EGU General Assembly 2020

Semantic harmonization of geoscientific data using Linked Data and project specific vocabularies

M. Schiegl

GBA - Geological Survey of Austria

Chat time: Monday, 4 May 2020, 08:30–10:15

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166

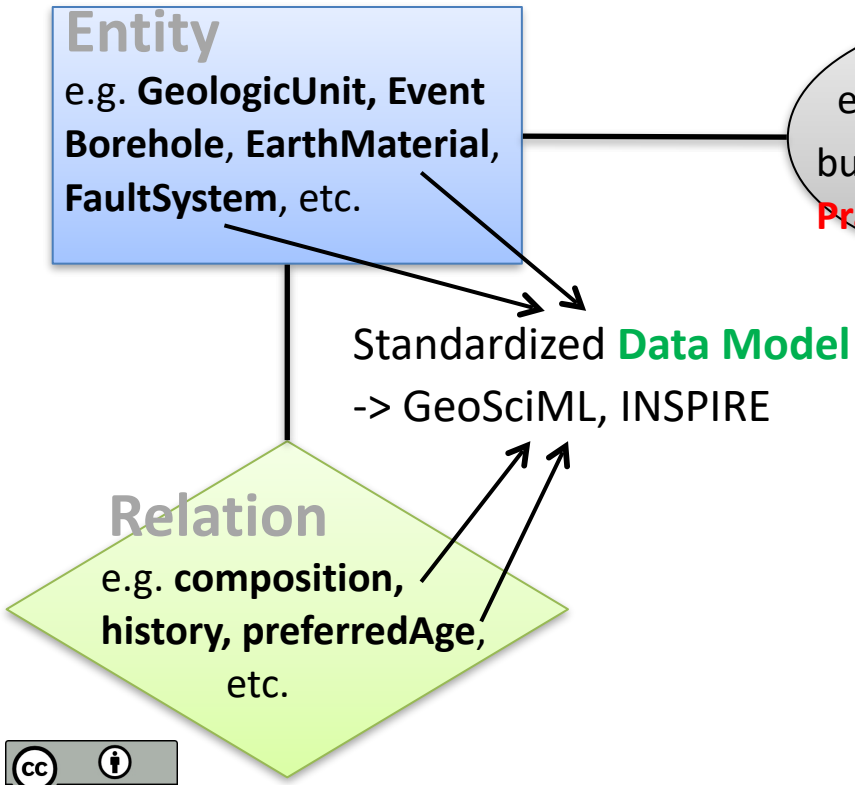
GeoERA INFORMATION PLATFORM

What is Semantic Harmonization?

- making datasets and their attribute **data consistent** and **compatible**, relating to the **meaning in language** and logic
- **Examples**: *Werfen Formation (Italy) – Werfener Schichten (Austria)* *(=quasi synonym)*, *or Zementmergelserie* *(=polysem)*, *silty sandy limestone* *(=combination term)*, *concept definitions of Ice Ages or the Paratethys Stage System, named groundwater bodies or fault systems*
- **Clarification of the meaning** -> bibliography, relations, links,..
- Linked context (knowledge base) -> semantic interoperability
- RDF triple store (Linked Data) + GIS/SQL database (relational)
- But use international standards like **INSPIRE codelists** when there is no need to clarify scientific terms!

harmonizing project data

and metadata !!



Semantic Harmonization

with **knowledge representations** (Controlled Vocabularies, Linked Data) for attribute labels with principle lack of international standards
e.g. national-language, regional and historical attribute labels

GeoERA Project, part GIP-P WP4

GeoERA **project vocabularies** are collections of (linguistic labeled) scientific concepts, stored in named graphs and concept schemes according to different modeling approaches. They are described in SKOS/RDF plus common metadata and properties with focus on scientific reusability.

GeoERA vocabulary concepts can be applied to following types of concepts:

1. Terms describing geoscientific feature types or properties (schema-level)
2. Terms named by classifications or prototype theory's
3. Combination terms for geologic features (like map legend items)
4. Located and named occurrences of geoscientific types or properties (instance-level)

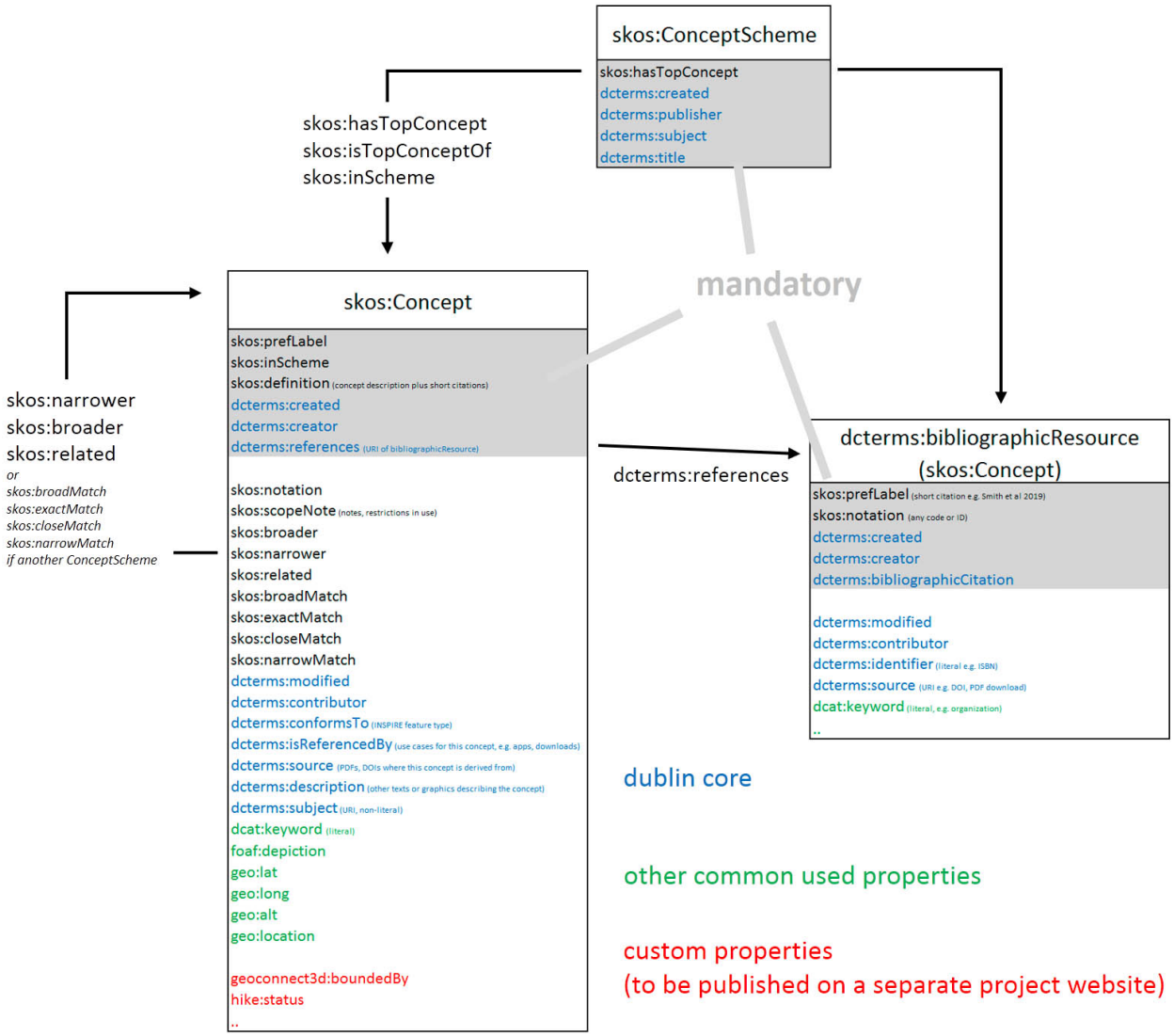
These kind of vocabularies can be used both to harmonize and consequentially connect datasets and to extend standards like INSPIRE codelists.

GeoERA Project: https://geoera.eu
WP4 documentation: https://github.com/GeoEra-GIP/WP4-Semantics
WP4 report: https://geoera.eu/wp-content/uploads/2019/11/D4.3-GeoERA-Project-Vocabularies.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166

GeoERA INFORMATION PLATFORM

Project Vocabularies – data model (GBA/SGU)



Project Vocabularies – Excel template (GBA)

scheme	concept	concept	concept	concept	concept
Tectonic Boundaries Classification Scheme					
	Tectonic Boundary				
		Fault Set			
		Nappe Boundary			
		Large-scale Fault System			
			Fault System		
				Fault	
					Subfault
				Shear zone	
		Subfault System			

Concept scheme label

Top concept - Uppermost concept hierarchy (level 1)

Narrower concept label – next hierarchy (level 2)

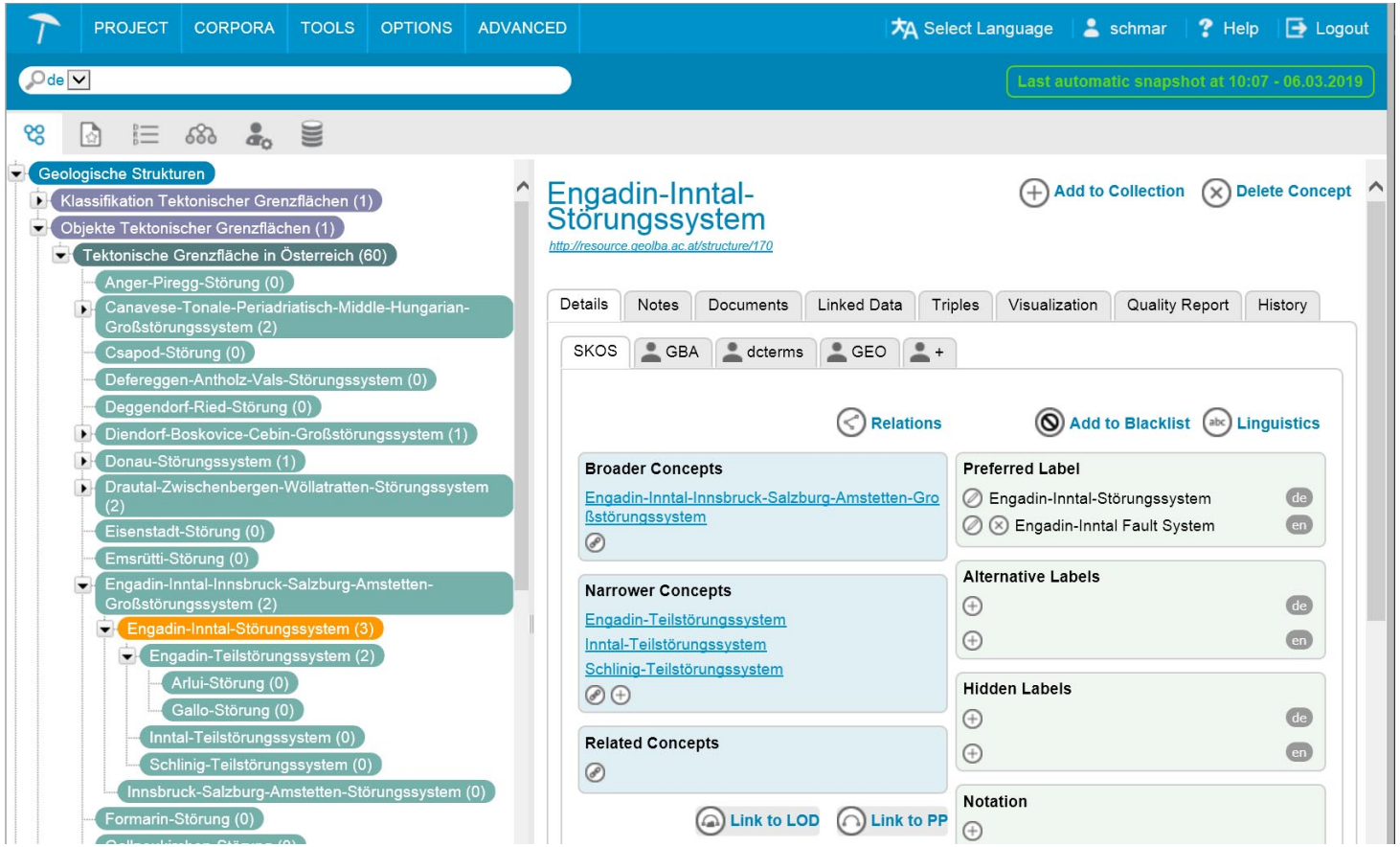
Narrower concept label – next hierarchy (level 3)

Narrower concept label – next hierarchy (level 4)

Narrower concept label – next hierarchy (level 5)

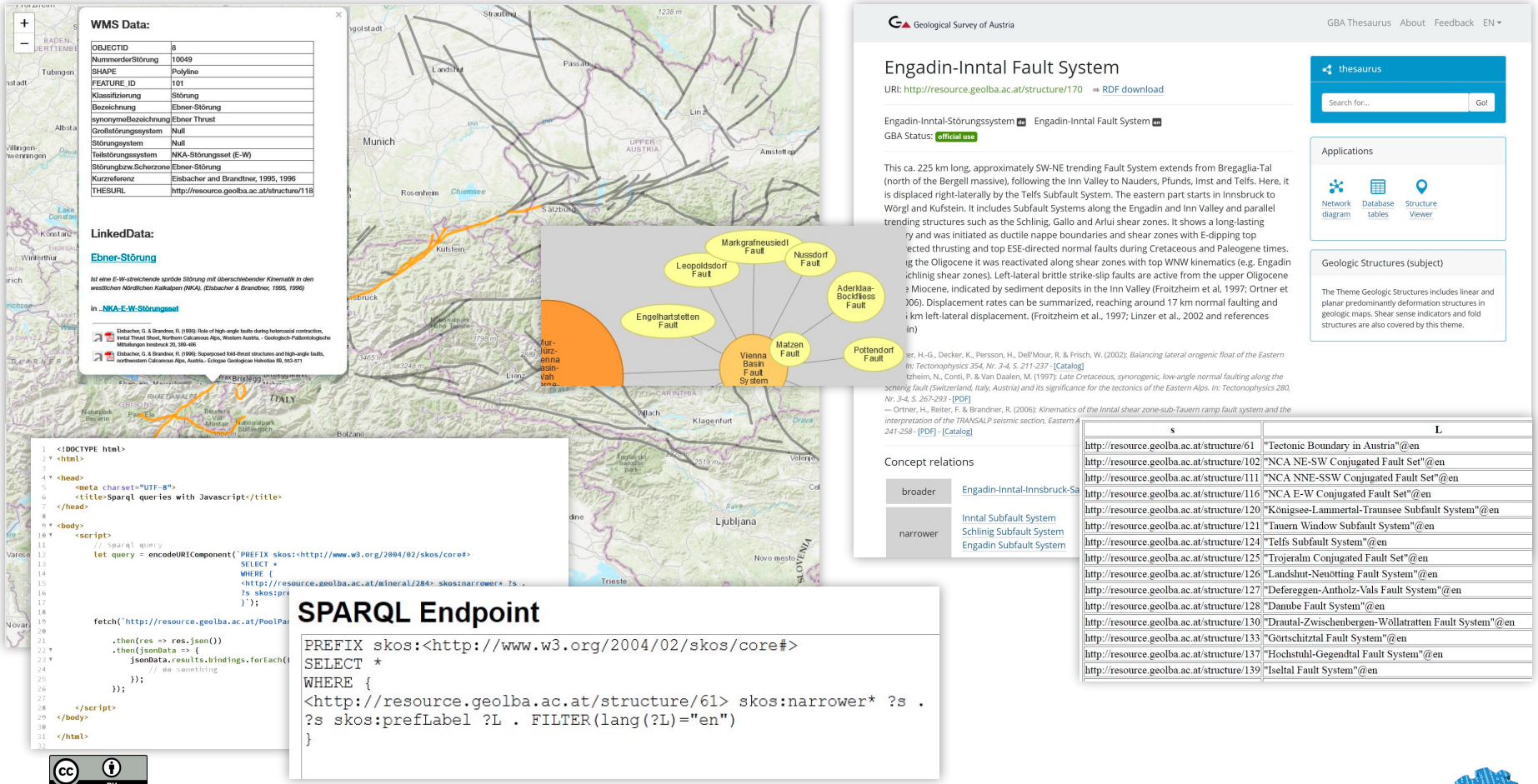
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166

Editing system



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166

Multifunctional representation



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166

GeoERA vocabularies workflow in comparison

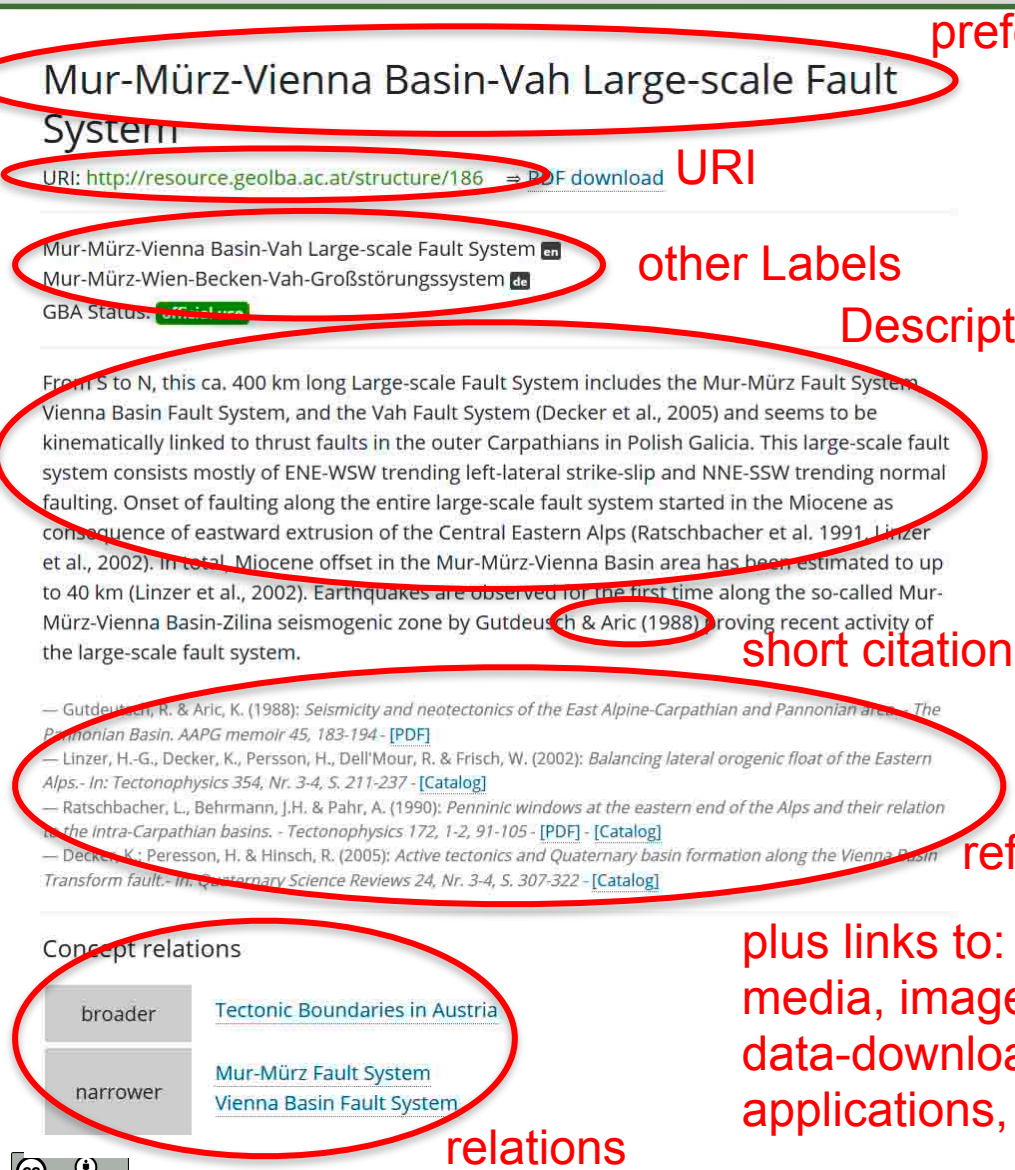
	Vocabulary draft	GeoERA Project Vocabularies	European Geoscience Register	INSPIRE register federation	INSPIRE codelists
Governance	project team	GeoERA project team	EGS expert groups	JRC INSPIRE registry team	JRC INSPIRE registry team
Publisher	project team	GIP-P WP4	BRGM	Relevant organization (BRGM)	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	

GeoERA project vocabularies

- **Harmonization** effects by using a **common project vocabulary**
- **Online availability** for web applications, EU **multilingual**ism, cross-linked bibliographic references
- **meaning of concepts defined** by provided context (eg bibliography)
- **scientific** traceability of information (**provenance**)
- Even **weakly structured** data/knowledge can be processed
- **Sustainable** knowledge management through **reusability** and expandability in the next project
- **interdisciplinary** usable, because of crosslinking
- **extend**/complete the harmonization efforts of **INSPIRE and GeoSciML**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166



From scientific term to scientific concept

http://resource.geolba.ac.at/structure/186

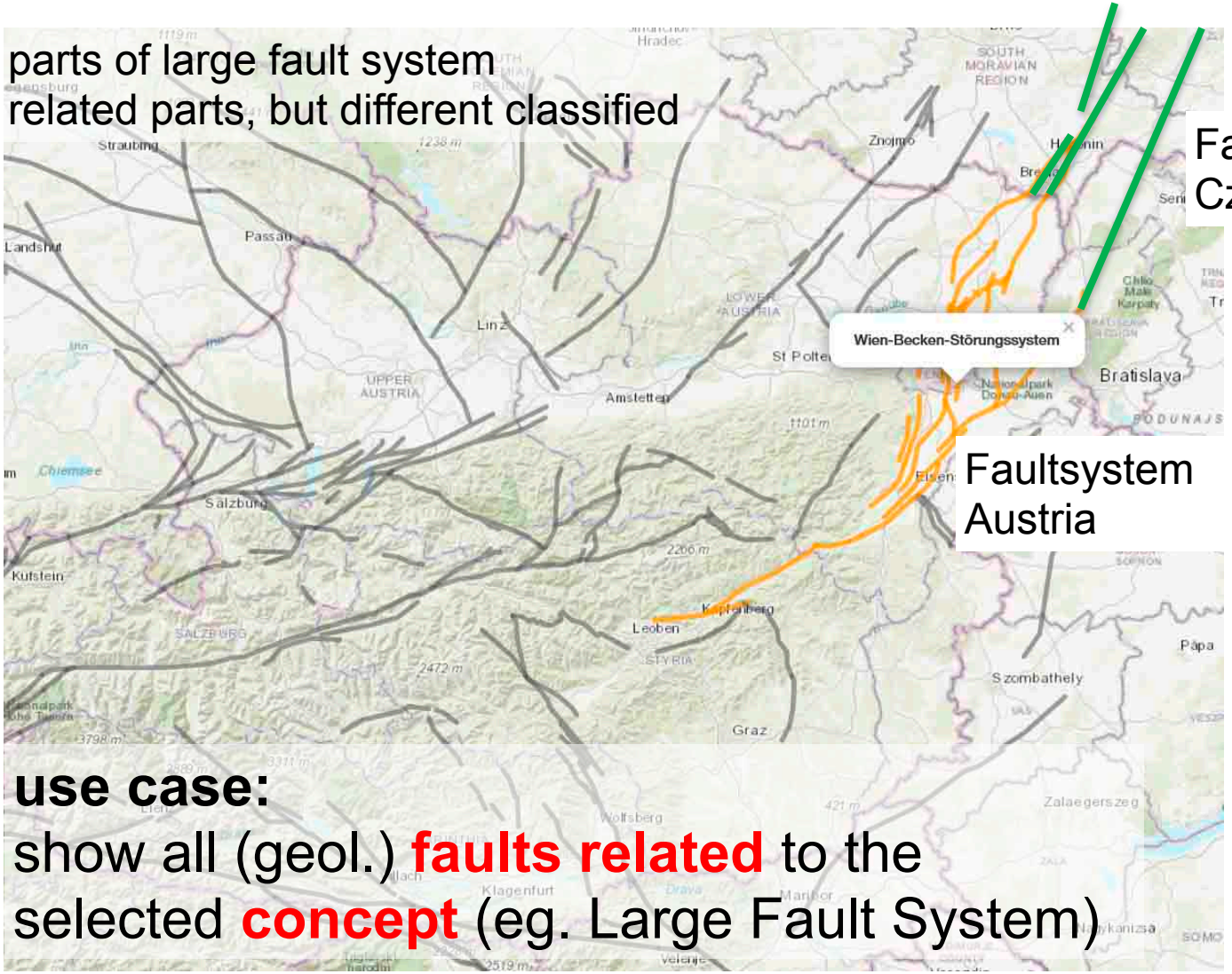
- example GBA Thesaurus (HIKE fault systems)
- scientific confirmability
- loosely structured information
- multilingualty

plus links to:
media, images, libraries
data-downloads, geonames
applications, wikipedia articles

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166

GeoERA INFORMATION PLATFORM

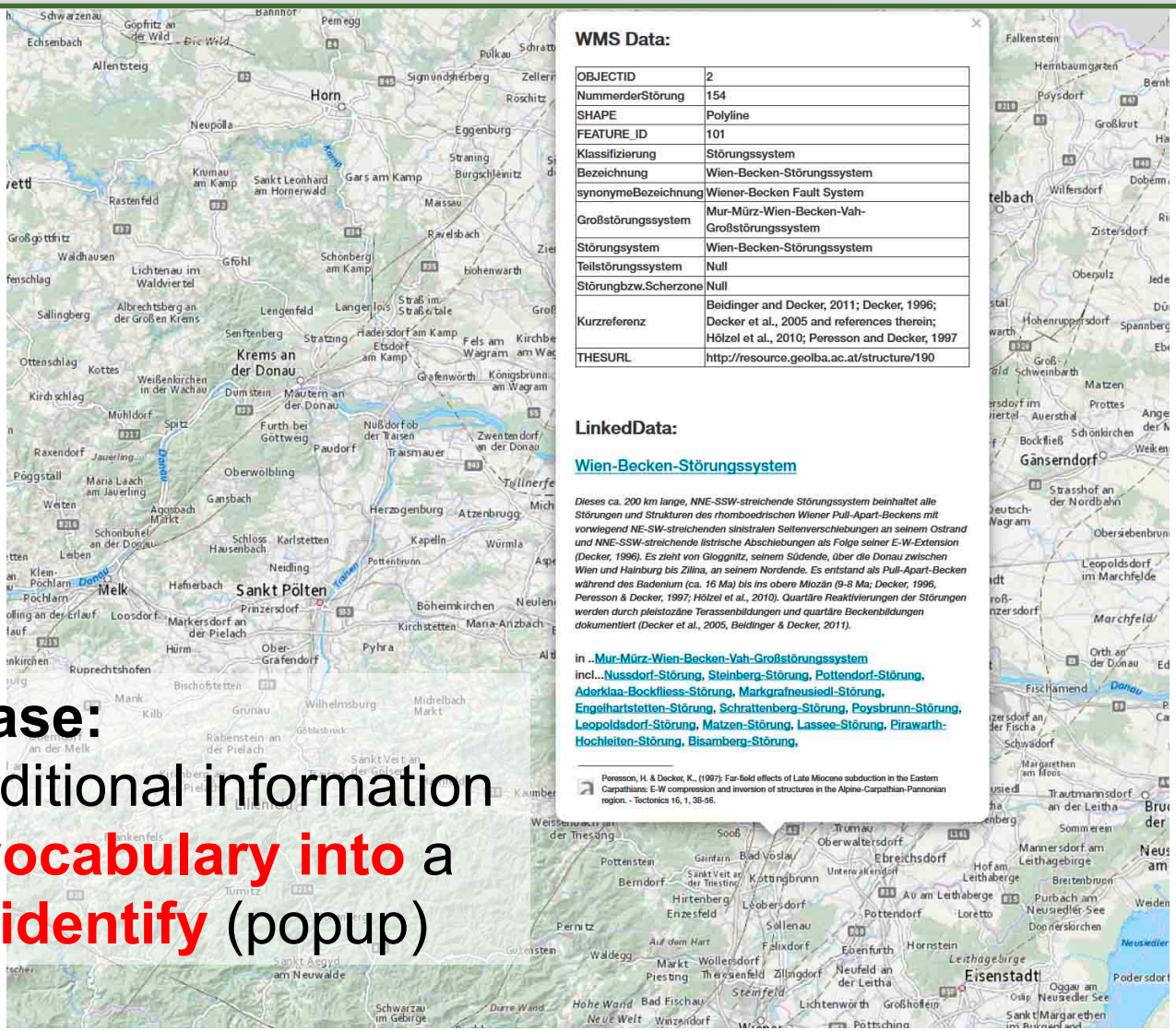
parts of large fault system
related parts, but different classified



use case:
show all (geol.) **faults related** to the selected **concept** (eg. Large Fault System)

https://thesaurus.geolba.ac.at/structureViewer.html?uri=http://resource.geolba.ac.at/structure/186&lang=en

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166



use case:
get additional information from **vocabulary into** a **WMS-identify** (popup)

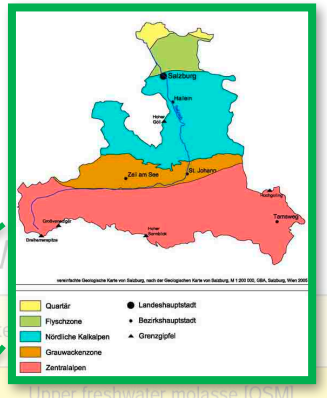
http://www.ce-gic.org/wms/GBA_structures_2.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166

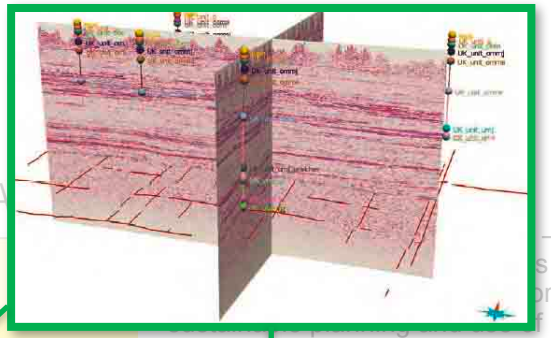
local 3D model

attributed with

Map dataset



Borehole dataset



attributed with

narrower terms

Formation B

Formation A

Formation C

related terms

Formation Z

narrower terms

Formation X

Formation Y

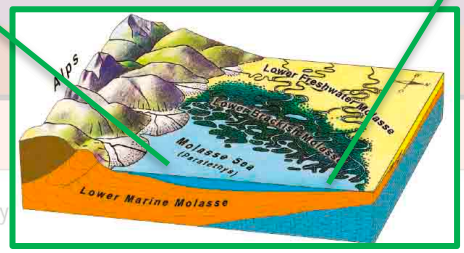
Swabian & Helvetic facies of Upper Jurassic

related terms

Germanic facies of Upper Jurassic

GeoERA project vocabulary

click&identify



3D model viewer

http://www.geomol.eu/geomol/report?lang=2

use case:
click on 3D view to **access** (bridge) **all source datasets** related to the selected 3D body

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166

Questions?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731166

GeoERA INFORMATION PLATFORM