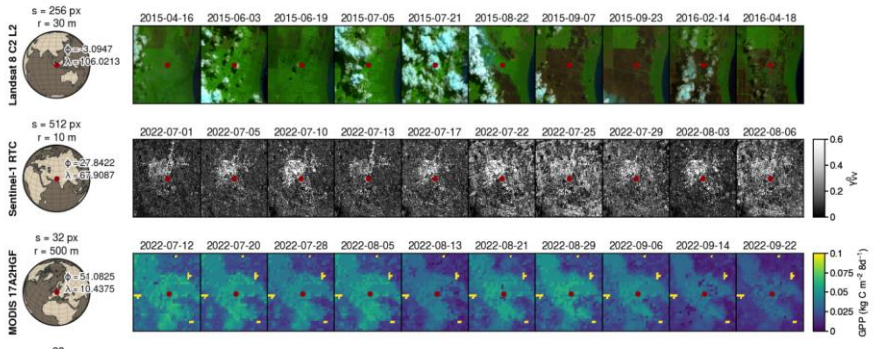
INDIAN INSTITUTE OF REMOTE SENSING, DEHRADUN

Cloud based Geospatial Data Processing in Python

Prasun Kumar Gupta
Scientist/Engineer-'SF', Geoinformatics Department
Geospatial Technology and Outreach Program Group
Indian Institute of Remote Sensing (ISRO), Dehradun
prasun@iirs.gov.in

Introduction

- STAC, COG (Cloud Optimized GeoTIFF), and Xarray are three technologies that work together in a powerful workflow for handling and analyzing geospatial data.



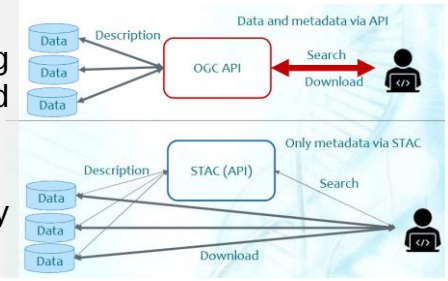
Montero *et al.*, 2024

STAC

- Started in 2018, rapidly developing
- New de facto metadata and search standard
- May become an OGC standard in the future.

- Describes datasets at the level of individual files.
- It is most commonly used for remote sensing data, but it is suitable for any data with **time** and **location** information.

- Users: ESA, USGS, Microsoft Planetary computer, Google Earth Engine, FMI and CSC



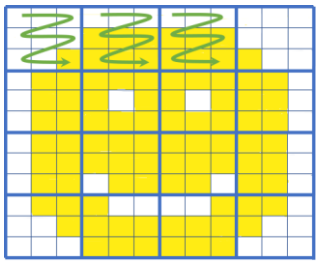
https://www.stacindex.org/

Cloud-Optimized GeoTiff (COG)

- 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.

- It does this by leveraging the ability of clients issuing HTTP GET range requests to ask for just the parts of a file they need.

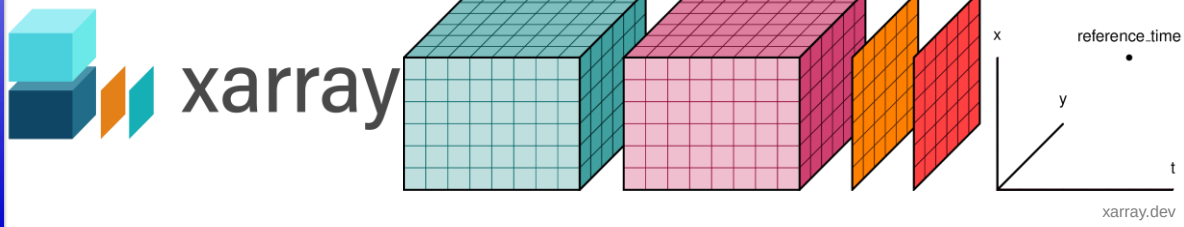
- Enables partial loading of data
- Only loads data for a specific area
- Generalized version of the data is available



element84.com

Xarray

- Introduces labels in the form of dimensions, coordinates, and attributes on top of raw NumPy-like arrays, which allows for more intuitive, more concise, and less error-prone user experience.
- Includes a large and growing library of domain-agnostic functions for advanced analytics and visualization with these data structures.



How do they work together?

- A STAC catalog provides links to individual COG files.
- A library like **pystac** reads the STAC catalog to find the relevant COGs for a specific area and time.
- Instead of downloading the full COG, these libraries use the STAC metadata and the COG's structure to read only the necessary parts.
- This data is then loaded into an Xarray DataArray, ready for analysis. This process is often done lazily, meaning the data is not actually loaded until you perform an operation on it.
- Once the data is in an Xarray object, you can easily filter it, perform calculations, and create composites without manually handling all the file merging.

Time for some hands-on using Python

Thank You

prasun@iirs.gov.in