  FILTER(?kspClass != skos:Concept && ?kspClass != hike:KSP)
  FILTER(lang(?sLe)="en")
OPTIONAL{?uri hike:label ?hL; hike:label ?hLe
  FILTER(lang(?hLe)="en")}
OPTIONAL{?uri hike:id ?i}
OPTIONAL{?uri hike:definition ?d}
OPTIONAL{?uri hike:citation ?c}
OPTIONAL{?uri skos:narrower+ ?n . ?n a ?kspClass; skos:prefLabel ?nL
  FILTER(lang(?nL)="en")} #all narrower concepts of same KSP class
OPTIONAL{?uri hike:seeAlso ?sa . ?sa skos:prefLabel ?saL
  FILTER(lang(?saL)="en")} #similar concepts specified by hike
OPTIONAL{?sdr skos:prefLabel ?sdrL
  FILTER(lang(?sdrL)="en")} #skos concepts within semantic distance of 1
OPTIONAL{?uri hike:related ?hr . ?hr rdf:type ?relClass; skos:prefLabel ?hrL
  FILTER(lang(?hrL)="en") FILTER(?relClass != ?kspClass)} #relations between KSP classes
specified by hike logic
}

GROUP BY ?uri ?kspClass
ORDER BY DESC (?kspClass)
```

2.2.2 Background

This paragraph was copied from the online EGDI documentation (http://egdi-public.gitlabpages.geus.dk/egdi-documentation/#/main-content/DeliveringDataToEGDI) to provide some background to the document repository:

2.2.2.1 Delivery

Several projects want to deliver unstructured data such as documents (with or without spatial reference), pictures / images and tabular data. For these data a document repository has been developed. It is possible to upload / register the following types of unstructured data to the EGDI platform:

- *PDF files*
- *DOI reference to documents*
- *Pictures (png, jpg, jpeg)*
- *Tabular data in CSV format*

The repository is intended to store files that the different projects want to make available for the end users. The files uploaded to the repository can be made available from the EGDI Web GIS through links to the file. The content of the files with tabular data will not be imported into the EGDI database and cannot be used in combination with the other data in the database.

Metadata for unstructured data are stored in the central EGDI database (not in the EGDI Metadata Catalogue as metadata for spatial data and services are). These metadata include title, author, keywords, abstract, (spatial extent). The possible metadata types will differ for the different types of data. Where possible the metadata will be extracted from the uploaded file (relevant for PDF files and pictures) and the user will be able to add and change the different metadata elements.

All the files uploaded to the document repository will return a URI (a static URL) that can be used to refer to the file. All files uploaded to the repository can be accessed through a homepage where all files of the different types can be listed. This list can be filtered by project, file type etc. For users logged into the EGDI administration module it will also be possible to edit the metadata / delete files.

2.2.2.2 Searching

The EGDI search system will be able to search through the metadata for all files uploaded to the platform. If PDF files are readable also the content of the files searched. If the uploaded file has been geocoded they can also be found based on their location. If a document is registered through a DOI only its metadata can be searched (these metadata can contain its location).

2.2.2.3 Availability

Besides being available through the platform all files uploaded to the platform will be available through an URL. You will get this URL when the document has been uploaded.

3 DATA USAGE

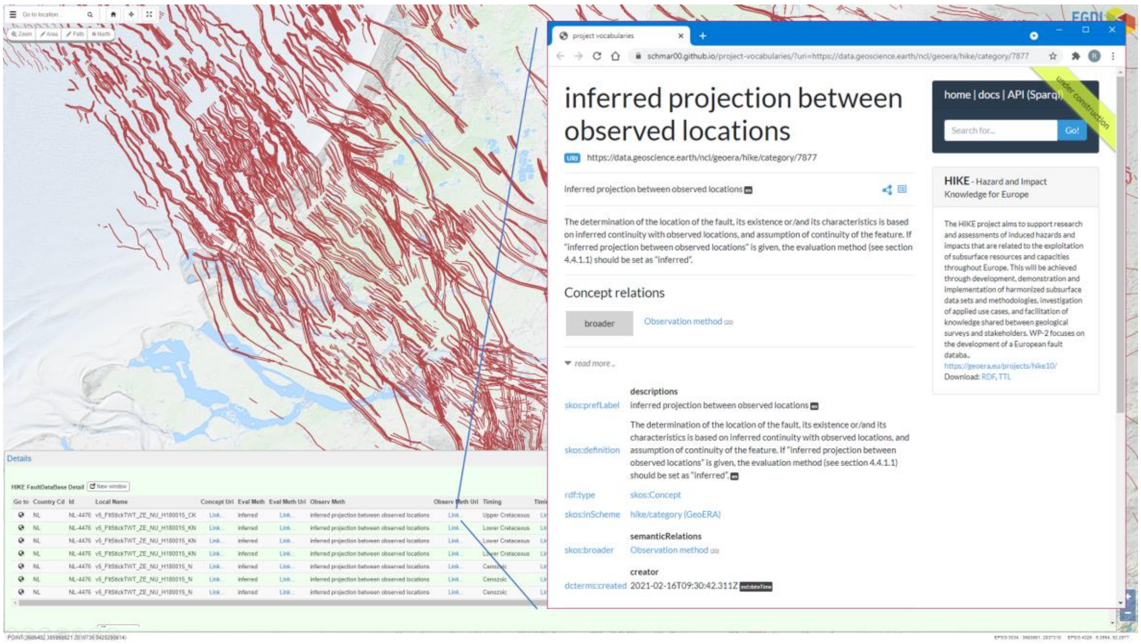
3.1 European Fault Database

The European Fault Database is available through the GeoERA / EGDI portal for online viewing as well as for downloading at https://geoera.eu/projects/hike10/faultdatabase/. The online viewer is set-up for quick and easy data discovery and for looking at the data in combination with pre-defined relevant other content. For more advanced analysis of the data, the complete underlying GeoPackage database is available for download. Instructions for use for both of these options are given below.

3.1.1 Online (EGDI viewer)

3.1.1.1 Map

The online map viewer is configured to show the data at two levels of detail: Overview and Detail. Depending on the zoom-level, the switch between these two levels of detail is made automatically at the 1:1million scale.



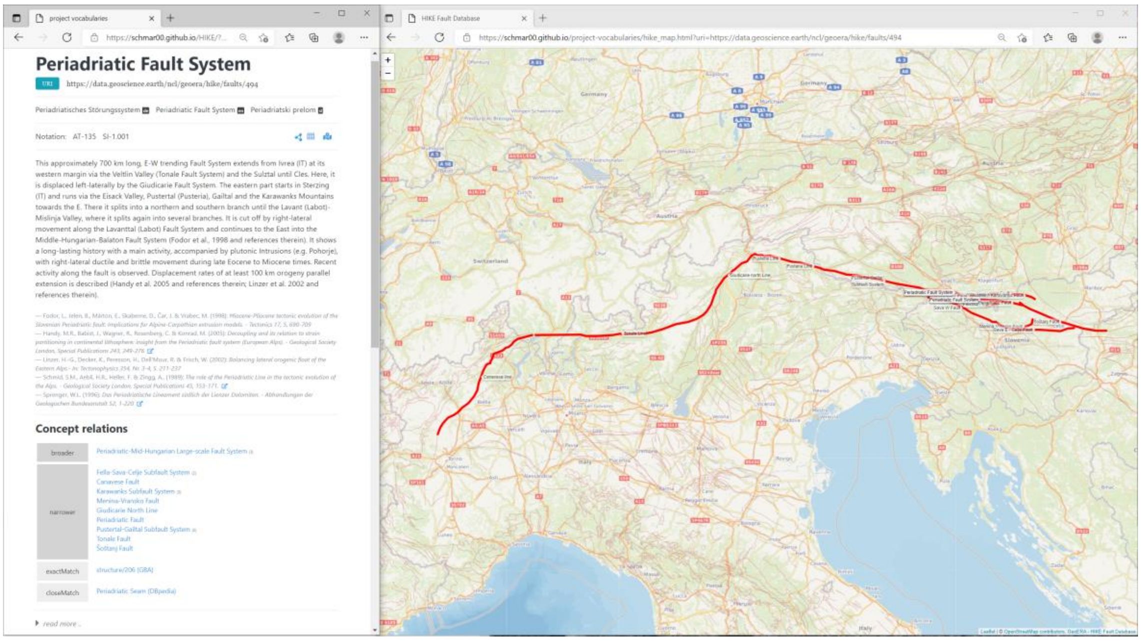
For ease-of-use, geometry- and attribute tables are already joined together, so that clicking on a fault on the map results in the opening of an attribute table that contains all the attribute information that is available for that particular fault.

Attributes that are populated based on code-list values have a neighbouring attribute “attributeName_URI”. This URI attribute holds a clickable link to the corresponding value and redirects the user to the official landing page for that value. Within HIKE, this is either to a CGI, INSPIRE or a HIKE-managed vocabulary. Through this link, the exact meaning of the value can be retrieved. In the same way, a link to the description of the named fault (or its hierarchically higher level) in the Project Vocabulary is available through the CONCEPT_URI attribute.

Finally, each feature has a Metadata attribute that gives access to the partner-specific metadata record and to the general one describing the whole European Fault Database in cases where no partner-specific metadata exists.

3.1.1.2 Semantic Network

Next to the map, the HIKE Project Vocabulary can also be used as an entry-point for exploring the contents of the European Fault Database: The faults, fault sets, large scale fault systems etcetera can be browsed on https://geoera.eu/projects/hike10/semanticnetwork/ . When landed on an object, it is possible to highlight it on the map by clicking the HIKE map icon. In this way, it is possible to see which faults are part of a certain large scale fault system.



In total, the HIKE project has created three different vocabularies that are interlinked via SKOS principles (https://www.w3.org/TR/skos-reference).

The first scheme contains the national/regional inventory of actual fault objects that form the European FDB, defined by their geographic extend and name the European FDB and delivered by each GSO partner. The description provides additional information to the structured fault data of the European FDB and have -in certain instances- been related to similar tectonic boundary objects in adjacent countries. In total, over 2500 fault descriptions from all partner GSOs have been included, making it the largest vocabulary of named fault objects worldwide. Polyhierarchy is possible, i.e., a fault can belong to several broader concepts, and vice versa, a fault system can consists of several narrower concepts. Under the “read more” arrow, technical information in form of all used triples is provided. For detailed explanation, please refer to GeoERA GIP-IP D4.3. The paperclip provides links to other online resources, e.g. the same fault represented in other European fault databases, or additional references. A detailed description of the interface elements is available in GeoERA HIKE D2.1b, § 5.3.

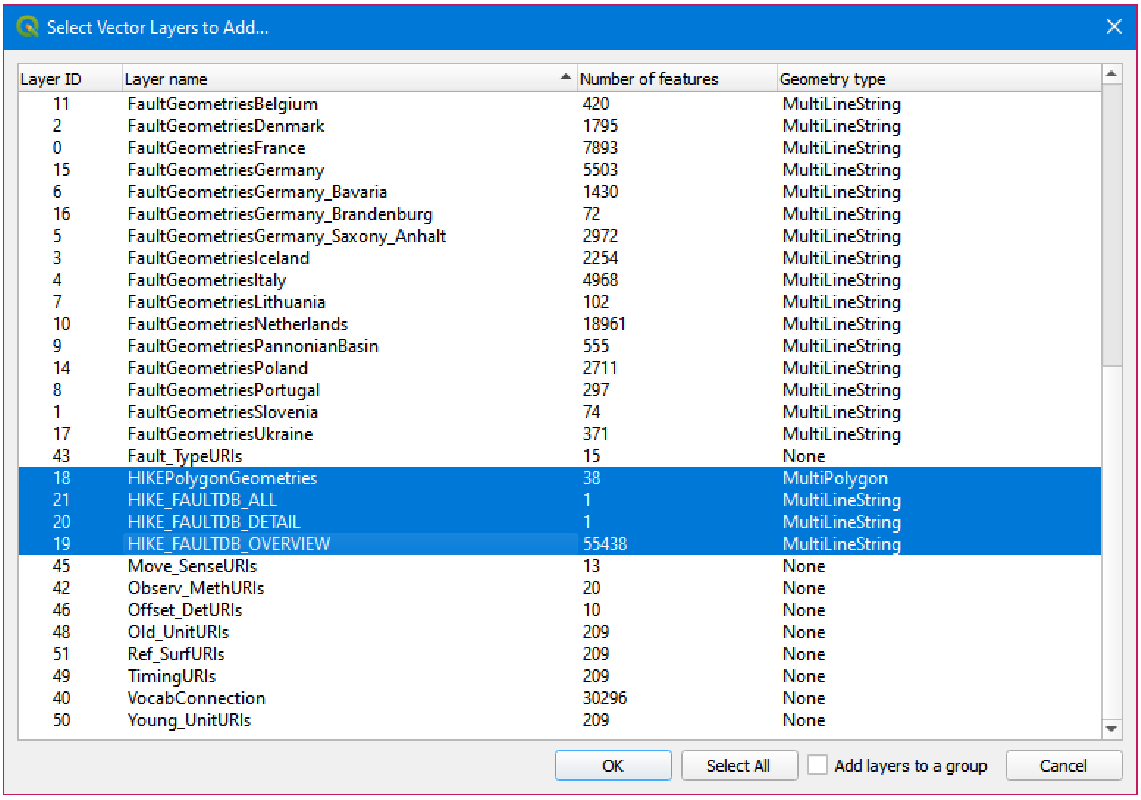
The second vocabulary defines the general terms used in the HIKE project for fault classification and ranking. The different ranking concepts are interchangeable, e.g. a fault domain can be subdivided either into fault sets, fault systems or fault subdomains, depending on the regional/local circumstances. They are also listed in GeoERA HIKE D2.1b, § 5.2. The third vocabulary part delivers a detailed list of all attribute parameters used in the European FDB with all options. Where possible, the parameter lists are semantically related to already existing vocabularies and codelists (e.g, CGI and INSPIRE). For a detailed description and background information, see GeoERA HIKE D2.1b, § 4. These vocabulary entries are directly linked through the _URI attributes in the European Fault Database.

3.1.2 Offline (GeoPackage download)

The downloadable GeoPackage (in fact a SQLite database) contains all fault data that was delivered to the HIKE project, in its original form: both a geometry- and attribute table per partner as well as several mapping tables and a number of views through which all parts are wired together: It is these views that join the geometry- and attribute tables from all partners together and take care of the links to the Project Vocabularies and code-lists through which the dataset is enriched. The structure is as follows (most important items **in bold**):

- **HIKE_FAULTDB_*VIEWS*** – Views with all fault data + Linked Data URI's combined. These views are intended as the main data layers for the European Fault Database
- **HIKEPolygonGeometries** – Optional polygon layer containing shear zones
- FaultAttributes*Partner* – The fault attribute table as delivered – one for each partner
- FaultGeometries*Partner* – The fault geometry table as delivered – one for each partner
- PropertyURIs – Helper mapping tables for linking attribute values to code list URI's
- VocabConnection – Helper mapping table for linking HIKE fault Id's to project vocabulary

When the views are added as layers to QGIS, the included styles (including scale dependency) are automatically applied. There is a view containing all faults at all scales, but there are also separate views for overview scale and detailed scale:

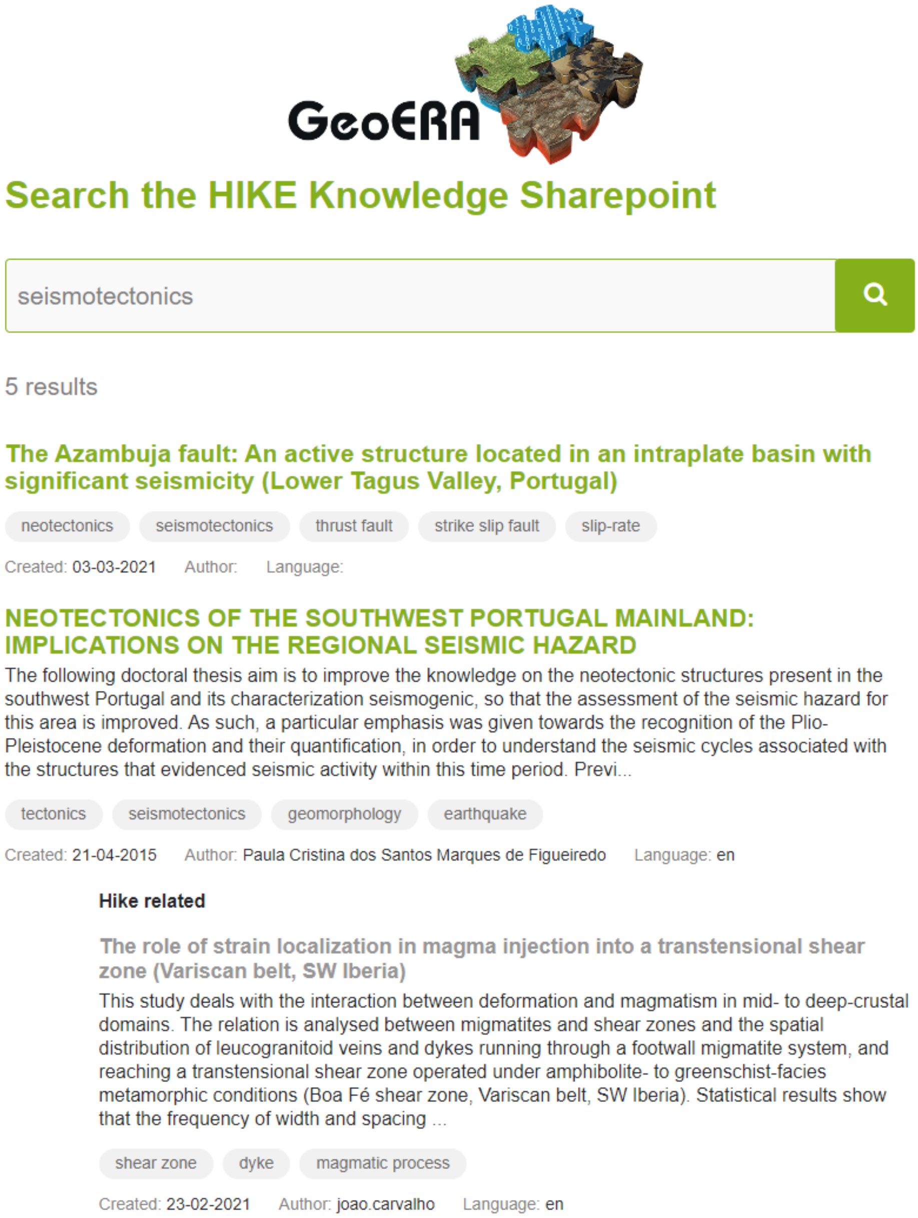


Layer ID	Layer name	Number of features	Geometry type
11	FaultGeometriesBelgium	420	MultiLineString
2	FaultGeometriesDenmark	1795	MultiLineString
0	FaultGeometriesFrance	7893	MultiLineString
15	FaultGeometriesGermany	5503	MultiLineString
6	FaultGeometriesGermany_Bavaria	1430	MultiLineString
16	FaultGeometriesGermany_Brandenburg	72	MultiLineString
5	FaultGeometriesGermany_Saxony_Anhalt	2972	MultiLineString
3	FaultGeometriesIceland	2254	MultiLineString
4	FaultGeometriesItaly	4968	MultiLineString
7	FaultGeometriesLithuania	102	MultiLineString
10	FaultGeometriesNetherlands	18961	MultiLineString
9	FaultGeometriesPannonianBasin	555	MultiLineString
14	FaultGeometriesPoland	2711	MultiLineString
8	FaultGeometriesPortugal	297	MultiLineString
1	FaultGeometriesSlovenia	74	MultiLineString
17	FaultGeometriesUkraine	371	MultiLineString
43	Fault_TypeURIs	15	None
18	HIKEPolygonGeometries	38	MultiPolygon
21	HIKE_FAULTDB_ALL	1	MultiLineString
20	HIKE_FAULTDB_DETAIL	1	MultiLineString
19	HIKE_FAULTDB_OVERVIEW	55438	MultiLineString
45	Move_SenseURIs	13	None
42	Observ_MethURIs	20	None
46	Offset_DetURIs	10	None
48	Old_UnitURIs	209	None
51	Ref_SurfURIs	209	None
49	TimingURIs	209	None
40	VocabConnection	30296	None
50	Young_UnitURIs	209	None

3.2 Knowledge Sharepoint

3.2.1 Documents

The entry point for the HIKE Knowledge Sharepoint is a free-text search field. It is available through https://geoera.eu/projects/hike10/knowledgesharepoint/ . The contents of all documents that are present in the system are searched for the words that are entered in the search field. Apart from the documents that are found directly, all documents related though the attached keywords are also returned and shown at an indent and under the heading "Hike related". The keywords on which the documents are related are also shown.
Clicking on a document starts a download, after which the document can be read.



3.2.2 Keywords

The HIKE Knowledge Sharepoint keywords are part of the GeoERA Keyword Thesaurus. They were loaded following the hierarchical context that was created specifically for the HIKE Knowledge Base: The relations from one keyword to the next, as well as definitions and accompanying citations are stored. Relation description, definition of the keywords and background information are provided in GeoERA HIKE D4.2 and GeoERA HIKE D4.3.