OPEN (WEB) SEARCH, A BOOSTER FOR OPEN SCIENCE?

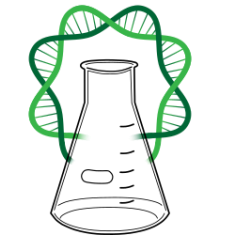
Stefan Voigt and Tobias Hecking

German Aerospace Center, Oberpfaffenhofen/Cologne, Germany

#ossym2023, CERN – Geneva, 4-6 Octobre 2023

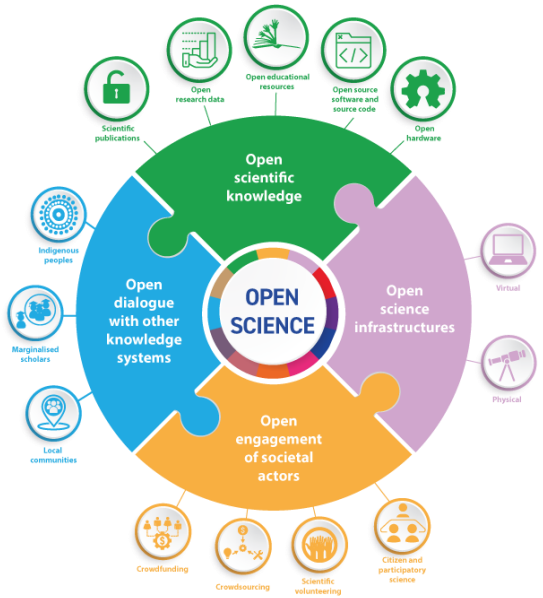
Openness, transparency, accessibility and reproducibility are fundamental prerequisites of science

- **‘Open science’ promotes the general openness of scientific methods, computer-source-code, data, publications, peer-review processes and educational resources**
- The **open access principle** for scientific ‘publications’ is of utmost importance to widely spread, share and debate, and in the end societally accept scientific findings.
- **Similarly, the findability and accessibility of (research-) data is of utmost importance for acceptance and reproducibility of scientific findings**
  - Transparent and reproducible finding of information therein is of fundamental importance for science in general
  - An open web index can be established as a publicly audited and constantly available data base for web resources, and if it can be accessed and searched in an open and unbiased manor, it may serve as a significant reference and booster for open science in the long run



open science

CC BY-SA 3.0



Source: UNESCO.org
Understanding open science

In this talk, we discuss

- How DLR is working to improve the synergistic searchability, findability of internal documents, data, publications and computer code in synergy and together with external web resources and web indices generated in the pan European Open Web Search ecosystem, currently under development.
- This synergistic use of state-of-the-art and innovative open search capacities, by linking DLR-internal and external scientific resources, aims at providing DLR scientists with up-do-date, objective and unbiased information and will allow DLR researchers to transparently search and access Web resources for open and reproducible open science.

Working towards open science search and supporting open science

Navigation in complex information spaces

Curated research datasets

PANGAEA.
Data Publisher for Earth & Environmental Science

terrabyte STAC API

EOWEB® GeoPortal

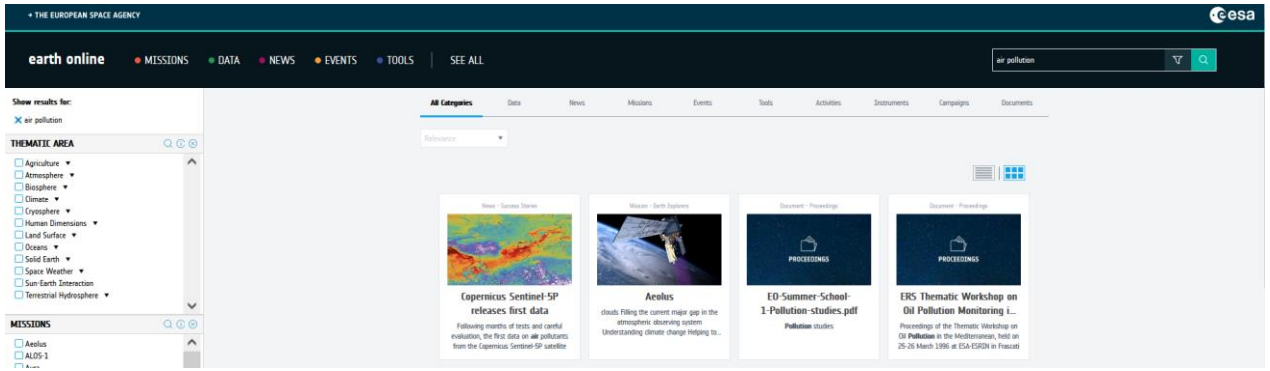
Open WebSearch .eu

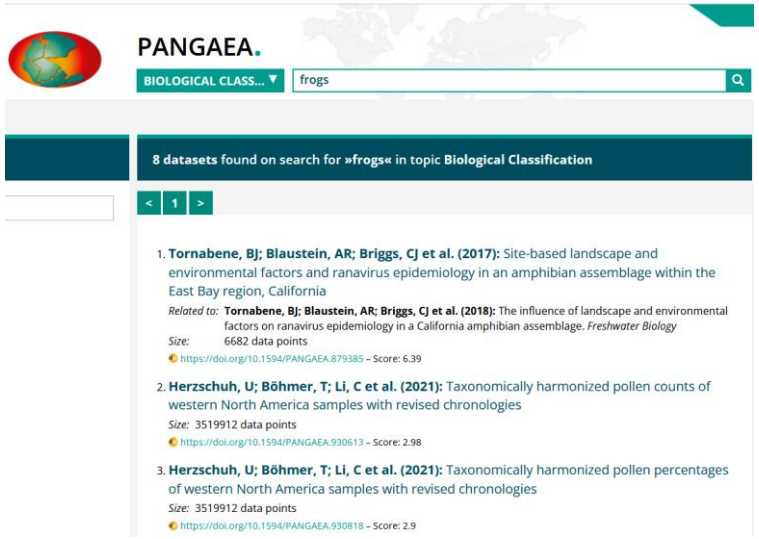
COMMON CRAWL

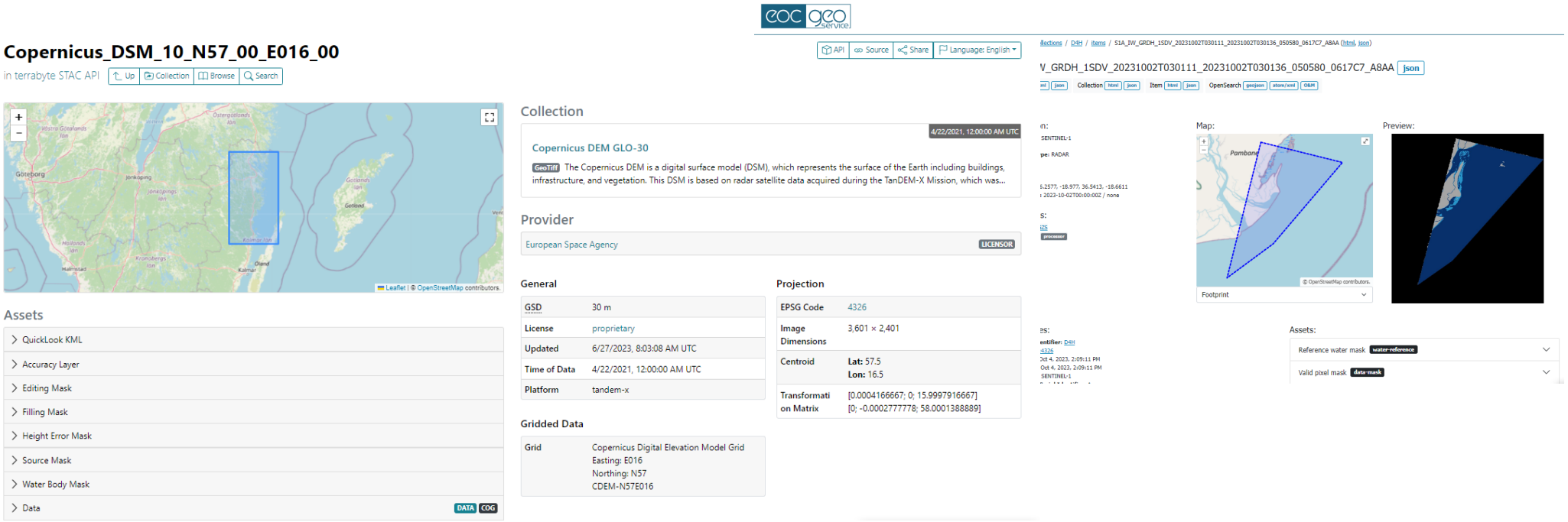
elib
Publikationen des DLR

OpenAlex

Data focussed search engines







Working towards open science search and supporting open science

Navigation in complex information spaces

PANGAEA.
Data Publisher for Earth & Environmental Science

terrabyte STAC API

EOWEB® GeoPortal

Open WebSearch .eu

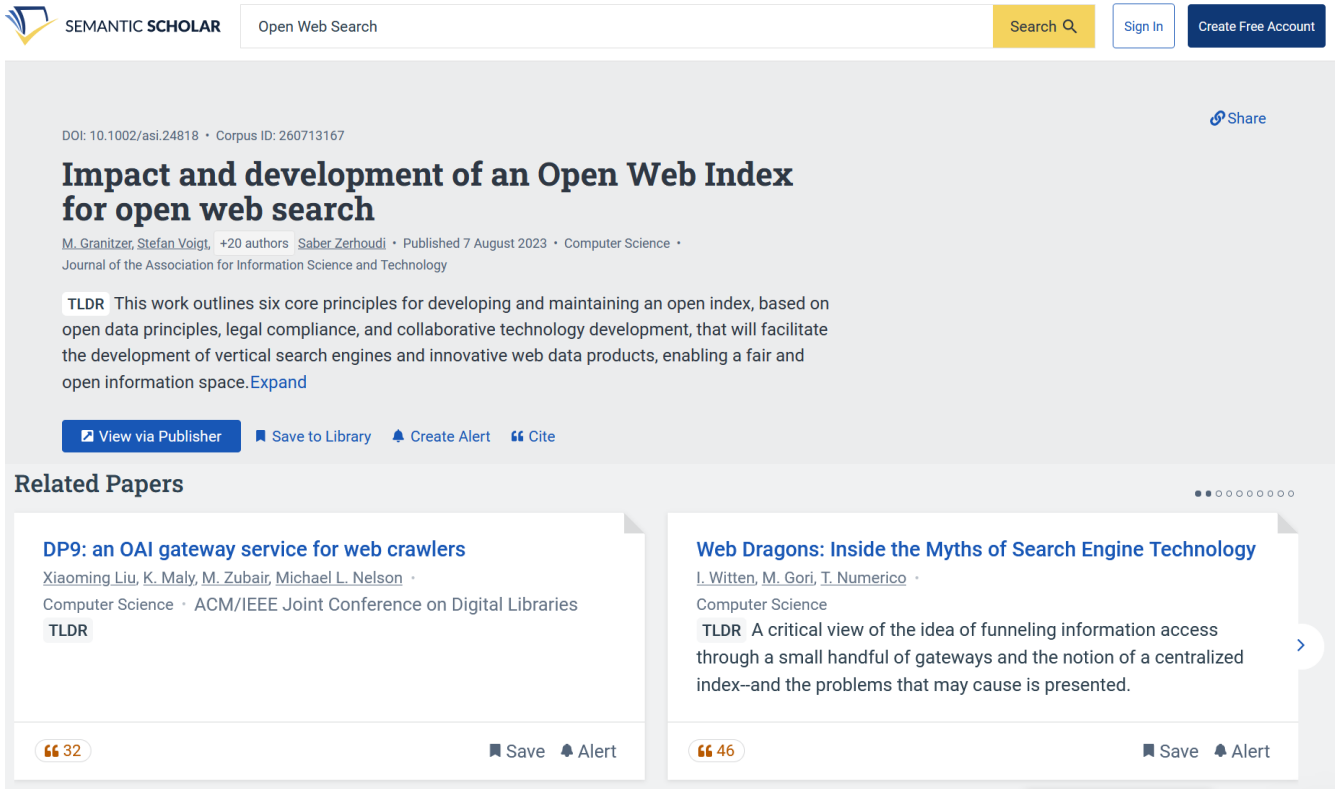
COMMON CRAWL

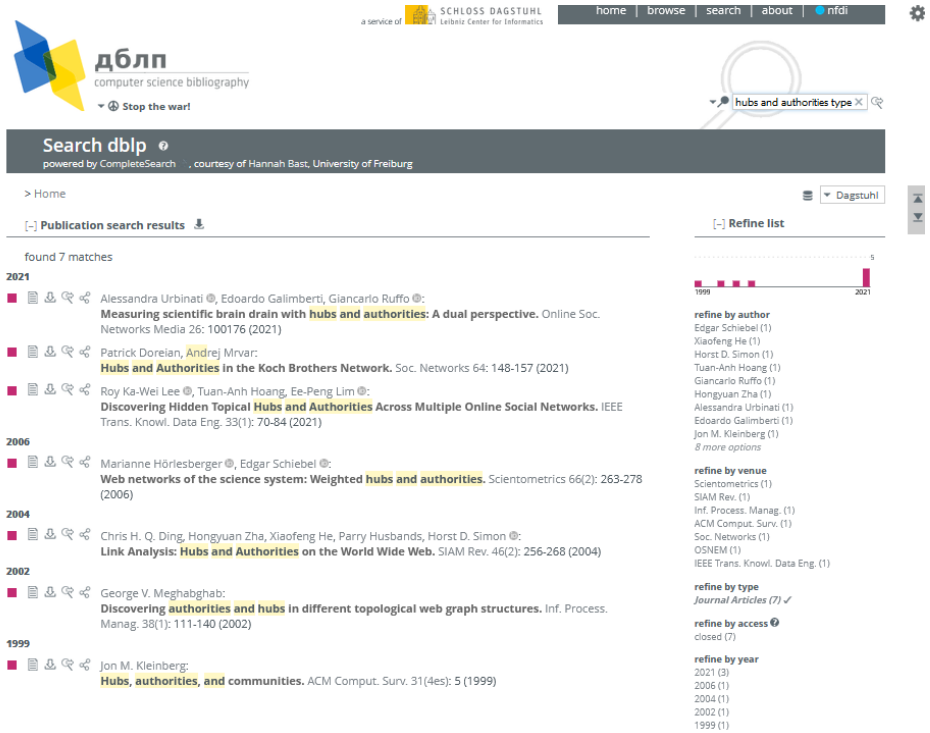
Formally published and (peer reviewed) knowledge

elib
Publikationen des DLR

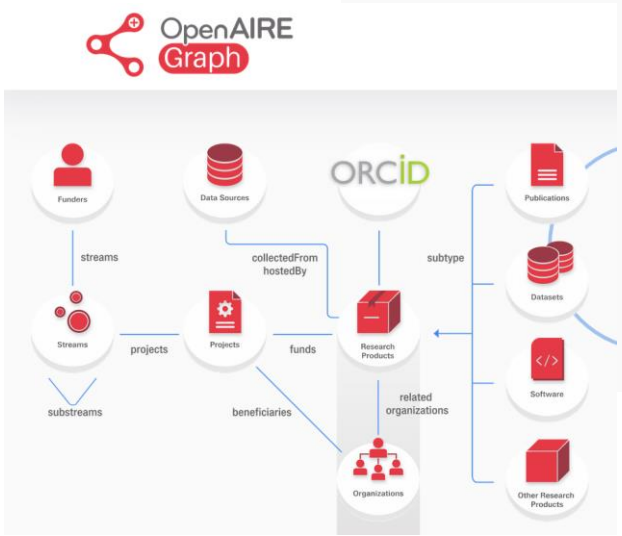
OpenAlex

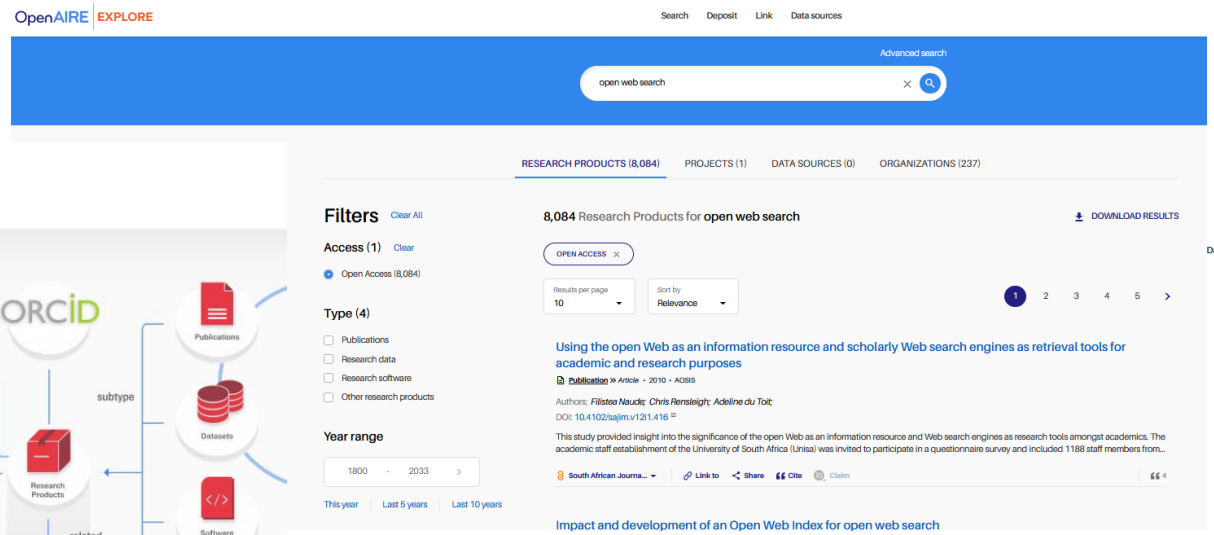
Publication focussed search engines

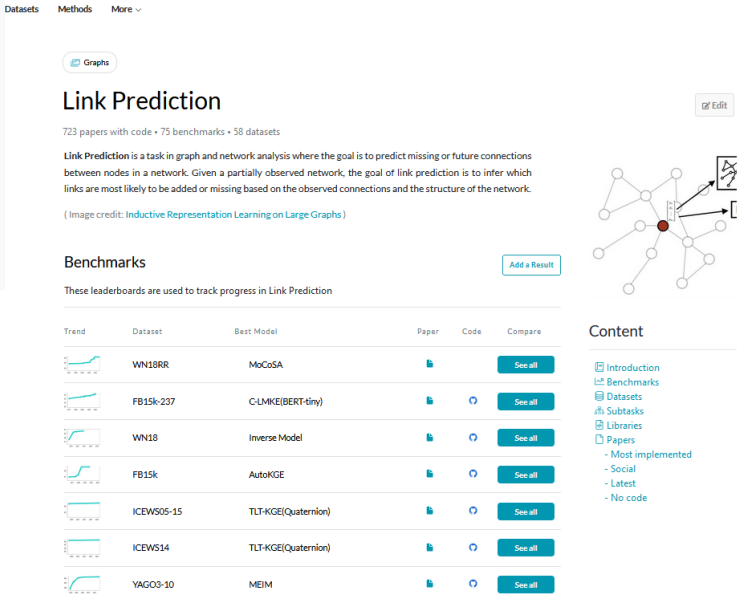




Integrated search of research outputs







Working towards open science search and supporting open science in OpenSearch@DLR 2.0

Navigation in complex information spaces

PANGAEA.
Data Publisher for Earth & Environmental Science

terrabyte STAC API

EOWEB®
GeoPortal

Unstructured knowledge on observations, reports, …

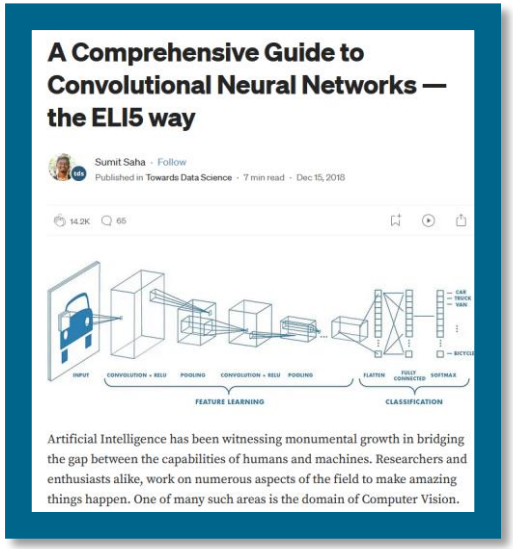
Open WebSearch .eu

COMMON CRAWL

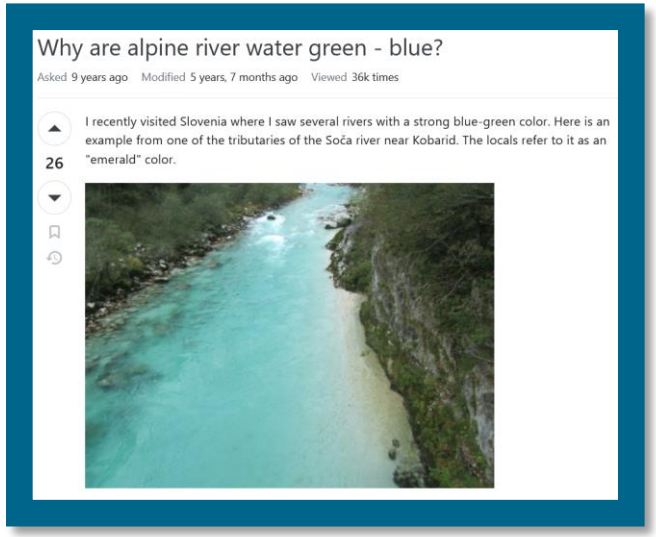
elib
Publikationen des DLR

OpenAlex

The web is full of scientifically relevant information



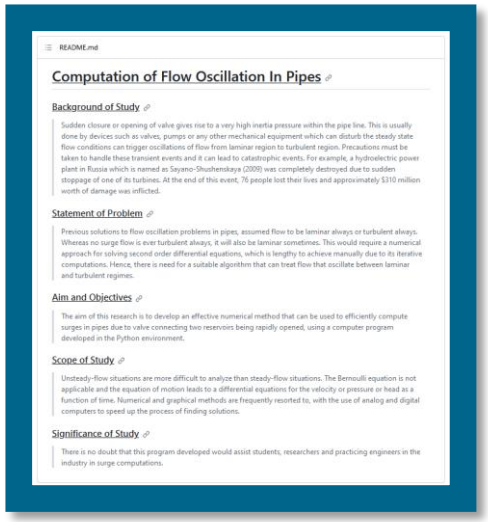
Gray literature (blogs, wikis, …)



Observation reports

Reports and brochures

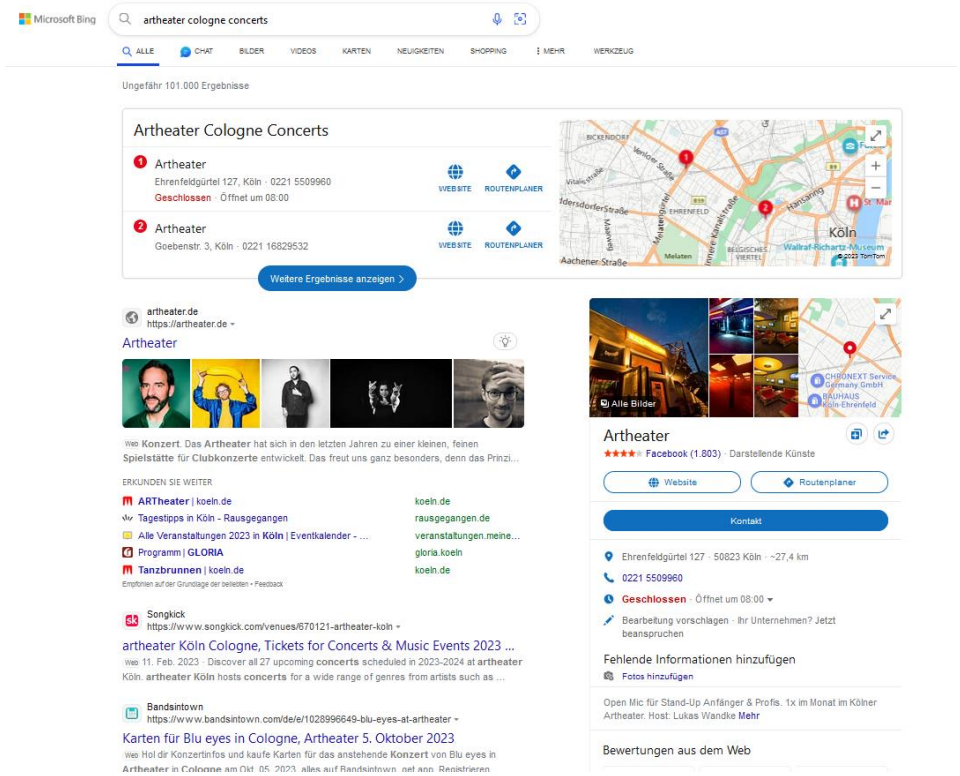
News

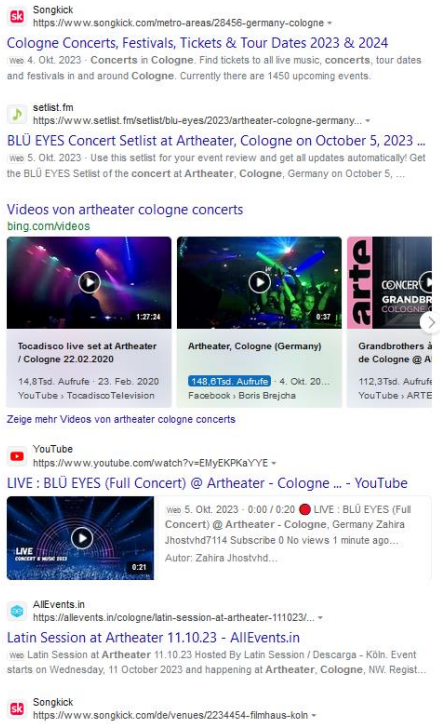


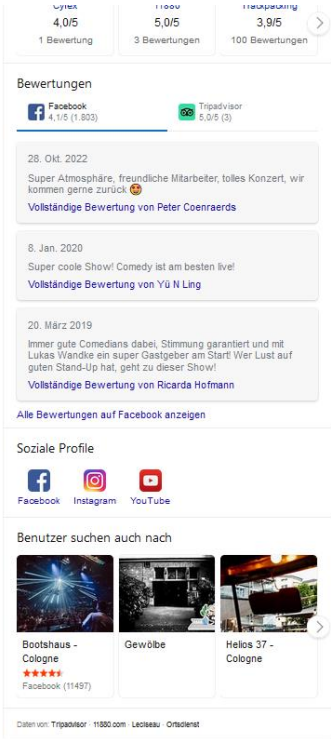
Software and data descriptions

Search result enrichment for science search

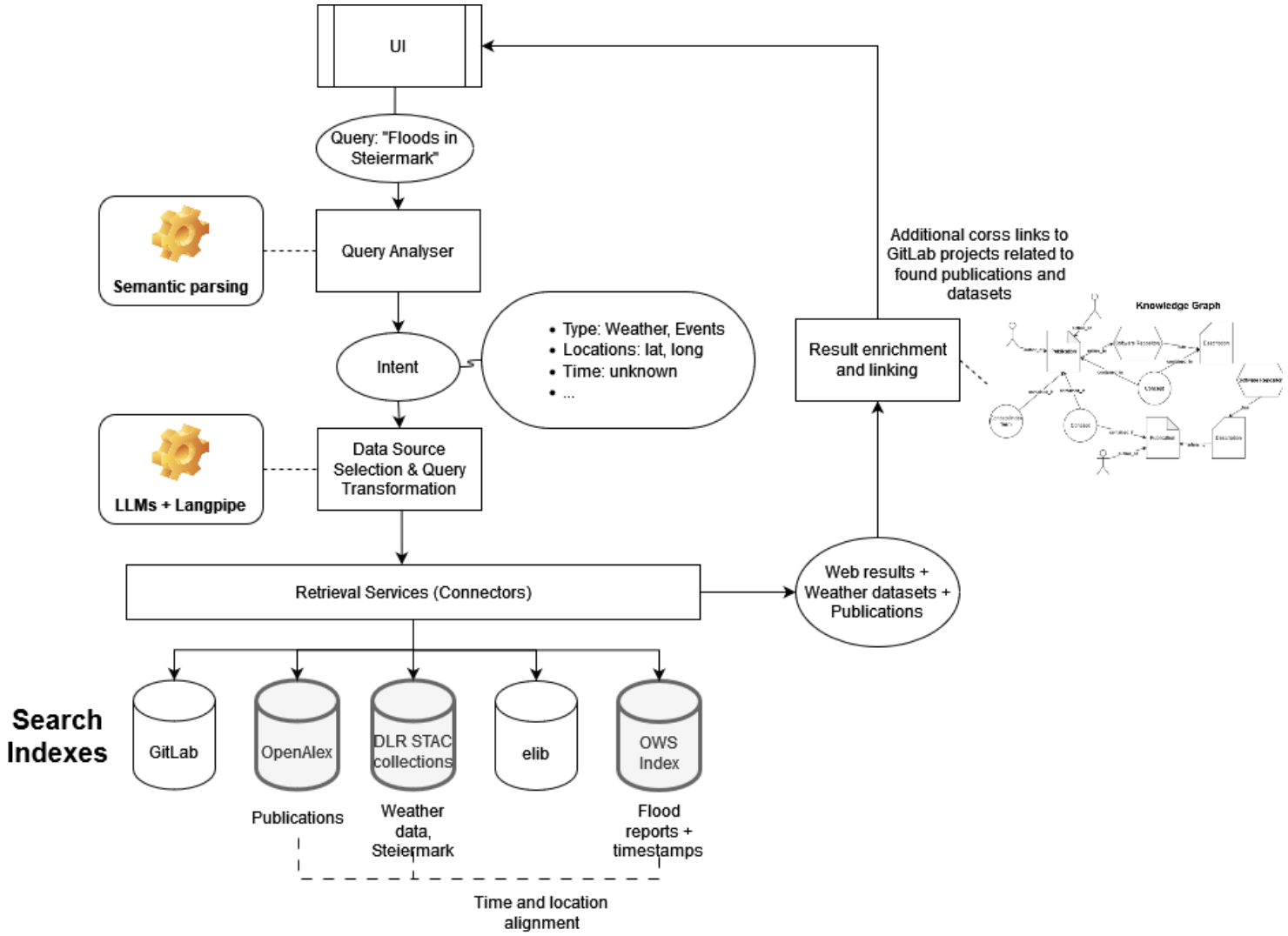
- Result enrichment is very common for general web search.
- Why not for scientific search?



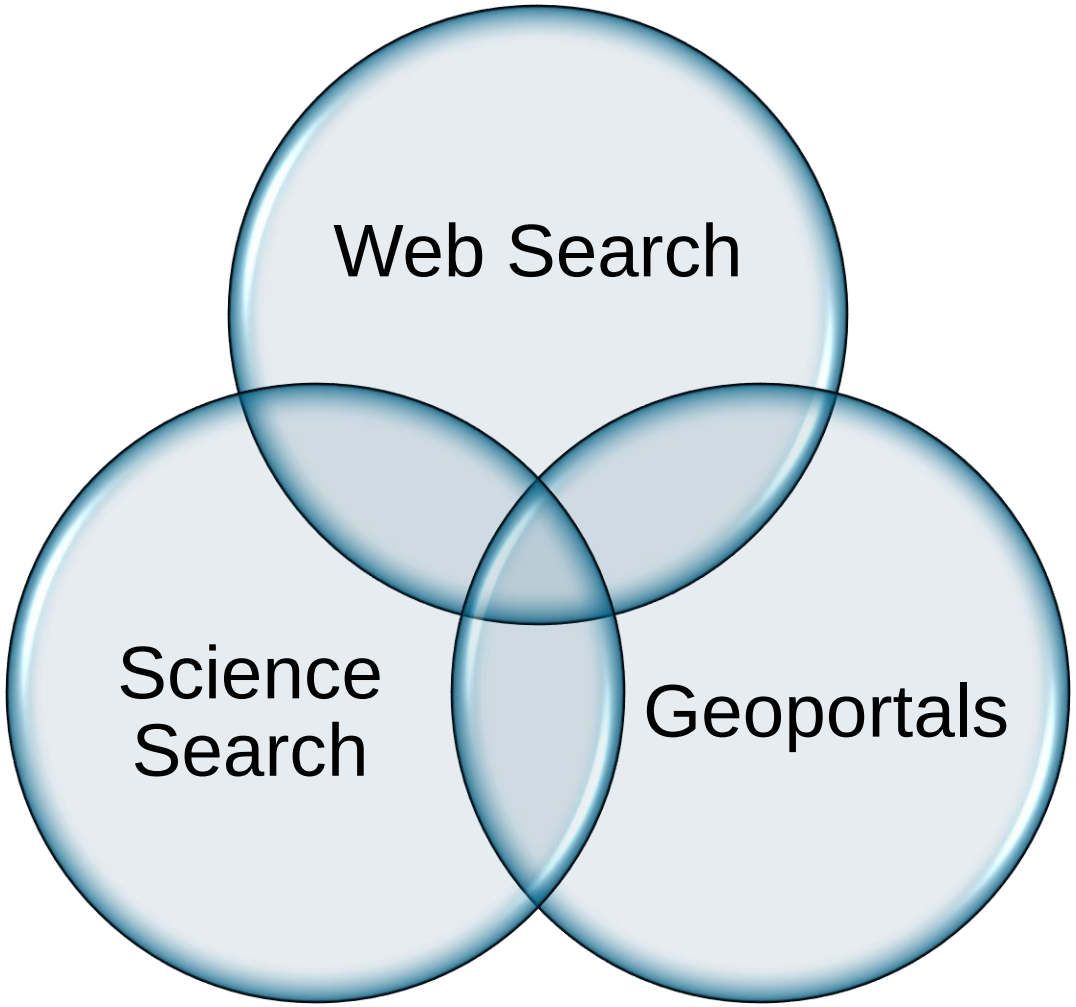


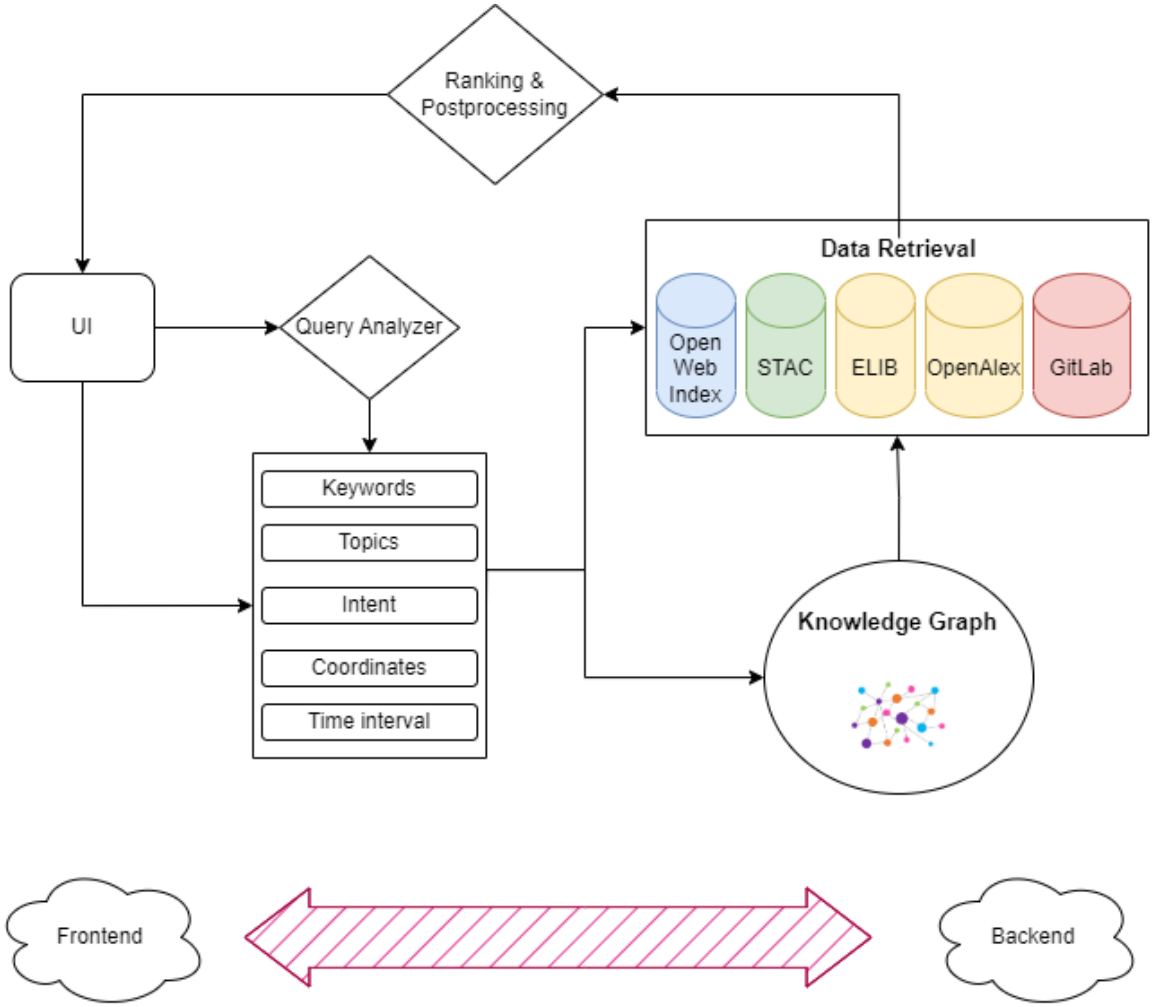


Pilot application: Search result enrichment for environmental science search



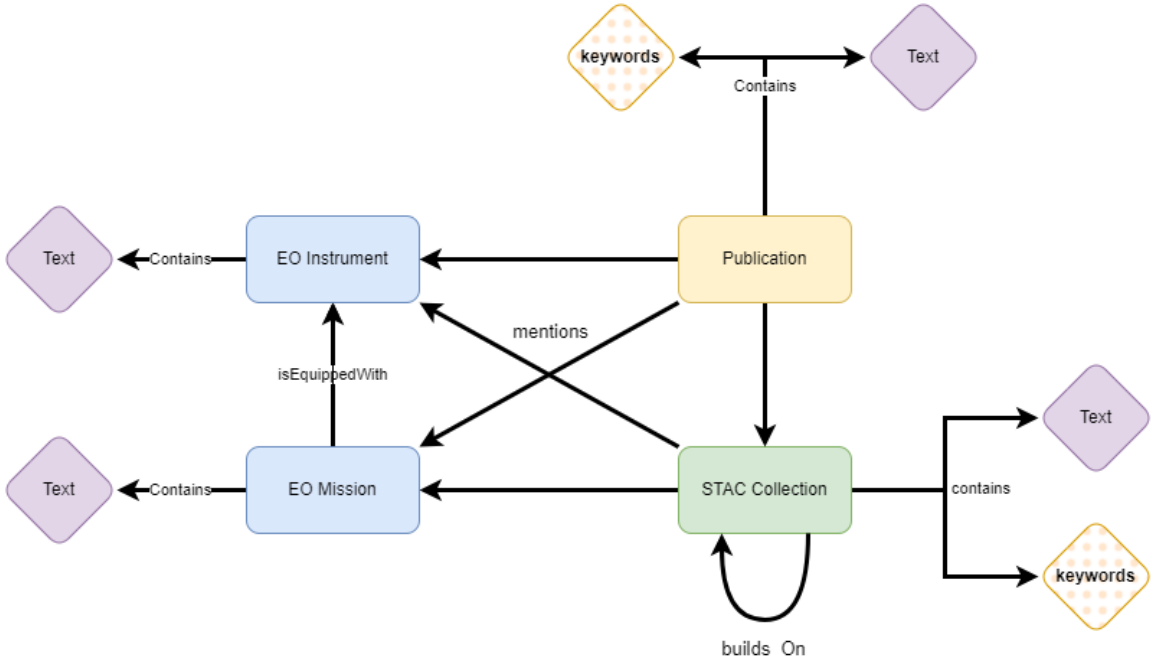
Pilot application: Search result enrichment for environmental science search



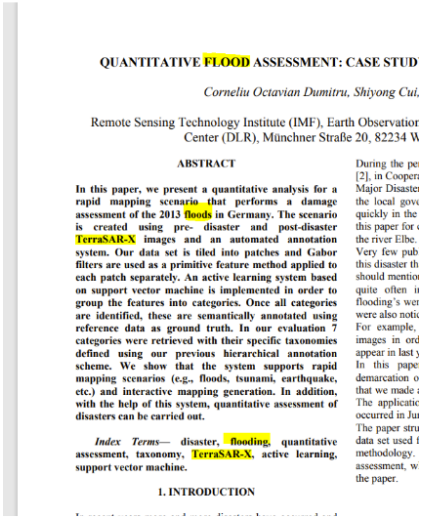


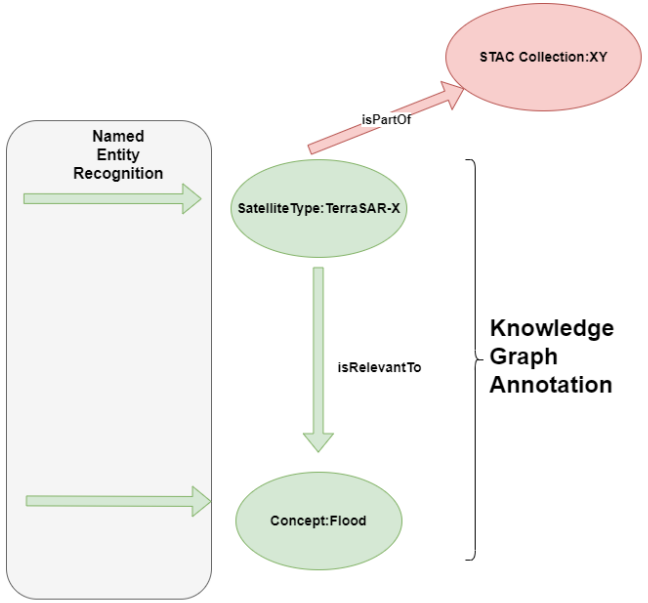
Pilot application: Search result enrichment for environmental science search

Structured ressource represenation using Corpus Analytics Builder*



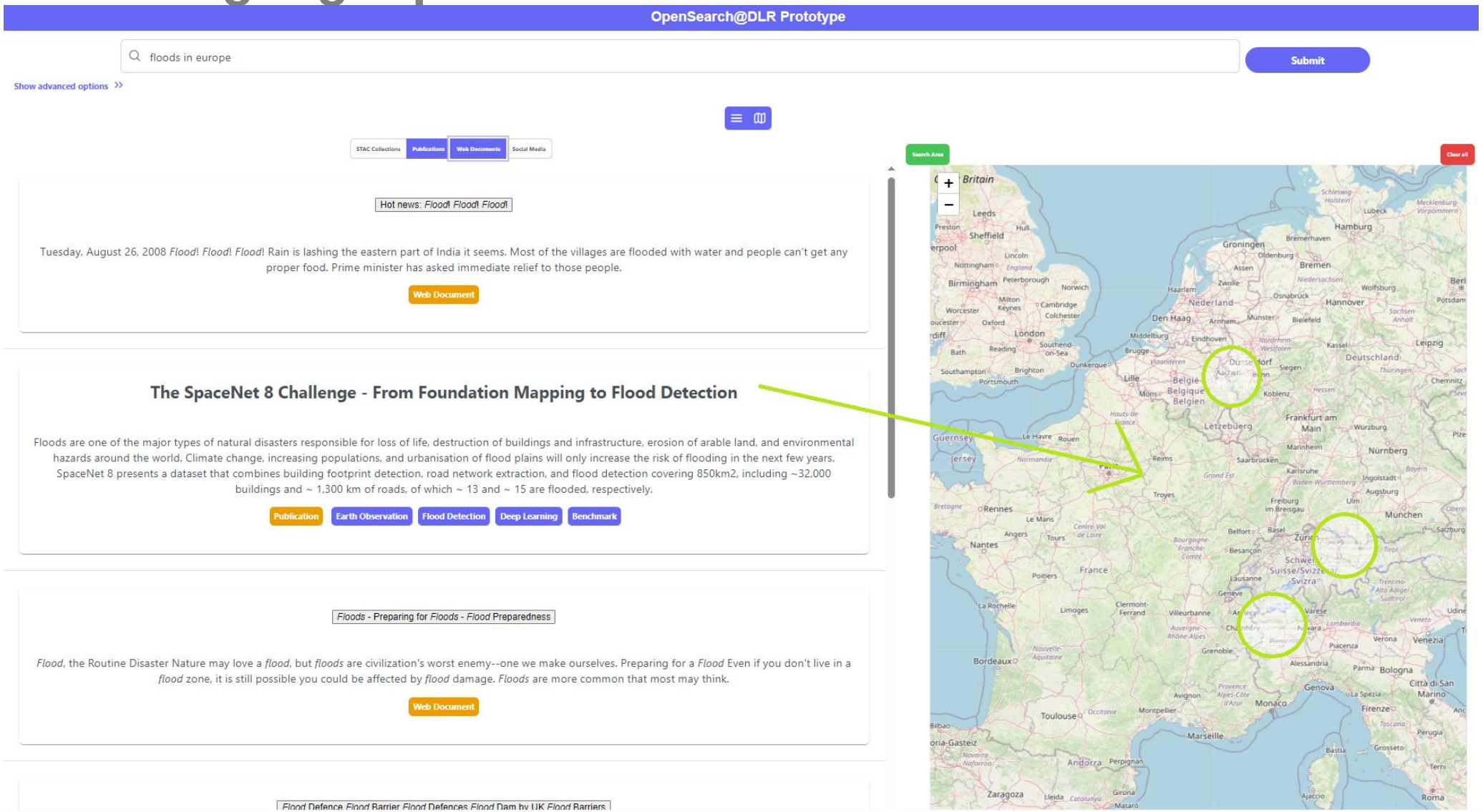
Publications as mediating artefacts



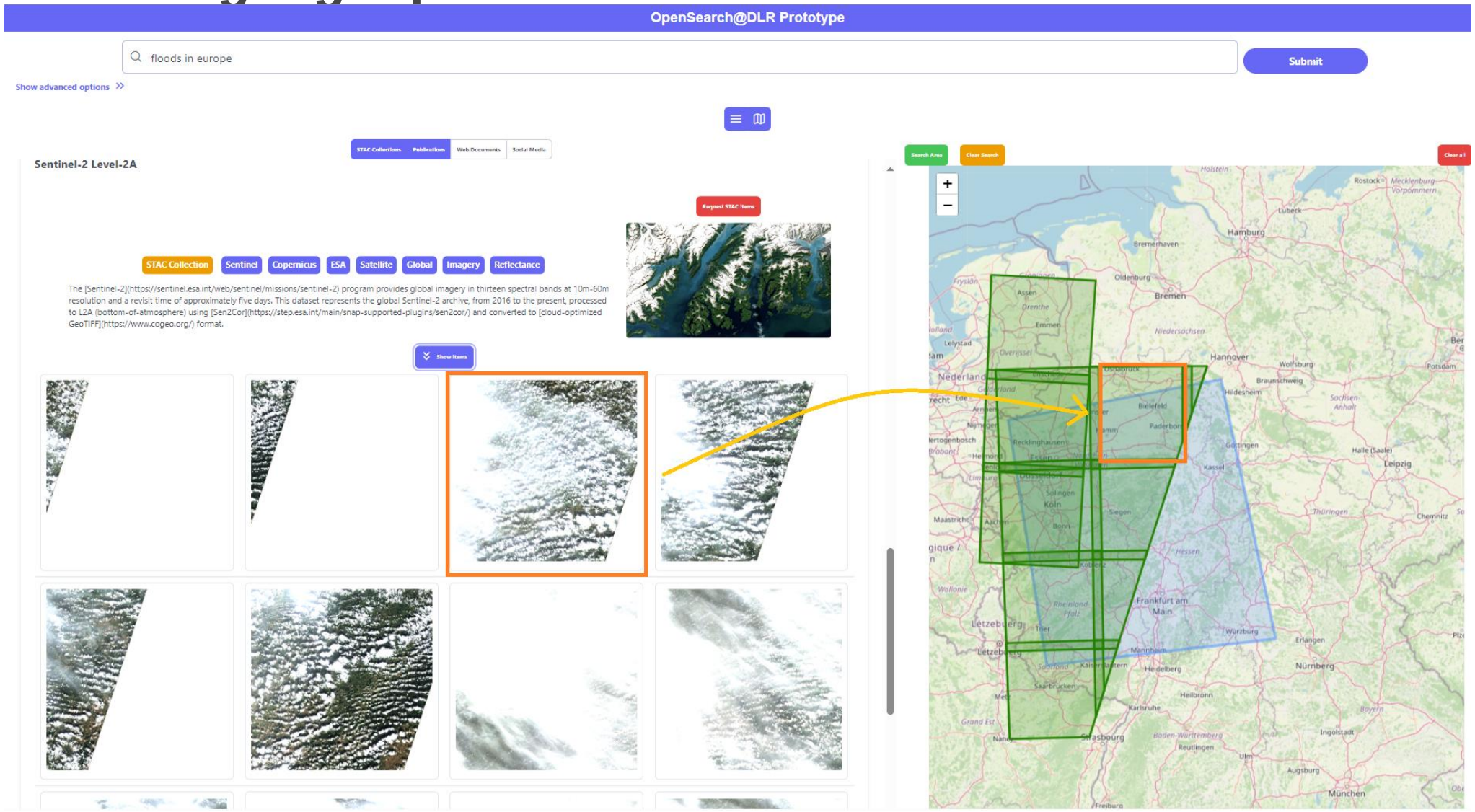


* El Baff, R., Hecking, T., Hamm, A., Korte, J. W., & Bartsch, S. (2023). Corpus Annotation Graph Builder (CAG): An Architectural Framework to Create and Annotate a Multi-source Graph. In *Proceedings of EACL'23: System Demonstrations* (pp. 248-255).

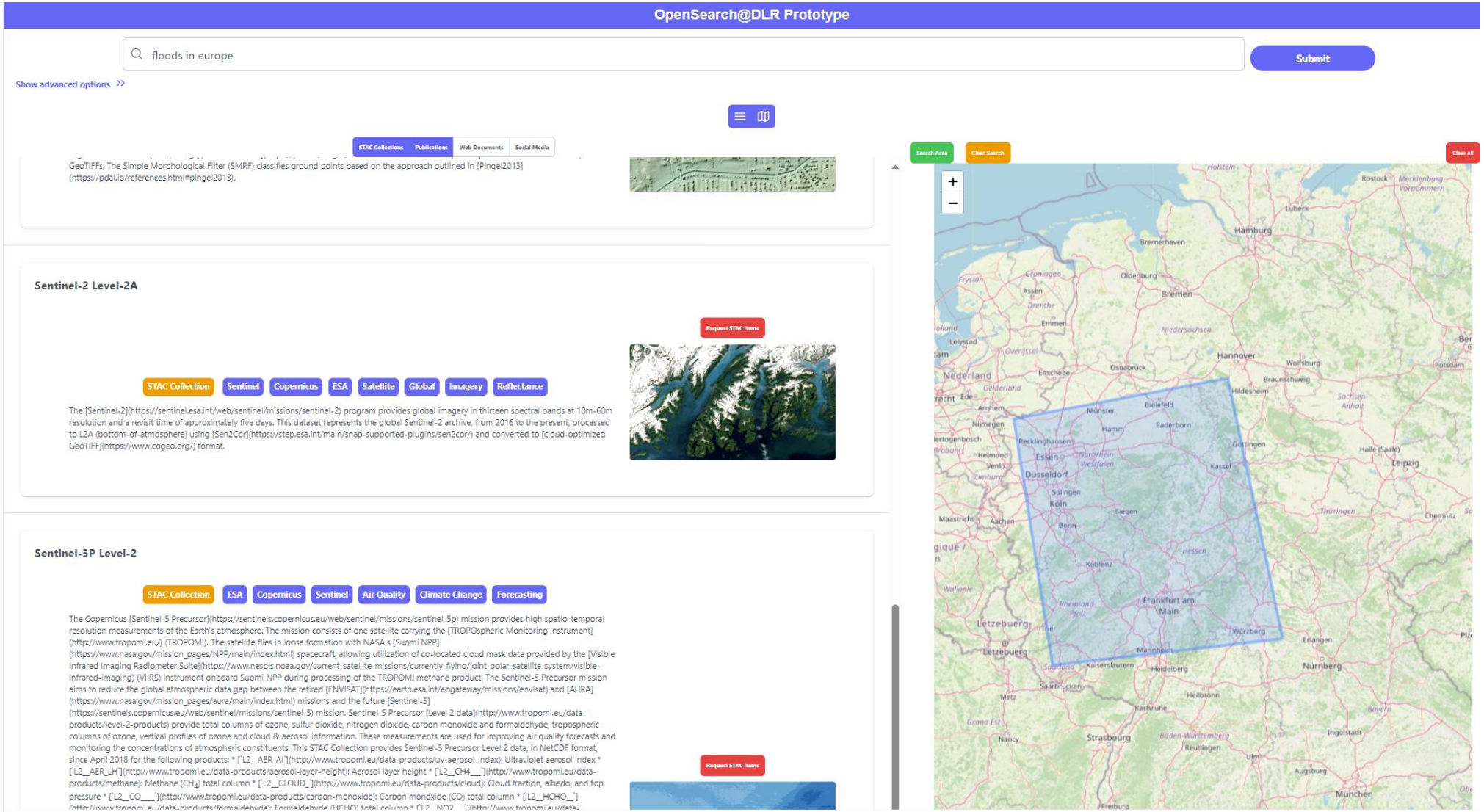
Relating remote sensing data, publications and the web: Natural language queries



Relating remote sensing data, publications and the web: Natural language queries



Relating remote sensing data, publications and the web: Spatio-temporal queries



Conclusions

- Tapping the web for scientific relevant information is still underexplored.
- Open (Web) Search is still in an early stage of its development. However, given the ever-increasing informational needs for open and objective science, Open Web Search needs to be implemented at its full potential and in a joint effort of science organisations in the years to come.
- Huge potential also beyond science applications:
  - Underpinning news, claims, etc. with data.
  - Identification of information biases.