Geo Engine: Harmonized data access for data analysis pipelines

Dr. Christian Beilschmidt
GWF 2023
About us

- Geo Engine GmbH est. 2021
- Start-up of University of Marburg, Germany
- Geo informatics researchers from biodiversity and remote sensing projects
- EXIST research transfer

Participation in NFDI4Biodiversity

Co-applicant in CropHype (EnMap)

ML training data pipelines (Sentinel 2 aggregates)
Geo Engine in a nutshell

- Platform for spatio-temporal processing
- Transparent ad-hoc integration of external data via Data Providers
- Process data as time-series instead of files  ➞ Transition from static to temporal analytics
Concept 1: All datasets are time series
Concept 1: All datasets are time series

Data Source

- Vector
- Raster

Time:
- t1
- t2
- t3
- t4

- Has a native resolution
- Handles formats and time-series abstraction
Concept 2: Internal and external data

**Internal datasets**
e.g. point observations

1. GeoPackage

2. Annotation

<table>
<thead>
<tr>
<th>Geom</th>
<th>Time</th>
<th>Text</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2023-05-04</td>
<td>This</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>2023-05-04</td>
<td>is</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>2023-05-04</td>
<td>Text</td>
<td>20.4</td>
</tr>
</tbody>
</table>

3. Integrated time series dataset

**External data**
e.g. Sentinel-2 data from STAC service

1. Sentinel-2 STAC

2. Instantiate STAC Provider
   - Implements listing
   - Implements resolving datasets

3. Integrated time series dataset
Concept 3: Workflows

Raster Source

Re-Project

e.g. A+B

Expression

change SRS (if needed)

automatic harmonization

Raster-Vector-Join

Vector Source

Raster Source
Concept 3: Workflows

Workflow Description

- Raster-Vector-Join
  - Expression
    - e.g. A+B
  - automatic harmonization
- Vector Source
  - Re-Project
  - Raster Source
  - change SRS (if needed)

Processing

register

Workflow ID
Concept 4: Query like a data cube

ID
Workflow + Time (Interval) + Bounding Box + Resolution

Processing

PNGs
Raster Tiles
Vector Chunks
Concept 5: Access from your favorite environment

Web UI WYSIWYG
Concept 5: Access from your favorite environment

Jupyter Notebooks
Programmatically
Example: Random Forest using custom point and Sentinel data

Retrieve S2 data via STAC

Monthly mean of cloud-free NDVI

Load sensor training points

Attach Sentinel-2 product values to points

Stream data in RF model using Python & Jupyter

Apply to all pixel values
Summary

- Geo Engine platform for geo processing & analysis
- Native time series
- Workflows
- Data harmonization
- Low-code
- Jupyter notebooks

Outlook

- Integrate machine learning framework(s) into Geo Engine
- ML ingestion inside
- ML as workflow operator

Call for Cooperation

- Data ingestion use case
- Data products from different EO sources
- Proof-of-concepts
Thank you!