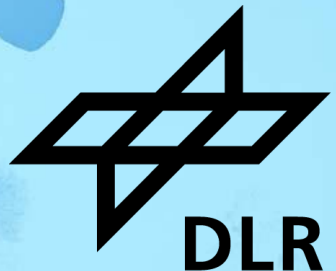


# OPEN (WEB) SEARCH, A BOOSTER FOR OPEN SCIENCE?

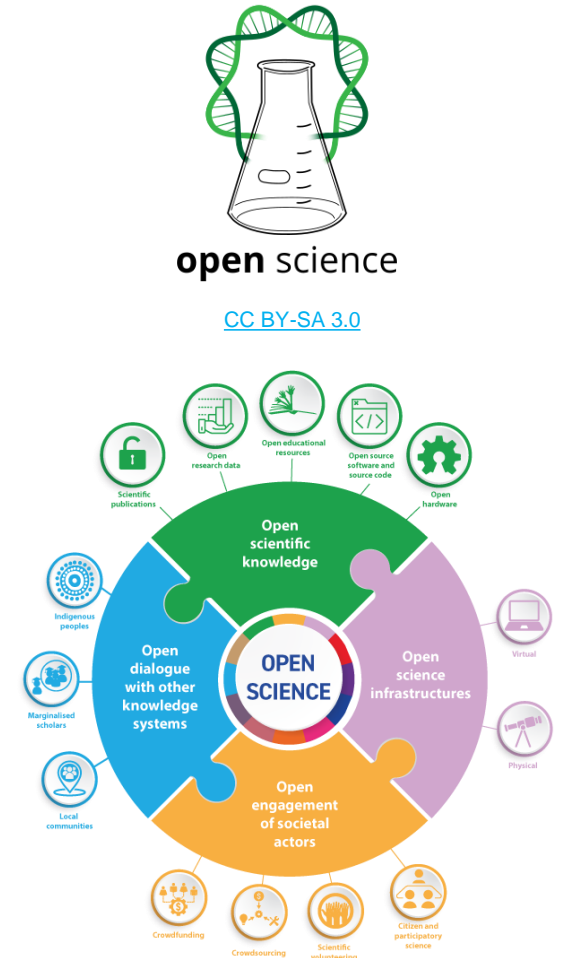
**Stefan Voigt and Tobias Hecking**

German Aerospace Center, Oberpfaffenhofen/Cologne, Germany



# 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



[CC BY-SA 3.0](#)

Source: [UNESCO.org](https://www.unesco.org/en/understanding-open-science)  
[Understanding open science](https://www.unesco.org/en/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

**terabyte STAC API**

EOWEB®  
GeoPortal



**elib**  
Publikationen des DLR



# Data focussed search engines



THE EUROPEAN SPACE AGENCY esa

earth online MISSIONS DATA NEWS EVENTS TOOLS SEE ALL

air pollution

Show results for: air pollution

**THEMATIC AREA**

- Agriculture
- Atmosphere
- Biosphere
- Climate
- Cryosphere
- Human Dimensions
- Land Surface
- Oceans
- Solid Earth
- Space Weather
- Sun-Earth Interaction
- Terrestrial Hydrosphere

**MISSIONS**

- Aeolus
- ALOS-1
- ...

**All Categories** | Data | News | Missions | Events | Tools | Activities | Instruments | Campaigns | Documents

Relevance

**Copernicus Sentinel-5P releases first data**

Following months of tests and careful evaluation, the first data on air pollutants from the Copernicus Sentinel-5P satellite

**Aeolus**

clouds: Filling the current major gap in the atmospheric observing system

**ED Summer School 1-Pollution-studies.pdf**

Pollution studies

**ERS Thematic Workshop on Oil Pollution Monitoring I...**

Proceedings of the Thematic Workshop on Oil Pollution in the Mediterranean, held on 25-26 March 1996 at ESA-ESRIN in Frascati

**PANGAEA.**

BIOLOGICAL CLASS...

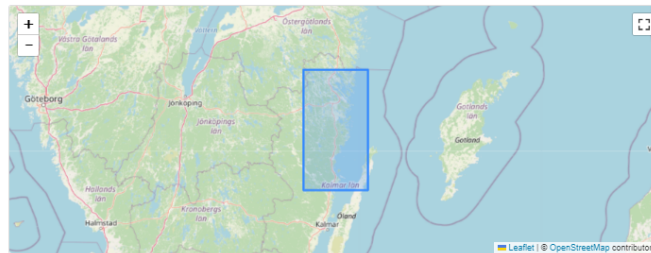
**8 datasets found on search for »frogs« in topic Biological Classification**

< 1 >

- Tornabene, BJ; Blaustein, AR; Briggs, CJ et al. (2017):** Site-based landscape and environmental factors and ranavirus epidemiology in an amphibian assemblage within the East Bay region, California  
*Related to:* **Tornabene, BJ; Blaustein, AR; Briggs, CJ et al. (2018):** The influence of landscape and environmental factors on ranavirus epidemiology in a California amphibian assemblage. *Freshwater Biology*  
 Size: 6682 data points  
<https://doi.org/10.1594/PANGAEA.879385> - Score: 6.39
- Herzschuh, U; Böhmer, T; Li, C et al. (2021):** Taxonomically harmonized pollen counts of western North America samples with revised chronologies  
 Size: 3519912 data points  
<https://doi.org/10.1594/PANGAEA.930613> - Score: 2.98
- Herzschuh, U; Böhmer, T; Li, C et al. (2021):** Taxonomically harmonized pollen percentages of western North America samples with revised chronologies  
 Size: 3519912 data points  
<https://doi.org/10.1594/PANGAEA.930818> - Score: 2.9

## Copernicus\_DSM\_10\_N57\_00\_E016\_00

in terrabyte STAC API



### Assets

- > QuickLook KML
- > Accuracy Layer
- > Editing Mask
- > Filling Mask
- > Height Error Mask
- > Source Mask
- > Water Body Mask
- > Data



[Sections](#) / [Data](#) / [Items](#) / [S1A\\_IW\\_GRDH\\_1SDV\\_20231002T030111\\_20231002T030136\\_050580\\_0617C7\\_A8AA](#) [\(html\)](#) [\(json\)](#)

V\_GRDH\_1SDV\_20231002T030111\_20231002T030136\_050580\_0617C7\_A8AA

### Collection

#### Copernicus DEM GLO-30

**GeoTIFF** The Copernicus DEM is a digital surface model (DSM), which represents the surface of the Earth including buildings, infrastructure, and vegetation. This DSM is based on radar satellite data acquired during the TanDEM-X Mission, which was...

4/22/2021, 12:00:00 AM UTC

### Provider

European Space Agency

### General

**GSD** 30 m

**License** proprietary

**Updated** 6/27/2023, 8:03:08 AM UTC

**Time of Data** 4/22/2021, 12:00:00 AM UTC

**Platform** tandem-x

### Projection

**EPSG Code** 4326

**Image Dimensions** 3,601 x 2,401

**Centroid** **Lat:** 57.5  
**Lon:** 16.5

**Transformati on Matrix** [0.0004166667; 0; 15.9997916667]  
[0; -0.0002777778; 58.0001388889]

### Gridded Data

**Grid** Copernicus Digital Elevation Model Grid  
**Eastings:** E016  
**Northings:** N57  
**CDEM-N57E016**

id:

SENTINEL-1

pe: RADAR

5.2577; -18.977; 36.5413; -18.6611

1.2023-10-02T00:00:00Z / none

S:

US

processor

processor

processor

processor

processor

processor

processor

processor

processor

processor

processor

processor

processor

processor

processor

processor

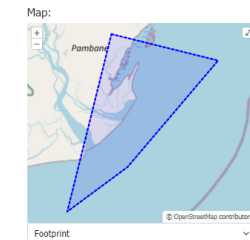
processor

processor

processor

processor

processor



Preview:



Assets:

- Reference water mask
- Valid pixel mask

# Working towards open science search and supporting open science



## Navigation in complex information spaces



**PANGAEA.**

Data Publisher for Earth & Environmental Science

**terabyte STAC API**

**EOWEB®  
GeoPortal**



**Formally published and (peer reviewed) knowledge**

**elib**  
Publikationen des DLR



# Publication focussed search engines



SEMANTIC SCHOLAR Open Web Search Search Q Sign In Create Free Account

DOI: 10.1002/asi.24818 • Corpus ID: 260713167

Share

## Impact and development of an Open Web Index for open web search

M. Granitzer, Stefan Voigt, +20 authors Saber Zehoudi • Published 7 August 2023 • Computer Science • Journal of the Association for Information Science and Technology

**TLDR** This work outlines six core principles for developing and maintaining an open index, based on open data principles, legal compliance, and collaborative technology development, that will facilitate the development of vertical search engines and innovative web data products, enabling a fair and open information space. [Expand](#)

View via Publisher Save to Library Create Alert Cite

### Related Papers

#### DP9: an OAI gateway service for web crawlers

Xiaoming Liu, K. Maly, M. Zubair, Michael L. Nelson • Computer Science • ACM/IEEE Joint Conference on Digital Libraries  
**TLDR**

32

Save Alert

#### Web Dragons: Inside the Myths of Search Engine Technology

I. Witten, M. Gori, T. Numerico • Computer Science  
**TLDR** A critical view of the idea of funneling information access through a small handful of gateways and the notion of a centralized index--and the problems that may cause is presented.

46

Save Alert

The screenshot shows the dblp search engine interface. At the top, there is a navigation bar with 'home', 'browse', 'search', 'about', and 'nfdi' links. The main search bar contains the query 'hubs and authorities type X'. Below the search bar, the results are displayed as a list of publications. The first result is 'Measuring scientific brain drain with hubs and authorities: A dual perspective' by Alessandra Urbanati, Edoardo Galimberti, and Giancarlo Ruffo, published in 2021. Other results include 'Hubs and Authorities in the Koch Brothers Network' (2021), 'Discovering Hidden Topical Hubs and Authorities Across Multiple Online Social Networks' (2021), 'Web networks of the science system: Weighted hubs and authorities' (2006), 'Link Analysis: Hubs and Authorities on the World Wide Web' (2004), 'Discovering authorities and hubs in different topological web graph structures' (2002), and 'Hubs, authorities, and communities' (1999). On the right side, there are several refinement options: 'refine by author', 'refine by venue', 'refine by type', 'refine by access', and 'refine by year'. A bar chart at the top right shows the distribution of results over time, with a peak in 2021.

# Integrated search of research outputs



OpenAIRE EXPLORE Search Deposit Link Data sources

Advanced search  
open web search

RESEARCH PRODUCTS (8,084) PROJECTS (1) DATA SOURCES (0) ORGANIZATIONS (237)

**Filters** Clear All

**Access (1)** Clear

- Open Access (8,084)

**Type (4)**

- Publications
- Research data
- Research software
- Other research products

**Year range**

1800 - 2033

This year | Last 5 years | Last 10 years

8,084 Research Products for open web search

OPEN ACCESS

Results per page: 10 | Sort by: Relevance

1 2 3 4 5

**Using the open Web as an Information resource and scholarly Web search engines as retrieval tools for academic and research purposes**

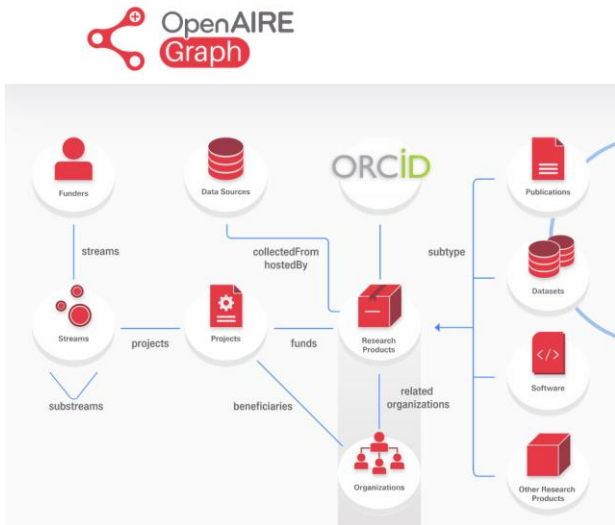
Publication Article - 2010 - ACSIS

Author: Filizta Naudic, Chris Rensleigh, Adeline du Toit  
DOI: 10.4102/sajim.v12i1.416

This study provided insight into the significance of the open Web as an information resource and Web search engines as research tools amongst academics. The academic staff establishment of the University of South Africa (Unisa) was invited to participate in a questionnaire survey and included 1188 staff members from...

South African Journa... | Link to | Share | Cite | Claim

Impact and development of an Open Web Index for open web search



Datasets Methods More

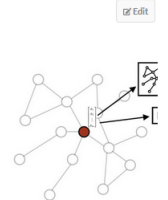
Graphs

## Link Prediction

723 papers with code · 75 benchmarks · 58 datasets

Link Prediction is a task in graph and network analysis where the goal is to predict missing or future connections between nodes in a network. Given a partially observed network, the goal of link prediction is to infer which links are most likely to be added or missing based on the observed connections and the structure of the network.

(Image credit: Inductive Representation Learning on Large Graphs)



## Benchmarks

These leaderboards are used to track progress in Link Prediction

Trend	Dataset	Best Model	Paper	Code	Compare
	WN18RR	MoCoSA			<a href="#">See all</a>
	FB15k-237	C-LMKE(BERT-tiny)			<a href="#">See all</a>
	WN18	Inverse Model			<a href="#">See all</a>
	FB15k	AutoKGE			<a href="#">See all</a>
	KCEW505-15	TLT-KGE(Quaternion)			<a href="#">See all</a>
	KCEW514	TLT-KGE(Quaternion)			<a href="#">See all</a>
	YAGO3-10	MEIM			<a href="#">See all</a>

## Content

- Introduction
- Benchmarks
- Datasets
- Subtasks
- Libraries
- Papers
  - Most implemented
  - Social
  - Latest
  - No code



# 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

Unstructured knowledge  
on observations, reports, ...



**elib**  
Publikationen des DLR

**terabyte STAC API**

EOWEB®  
GeoPortal

**COMMON  
CRAWL**

**OpenAlex**

# The web is full of scientifically relevant information



### A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way

Sumit Saha · Follow  
Published in Towards Data Science · 7 min read · Dec 15, 2018

14.2K · 65

Artificial Intelligence has been witnessing monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision.

Gray literature (blogs, wikis, ...)

### Why are alpine river water green - blue?

Asked 9 years ago · Modified 5 years, 7 months ago · Viewed 36k times

26

I recently visited Slovenia where I saw several rivers with a strong blue-green color. Here is an example from one of the tributaries of the Soča river near Kobarid. The locals refer to it as an "emerald" color.

Observation reports

### IPCC | Climate Change 2023: Synthesis Report

Feasible, effective options to reduce greenhouse gas emissions and adapt to human-caused climate change exist and need to be put into action now.

[Read the report](#)

Reports and brochures

### Tenerife: Evacuations as fire flares up again in high temperatures

4 hours ago

News

### Computation of Flow Oscillation In Pipes

**Background of Study**

Sudden closure or opening of valve gives rise to a very high inertia pressure within the pipe line. This is usually done by devices such as valves, pumps or any other mechanical equipment which can disturb the steady state flow conditions can trigger oscillations of flow from laminar region to turbulent region. Precautions must be taken to handle these transient events and it can lead to catastrophic events. For example, a hydroelectric power plant in Russia which is named as Sayano-Shuentskaya (2009) was completely destroyed due to sudden stoppage of one of its turbines. At the end of this event, 76 people lost their lives and approximately \$110 million worth of damage was inflicted.

**Statement of Problem**

Precious solutions to flow oscillation problems in pipes, assumed flow to be laminar always or turbulent always. Whereas no surge flow is ever turbulent always, it will also be laminar sometimes. This would require a numerical approach for solving second order differential equations, which is lengthy to achieve manually due to its iterative computations. Hence, there is need for a suitable algorithm that can treat flow that oscillate between laminar and turbulent regimes.

**Aim and Objectives**

The aim of this research is to develop an effective numerical method that can be used to efficiently compute surges in pipes due to valve connecting two reservoirs being rapidly opened, using a computer program developed on the Python environment.

**Scope of Study**

Unsteady-flow situations are more difficult to analyze than steady-flow situations. The Bernoulli equation is not applicable and the equation of motion leads to a differential equations for the velocity or pressure or head as a function of time. Numerical and graphical methods are frequently resorted to, with the use of analog and digital computers to speed up the process of finding solutions.

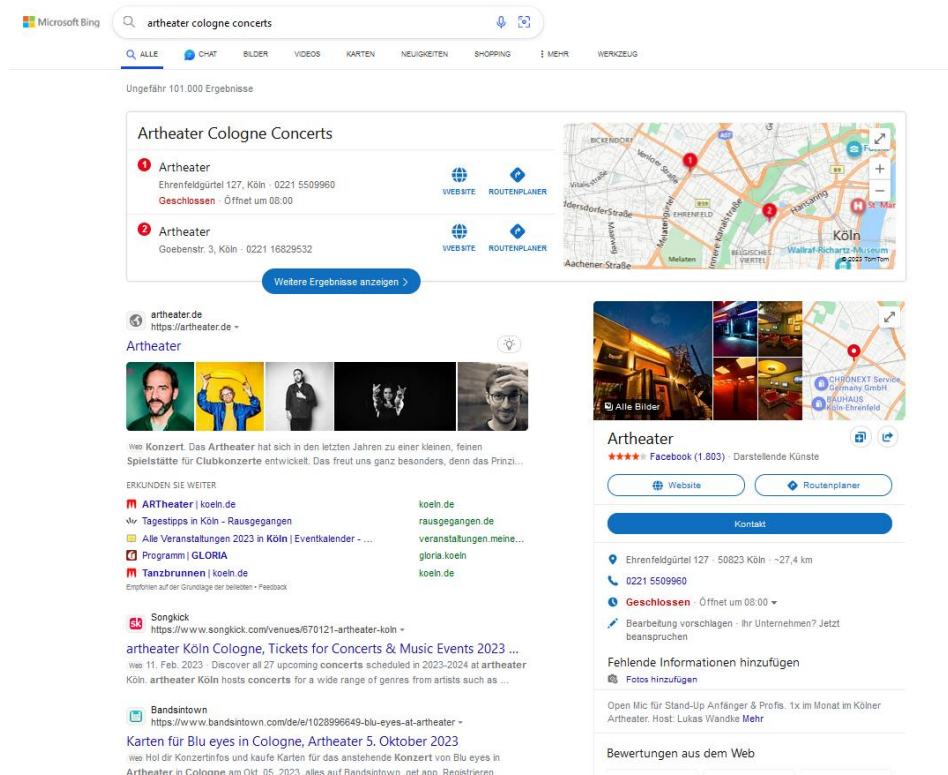
**Significance of Study**

There is no doubt that this program developed would assist students, researchers and practicing engineers in the industry in surge computations.

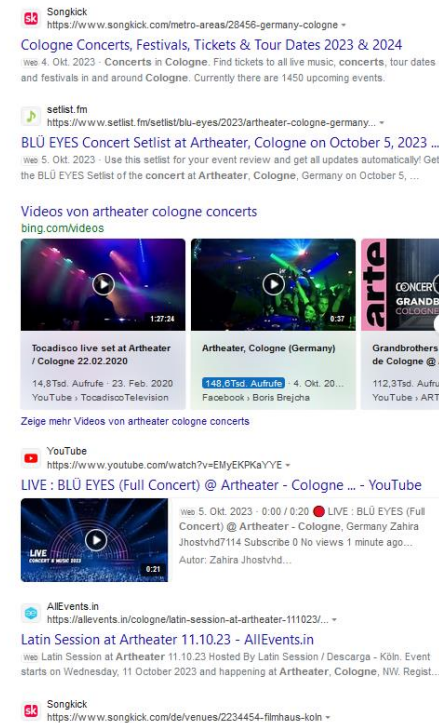
Software and data descriptions

# Search result enrichment for science search

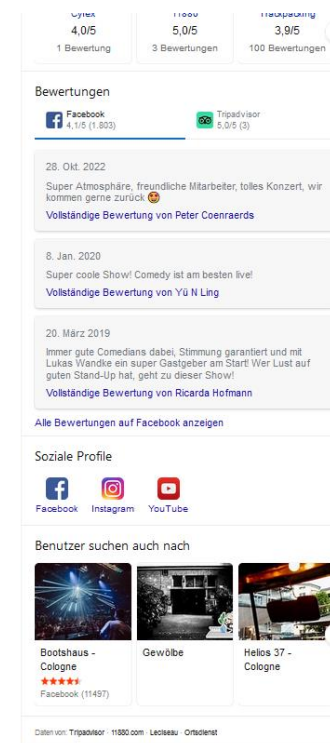
- Result enrichment is very common for general web search.
- Why not for scientific search?



The screenshot shows a Bing search for "arthater cologne concerts". The results are highly enriched with a map, multiple listings for Artheater, and various links to websites and social media. The map shows the location of Artheater in Cologne. The listings include details like "Ehrenfeldgürtel 127, Köln" and "Geschlossen". There are also links to "Website" and "Routenplaner".

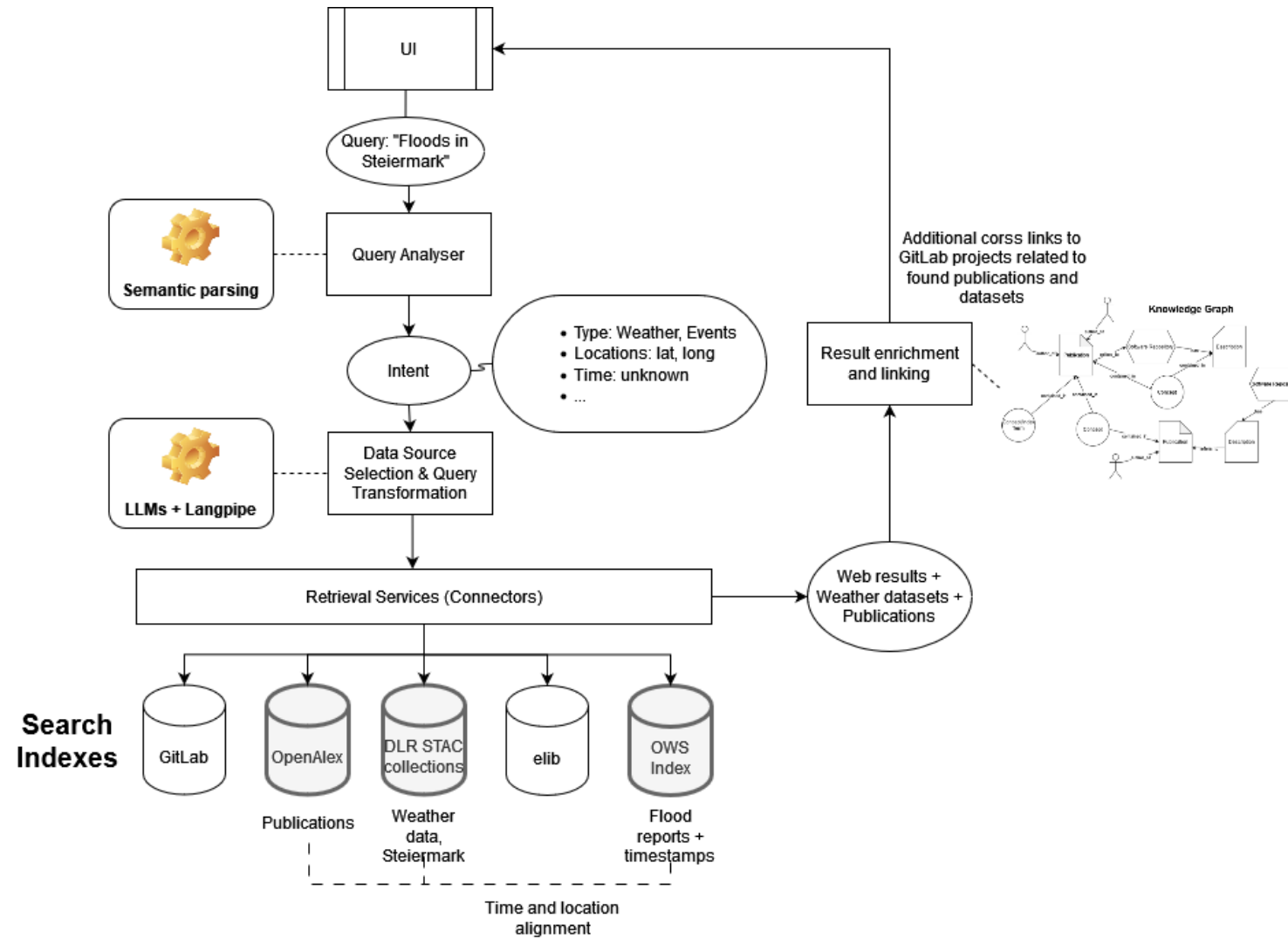


The screenshot shows Songkick search results for "Cologne Concerts, Festivals, Tickets & Tour Dates 2023 & 2024". It features a list of upcoming events, including "BLÜ EYES Concert Setlist at Artheater, Cologne on October 5, 2023". There are also video thumbnails and social media links for the events.

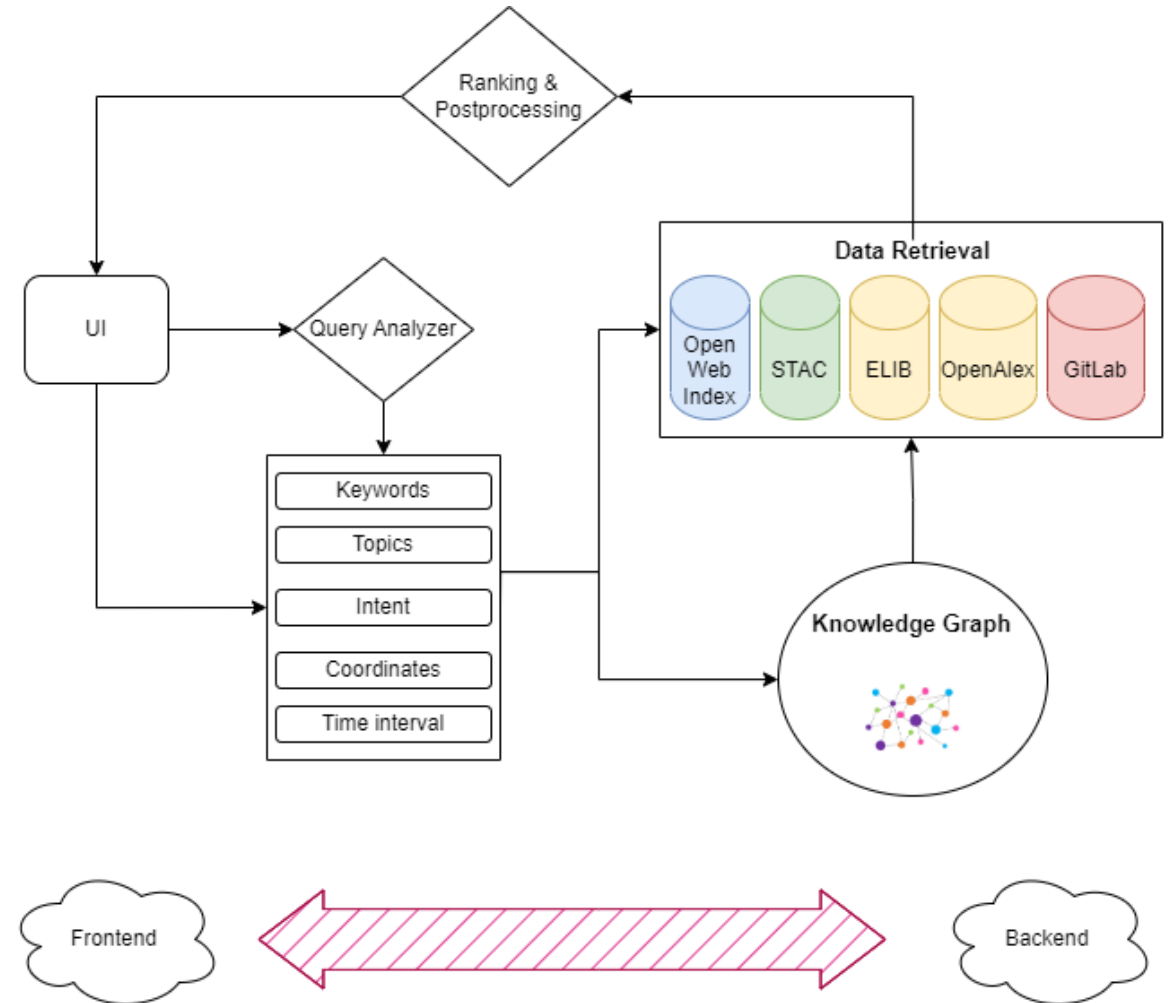
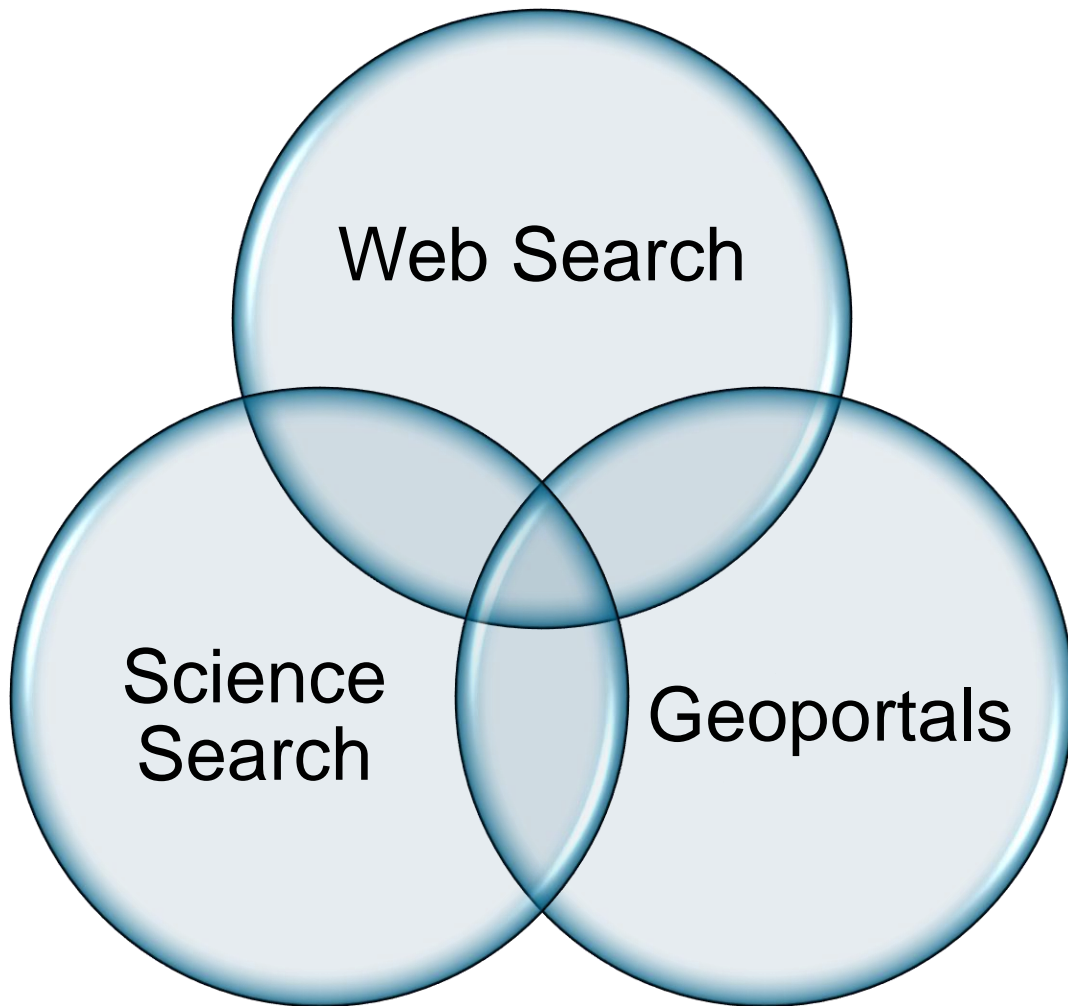


The screenshot shows a review page for Artheater concerts. It includes a section for "Bewertungen" (Reviews) with a Facebook rating of 4.1/5 and a TripAdvisor rating of 5.0/5. There are also sections for "Videos von artheater cologne concerts" and "Soziale Profile" (Social Profiles) for Facebook, Instagram, and YouTube. The page also lists "Benutzer suchen auch nach" (Users also search for) with thumbnails for "Bootshaus - Cologne", "Gewölbe", and "Helios 37 - Cologne".

# Pilot application: Search result enrichment for environmental science search



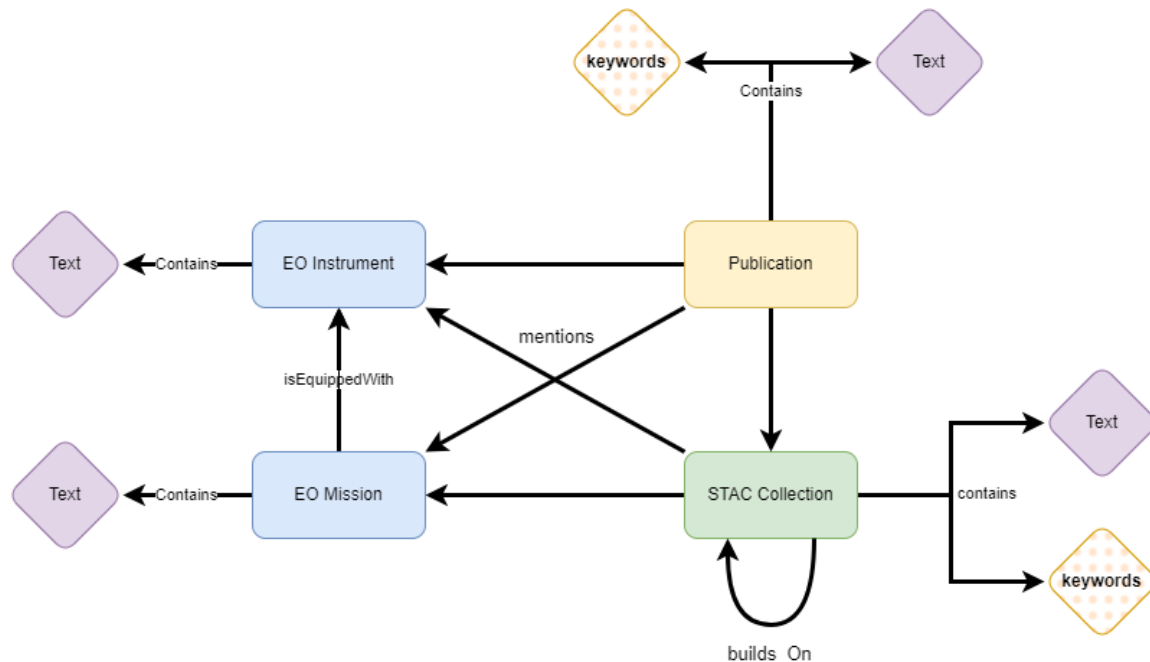
# Pilot application: Search result enrichment for environmental science search



# Pilot application: Search result enrichment for environmental science search



## Structured resource representation using Corpus Analytics Builder\*



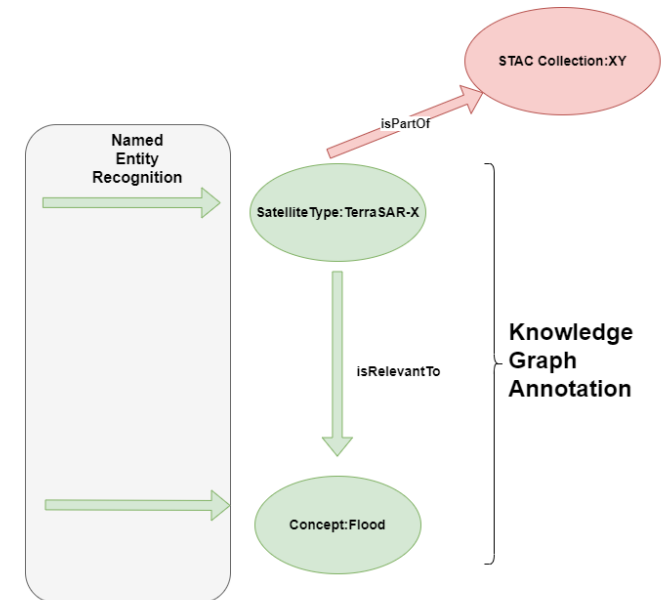
## Publications as mediating artefacts

**QUANTITATIVE FLOOD ASSESSMENT: CASE STUDY**  
 Corneliu Octavian Dumitru, Shiyong Cui,  
 Remote Sensing Technology Institute (IMF), Earth Observer  
 Center (DLR), Münchner Straße 20, 82234 W

**ABSTRACT**  
 During the paper [2], in Cooper Major Disaster the local gov quickly in the this paper for the river Elbe. Very few pub this disaster th should mention quite often it flooding's wer were also not. For example, images in end appear in list y In this pape demarcation o that we made i The applicat occurred in Ju The paper stru data set used i methodology, assessment, w the paper.

**Index Terms**— disaster, flooding, quantitative assessment, taxonomy, TerraSAR-X, active learning, support vector machine.

**1. INTRODUCTION**



\* 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 EAACL'23: System Demonstrations* (pp. 248-255).

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

OpenSearch@DLR Prototype

Q floods in europe

Submit

Show advanced options >>



STAC Collections Publications Web Documents Social Media

Hot news: *Flood! Flood! Flood!*

Tuesday, August 26, 2008 *Flood! Flood! Flood!* Rain is lashing the eastern part of India it seems. Most of the villages are flooded with water and people can't get any proper food. Prime minister has asked immediate relief to those people.

Web Document

## The SpaceNet 8 Challenge - From Foundation Mapping to Flood Detection

Floods are one of the major types of natural disasters responsible for loss of life, destruction of buildings and infrastructure, erosion of arable land, and environmental hazards around the world. Climate change, increasing populations, and urbanisation of flood plains will only increase the risk of flooding in the next few years. SpaceNet 8 presents a dataset that combines building footprint detection, road network extraction, and flood detection covering 850km2, including ~32,000 buildings and ~ 1,300 km of roads, of which ~ 13 and ~ 15 are flooded, respectively.

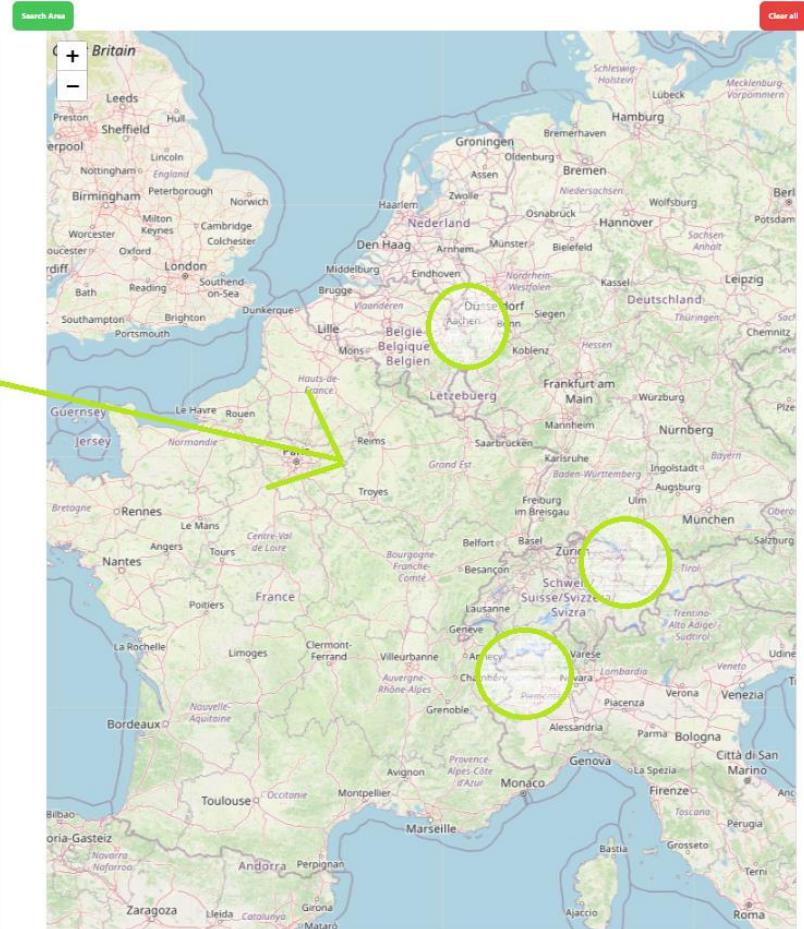
Publication Earth Observation Flood Detection Deep Learning Benchmark

Floods - Preparing for Floods - Flood Preparedness

*Flood*, the Routine Disaster Nature may love a *flood*, but *floods* are civilization's worst enemy--one we make ourselves. Preparing for a *Flood* if you don't live in a *flood* zone, it is still possible you could be affected by *flood* damage. *Floods* are more common that most may think.

Web Document

Flood Defence Flood Barrier Flood Defences Flood Dam by UK Flood Barriers



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

floods in europe

Submit

Show advanced options >>

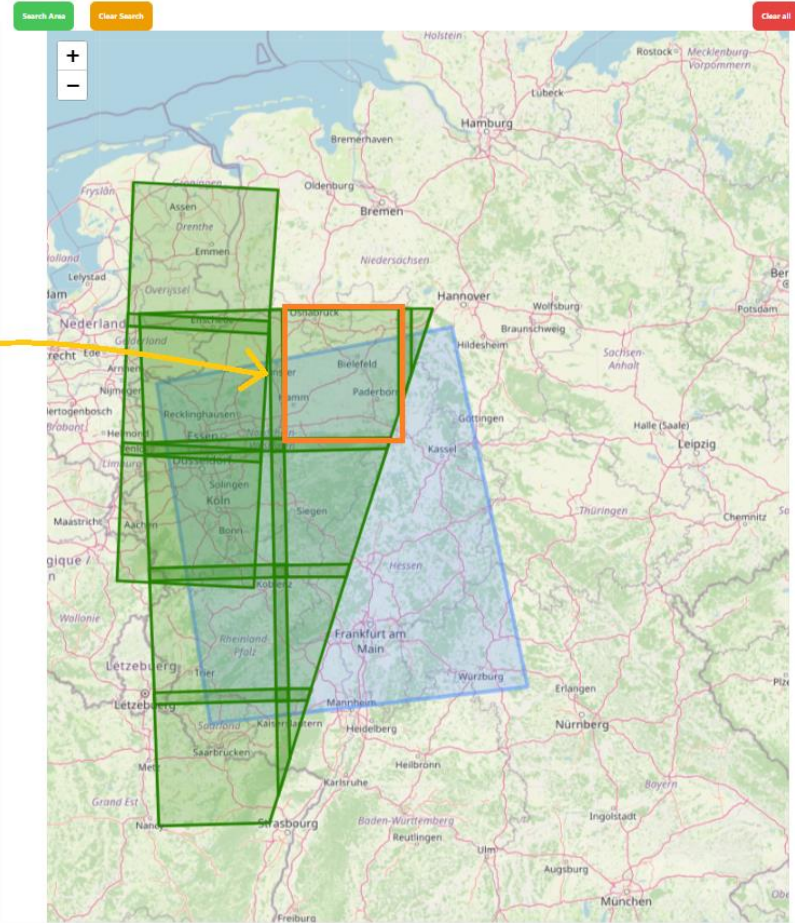
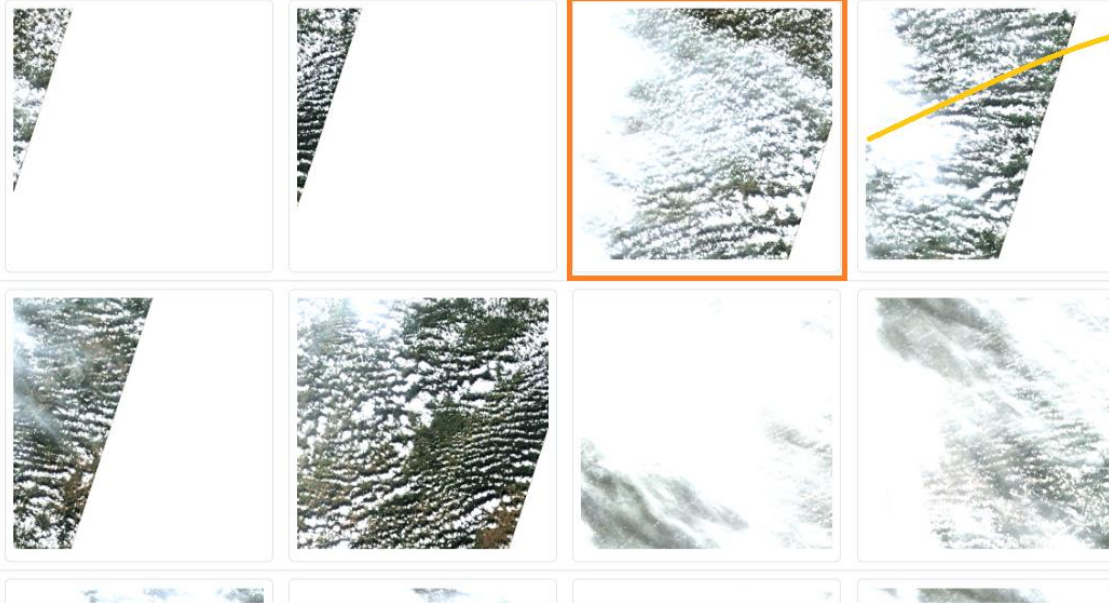
Sentinel-2 Level-2A

STAC Collections Publications Web Documents Social Media

STAC Collection Sentinel Copernicus ESA Satellite Global Imagery Reflectance

The [Sentinel-2](https://sentinel.esa.int/web/sentinel/missions/sentinel-2) program provides global imagery in thirteen spectral bands at 10m-60m resolution and a revisit time of approximately five days. This dataset represents the global Sentinel-2 archive, from 2016 to the present, processed to L2A (bottom-of-atmosphere) using [Sen2Cor](https://step.esa.int/main/snap-supported-plugins/sen2cor/) and converted to [cloud-optimized GeoTIFF](https://www.coge.org/) format.

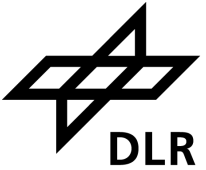
Request STAC Items







# 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.