



Nationales Zentrum
für Umwelt- und
Naturschutzinformationen

Umwelt
Bundesamt

umwelt.info

search engine for environmental & nature knowledge

Johannes Frey - Scientific Officer (Wiss. Mitarbeiter), Umweltbundesamt





Agenda

- Who are we?
- What is umwelt.info?
- The role of Ontologies / Vocabularies & KGs in umwelt.info
 - UMTHES and SNS
 - Wikidata



Who are we?

The National Centre for Environment and Nature Conservation Information

- Team of ~25 people
 - data scientists, researchers, software & web developers, and content & PR specialists
- Located in Merseburg as branch of Umweltbundesamt (German Environment Agency)
- We design, develop, and maintain **umwelt.info** as our primary service
 - mostly in-house with cross-functional SCRUM teams

Motivation and Vision

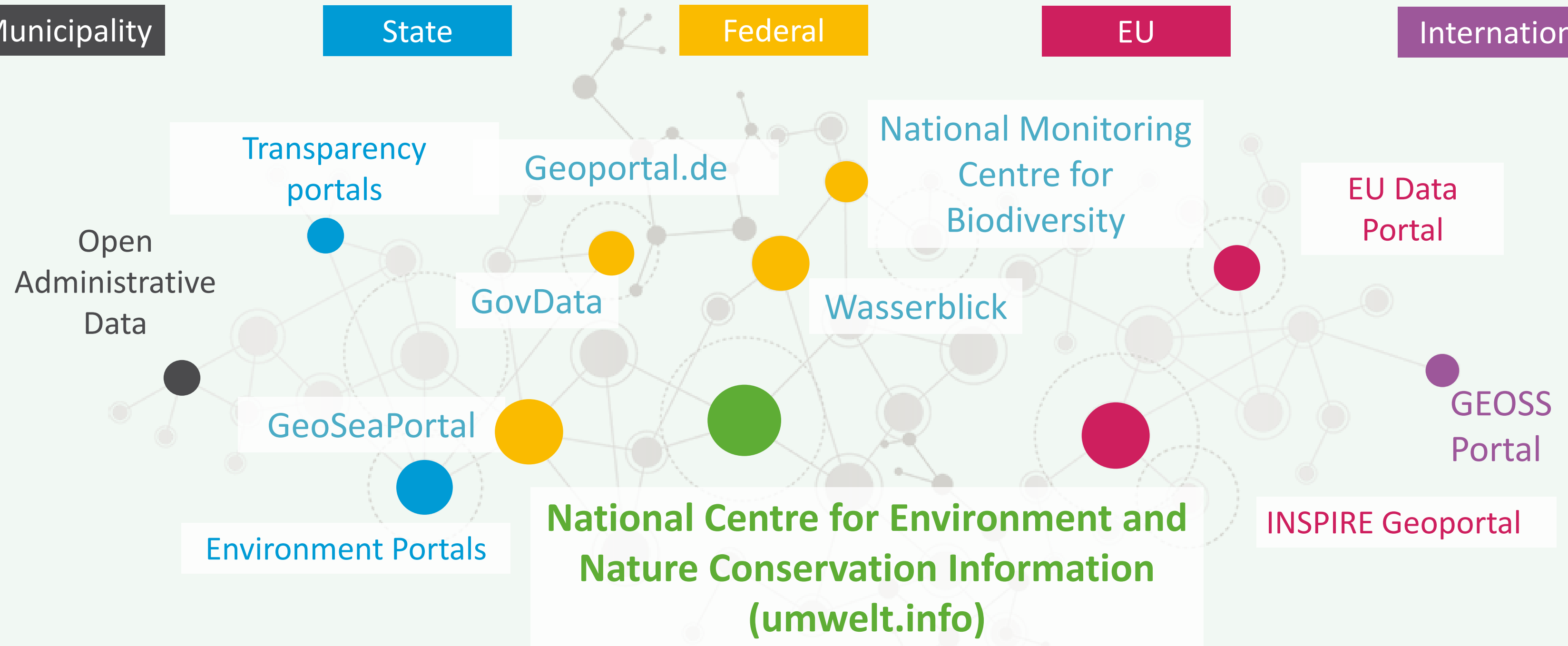
Municipality

State

Federal

EU

International

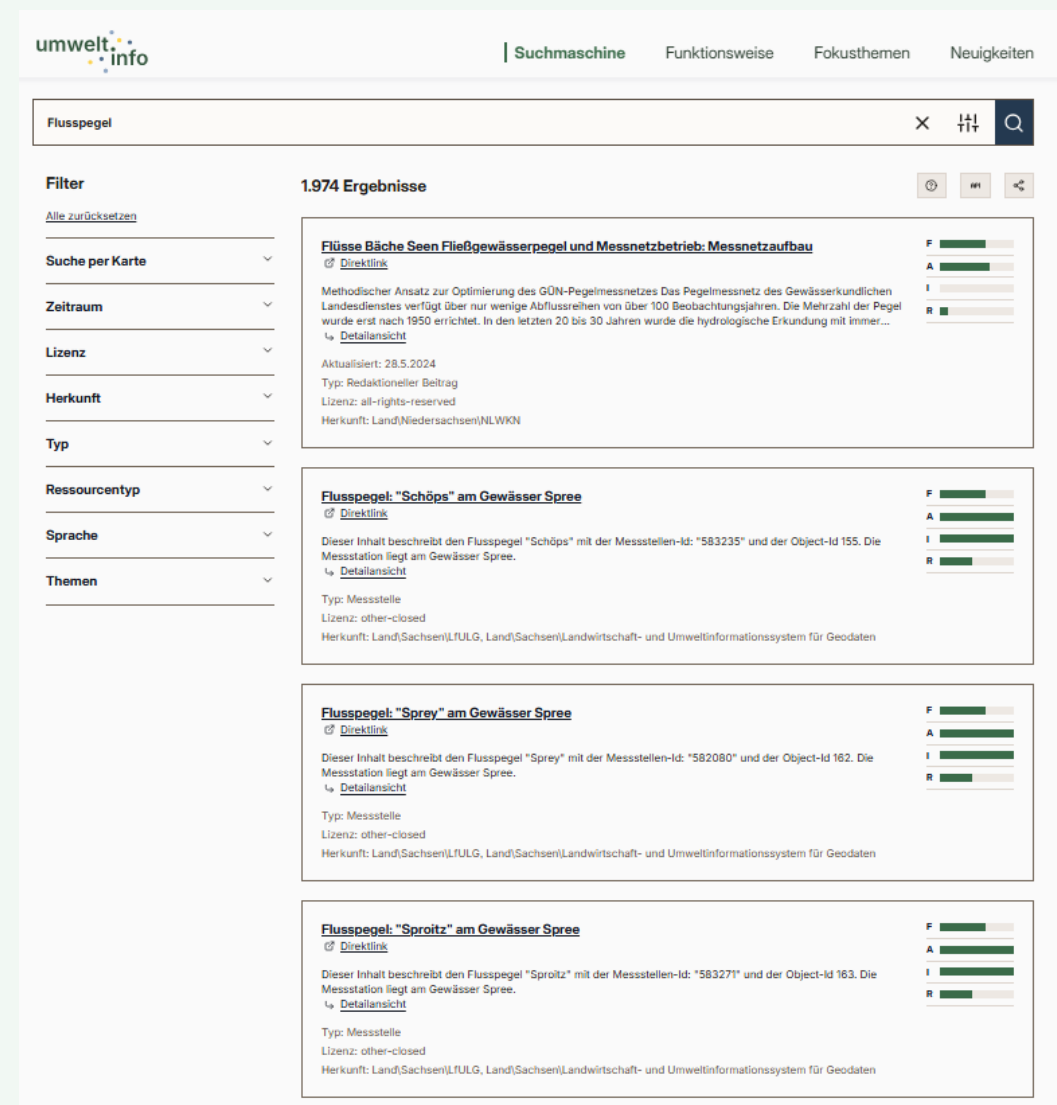


Vision: Allow central digital access to all nationwide publicly available data and information on the environment and nature conservation

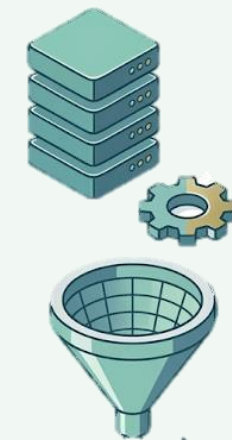
Current umwelt.info approach in a nutshell:

- We perform crawling, harvesting, and indexing of domain relevant information from official and other sources into a domain-scoped **metadata-driven** search system
- Provide a flexible and powerful faceted search UI (and API) on top of that index

Umwelt.info Portal Search UI (and API)



Metadata Aggregation, Consolidation & Enrichment



umwelt.info Metadata index



Metadata Harvesting & Extraction

Selected (Meta)Data Sources

govdata.de

open.nrw

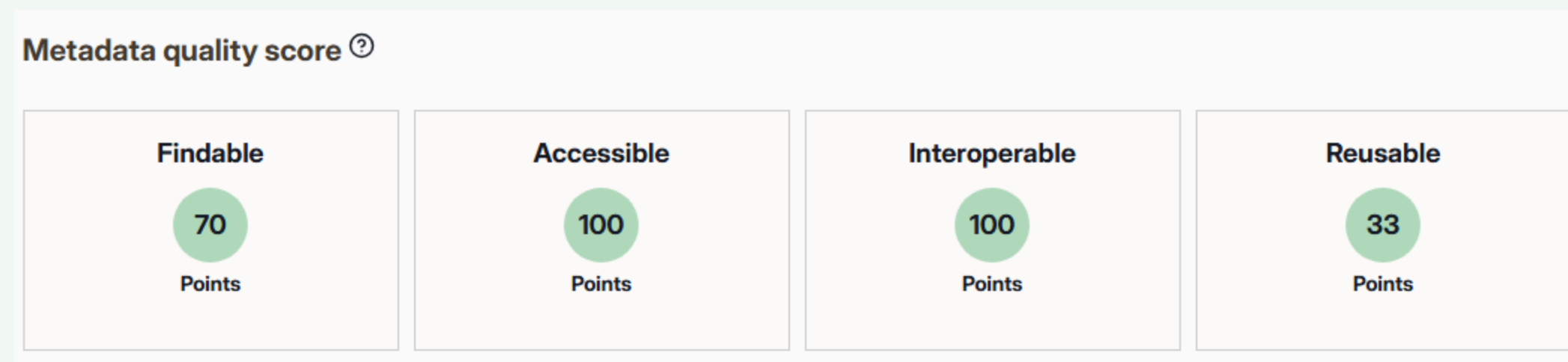
Sachsenforst

Wasser-DE

Current umwelt.info approach (continued)

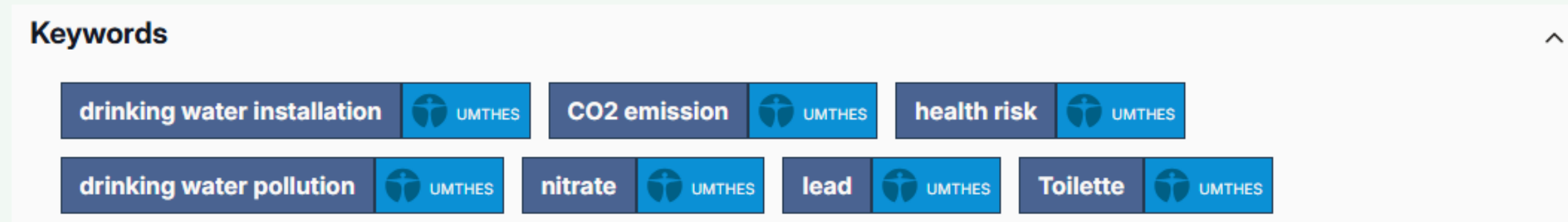
But there is more to it than just plain metadata aggregation:

- Harmonization of metadata fields and identifiers for unified search
- Deduplication of datasets(records)
- more fine-grained metadata w.r.t. domain interest and demands
 - indexing important aspects/content of the dataset itself
 - **enrichment** with additional information, KG data and semantics
- evaluating adherence of published data to FAIR principles with umwelt.info FAIR score



Semantic Metadata Enrichment using UMTHESES and SNS

- Automated annotation of keywords and topics from metadata description
- Then linked to UMTHESES concepts (German environmental thesaurus)



UMTHES and SNS

- UMTHES is SKOS ConceptScheme maintained at Umweltbundesamt
- contains 14,000 SKOS concepts connected via broader/narrower relations
- Polyhierarchical: air pollution ($\Rightarrow$ air; environmental pollution)
- Thematically related terms are linked (waste $\Leftrightarrow$ waste law)
- 31,000 German synonyms + 35,000 English synonyms
- Linked Data Deployment of UMTHES via SNS (Semantic Network Service)

Linking to Knowledge Graphs

- Organisations involved with the data are matched and linked to Wikidata entities if possible




Stream measurement site Neu-Ulm, Donau

The measurement site Neu-Ulm (ID: 10026293) is located in the waterbody Donau in Bavaria. The site is surveilling the flow rate.

Direct link:
<https://www.gkd.bayern.de/en/rivers/discharge/kelheim/neu-ulm-10026293>

 **Provide feedback**

Organisations

[Bavarian Ministry of the Environment and Consumer Protection \(Provider\)](#) 

 [Gewässerkundlicher Dienst Bayern \(Publisher\)](#) 

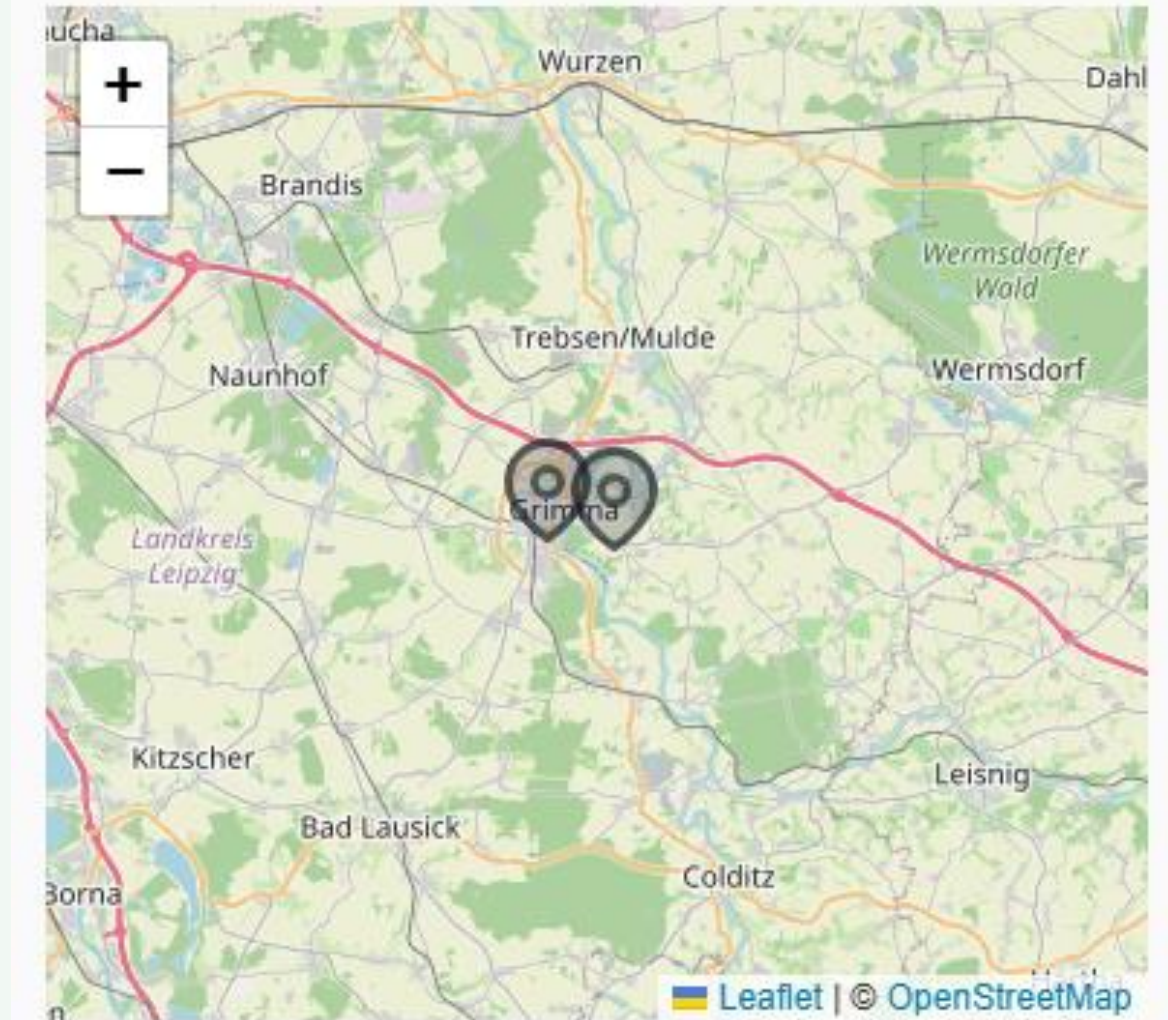
[Wasserwirtschaftsamt Donauwörth \(Operator\)](#)  

Linking to Knowledge Graphs and Vocab

- Geocoding to EIONET spatial Units, DCAT-AP.de regions, and Geonames entities

```
"regions": [
  {
    "Watershed": {
      "id": "DE5400",
      "label": "MULDEELBESCHWARZE ELSTER",
      "url": "http://dd.eionet.europa.eu/vocabulary/wise/SpatialUnit/euSubUnitCode.DE5400"
    }
  },
  {
    "RegionalKey": {
      "id": 1400000000000,
      "label": "Sachsen",
      "url": "https://www.dcat-ap.de/def/politicalGeocoding/regionalKey/#1400000000000"
    }
  },
  {
    "GeoName": {
      "id": 6550866,
      "label": "Grimma",
      "url": "https://www.geonames.org/6550866/"
    }
  }
],
```

Region



MULDEELBESCHWARZE ELSTER

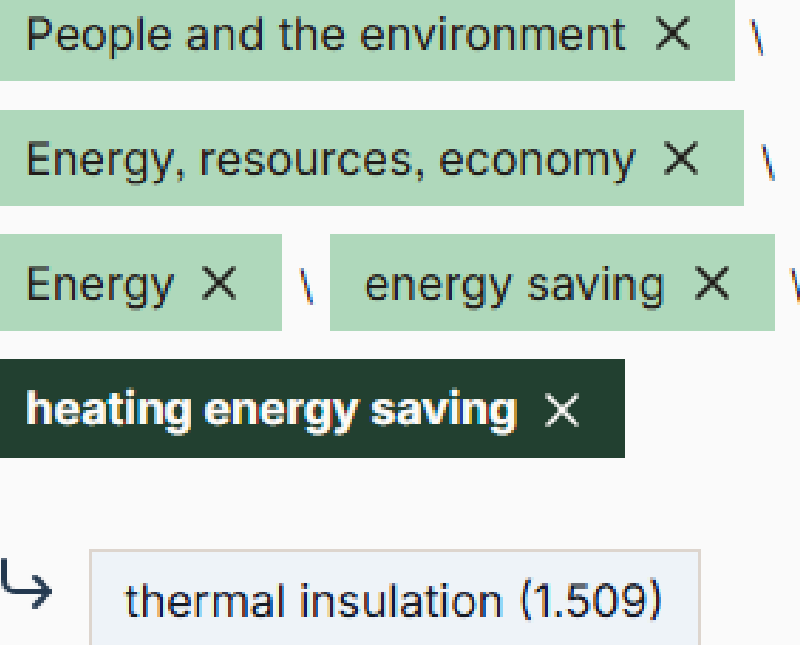
Sachsen

Grimma

Semantic Enrichments allow for better UX

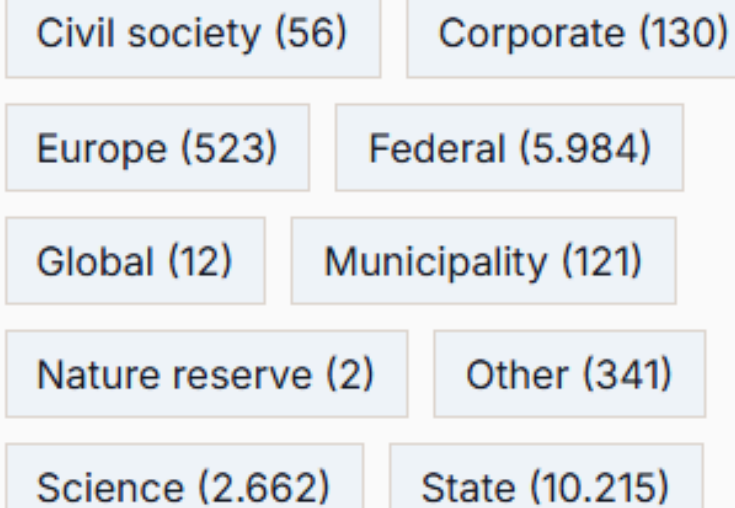
Facets leveraging information of those linked resources allow for structured filtering of search results e.g. by...

Topics



UMTHES topics

Organisation



Organisation (types)

Search by map



GeoLocation

Danke!

Please reach out to us for any kind of feedback.



Nationales Zentrum für Umwelt-
und Naturschutzinformationen

Gotthardstraße 37
06217 Merseburg



<https://umwelt.info/en/contact>



umwelt.info@uba.de

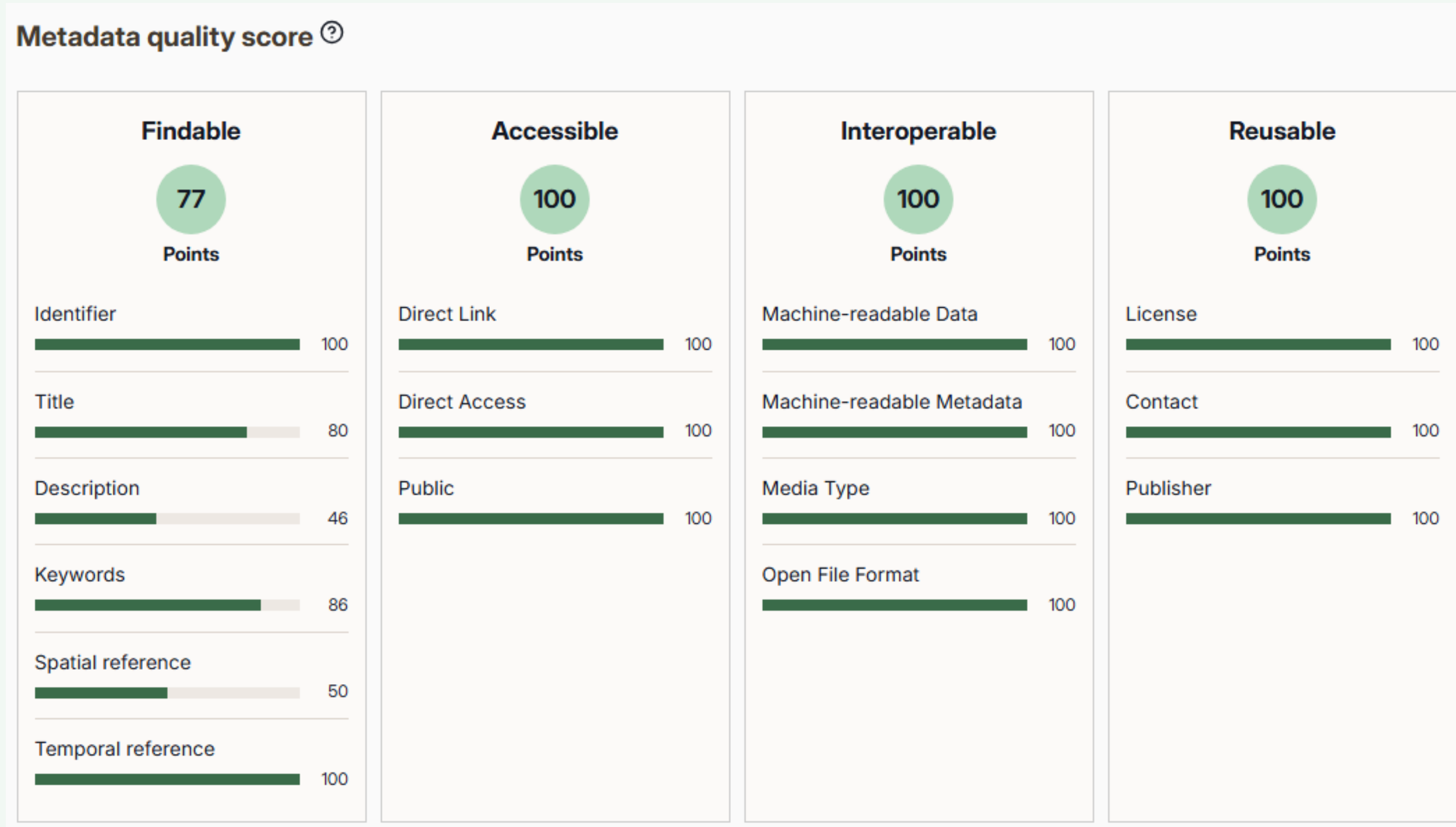


www.umwelt.info

Linked Data benefits & potential

- API standards and standardized metadata schemas allow for easier harvesting of new (meta)data sources into our index
- Easier aggregation & consolidation of information from different and external sources into our index
- More powerful description of the data and its context with rich Ontologies / Vocab / Thesauri => allow for better search (recall and precision)
- If metadata is consumed directly into information systems (or AI agents) and well connected to LOD => ample opportunities to fetch additional data needed to your specific use case
- However there is still a lack of Linked Data efforts when it comes to (German) open gov data
- LOD Core Infrastructure for administrative / authoritative Data (e.g. Rivers, Streets, Regions, Organisations, Measurement stations, etc.) could be beneficial

FAIRness Score



The case of licenses:

License ^

Open X \ **Creative Commons** X

↳ Attribution (156.588)

Attribution-NoDerivs (57)

Attribution-NoDerivs-NonCommercial (194.960)

Attribution-NonCommercial (1.068)

Attribution-NonCommercial-ShareAlike (39)

Attribution-ShareAlike (9.732)

Public Domain Mark (28)

zero (5.814)