

opendata.swiss O

Open Data Navigator & opendata.swiss

OA publication recommendations for datasets – Concept tagging with Wikidata Identifiers

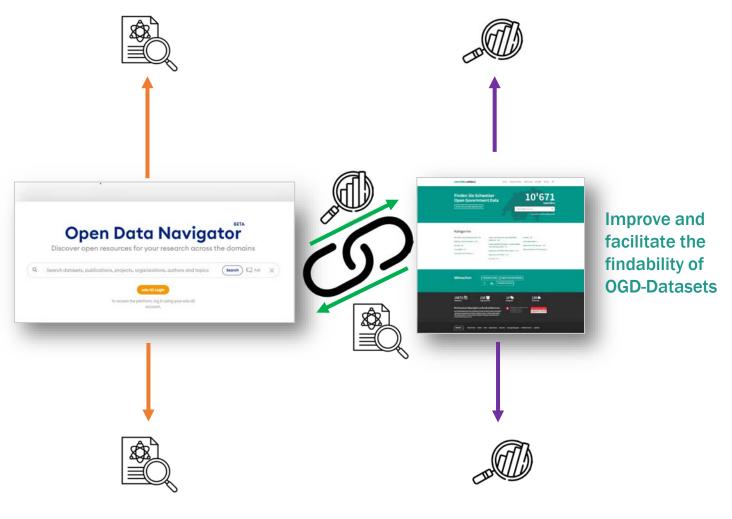
8. OGD FORUM, NEUCHÂTEL 22.04.2024 MICHÈLE SPICHTIG, OGD OFFICE, FSO
TOBIAS SCHWEIZER, SWITCH

Agenda

- 1. Who we are, what we do
- 2. Goal of the project
- 3. Tagging concepts
- 4. Integration
- 5. Key Take Aways

Facilitating the discovery of Open Data and resources in the Swiss Open Data ecosystem

Improve and facilitate the findability of research projects, publications, ORD



Who we are what we do

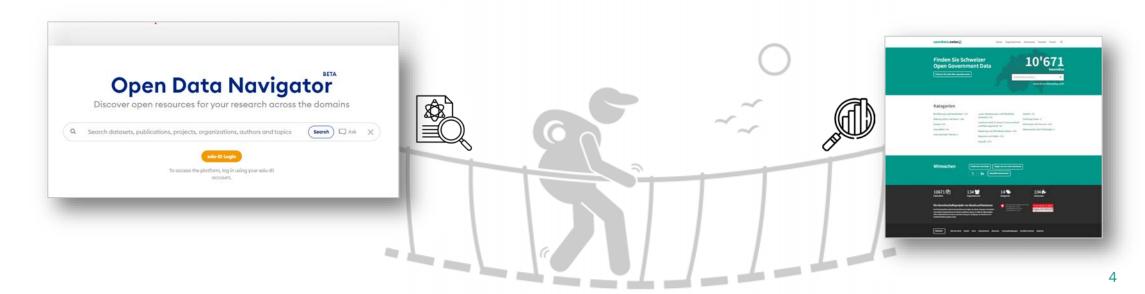
OA Publications

Keywords

Open Government Data

Goal of the project

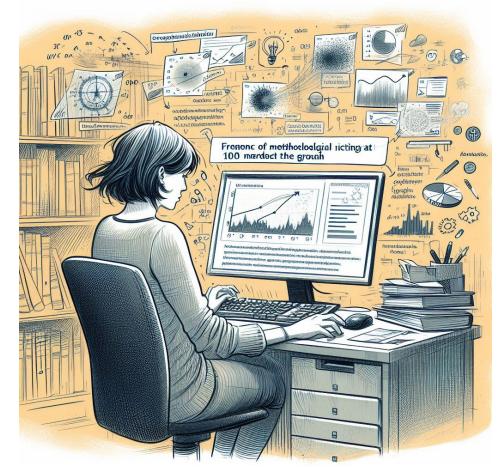
- Improvement of the Open Government Data exploration journey for users with a research focus
- PoC of a recommendation system: show relevant resources for delving into the topics and data on opendata.swiss
- Build Bridges: develop a feature that displays Open Access (OA) publications related to each dataset on opendata.swiss



User Story: How can we help Paula to find datasets and publications?

Problem statement:

- Paula searches for open government data on "Frequency of meteorological icing at 100 metres above the ground" and relevant OA publications
- She goes to the portal opendata.swiss and finds a relevant dataset, and now she has to go to the open data navigator to find relevant OA publications
- Wouldn't it be nice if she could get recommendations for her search on openata.swiss?



Idea for a solution to help Paula find recommended OA publications for her datasets

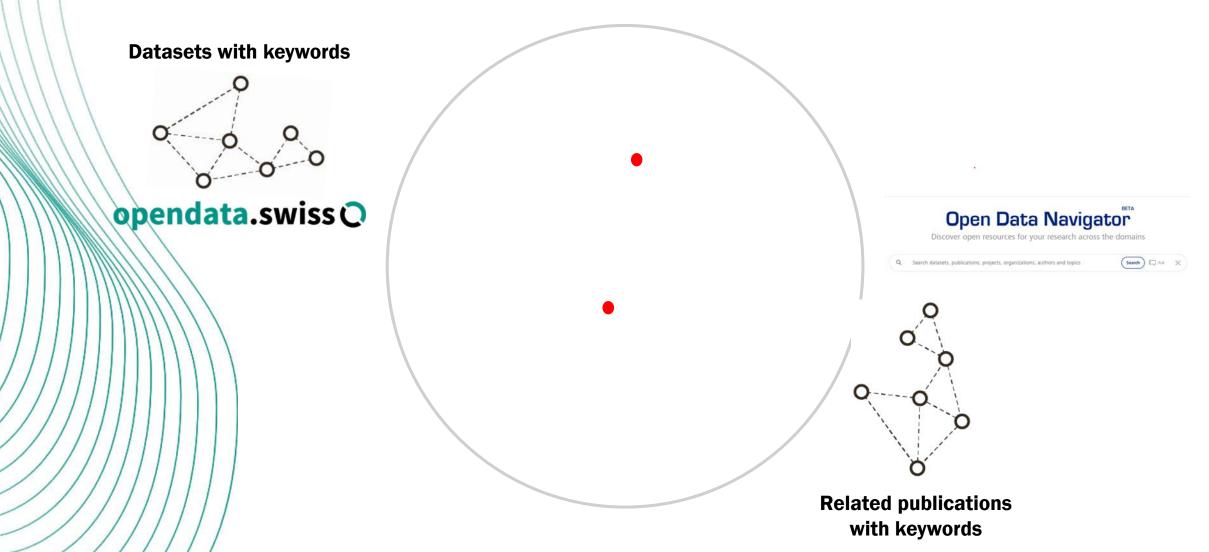
Idea:

 Build a recommendation system for OA publications based on linking common keywords that are captured in the metadata

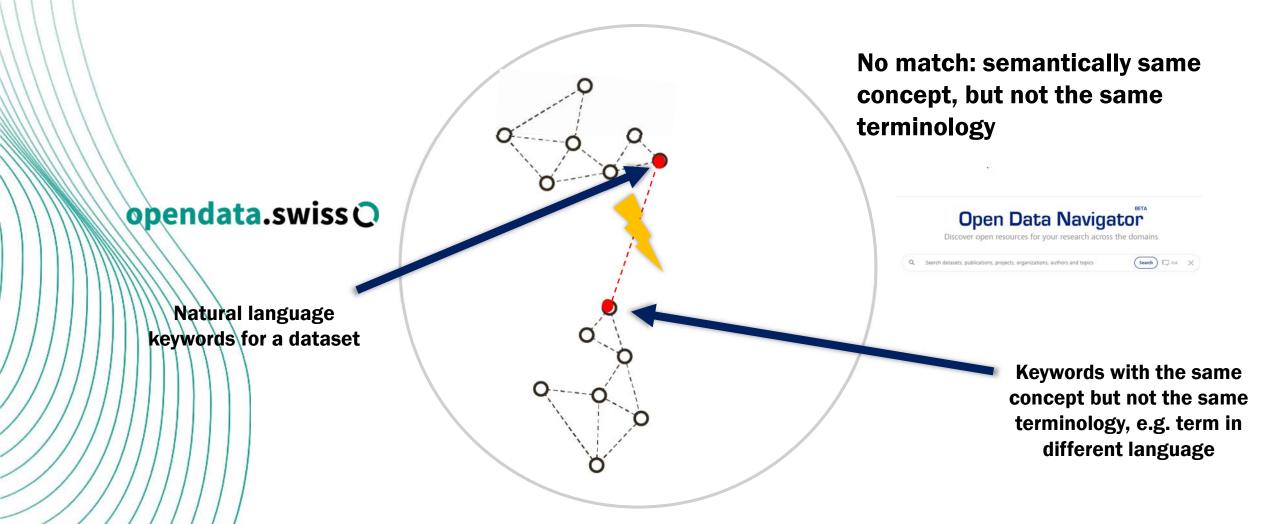
Keywords

Environmental science, Engineering, Computer science, Physics, Electrical engineering, Power (physics), Quantum mechanics, Electricity generation, Wind power, Meteorology, Photovoltaic system, Electric power system, Hybrid power

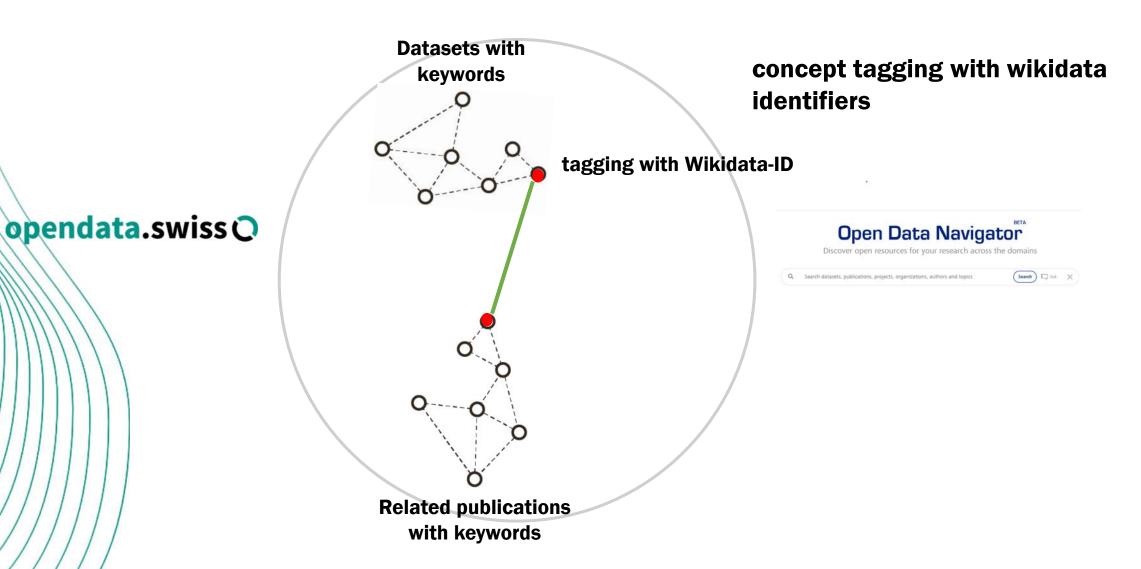
Idea: Linking with common keywords



Problem: not the same terminology



Idea: concept tagging with wikidata identifiers



Natural language keywords on opendata.swiss

Find relevant publications for a given opendata.swiss dataset based on the given keywords Keywords are entered manually by datapublishers (mostly experts but not necessarily)

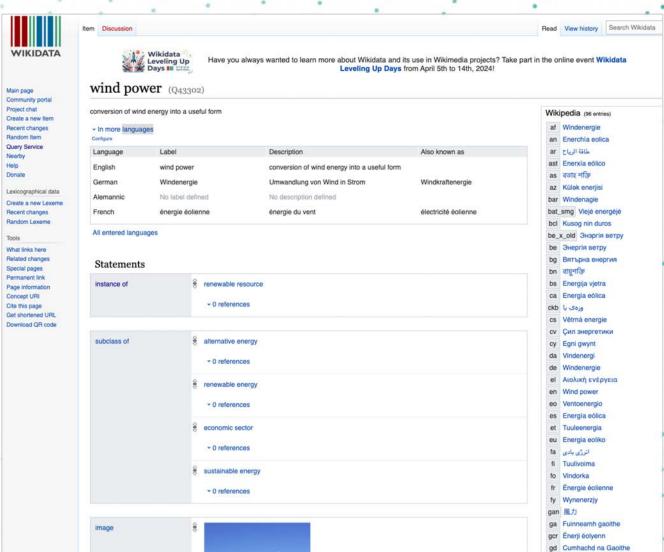
EN Keywords:	EN Keywords:
× meteorology × energy × wind-power × wind-power-station	
× fsdi-federal-spatial-data-infrastructure	
DE Keywords:	DE Keywords:
× meteorologie x energie x windenergie x windenergieanlage	x vektordaten x geoportal x polygondaten x ausbauetappen x geodaten x stzh x fernwarme
× bgdi-bundesgeodaten-infrastruktur	
FR Keywords:	FR Keywords:
FR Keywords: x meteorologie x energie x energie-eolienne x centrale-a-energie-eolienne	FR Keyworas:
# C 100 C 10	FR Keywords:
x meteorologie x energie x energie-eolienne x centrale-a-energie-eolienne	IT Keywords:
x meteorologie x energie x energie-eolienne x centrale-a-energie-eolienne x ifdg-linfrastructure-federale-de-donnees-geographiques	

Problem: It is difficult to search for international publications (English) with these keywords **Solution:** Use concepts instead of natural language keywords

Image source: https://www.wikidata.org/wiki/043302

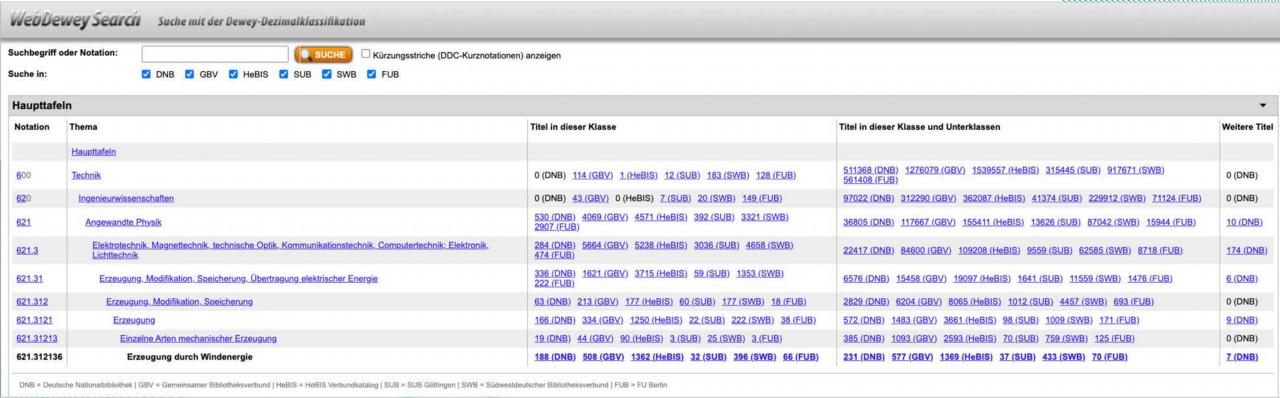
Wikidata

- Wikidata contains a lot of information about concepts, persons, locations etc.
- A Wikidata entry is uniquely identified by a Wikdiata ID
- As Wikidata IDs are very common, they can be used as keywords instead of natural language terms
 - For example, Q43302 stands for the concept of "wind power".
- Labels are given in many different languages.
- In addition, the concept is related to other concept (Wikidata IDs) such as "alternative energy" or "renewable energy"



Controlled Vocabularies

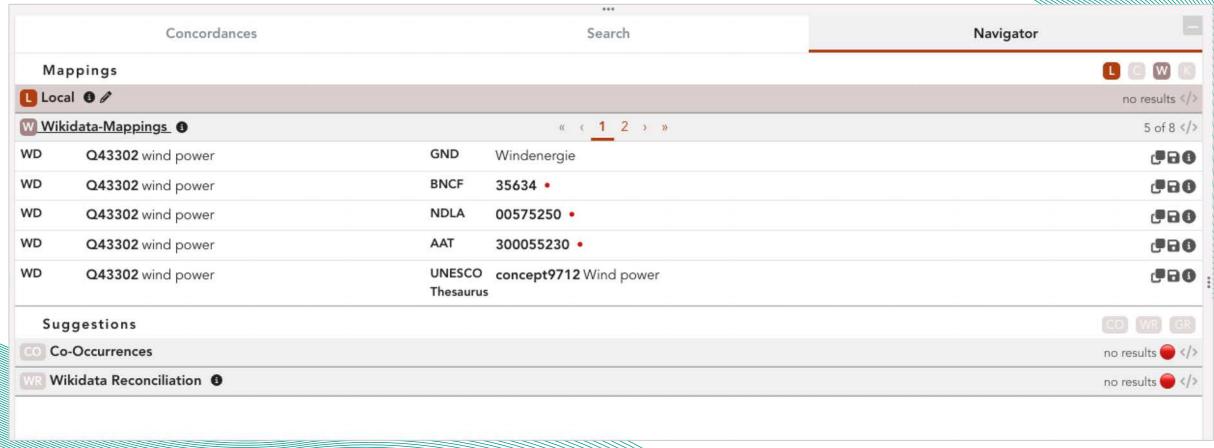
- Wikidata is one example of controlled vocabularies
- There are many others such as the Dewey Decimal Classification (DDC):



Source: https://deweysearchde.pansoft.de/webdeweysearch

Which Controlled Vocabulary?

 Mappings between different vocabularies can be queried (correspondence, equivalence of entries)



Source: https://coli-conc.gbv.de/cocoda/

Named Entity Recognition and Linking

"La vittoria di **Ferdy Kubler**. A **Ginevra** , i 38 concorrenti rimasti in lizza iniziano la terza tappa del Giro di Romandia ..."

```
"text": "Ferdy Kübler",
"kbid": "Q458713",
"label": "PER",
"start": 15.
"end": 27
"text": "Ginevra",
"kbid": "Q71",
"label": "LOC",
"start": 31.
"end": 37
```

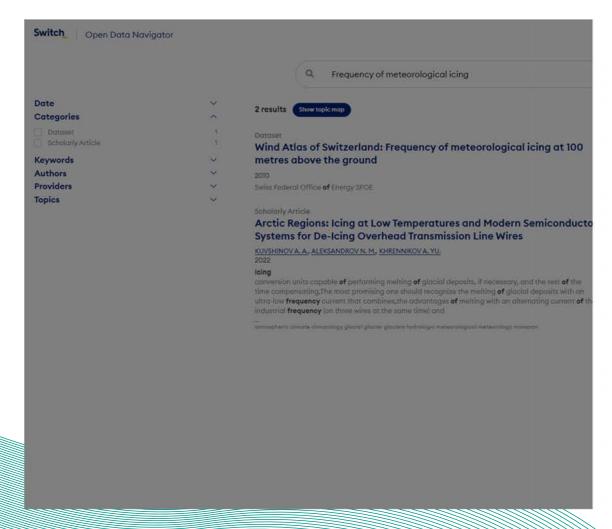




Source: https://www.wikidata.org/

Integration

 Users browse on opendatanavigator.switch.ch and discover datasets from opendata.swiss



Dataset

Wind Atlas of Switzerland: Frequency of meteorological icing at 100 metres above the ground

Publisher

SWISS FEDERAL OFFICE OF ENERGY SFOE

Date Published

2010

In Switzerland, the mountain regions (Jura, Alps and Alpine foothills) are the most suitable regions for exploiting wind energy potential. But here, the air temperature during the winter months is frequently below 0° C for lengthy periods, and this can cause measuring instruments and rotor blades to ice up, Icing influences the planning and operation of a wind power plant in a variety of ways: ice formation on the rotor blades interferes with their aerodynamics and results in production losses. The additional weight and imbalance associated with ice formation also burden the structural elements of a wind turbine. For safety reasons (ice fall-off), and in view of the additional burden placed on their components, wind turbines have to be turned off when they ice up. Ice formation on a wind turbine's measuring instruments can also give rise to faulty operation. Calculations of icing frequency are based on comprehensive data relating to cloud water, temperature and wind obtained from analyses of the MeteoSwiss COSMO-2 weather forecasting model. The icing model calculates the ice load on a cylindrical, freely rotating structure. The icing frequency is depicted on a 2.2-kilometre raster and has been verified on the basis of measurements carried out at IMIS stations in the Alps and readings recorded in the Jura range. The map depicts the frequency of meteorological icing at a height of 100 metres above the ground for the period from August 2007 to July 2009. The 10-year average is around 5 percent lower. The frequency data are based on calculations made with the aid of models, and they indicate where icing can occur in Switzerland. However, the data cannot be regarded as absolutely precise for a given location, because deviations can occur in confined spaces such as those frequently encountered in the central Alps. The frequency of icing is higher in exposed areas and lower in protected areas than the levels shown on the map. The map depicts meteorological icing, while the periods of measured icing tend to be longer on average. To assess the icing conditions at a specific location, measurement on site is thus essential.

Link to Source

94cedad7-6d19-46b2-9ecf-0d1d1a58bea8@bundesamt-fur-energie-bfe

Keywords

energy, fsdi-federal-spatial-data-infrastructure, meteorology, wind-power, wind-

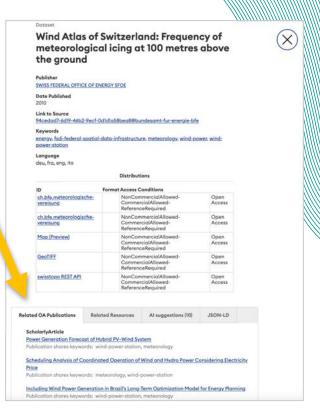
Integration

1. Users browse on opendatanavigator.switch.ch and disconfrom opendata.swiss

2. <u>In development:</u> opendata.swiss implements a button on a dataset's detail view that leads the user to the corresponding page on

opendatanavigator.switch.ch (Connectome resolver)





corresponding page on opendatanavigator. switch.ch

User leaves the page to the corresponding page

ataset

Wind Atlas of Switzerland: Frequency of meteorological icing at 100 metres above the ground



Publisher

SWISS FEDERAL OFFICE OF ENERGY SFOR

Date Published

2010

Link to Source

94cedad7-6d19-46b2-9ecf-0d1d1a58bea8@bundesamt-fur-energie-bfe

Keywords

energy, fsdi-federal-spatial-data-infrastructure, meteorology, wind-power, windpower-station

Language

deu, fra, eng, ita

Distributions

ID Fo	ormat Access Conditions	
ch.bfe.meteorologische- vereisung	NonCommercialAllowed- CommercialAllowed- ReferenceRequired	Open Access
ch.bfe.meteorologische- vereisung	NonCommercialAllowed- CommercialAllowed- ReferenceRequired	Open Access
Map (Preview)	NonCommercialAllowed- CommercialAllowed- ReferenceRequired	Open Access
GeoTIFF	NonCommercialAllowed- CommercialAllowed- ReferenceRequired	Open Access
swisstopo REST API	NonCommercialAllowed- CommercialAllowed- ReferenceRequired	Open Access

Related OA Publications	Related Resources	Al suggestions (10)	JSON-LD
ScholarlyArticle			
Power Generation Foreco	ast of Hybrid PV-Wind Sys	tem	
m 1 11 11 11 1		700	
Publication shares keywo	ords: wind-power-station	meteorology	
Publication shares keywo	ords: wind-power-station	meteorology	
	oordinated Operation of		onsidering Electric
			onsidering Electric
Scheduling Analysis of C Price		Wind and Hydro Power C	onsidering Electric
Scheduling Analysis of C Price	oordinated Operation of	Wind and Hydro Power C	onsidering <mark>Electric</mark>
Scheduling Analysis of C Price Publication shares keywo	oordinated Operation of	Wind and Hydro Power Co	V

Integration

- Users browse on opendatanavigator.switch.ch and discove from opendata.swiss
- 2. In development: opendata.swiss implements a button on a datased detail view that leads the user to the corresponding page on opendatanavigator.switch.ch (Connectome resolver)
- 3. Possible future development: **opendata.swiss queries the Connectome GraphQL endpoint** to get related publications for a given dataset (using its identifier)

Possible integration of option 3 – Mock up

Advantage: Users see related OA publications without having to leave the page



Messinstrumenten und Rotorblättern. Die Vereisung beeinflusst die Planung und den Betrieb einer Windenergieanlage (WEA) auf unterschiedliche Weise: Eisansatz an den Rotorblättern führt zu einer gestörten Aerodynamik der Flügel und verursacht Produktionsverluste. Die mit der Vereisung verbundenen Zusatzlasten und Unwuchten belasten die Strukturelemente der WEA. Aus Sicherheitsgründen (Eiswurf) und auf Grund der zusätzlichen Materialbelastung werden die WEA bei Vereisung abgeschaltet. Weiter kann Eisansatz an den Windmessgeräten einer WEA zu einem fehlerhaften Betrieb führen. Die Berechnungen der Vereisungshäufigkeit basieren auf flächendeckenden Daten zu Wolkenwasser, Temperatur und Wind aus der Analyse des Wettervorhersagemodells COSMO-2 der MeteoSchweiz. Das Vereisungsmodell berechnet die Eislast an einer zylindrischen, frei rotierenden Struktur. Die Vereisungshäufigkeit liegt auf einem 2.2-km-Raster vor und wurde anhand von Messungen von IMIS-Stationen in den Alpen und Messungen im Jura verifiziert. Die Karte zeigt die Häufigkeit meteorologischer Vereisung auf 100 m über Grund für den Zeitraum August 2007 bis Juli 2009. Im 10-Jahres-Mittel liegt die mittlere Vereisung etwa 5% tiefer. Die Häufigkeitsangaben sind modellierte Werte. Sie geben Hinweise, wo in der Schweiz Vereisungen auftreten können. Die Häufigkeitsangaben können nicht als punktgenaue, exakte Werte interpretiert werden. In kleinräumigem Gelände, wie es häufig in den Zentralalpen anzutreffen ist, gibt es Abweichungen. Bei stark exponierten Lagen liegt der Vereisungswert höher, bei abgeschatteter Lage liegt der Vereisungswert höher, bei abgeschatteter Lage liegt der Vereisungswert höher, bei abgeschatteter Lage liegt der Vereisungswert tiefer als in der Karte dargestellt. Auf der Karte ist die meteorologische Vereisung dargestellt, die Perioden von instrumenteller Vereisung dauern im Durchschnitt länger. Bei Interesse an einer Standortentwicklung ist eine Messung vor Ort unumgänglich.

meteorologie energie windenergie windenergieanlage bgdi-bundesgeodaten-infrastruktur

Related OA Publications

ScholarlyArticle

Power Generation Forecast of Hybrid PV-Wind System

Publication shares keywords: wind-power-station, meteorology

Scheduling Analysis of Coordinated Operation of Wind and Hydro Power Considering Electricity
Price

Publication shares keywords: meteorology, wind-power-station

Including Wind Power Generation in Brazil's Long-Term Optimization Model for Energy Planning

Publication shares keywords: wind-power-station, meteorology

opendata.swiss queries Connectome GraphQL Endpoint – Developer Interface

```
GetRelatedCreativeWork
                                                                                                                                                                       STATUS 200 295ms 08
Operation
     query GetRelatedCreativeWork($identifier: String!, $provider: Provider!) {
                                                                                                                                                                                     自由
       getRelatedCreativeWork(identifier: $identifier, provider: $provider) {
                                                                                                          "data": {
                                                                                                            "getRelatedCreativeWork": {
           ... on Thing {
                                                                                                              "nodes": [
             __typename
             name
                                                                                                                  "__typename": "Book",
                                                                                                                  "name": "Wind Solar Hybrid Renewable Energy System",
             sameAs
                                                                                                                    "https://doi.org/10.5772/intechopen.77440"
                                                                                                                  "name": "Clean Water Using Solar and Wind: Outside the Power Grid
                                                                                                         (Persian Translation)",
                                                                                                                    "https://doi.org/10.2166/9781789062953"
                                                                                                                  "__typename": "Book",
          Headers Pre-Operation Script Post-Operation Script
                                                                                                                  "name": "Making Sense of Natural Disasters",
                                                                                                                    "https://doi.org/10.1007/978-3-030-94778-1"
       "identifier": "94cedad7-6d19-46b2-9ecf-0d1d1a58bea8@bundesamt-fur-energie-bfe",
       "provider": "opendata"
                                                                                                                  "__typename": "ScholarlyArticle",
                                                                                                                  "name": "The Role of Cyclone Activity in the Interannual Variability of
                                                                                                        the Summertime Beaufort High",
                                                                                                                  "sameAs": [
                                                                                                                    "https://doi.org/10.2151/sola.2015-025"
                                                                                                                  "_typename": "ScholarlyArticle".
                                                                                                                  "name": "Motion Synthesis of a Planar Watt II Type Six-Bar Mechanism
                                                                                                        with Two End-Effectors",
                                                                                                                    "https://doi.org/10.1007/978-3-319-17067-1 10"
```

Different Ways of Accessing the Connectome Knowledge Graph (KG)

- End users use Uls, not APIs
- Platform interoperability can be achieved through
 - **linking between Uls** (user switches between platforms in browser, e.g., different tabs)
 - platform queries the Connectome API and displays the data to the user (user remains on the same platform, e.g. opendata.swiss)



Key Take Aways

 Information in Metadata (e.g. keywords) helps to build a recommendation system

• Using controlled vocabularies enhances Metadata Quality (e.g. consistency)

 Wikidata serves as an introduction to a conceptual world that simplifies many things for us (e.g. multilingualism, related concepts)

 Concepts can be mapped with each other (differences in specificity)







Try it out:

Open Data Navigator:

https://opendatanavigator.switch.ch/ (Stand Q1 2024)

Button on Abnahme-Environment opendata.swiss categories energy and health

Energy: https://ogdch-

abnahme.clients.liip.ch/en/dataset?groups=ener

Health: https://ogdch-

abnahme.clients.liip.ch/en/dataset?groups=heal

opendata.swiss O

Email

tobias.schweizer@switch.ch michele.spichtig@bfs.admin.ch

LinkedIn Switch

https://www.linkedin.com/company/switch-switzerland/mycompany/

LinkedIn opendata.swiss

https://www.linkedin.com/showcase/92970995

Questions about the PoC:

opendata@bfs.admin.ch

Get In Touch