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

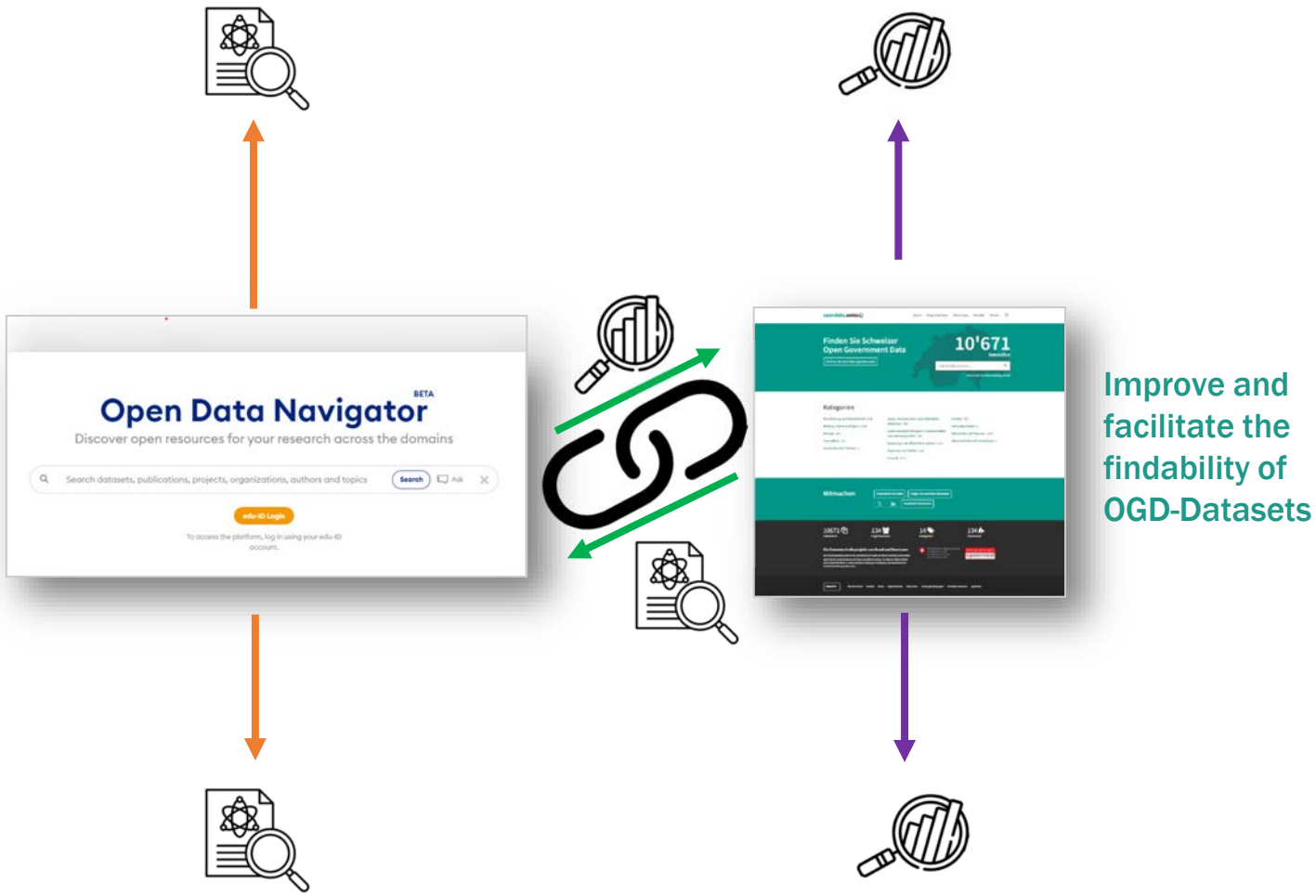
TOBIAS SCHWEIZER, SWITCH

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



Improve and facilitate the findability of OGD-Datasets

Who we are **what we do**

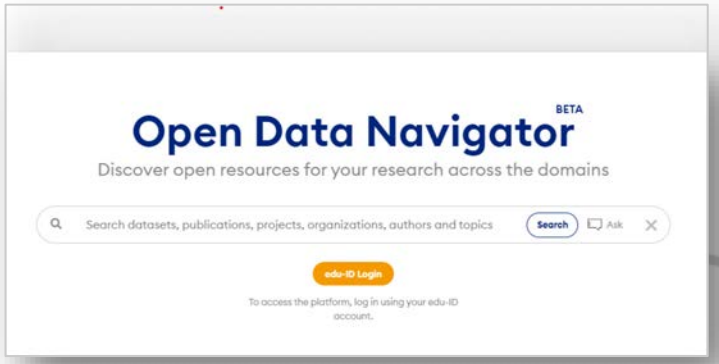
OA Publications

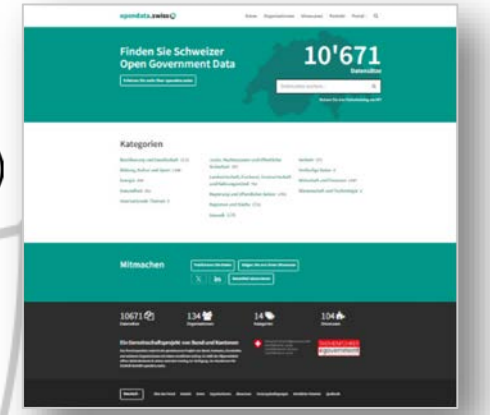
Keywords

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:

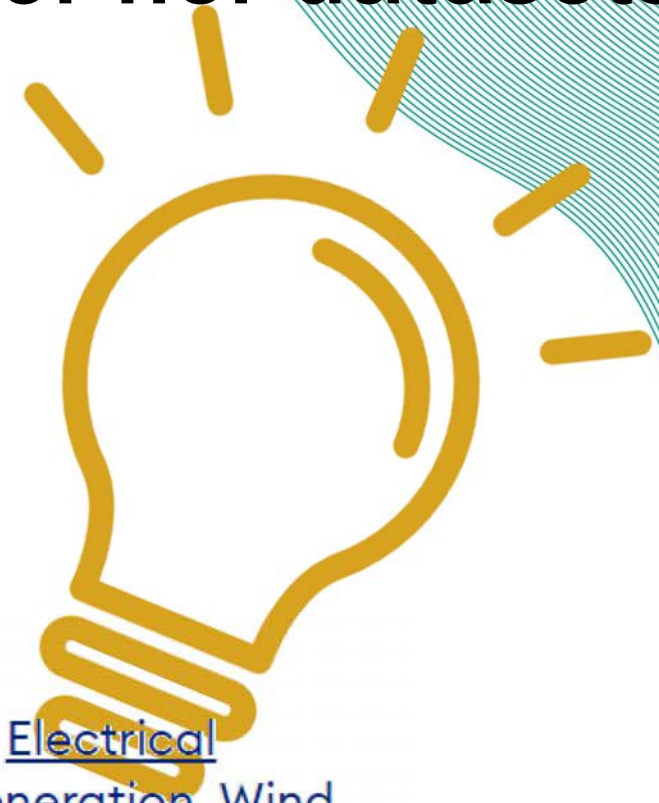
- Paula searches for open government data on *"Frequency of meteorological icing at 100 metres above the ground"* and relevant OA publications
- She goes to the portal opendata.swiss and finds a relevant dataset, and now she has to go to the open data navigator to find relevant OA publications
- Wouldn't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

Eigene Darstellung mit MS Copilot, prompt: Draw me a serious but scetchy picture of this: Paula searches open government data to the topic «Frequency of meteorological icing at 100 metres above the ground" and relevant Open Access (OA) publications

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**



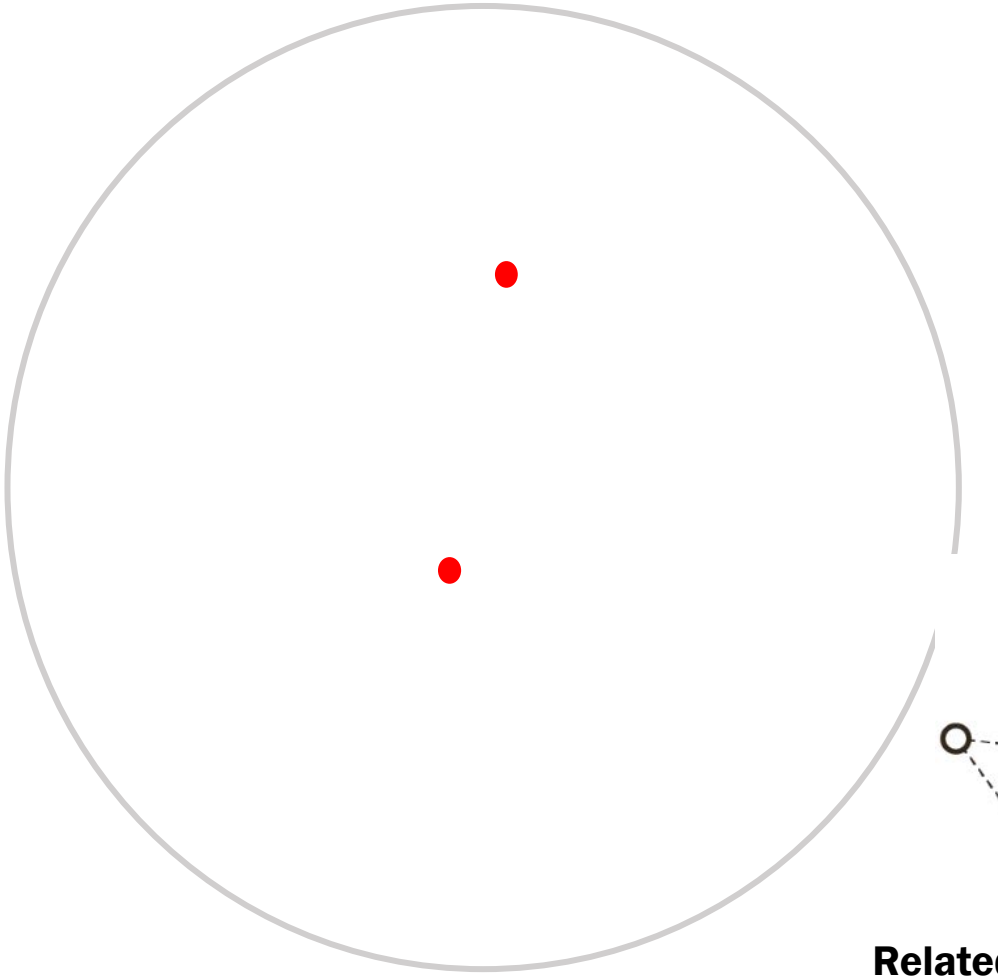
Keywords

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

Datasets with keywords

opendata.swiss

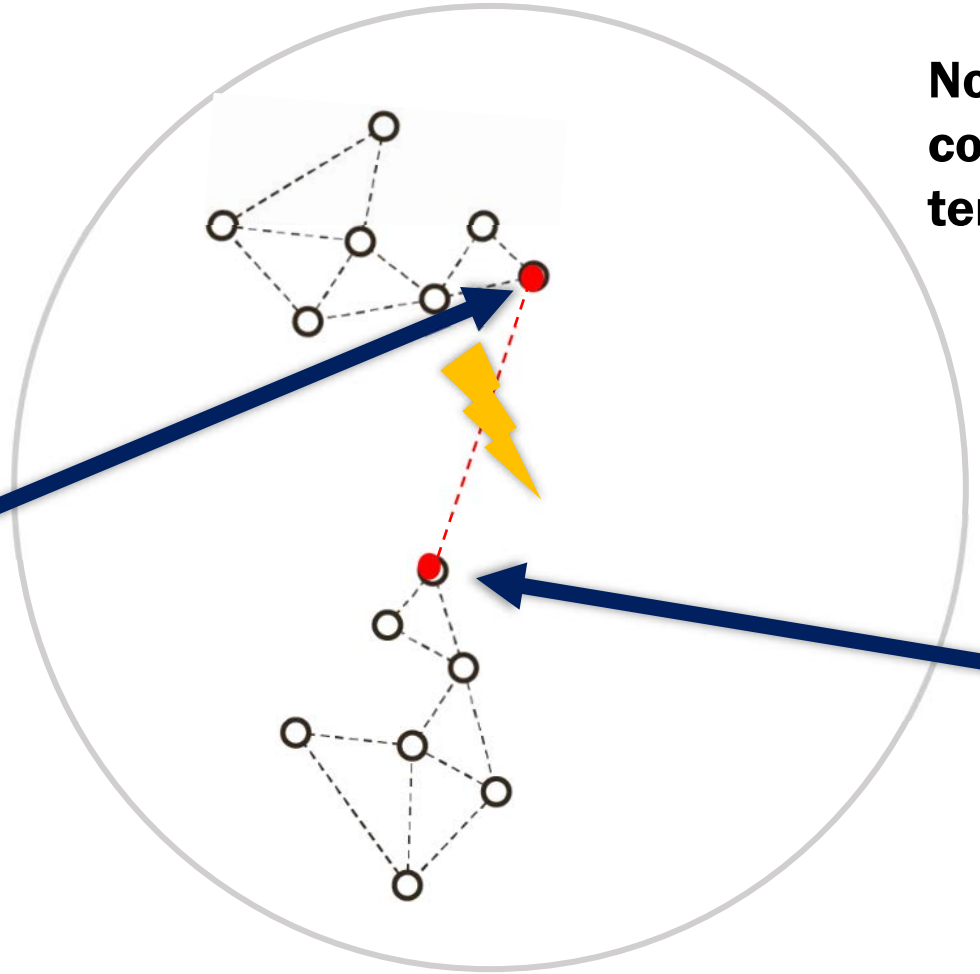


Related publications with keywords

Problem: not the same terminology

opendata.swiss

Natural language keywords for a dataset

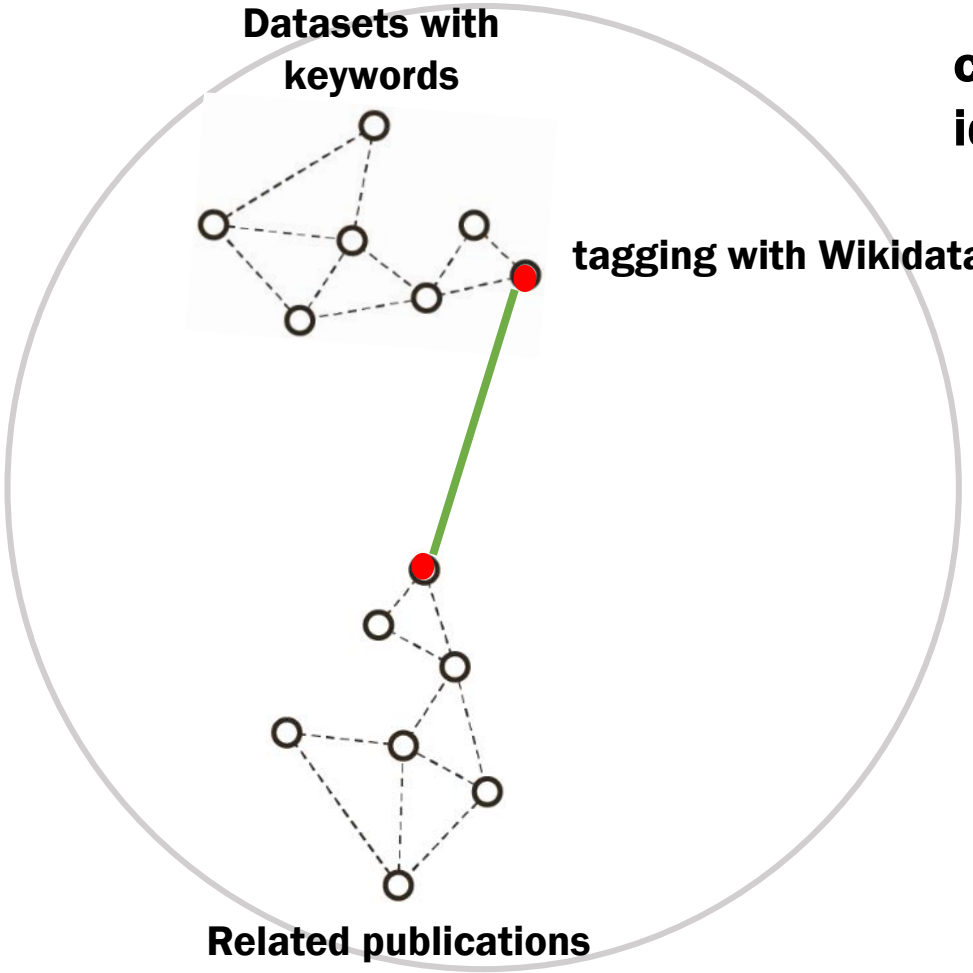


No match: semantically same concept, but not the same terminology

Keywords with the same concept but not the same terminology, e.g. term in different language

Idea: concept tagging with wikidata identifiers

opendata.swiss

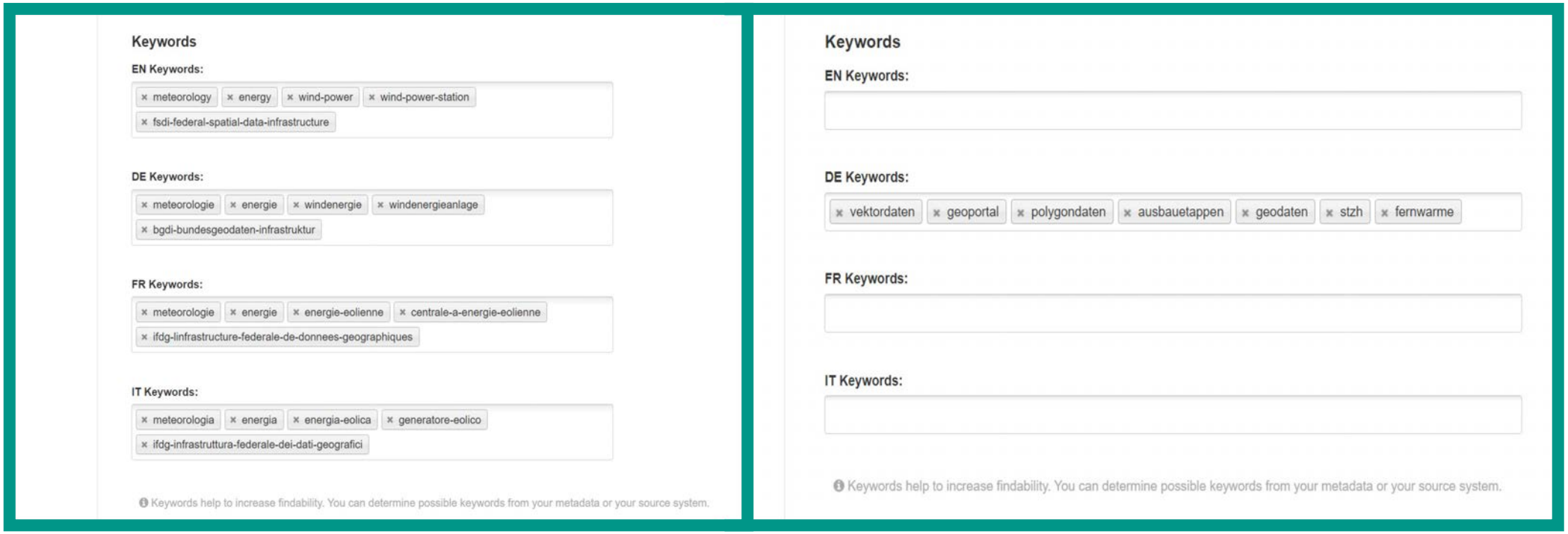


concept tagging with wikidata identifiers

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)



Images: Input mask for keywords in the CKAN-Backend opendata.swiss

Problem: It is difficult to search for international publications (English) with these keywords

Solution: Use concepts instead of natural language keywords

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"

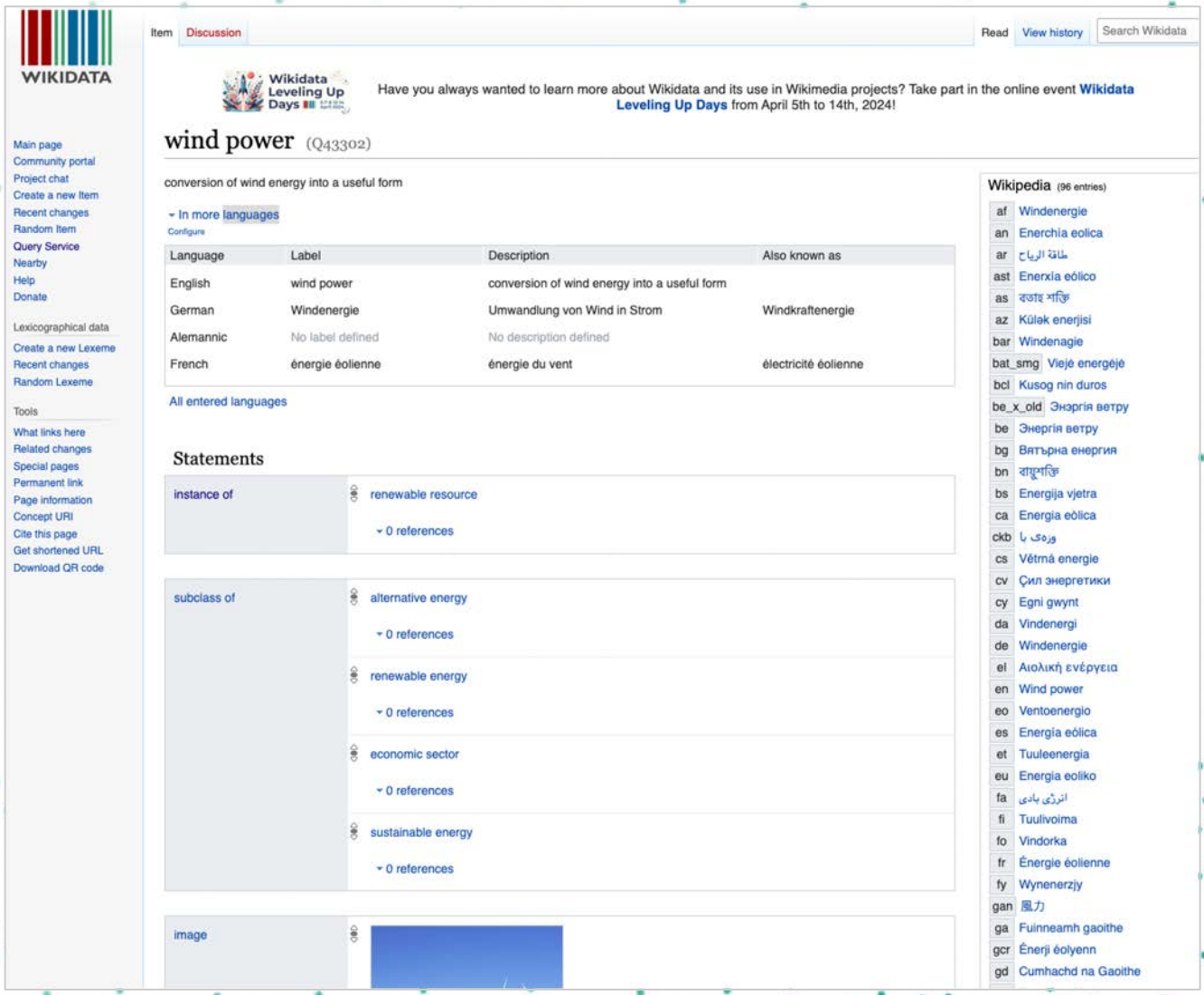
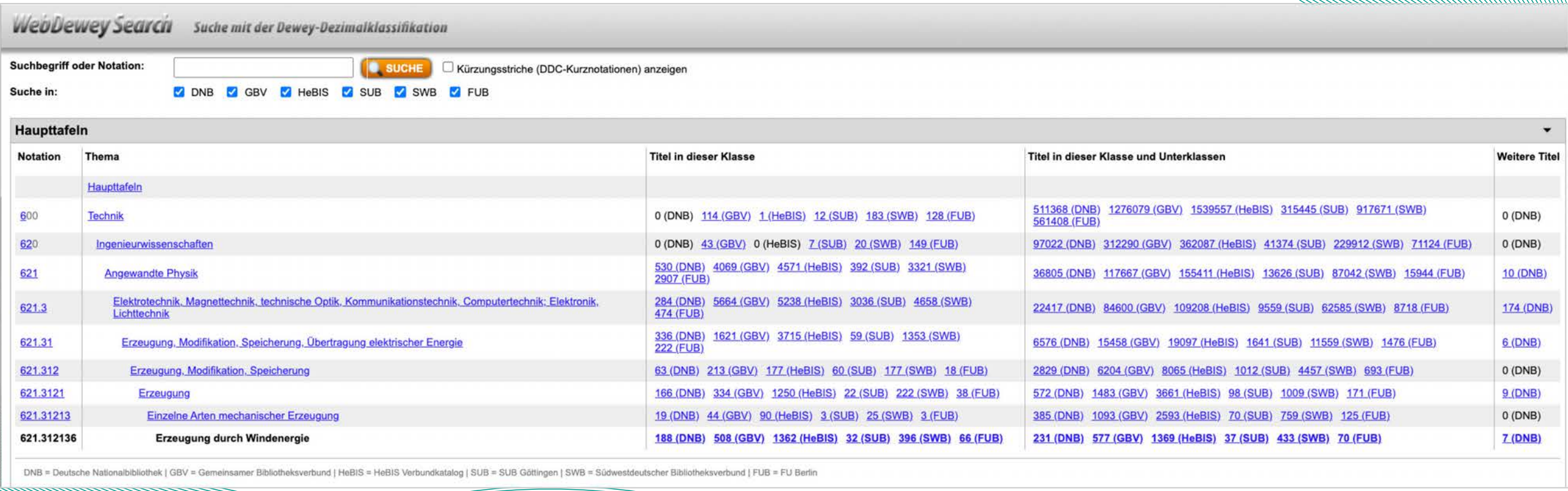


Image source: https://www.wikidata.org/wiki/Q43302

Controlled Vocabularies

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



WebDewey Search — Suche mit der Dewey-Dezimalklassifikation

Suchbegriff oder Notation: SUCHE ☐ Kürzungsstriche (DDC-Kurznotationen) anzeigen

Suche in: ☑ DNB ☑ GBV ☑ HeBIS ☑ SUB ☑ SWB ☑ FUB

Haupttafeln

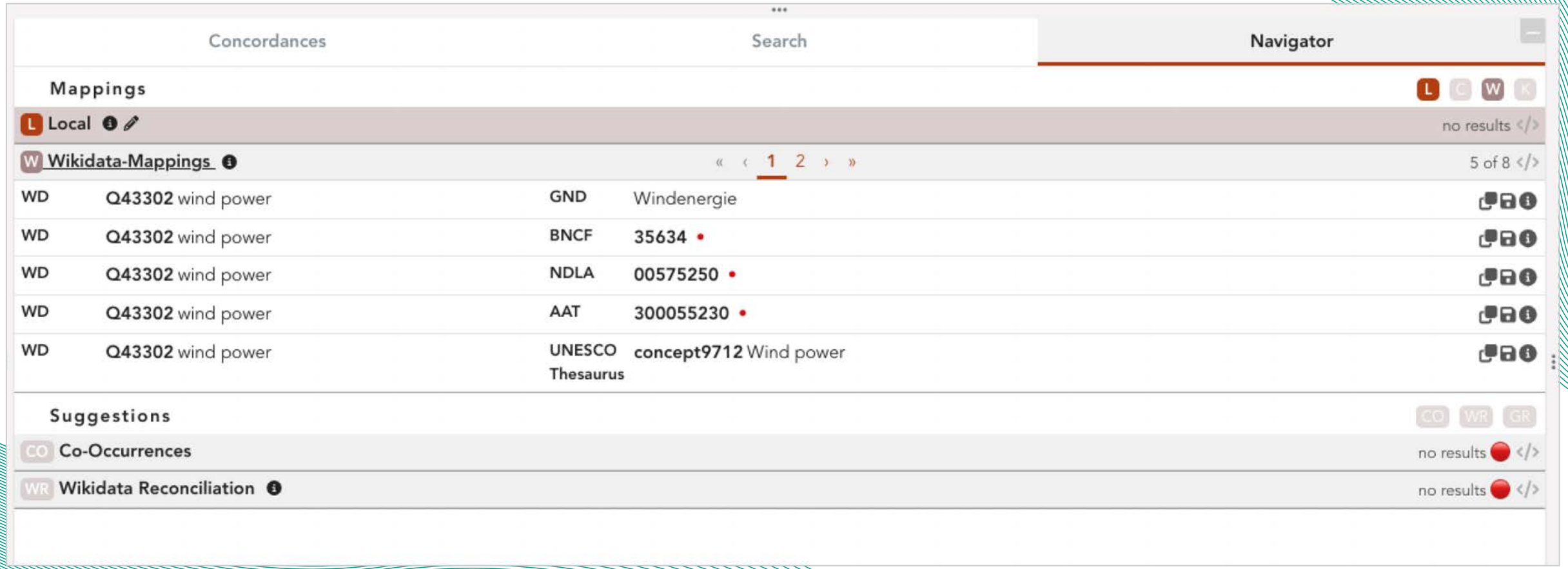
Notation	Thema	Titel in dieser Klasse	Titel in dieser Klasse und Unterklassen	Weitere Titel
	Haupttafeln			
600	Technik	0 (DNB) 114 (GBV) 1 (HeBIS) 12 (SUB) 183 (SWB) 128 (FUB)	511368 (DNB) 1276079 (GBV) 1539557 (HeBIS) 315445 (SUB) 917671 (SWB) 561408 (FUB)	0 (DNB)
620	Ingenieurwissenschaften	0 (DNB) 43 (GBV) 0 (HeBIS) 7 (SUB) 20 (SWB) 149 (FUB)	97022 (DNB) 312290 (GBV) 362087 (HeBIS) 41374 (SUB) 229912 (SWB) 71124 (FUB)	0 (DNB)
621	Angewandte Physik	530 (DNB) 4069 (GBV) 4571 (HeBIS) 392 (SUB) 3321 (SWB) 2907 (FUB)	36805 (DNB) 117667 (GBV) 155411 (HeBIS) 13626 (SUB) 87042 (SWB) 15944 (FUB)	10 (DNB)
621.3	Elektrotechnik, Magnettechnik, technische Optik, Kommunikationstechnik, Computertechnik; Elektronik, Lichttechnik	284 (DNB) 5664 (GBV) 5238 (HeBIS) 3036 (SUB) 4658 (SWB) 474 (FUB)	22417 (DNB) 84600 (GBV) 109208 (HeBIS) 9559 (SUB) 62585 (SWB) 8718 (FUB)	174 (DNB)
621.31	Erzeugung, Modifikation, Speicherung, Übertragung elektrischer Energie	336 (DNB) 1621 (GBV) 3715 (HeBIS) 59 (SUB) 1353 (SWB) 222 (FUB)	6576 (DNB) 15458 (GBV) 19097 (HeBIS) 1641 (SUB) 11559 (SWB) 1476 (FUB)	6 (DNB)
621.312	Erzeugung, Modifikation, Speicherung	63 (DNB) 213 (GBV) 177 (HeBIS) 60 (SUB) 177 (SWB) 18 (FUB)	2829 (DNB) 6204 (GBV) 8065 (HeBIS) 1012 (SUB) 4457 (SWB) 693 (FUB)	0 (DNB)
621.3121	Erzeugung	166 (DNB) 334 (GBV) 1250 (HeBIS) 22 (SUB) 222 (SWB) 38 (FUB)	572 (DNB) 1483 (GBV) 3661 (HeBIS) 98 (SUB) 1009 (SWB) 171 (FUB)	9 (DNB)
621.31213	Einzelne Arten mechanischer Erzeugung	19 (DNB) 44 (GBV) 90 (HeBIS) 3 (SUB) 25 (SWB) 3 (FUB)	385 (DNB) 1093 (GBV) 2593 (HeBIS) 70 (SUB) 759 (SWB) 125 (FUB)	0 (DNB)
621.312136	**Erzeugung durch Windenergie**	**188 (DNB) 508 (GBV) 1362 (HeBIS) 32 (SUB) 396 (SWB) 66 (FUB)**	**231 (DNB) 577 (GBV) 1369 (HeBIS) 37 (SUB) 433 (SWB) 70 (FUB)**	**7 (DNB)**

DNB = Deutsche Nationalbibliothek | GBV = Gemeinsamer Bibliotheksverbund | HeBIS = HeBIS Verbundkatalog | SUB = SUB Göttingen | SWB = Südwestdeutscher Bibliotheksverbund | FUB = FU Berlin

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

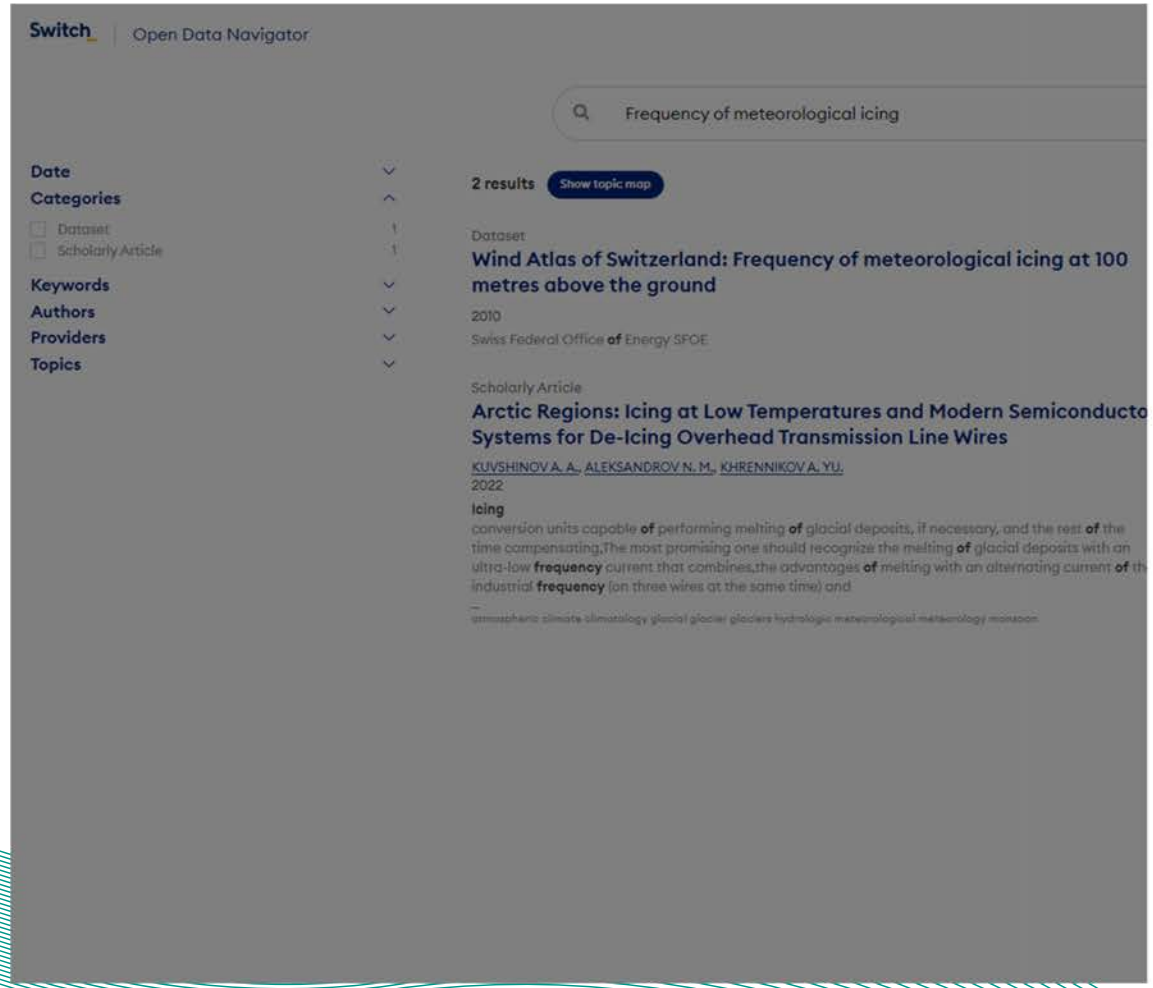
"La vittoria di **Ferdy Kubler**. A **Ginevra** , i
38 concorrenti rimasti in lizza iniziano la terza tappa del Giro di Romandia ..."


```
{
      "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

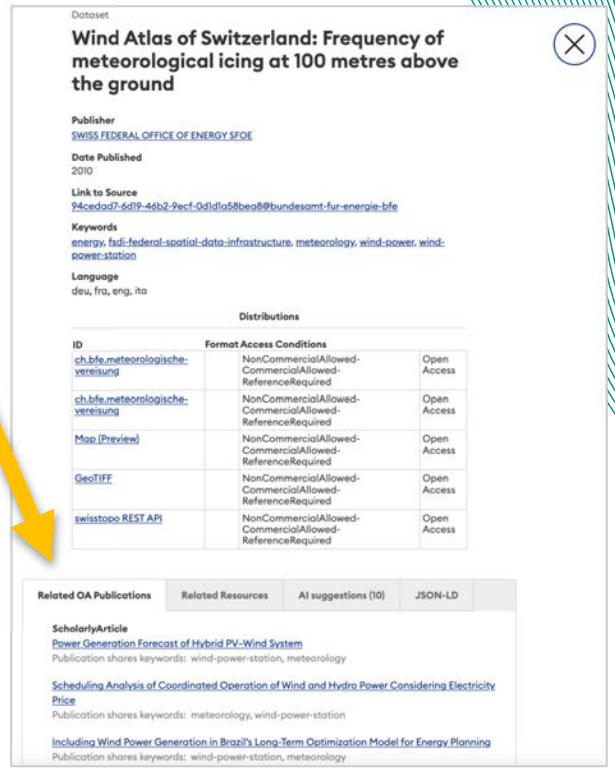
1. Users browse on opendatanavigator.switch.ch and discover datasets from opendata.swiss



Integration

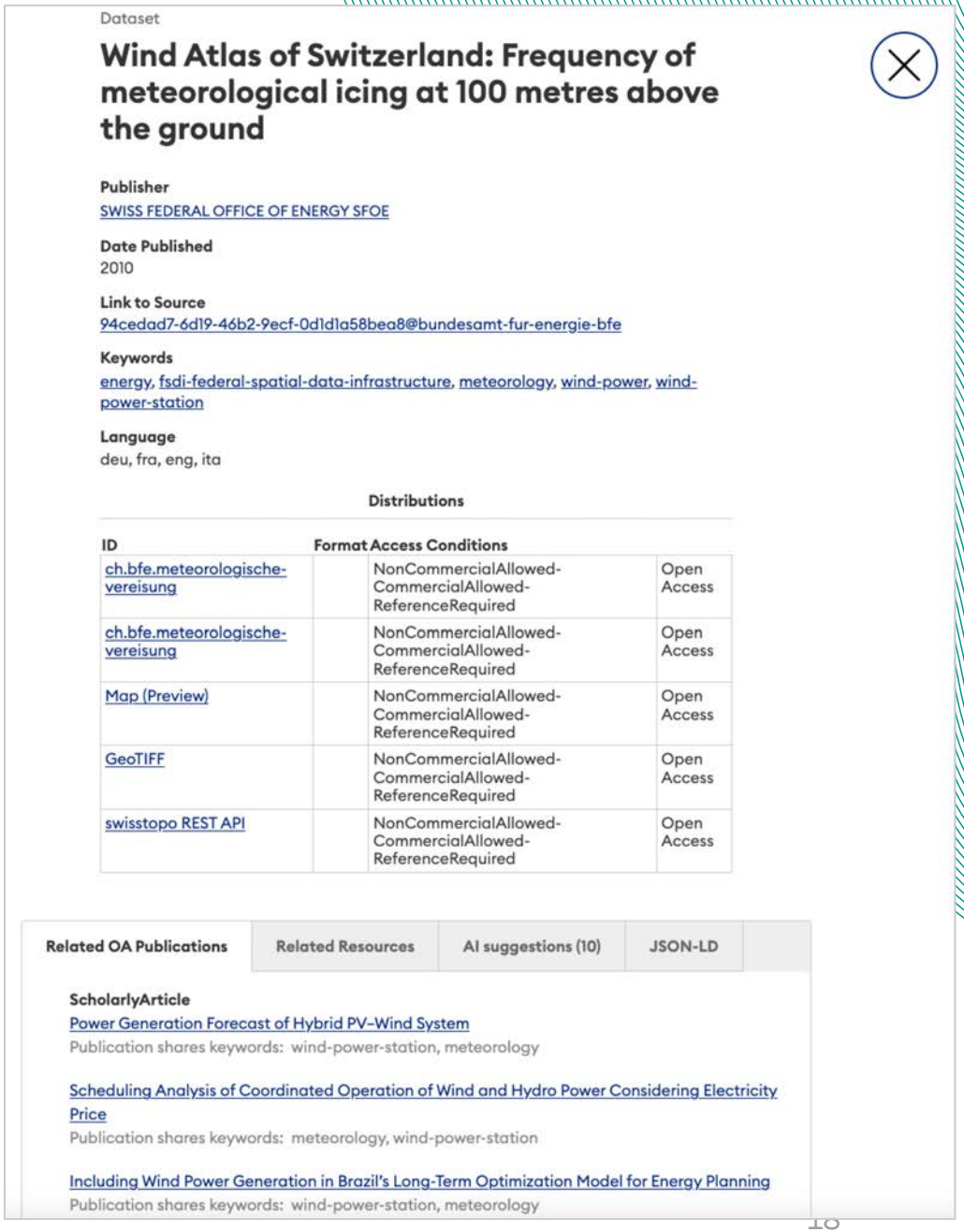
1. Users browse on opendatanavigator.switch.ch and discover datasets from opendata.swiss

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



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

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

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

Language
deu, fra, eng, ita

Distributions

ID	Format	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 | AI suggestions (10) | JSON-LD

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

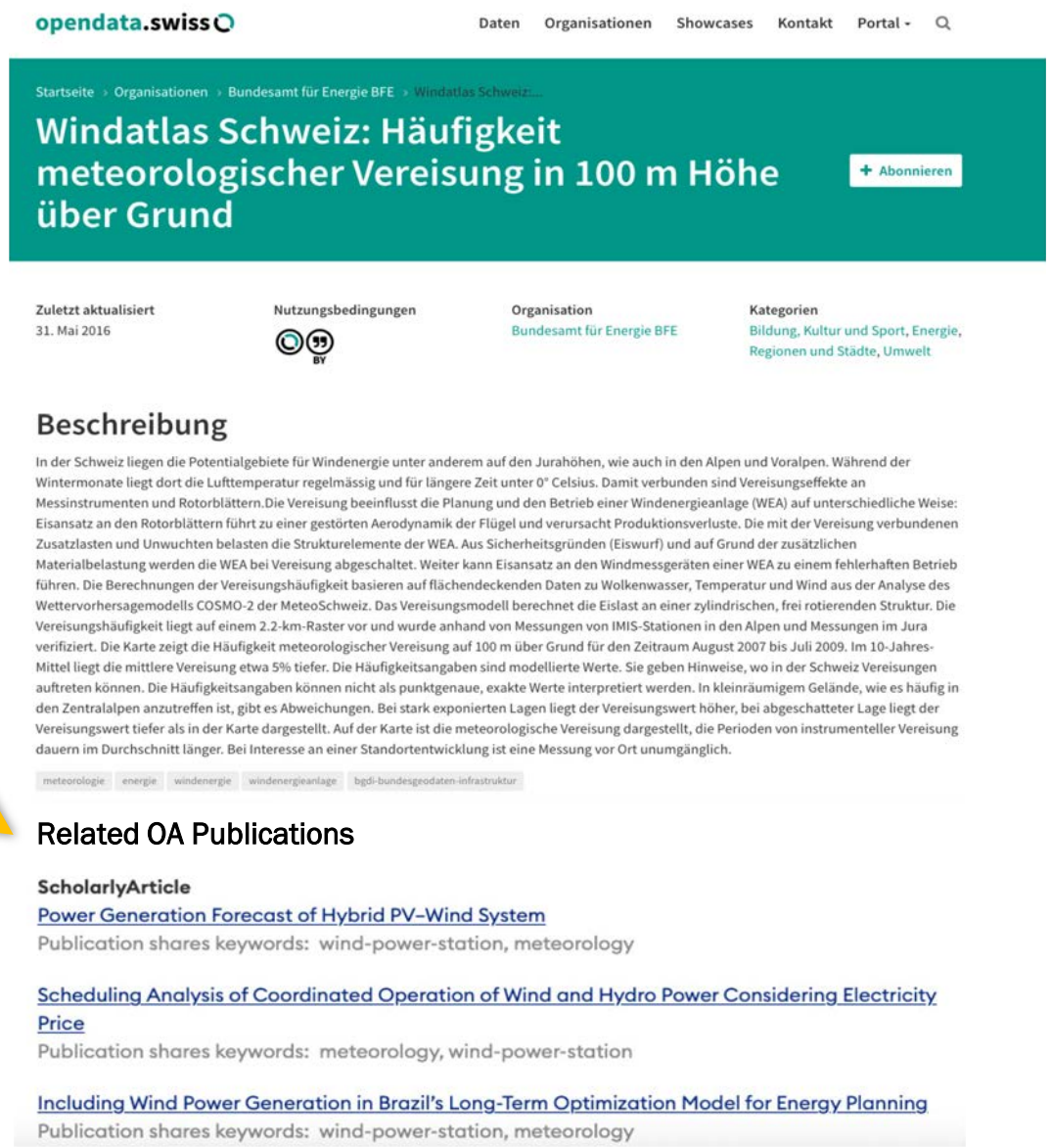
Including Wind Power Generation in Brazil's Long-Term Optimization Model for Energy Planning
Publication shares keywords: wind-power-station, meteorology

Integration

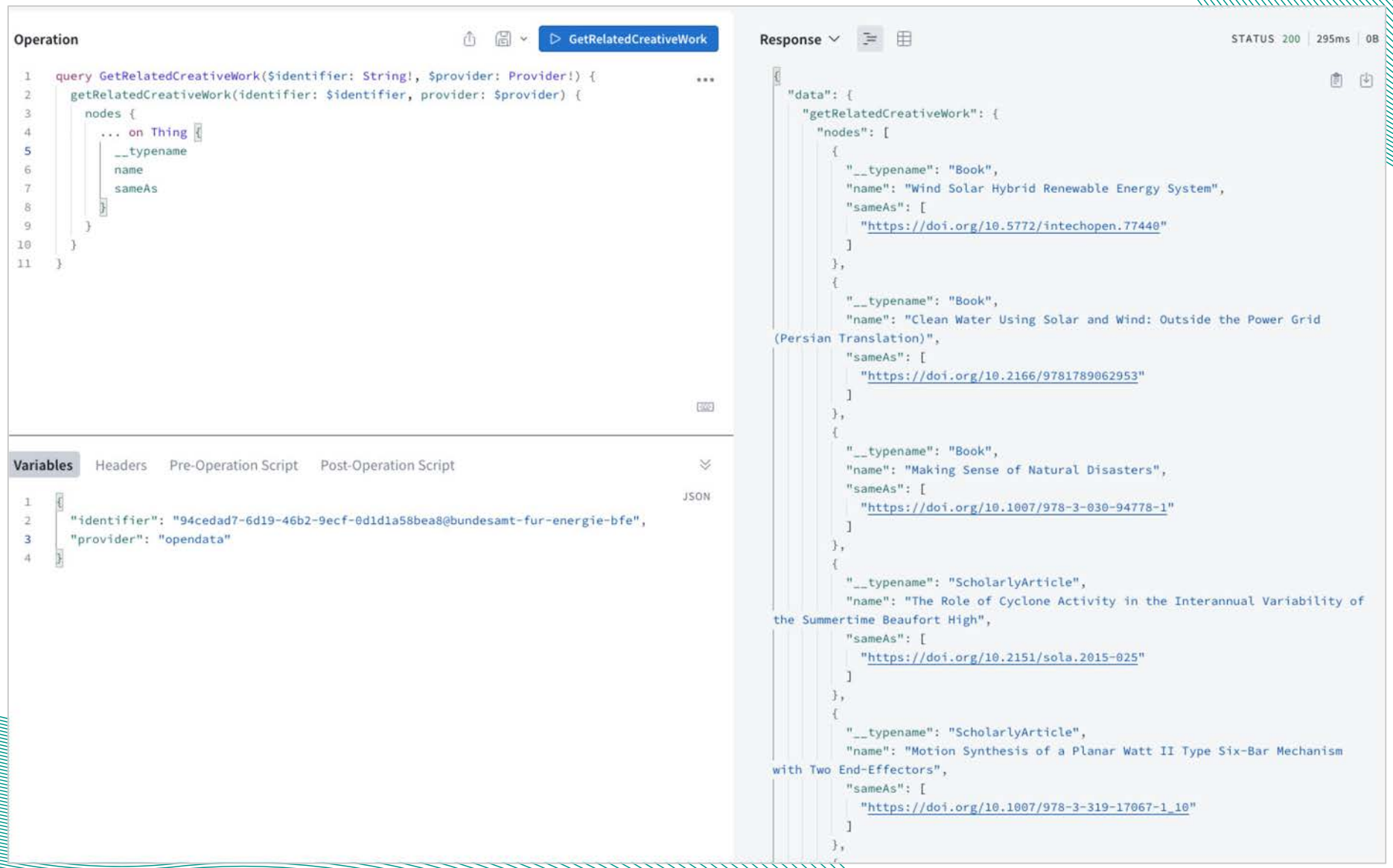
1. Users browse on opendatanavigator.switch.ch and discover datasets from opendata.swiss
2. In development: opendata.swiss implements a button on a dataset's detail view that leads the user to the corresponding page on opendatanavigator.switch.ch (Connectome resolver)
3. Possible future development: **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

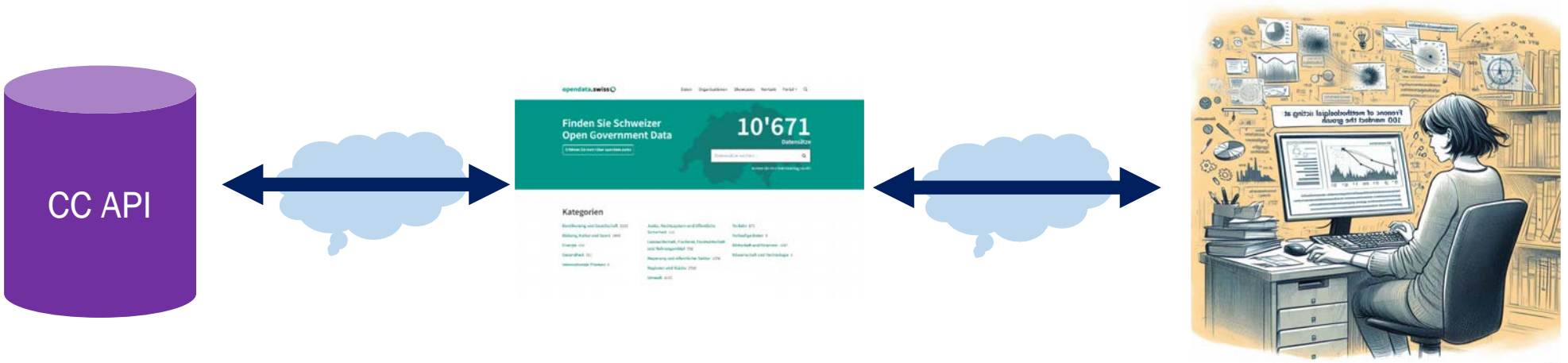


opendata.swiss queries Connectome GraphQL Endpoint – Developer Interface



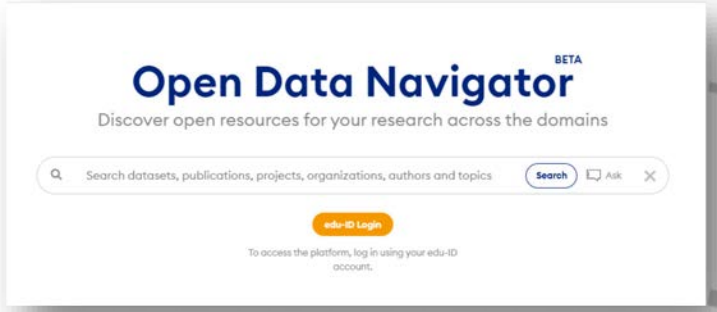
Different Ways of Accessing the Connectome Knowledge Graph (KG)

- End users use UIs, not APIs
- **Platform interoperability** can be achieved through
  - **linking between UIs** (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

Get In Touch

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