Philipps Universität Marburg

Geo Engine

Demonstrations in the biodiversity domain

Nikolaus Glombiewski,
Christian Beilschmidt &
Michael Mattig

FAIR-DS Final Event

Quick introduction

Nikolaus Glombiewski

University of Marburg

Bernhard Seeger

University of Marburg

Christian Beilschmidt

Geo Engine GmbH

Michael Mattig

Geo Engine GmbH

Johannes Drönner

Geo Engine GmbH

Demonstrator

NFDI4Biodiversity
RDC

Extension 1

Copernicus Data Space
Machine Learning

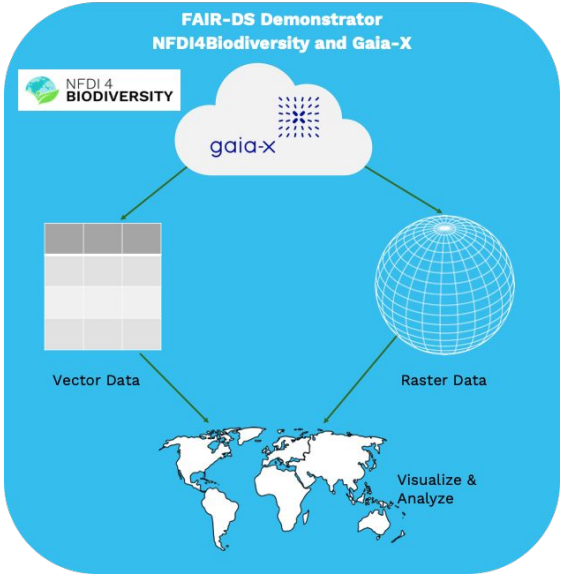
Extension 2

Industry Data & ESG
Virtual Data Trustee

NFDI4Biodiversity Demonstrator

Overview

- **Application**
  - Spatio-Temporal Data Analysis
  - Heterogeneous Data Sources
- **Geo Engine**
- **Connection to Research Data Infrastructure**
  - NFDI4Biodiversity Research Data Commons
- **Connection to Gaia-X**



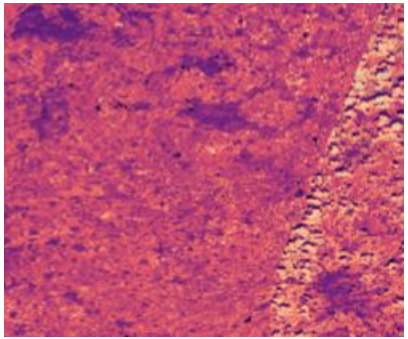
Demonstrator Milestones

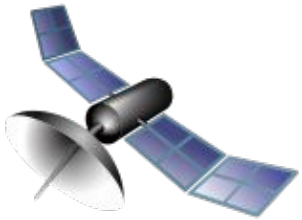
- Cloud Deployment
- User Stories for Connecting Data Spaces
- Authentication and Authorization
- Automatic Data Deletion
- **Data Provider via Spatio-Temporal Data Exchange Protocol**
- **Service Offering in a Federated Catalogue**

Spatio-Temporal Data Analysis

Species	Coordinates	Date
Bird	48.856614, 2.352221	10.05.1977
Cat	41.8933203, 12.4829321	18.04.1980
Elephant	52.517037, 13.38886	21.10.2015

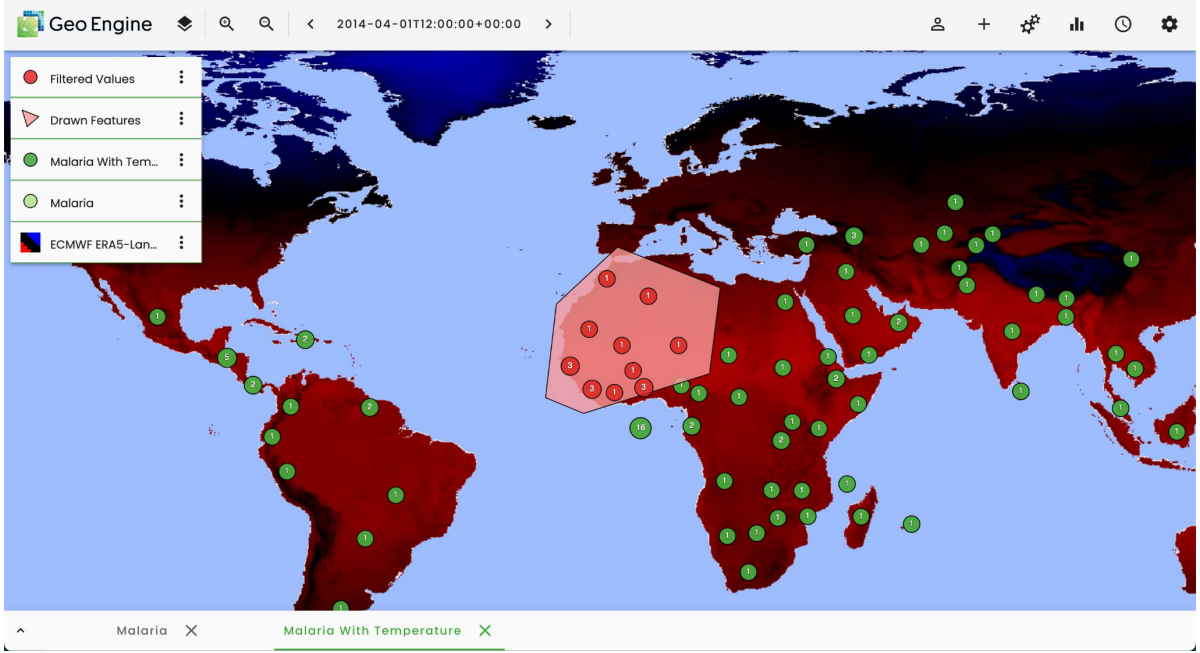
Example 1:
Vector Data
(e.g. manual recordings)



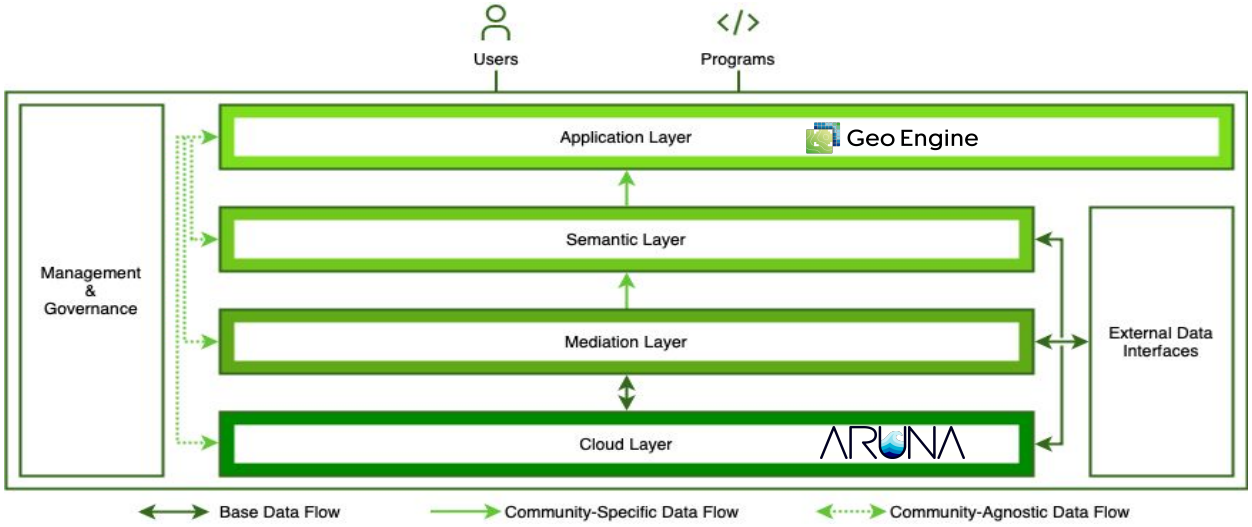


Example 2:
Raster Data
(e.g., images from satellites)

Geo Engine - Query



NFDI4Biodiversity - Research Data Commons

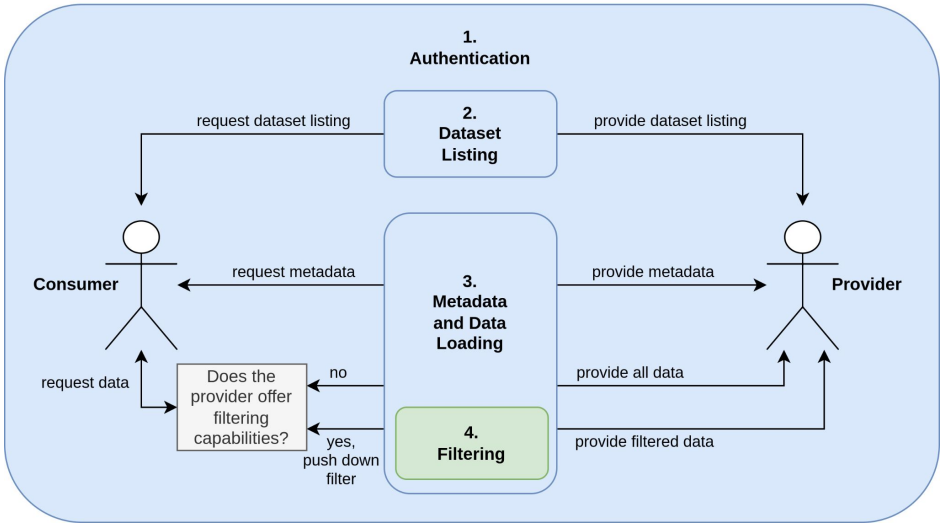


Spatio-Temporal Data Exchange Layer

- Problems in data providers in research data
  - Lack of spatio-temporal support
  - Not FAIR enough

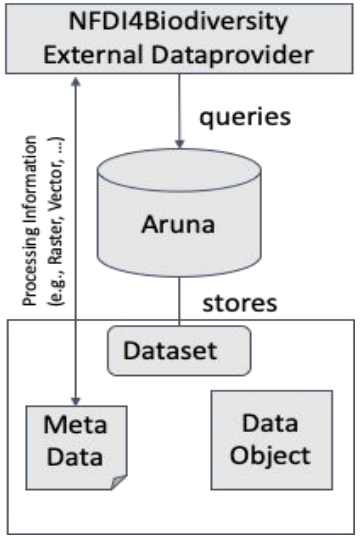
- Our approach:
  Intermediary spatio-temporal data exchange layer within Geo Engine based on the FAIR principles

SDExL: Required and optional steps



Connection to Aruna Object Storage

1. Authentication
   - Solved through NFDI4Biodiversity SSO
2. Listing
   - Geo Engine dataset conform to Aruna dataset
   - Specific tags
   - Verify Aruna dataset format (2 objects)
3. Dataset Loading
   - Meta object: Processing information
   - Data object: Actual data
4. Currently no filtering option available



Connection to Time Series Database

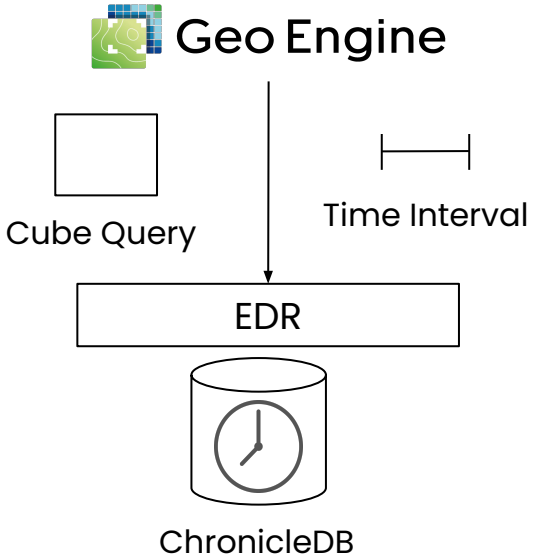
- Accessing time series queries (e.g., pattern matching)
- Challenges:
  - Filter Pushdown (Spatial and Temporal)
  - Queries are persistent
  - Results are changing
    => Avoid result materialization

- Choosing the correct protocol is critical

- OGC Environmental Data Retrieval (EDR) Protocol solves most of these challenges:
  - Support for common spatio-temporal queries (position, area, cube, corridor, …)
  - Persistent identifier for queries, optional for items

Environmental Data Retrieval (EDR) in Geo Engine

1. Authentication: Solved
2. Data Listing: Solved
3. Data Loading: Some problems with file formats require custom solutions
4. Filtering:
    - Cube query for spatial filtering
    - (Optional) time interval for temporal filtering

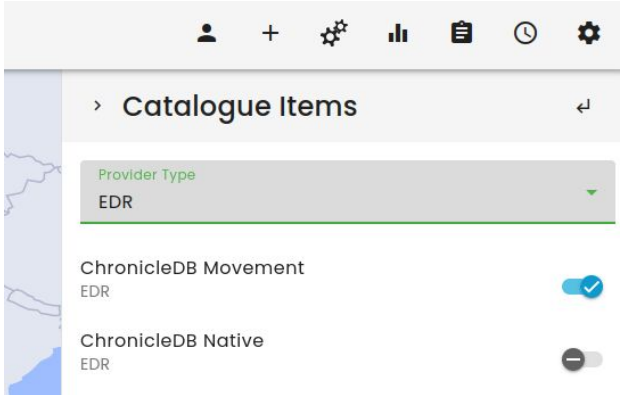


Geo Engine and Gaia-X Federated Catalogue

- **Technological Basis:**
  Gaia-X compliant catalogue developed by Eclipse XFSC (Cross Federation Services Components)

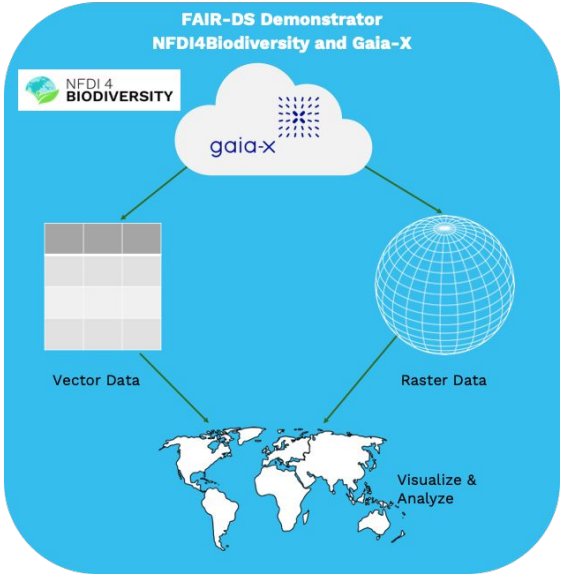
- **Geo Engine Self-Description:**
  - A Service Offering in JSON-LD format
  - Verifiable Credentials: Set of claims or attributes, digitally signed by a trusted entity
    Who? What? Where? Which standard?

- **Adding Data from supported dcat:DataService**
  - Aruna Object Storage
  - OGC Environmental Data Retrieval (EDR)



Summary

- **Application**
  - Spatio-Temporal Data Analysis
  - Heterogeneous Data Sources
- **Connection to Research Data Infrastructure**
  - Geo Engine as a Spatio-Temporal Exchange Layer
  - Aruna
  - OGC EDR
- **Connection to Gaia-X**
  - Connection to a Federated Catalogue



Combining Data Spaces for ML: NFDI4Biodiversity and Copernicus Data Spaces

Motivation

Space is important driver for analytics

- Huge pool of data, e.g., Sentinel missions
- Data integration issues
- Combining diverse data sources (NFDI, Copernicus, user data)

Leveraging ML for advanced data analysis

- Define complex workflows
- Process for training and application

FAIR and easy to use

- APIs, standards, authentication
- Dashboard

Goal

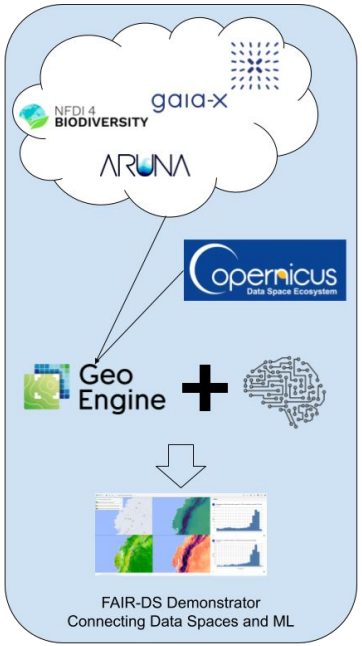
1. Connection of NFDI4Biodiversity and Copernicus Data Spaces
2. Machine Learning pipeline and use cases (land classification)

Development Steps

- Extension of the Aruna Object Store connector
- Copernicus Data Space Ecosystem connector
- Machine Learning Workflows

Value Creation

- Combination of research and space data
- FAIR principles via Connectors, Workflows and APIs
- New insights using Machine Learning



Machine Learning Pipeline

Access S2 Data → Access & Rasterize LUCAS Data → Feature Engineering → Train Modell

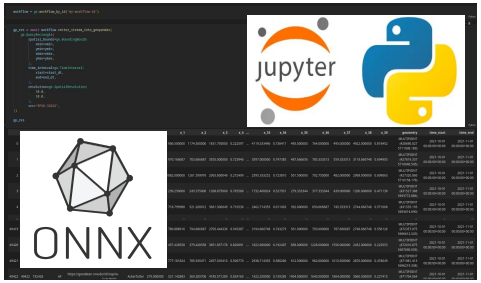
NFDI 4 BIODIVERSITY

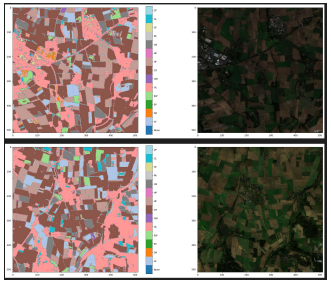
eurostat

Copernicus Data Space Ecosystem

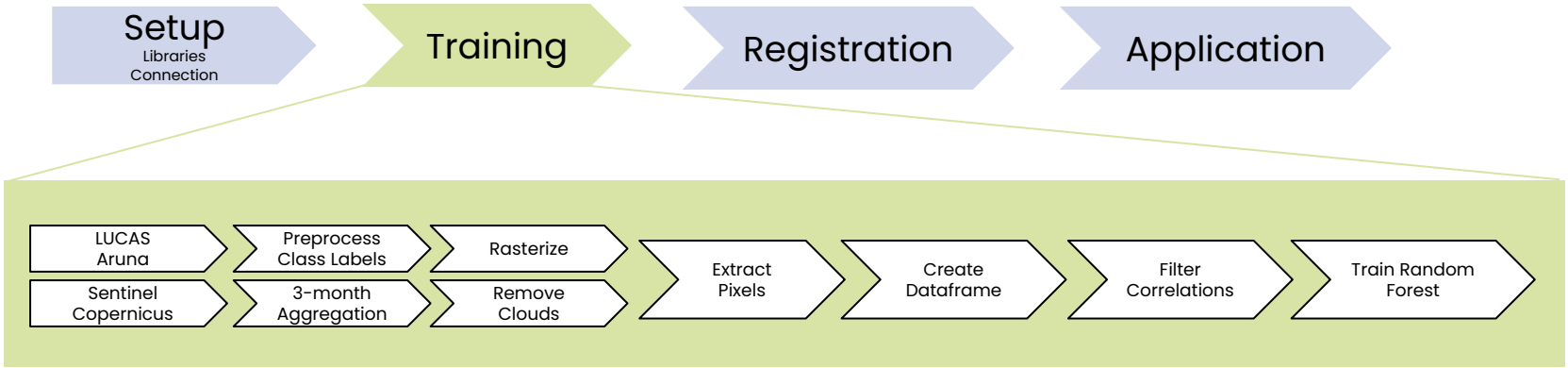
ARUNA

Geo Engine

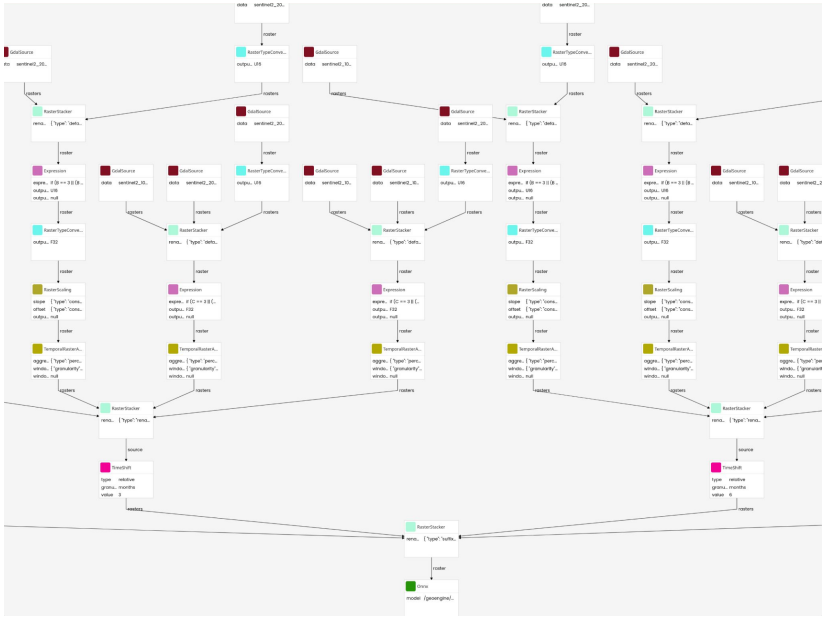


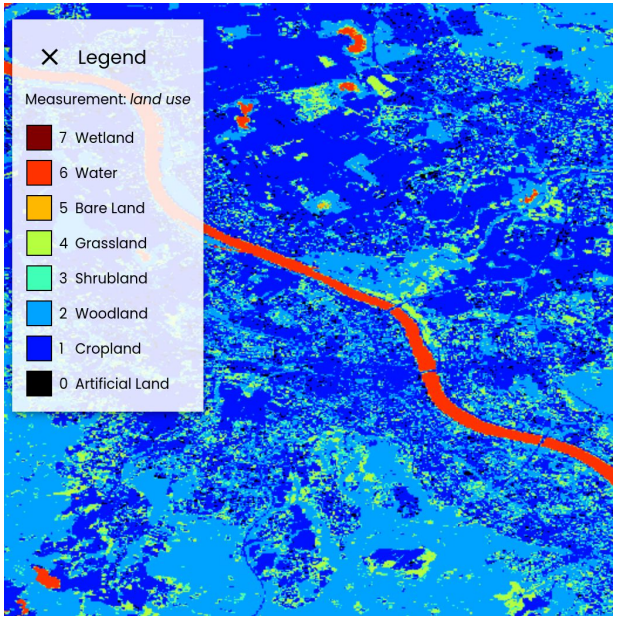


ML Training Process

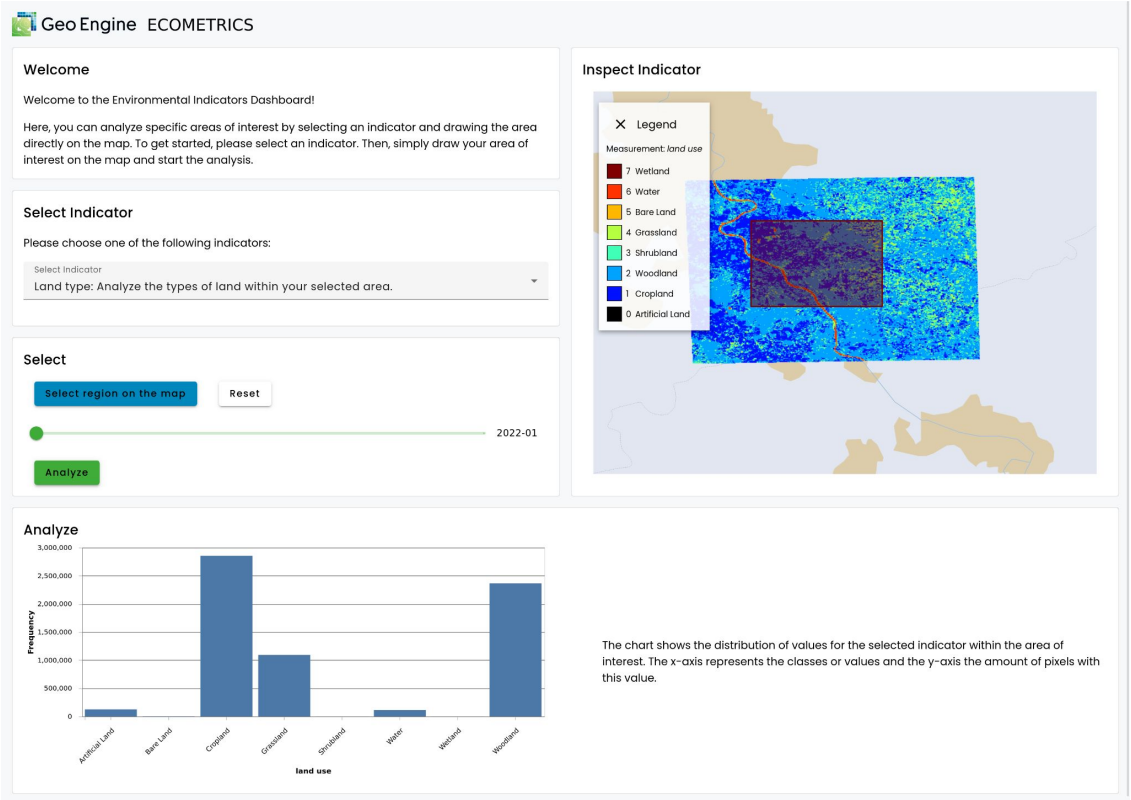


ML Application





Dashboard


Geo Engine ECOMETRICS
Welcome
Welcome to the Environmental Indicators Dashboard!
Here, you can analyze specific areas of interest by selecting an indicator and drawing the area directly on the map. To get started, please select an indicator. Then, simply draw your area of interest on the map and start the analysis.
Select Indicator
Please choose one of the following indicators:
Select Indicator
Land type: Analyze the types of land within your selected area.
Select
Select region on the map
Reset
2022-01
Analyze
Inspect Indicator
Legend
Measurement: land use
7 Wetland
6 Water
5 Bare Land
4 Grassland
3 Shrubland
2 Woodland
1 Cropland
0 Artificial Land
Analyze
3,000,000
2,500,000
2,000,000
1,500,000
1,000,000
500,000
0
Frequency
Artificial Land
Bare Land
Cropland
Grassland
Shrubland
Water
Wetland
Woodland
land use
The chart shows the distribution of values for the selected indicator within the area of interest. The x-axis represents the classes or values and the y-axis the amount of pixels with this value.

ESG Indicators using Virtual Data Trustees across Data Spaces

Motivation

ESG Reporting Gains Importance

- Regulatory Landscape (CSRD, CSDDD, EUDR, SFDR)
- Mandatory ESG reporting for 60k companies in the EU

Biodiversity as a Key ESG Factor (ESRS E4)

- Global biodiversity crisis
- Growing investor and consumer demand for sustainable practices
- Need for quantitative data on biodiversity impact

Challenges for Companies

- Data Availability and Quality
- Complexity of Biodiversity Assessments
- Limited budgets and expertise for biodiversity monitoring and analysis
- Data Protection

Bahn Use-Case

- Large land holding
- Biodiversity score for properties
- Compensation areas

Goal: A Virtual Data Trustee

Leveraging FAIR-DS

- Facilitating data sharing and interoperability
- Ensuring data privacy and security

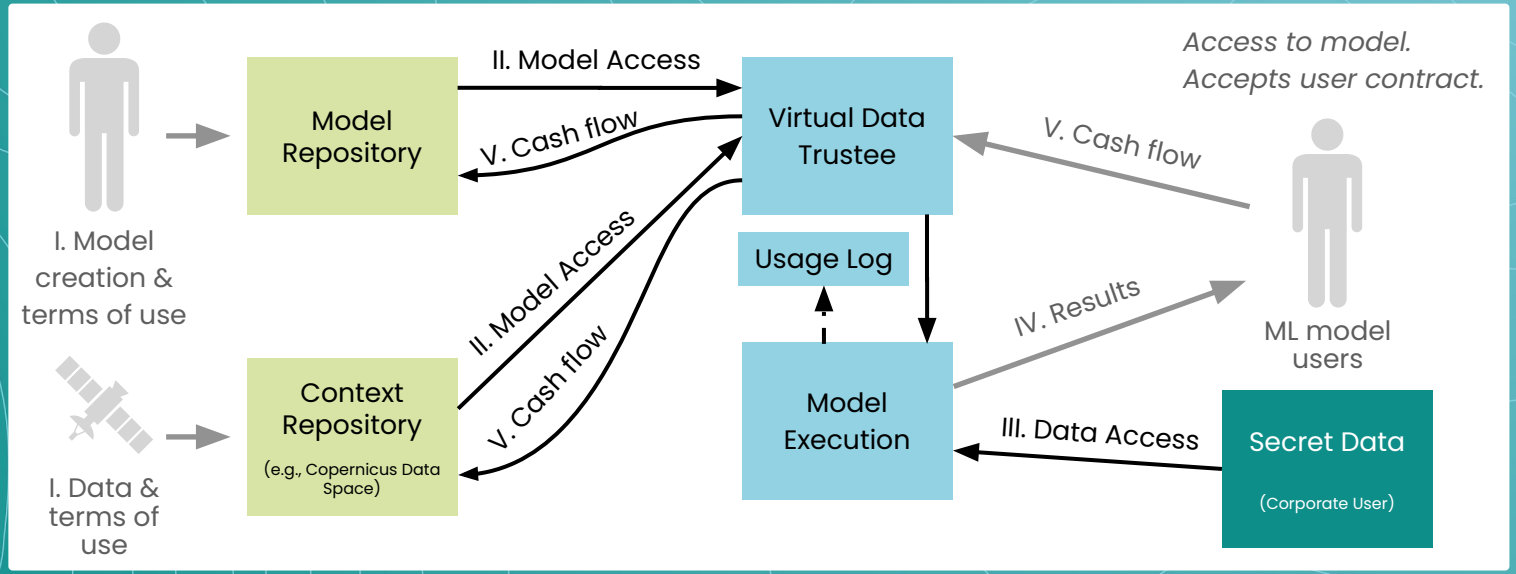
Key Features

- Access to diverse biodiversity indicator models
- Data processing and analysis for reports
- Usage logging for participants (“billing”)

Benefits for Companies

- Use models from research to evaluate own property areas

Design



1. The virtual data trustee mediates between the demand side (right) and the supply side (left).
2. The trustee is responsible for the usability of the demand side.
3. Utilisation of models and data is logged in the usage log for billing purposes, among other things

Why this architecture?

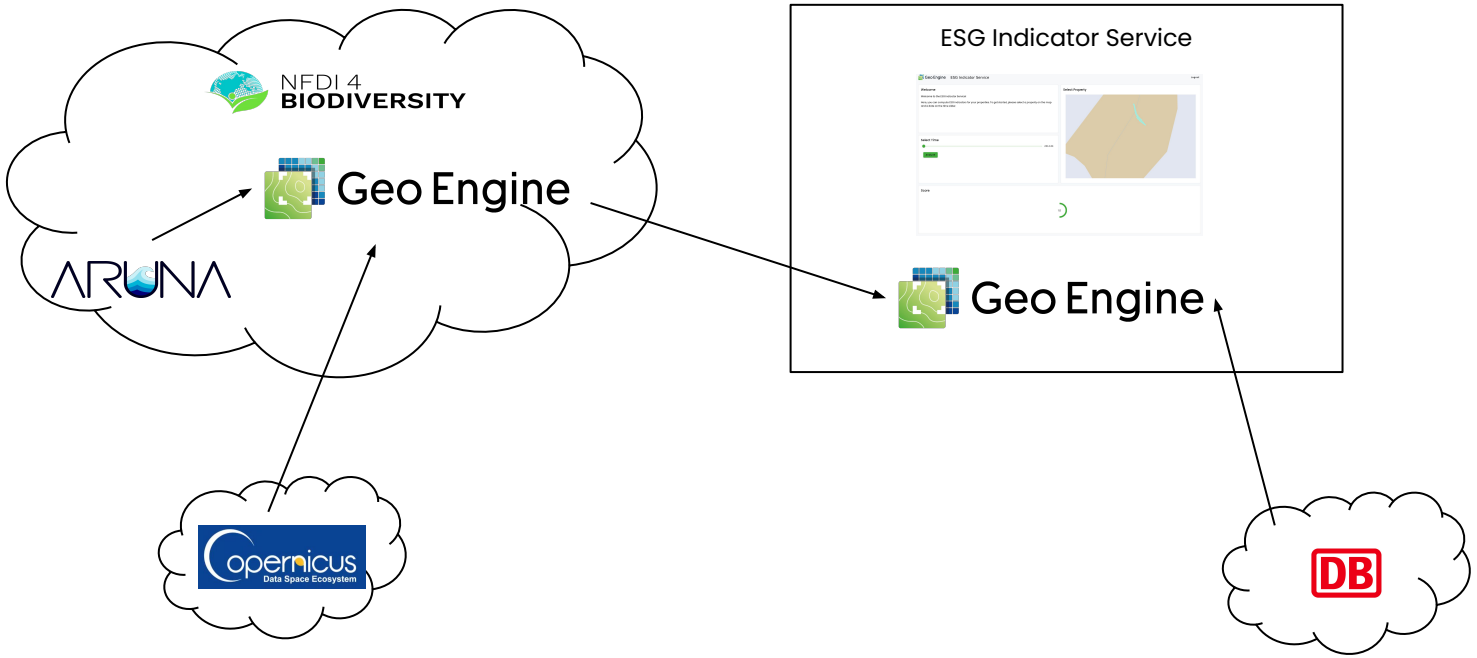
Strategy, incentives, etc.

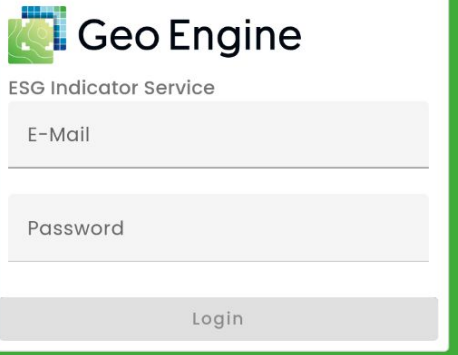
1. The data trustee takes on the role of an orchestrator [1].
2. For the supply side (models and accompanying information repositories), it creates value by mediating demand and simplifying billing.
3. On the demand side, it reduces the costs of model deployments through automation and integration of various models and data sources.
4. The trustee is additionally incentivised
   a. To provide models of the highest possible quality [2] in order to maximize revenue.
   b. Provide as many (existing) data sources and suitable models as possible in order to maximise turnover

[1] https://www.sitra.fi/en/articles/eight-lessons-from-building-data-spaces/

[2] Akerlof, George A: The Market for Lemons

Implementation (as is)




Geo Engine
ESG Indicator Service
E-Mail
Password
Login

Geo Engine ESG Indicator Service

Logout

Welcome

Welcome to the ESG Indicator Service!

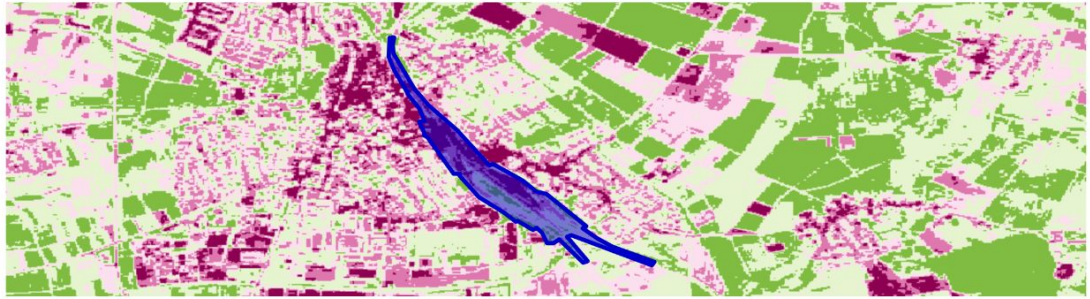
Here, you can compute ESG indicators for your properties. To get started, please select a property on the map and a date on the time slider.

Select Time

2022-05

Analyze

Select Property



Score

Please select a property and a date.

Usage log

Timestamp	Computation	Workflow	Count	Details
2024-11-29T10:42:31.973Z	65b12dd7-da10-4f24-a0c2-5f31d6562c0e	0ab0cb2c-2888-5536-a465-6bd5e60b1271	7	
2024-11-29T10:42:26.241Z	d8fd3168-d84d-442c-88fa-760c39dd268e	98e28de8-088a-5469-82c9-23fdbb4a5312	9	
2024-11-29T10:42:23.393Z	40fb4b23-7239-4995-9513-3cf167869ffc	890d41ec-7e4c-5000-8ab8-7394a758a86f	4	
2024-11-29T10:42:23.388Z	4b637e0d-930b-420e-b361-bff3426385ba	890d41ec-7e4c-5000-8ab8-7394a758a86f	2	
2024-11-29T10:42:23.354Z	6b827c64-33da-422e-8d43-27b9ec201b03	890d41ec-7e4c-5000-8ab8-7394a758a86f	1	

Refresh

Items per page: 5

0 of 0

Geo Engine ESG Indicator Service

Logout

Welcome

Welcome to the ESG Indicator Service!

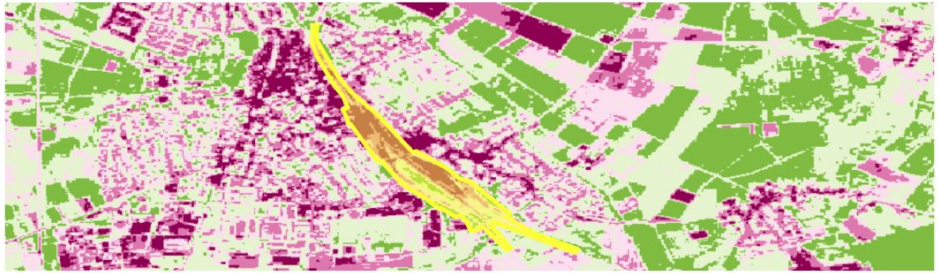
Here, you can compute ESG indicators for your properties. To get started, please select a property on the map and a date on the time slider.

Select Time

2022-05

Analyze

Select Property



Score

0.5

Usage log

Timestamp	Computation	Workflow	Count	Details
2024-11-29T16:14:56.094Z	8193f950-27b4-4c45-b774-cbd800eed576	98e28de8-088a-5469-82c9-23fdbb4a5312	9	
2024-11-29T16:14:34.671Z	6933ba66-3c5c-495c-af4a-3bd408cf7a02	890d41ec-7e4c-5000-8ab8-7394a758a86f	2	
2024-11-29T16:14:34.654Z	3d5c4db9-6b80-4a89-b103-1e1fe04eadbd	890d41ec-7e4c-5000-8ab8-7394a758a86f	2	
2024-11-29T16:14:34.584Z	5617cb3e-d1fa-4ab9-b223-fb028163a4b9	890d41ec-7e4c-5000-8ab8-7394a758a86f	1	
2024-11-29T16:14:34.581Z	45188f2d-d1cb-436e-bfed-5eddbd3cc778	ca978972-e040-596f-8860-5dfec494028a	3	

Refresh

Items per page: 5

1 – 5 of 6

Access Usage Log as Data Producer

Dependencies

[1]:
```
#%pip install git+https://github.com/geo-engine/geoengine-python.git@esg_usage
```

[2]:
```
import geoengine as ge
import matplotlib.pyplot as plt
```

Connect to the Virtual Data Trustee

[3]:
```
ge.initialize("https://trustee.app.geoengine.io/api") # loads credentials from .env file
```

Access Usage Log

[4]:
```
usage = ge.data_usage(0, 10)
usage
```

[4]:

	computationId	count	data	timestamp	userId
0	8193f950-27b4-4c45-b774-cbd800eed576	6	GdalSource	2024-11-29 16:14:56.114000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb
1	8193f950-27b4-4c45-b774-cbd800eed576	1	OgrSource	2024-11-29 16:14:56.094000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb
2	6933ba66-3c5c-495c-af4a-3bd408cf7a02	2	GdalSource	2024-11-29 16:14:34.671000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb
3	3d5c4db9-6b80-4a89-b103-1e1fe04eadbd	2	GdalSource	2024-11-29 16:14:34.654000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb
4	5617cb3e-d1fa-4ab9-b223-fb028163a4b9	1	GdalSource	2024-11-29 16:14:34.584000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb
5	45188f2d-d1cb-436e-bfed-5eddbd3cc778	1	OgrSource	2024-11-29 16:14:34.581000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb
6	3900881e-b894-4602-9dd8-bfaddf66cc13	1	GdalSource	2024-11-29 16:14:34.578000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb
7	7b8de3d8-65b2-4a44-9468-ca6f3c5dbe6a	2	GdalSource	2024-11-29 16:14:33.754000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb
8	f9365f25-3da3-46b6-bd08-5318c5ec450b	2	GdalSource	2024-11-29 16:14:33.738000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb
9	053481b8-d2d1-48a2-8fd6-7091c1fa9f3c	4	GdalSource	2024-11-29 16:14:33.690000+00:00	3aed16e5-5e6b-4a21-93ad-b147f01cf9bb

Access Usage Summary

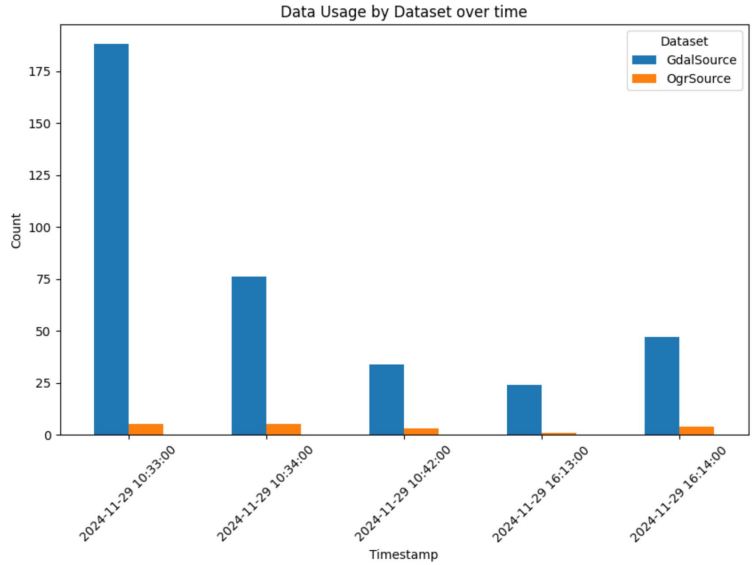
[5]:
```
usage_summary = ge.data_usage_summary(ge.UsageSummaryGranularity.MINUTES)
usage_summary
```

[5]:

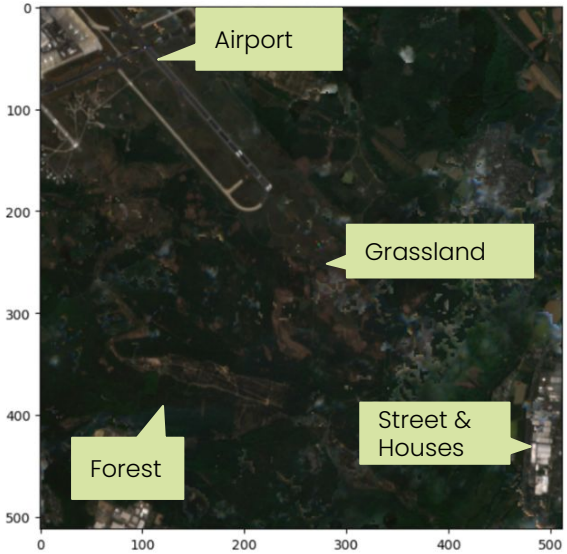
	count	dataset	timestamp
0	47	GdalSource	2024-11-29 16:14:00+00:00
1	4	OgrSource	2024-11-29 16:14:00+00:00
2	24	GdalSource	2024-11-29 16:13:00+00:00
3	1	OgrSource	2024-11-29 16:13:00+00:00
4	34	GdalSource	2024-11-29 10:42:00+00:00
5	3	OgrSource	2024-11-29 10:42:00+00:00
6	76	GdalSource	2024-11-29 10:34:00+00:00
7	5	OgrSource	2024-11-29 10:34:00+00:00
8	188	GdalSource	2024-11-29 10:33:00+00:00
9	5	OgrSource	2024-11-29 10:33:00+00:00

Plot usage summary

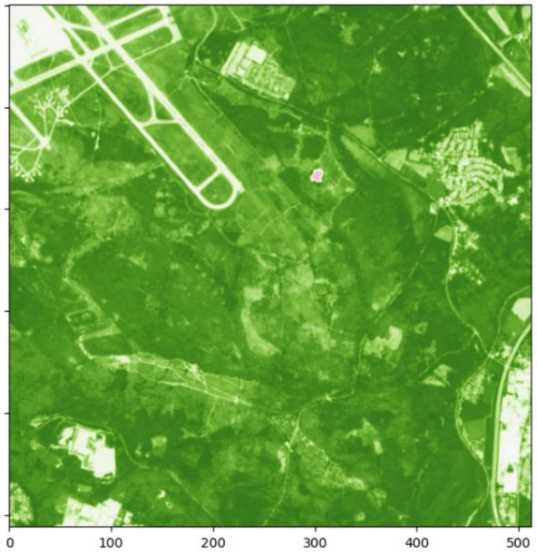
[6]:
```
ge.plot_data_usage_summary(ge.UsageSummaryGranularity.MINUTES)
```



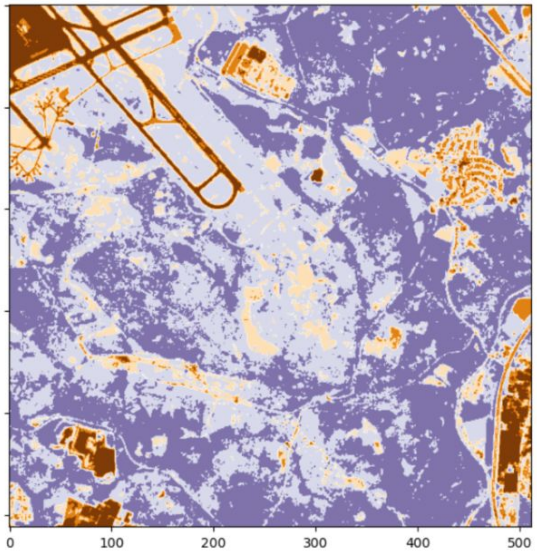
Image



Vegetation-Index



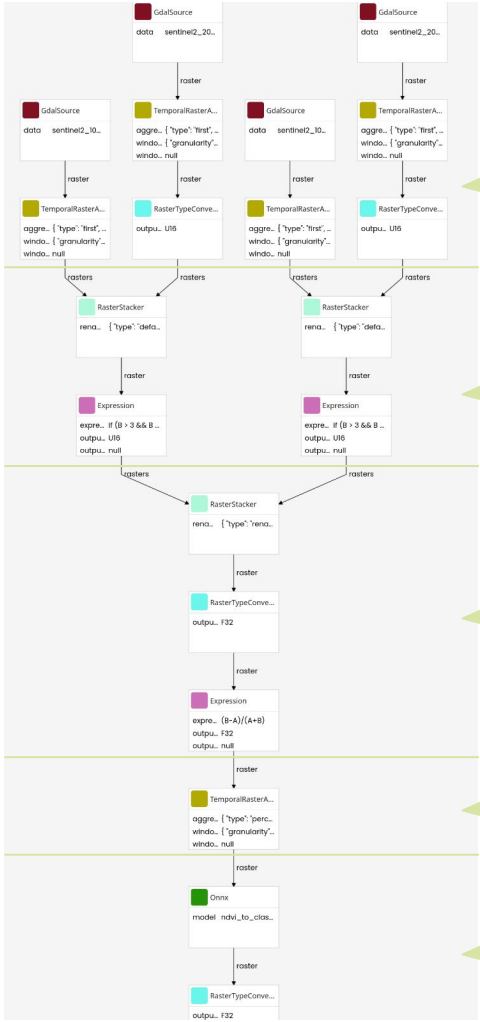
ESG-Indicator



- ESG-Indicator based on vegetation development over the year (first iteration)

- Build-up areas and streets - support no vegetation → white in vegetation-index, red in ESG-indicator
- Grasslands (agriculture) - change over time → light green in vegetation-index, light blue in ESG-indicator
- Forests - no change in vegetation → green in vegetation-index, purple in ESG-indicator

Workflow



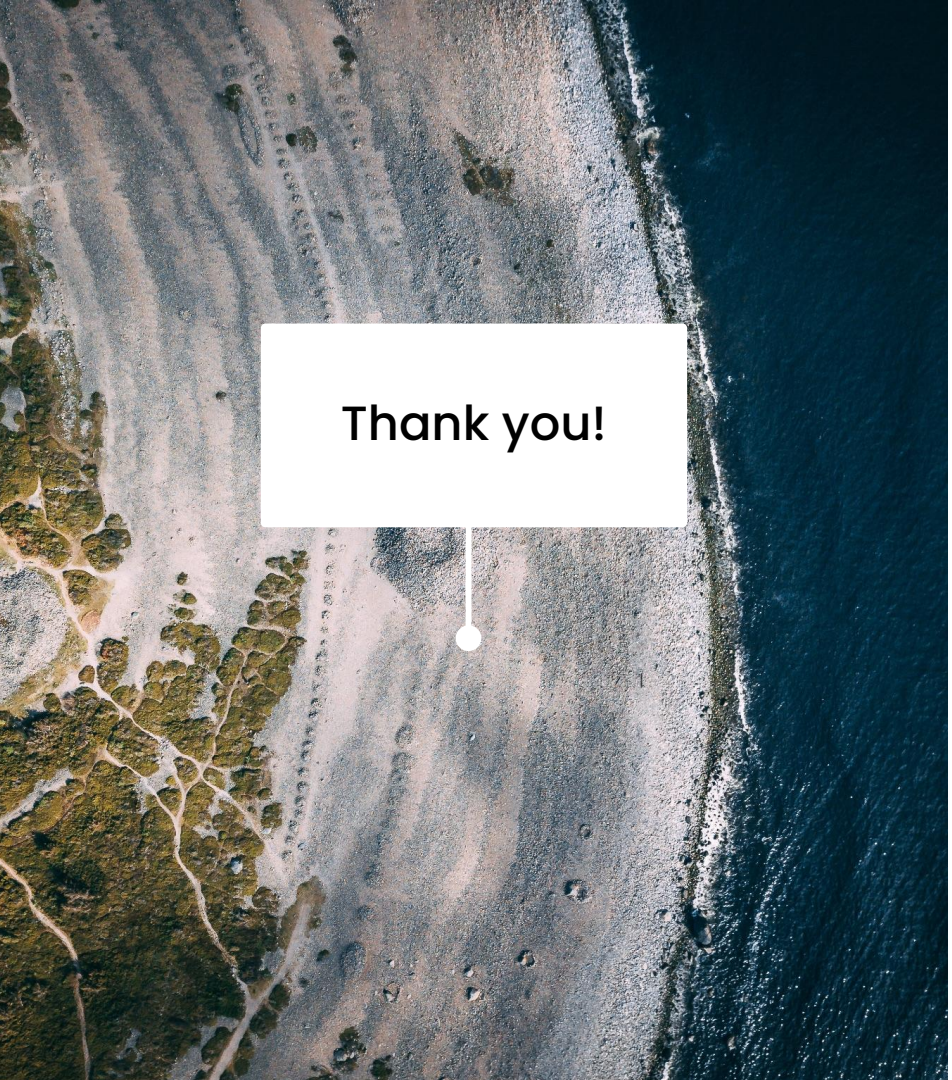
Collect Sentinel-2 data & aggregate to regular daily time-series

Create cloud-free daily data

Combine bands and calculate NDVI

Temporal aggregation to 3 month median

Apply trained ESG model (ONNX)



Geo Engine

Geo Engine GmbH
Am Kornacker 68
35041 Marburg

info@geoengine.de

geoengine-de

@teamgeoengine

www.geoengine.de