Working with Cloud-Optimized GeoTIFFs in Python

Sept 7, 2021: 13:30 - 15:00

Leandro Parente

leandro.parente@opengeohub.org

https://opengeohub.org

Luka Antonić

lantonic@multione.hr

https://multione.hr

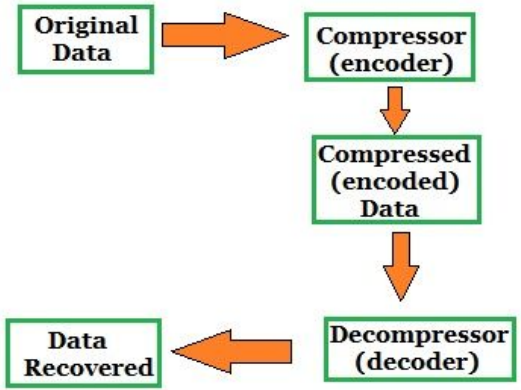
Co-financed by the Connecting Europe Facility of the European Union

Outline

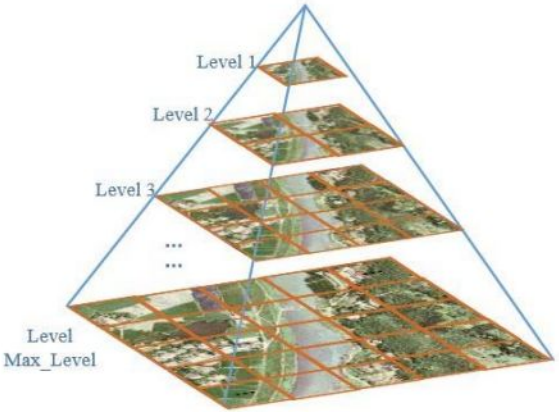
- Introduction to Cloud-Optimized GeoTIFF (COG)
- Accessing COG in QGIS, GDAL and python
- Clipping by a region of interest
- Time series access
- STAC

Introduction to Cloud-Optimized GeoTIFF (COG)

A COG is a regular GeoTIFF file, aimed at being hosted on a HTTP file server, with an internal organization that enables more efficient workflows on the cloud.



Lossless compression
(DEFLATE, LZW)



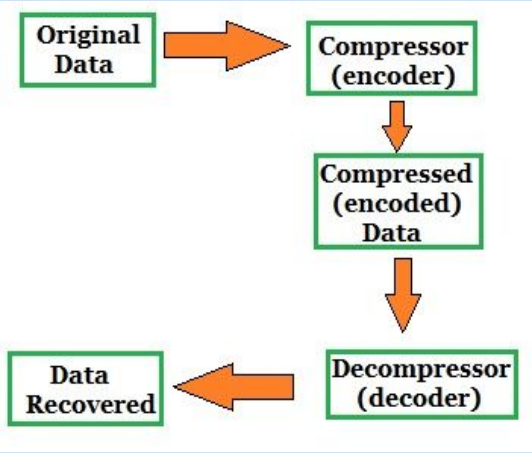
Tilling and overviews
(pyramid structure)

HTTP Get Range requests
(Accept-Ranges: bytes)

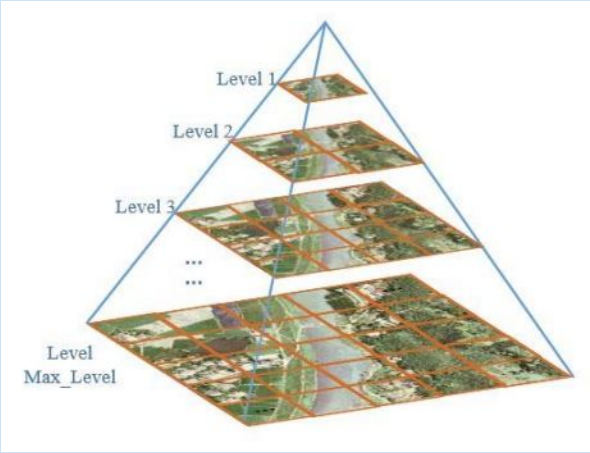
Source: Cloud Optimized GeoTIFF in depth and Guo et al., 2016

Introduction to Cloud-Optimized GeoTIFF (COG)

Cloud Object Storage



Lossless compression
(DEFLATE, LZW)



Tilling and overviews
(pyramid structure)

amazon S3

Google Cloud Storage

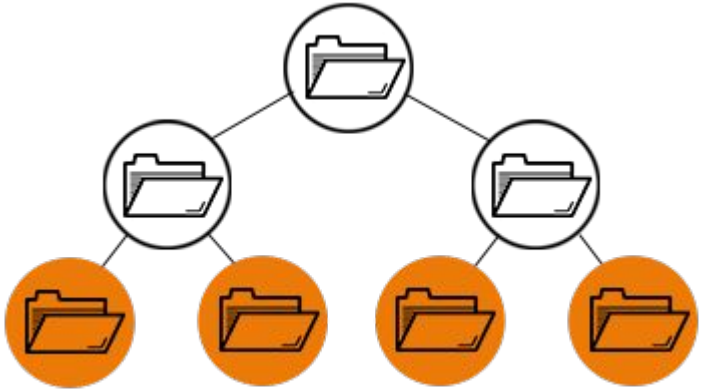
HTTP

HTTP Get Range requests
(Accept-Ranges: bytes)

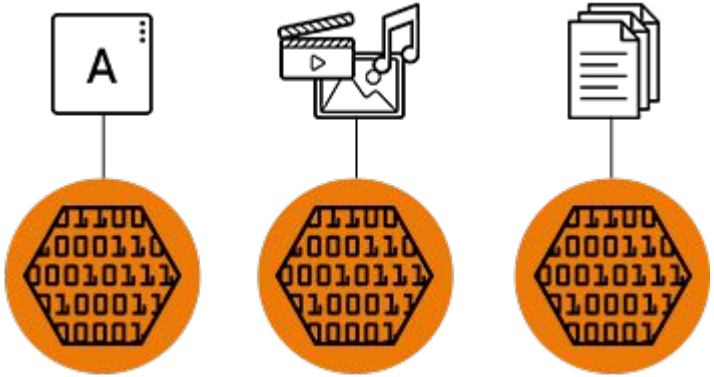
Source: Cloud Optimized GeoTIFF in depth and Guo et al., 2016

Introduction to Cloud-Optimized GeoTIFF (COG)

Object storage is a data storage architecture for large stores of **unstructured data**, which designates each piece of data as an **object**, keeps it in a separate storehouse, and bundles it with **metadata** and a **unique identifier** for easy access and retrieval.



File storage



Object Storage

Source: What is object storage? and File storage, block storage, or object storage?

Introduction to Cloud-Optimized GeoTIFF (COG)

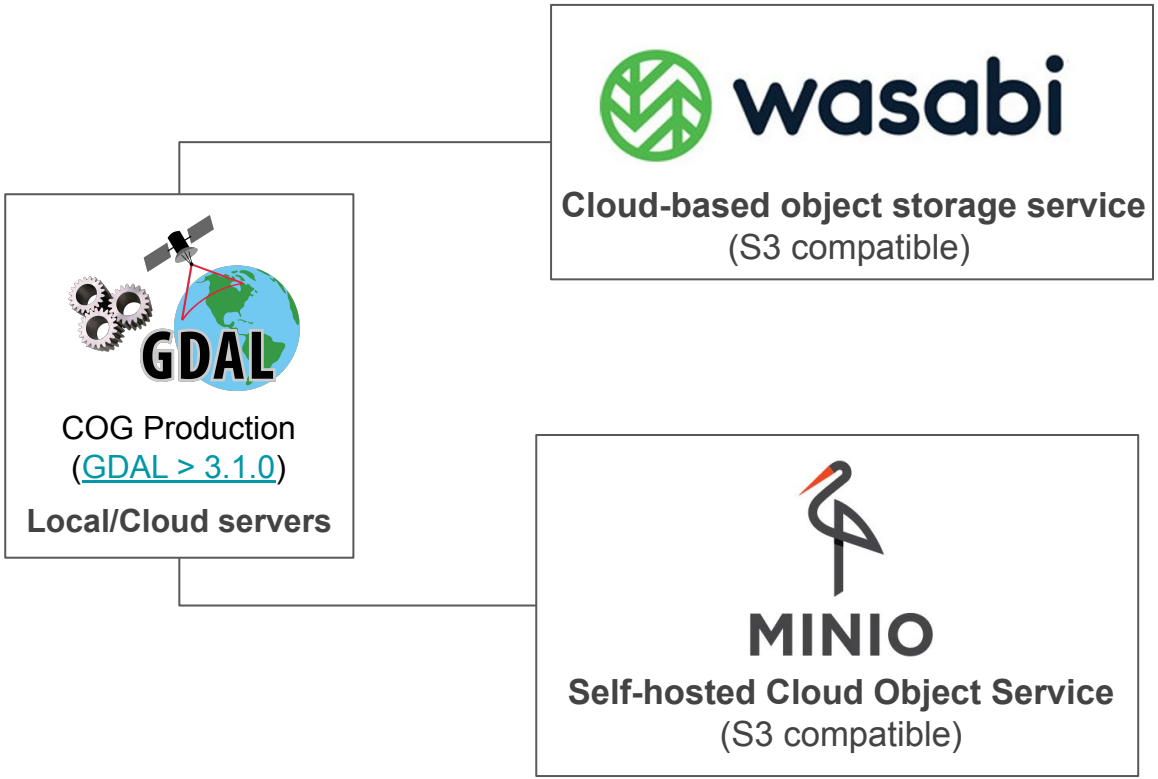
Landsat Cloud Optimized GeoTIFF Data Format Control Book

Cloud Storage public datasets

Sentinel-2, Landsat 8 and CBERS

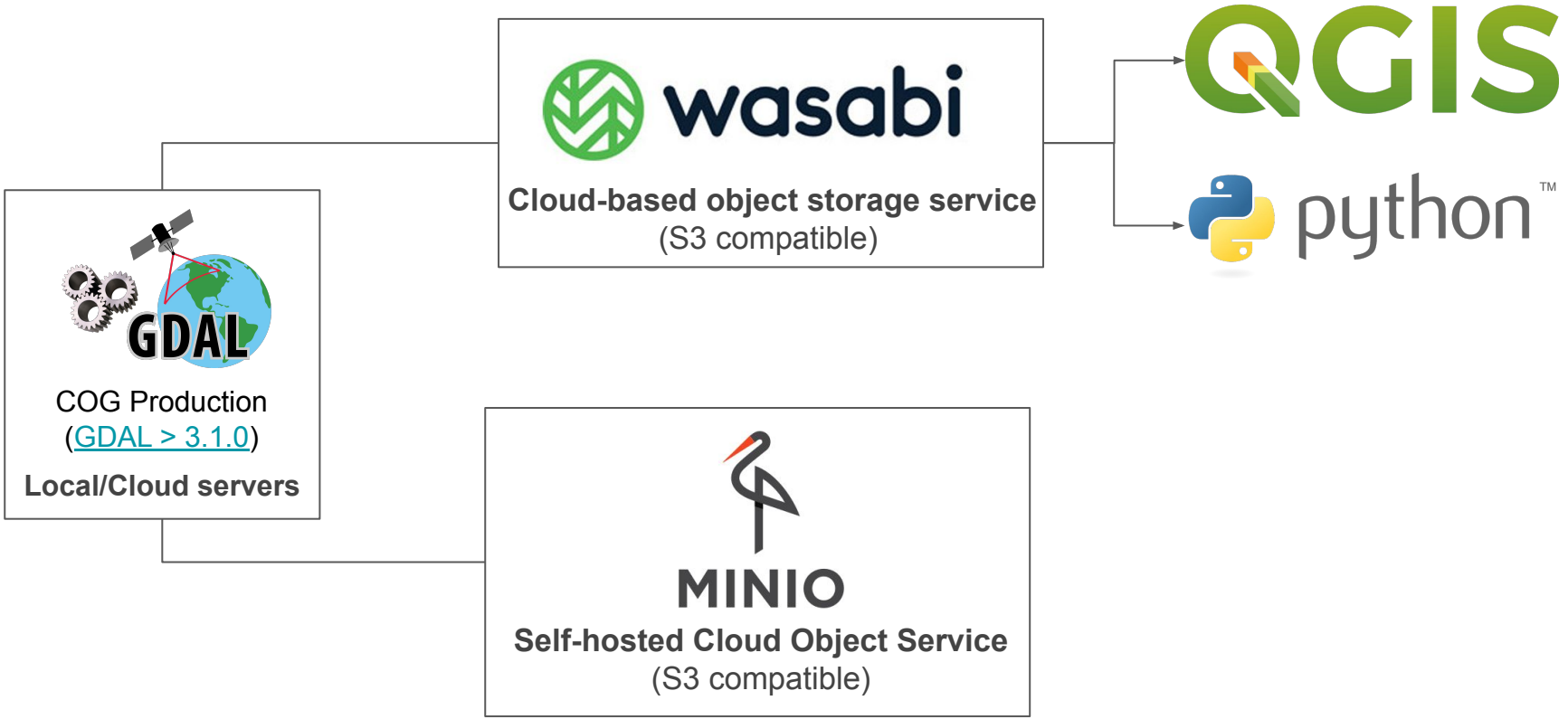
Source: What is object storage? and File storage, block storage, or object storage?

Accessing COG in QGIS, GDAL and python



Source: What is object storage? and File storage, block storage, or object storage?

Accessing COG in QGIS, GDAL and python



Source: What is object storage? and File storage, block storage, or object storage?

Accessing COG in QGIS, GDAL and python

RGB Mosaics

http://s3.eu-central-1.wasabisys.com/eumap/lcv/lcv_rgb_landsat.glad.ard_p50_30m_0..0cm_20110625..20110912_eumap_epsg3035_v1.0.tif

Hand-on

https://gitlab.com/geoharmonizer_inea/odse-workshop-2021

07_WORKING_WITH_COG_IN_PYTHON.IPYNB

- Working with Cloud-Optimized GeoTIFFs in Python
  - Reading and Writing
    - GDAL
  - Time Series Access

06_high_performance_c | 07_working_with_cog_in | opengeohub@odse-vbx:

Markdown — Python 3 (ipykernel)

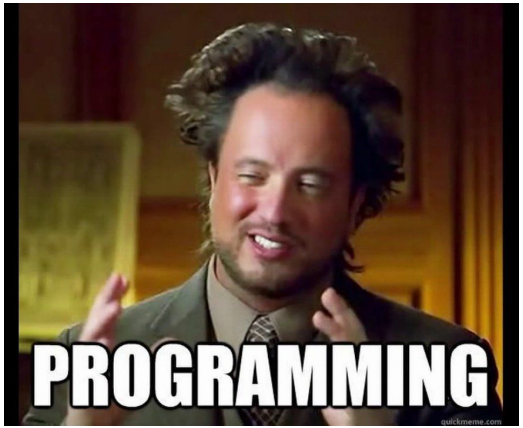
Working with Cloud-Optimized GeoTIFFs in Python

[11]:
```
from pathlib import Path
import rasterio
import numpy as np
import math

import warnings
warnings.filterwarnings('ignore')
```

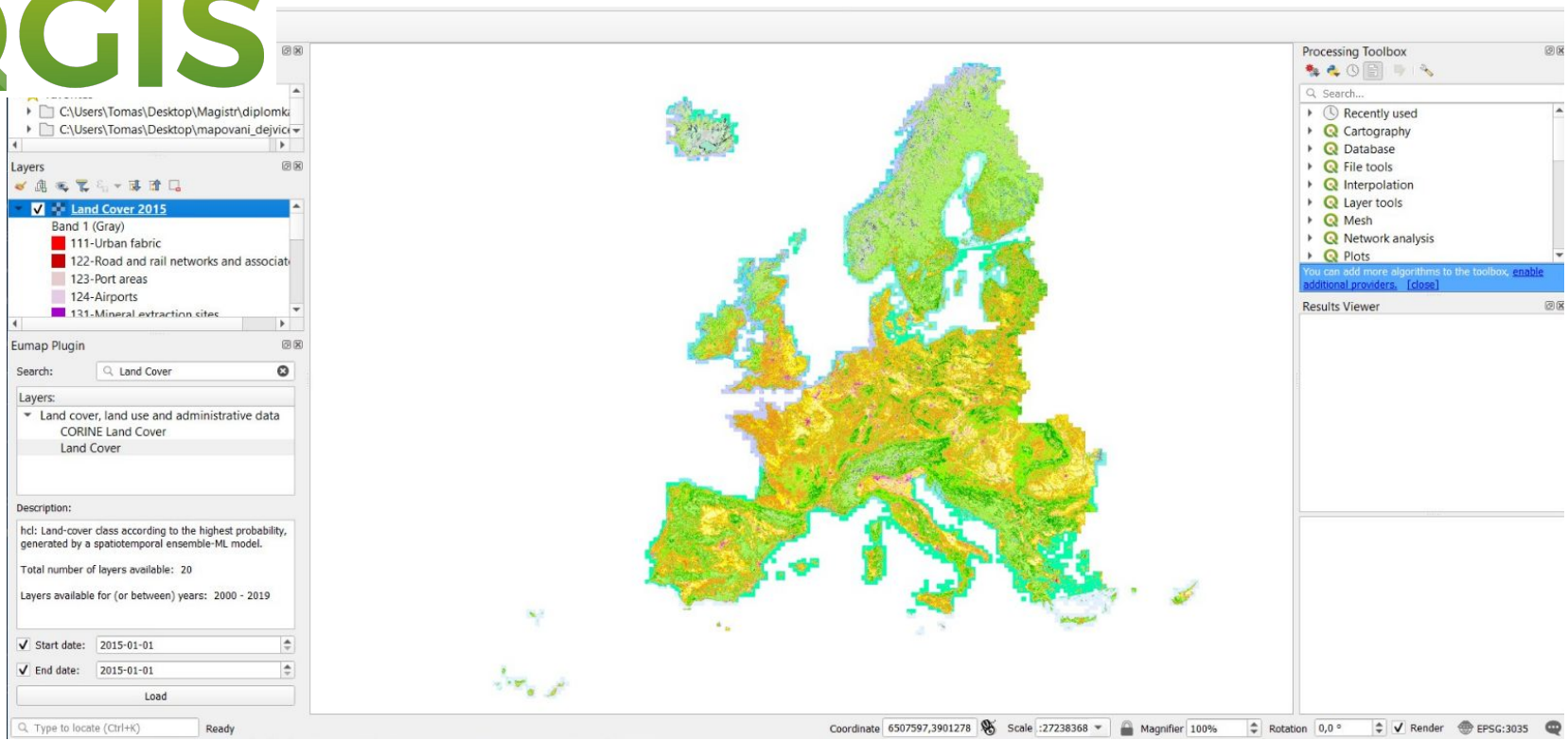
[12]:
```
ODSE_DATA_DIR = Path.home().joinpath('ODSE_workdir').joinpath('data')
raster_dir = ODSE_DATA_DIR.joinpath(f'cog_output')

print(f'ODSE data dir: {ODSE_DATA_DIR}')
print(f'Raster dir: {raster_dir}')
```

```
ODSE data dir: /home/opengeohub/ODSE_workdir/data
Raster dir: /home/opengeohub/ODSE_workdir/data/cog_output
```

[2]:
```
cog_url = 'https://s3.eu-central-1.wasabisys.com/eumap/lcv/lcv_red_landsat.glad.ard_p50_3

with rasterio.open(cog_url) as src:
    # List of overviews from biggest to smallest
    oviews = src.overviews(1)
    print(oviews)
    oview = oviews[-1]

    thumbnail = src.read(1, out_shape=(1, src.height // oview, src.width // oview))
    thumbnail = thumbnail.astype('float32') / 255
```



QGIS Eumap Plugin



What is STAC?

The SpatioTemporal Asset Catalog (STAC) specification provides a common language to describe a range of geospatial information, so it can more easily be indexed and discovered. A 'spatiotemporal asset' is any file that represents information about the earth captured in a certain space and time.

The goal is for all providers of spatiotemporal assets (Imagery, SAR, Point Clouds, Data Cubes, Full Motion Video, etc) to expose their data as SpatioTemporal Asset Catalogs (**STAC**), so that new code doesn't need to be written whenever a new data set or API is released.

Learn More

https://stacindex.org/catalogs

Catalogs

A list of STAC APIs and Static Catalogs.

Filter by Type

All APIs Static Catalogs

Filter by Access Level

All Public only Public & Protected only

Astraea Earth OnDemand

API Public

Astraea Earth OnDemand geospatial imagery query and analysis tool

https://eod-catalog-svc-prod.astraea.earth/

California Forest Observatory

Catalog Public

The Forest Observatory is a data-driven forest monitoring system that maps the drivers of wildfire behavior across the state with a focus on vegetation fuels. Data are available for non-commercial use under the Forest Observatory terms of use.

https://storage.googleapis.com/cfo-public/catalog.json

CBERS

Catalog Public

Imagery acquired by the China-Brazil Earth Resources Satellite (CBERS). The image files are recorded and processed by INPE and are converted to Cloud Optimized Geotiff format in order to optimize its use for cloud based applications. The repository contains all CBERS-4 MUX, AWFI, PAN5M and PAN10M scenes acquired since the start of the satellite mission and is daily updated with new scenes.

Conclusions

- “Cloud-Optimized GeoTIFF” (COG) is state-of-the-art standard to distribute data
- If you know how to use it, it functions as a geospatial database = you enable users to program spatial data analysis on top of your data;
- To further enhance use of your data, consider also adding all metadata into STAC;